Biochemical Thermodynamics: Applications of Mathematica

Robert A. Alberty

Massachusetts Institute of Technology
Cambridge, MA

This Page Intentionally Left Blank

Biochemical Thermodynamics:
Applications of Mathematica

**METHODS OF
BIOCHEMICAL ANALYSIS**

Biochemical Thermodynamics: Applications of Mathematica

Robert A. Alberty

Massachusetts Institute of Technology
Cambridge, MA

WILEY-
INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION

Contents

This Page Intentionally Left Blank

Preface

This book is about calculations on the thermodynamics of biochemical reactions that is based on Legendre transforms of the Gibbs energy that bring in the pH, pMg, and concentrations of coenzymes and ligands as independent variables. Chemical reactions are studied under the constraints of constant temperature and constant pressure, but biochemical reactions are studied under the addional constraints of pH and, perhaps, pMg or free concentrations of other metal ions. In considering systems of biochemical reactions, it may be useful to constrain concentrations of various coenzymes. In considering macro-molecule-ligand binding it may be useful to constrain the ligand concentration. For example, the binding of oxygen by hemoglobin can be treated at specified concentrations of molecular oxygen. As more intensive variables are specified, more thermodynamic properties of a system are defined, and the equations that represent thermodynamic properties as functions of independent variables become more complicated. Since more independent variables are involved in biochemical thermody-namics than in chemical thermodynamics, the equations for calculating properties become more complicated. Thanks to the development of mathematical applications for personal computers, these complicated calculations can be carried out much more easily. Thermodynamic calculations on such systems require the use of computers, and *Mathematica* is a very conve-nient application because of its symbolic capabilities, its calculation of partial derivatives, and its facilities for construction of databases, tables, and plots. The fact that *Mathematica* can be used to derive these very complicated functions and take partial derivatives to obtain other properties makes it possible to make calculations that would previously have been consid-ered impractical. Therefore, this book has been written in *Mathematica*. *Mathematica* makes it possible to intermingle text with calculations, as illustrated by a number of recent books. In this book, the calculations of biochemical thermodynamics are described in words, equations are derived by use of *Mathematica*, and then evaluated for specified values of independent variables. All the *Mathematica* calculations are shown, and so the data, equations and programs can be used for the calcula-tion of properties of reactants at other specified values of independent variables.

The main question that thermodynamics deals with is the direction of spontaneous change when a system is initially in a specified state. The state of a biochemical reaction system is described by specifying concentrations of reactants, temperature, pH, ionic strength, and concentrations of free metal ions that are bound by reactants. Concentrations of coen-zymes and ligands like molecular oxygen can also be specified. Thermodynamics is important in biochemistry because it can tell us whether a given enzyme-catalyzed reaction or ligand binding will go to the right or to the left under specified condi-tions. It can also give us the equilibrium composition. Enzymes determine the reactions that are catalyzed in a given system and their rates, but enzymes do not determine the directions of reactions or the amount of energy that is stored, transferred, or is required to synthesize a needed reactant. In a cell, certain reactions are needed to store energy and other reactions are needed to use this energy for necessary purposes of life. If we want to understand how energy is stored and used, we need to know the apparent equilibrium constants K' of the reactions involved under the ambient conditions and also heats of reaction. Other biochemical reactions, like the binding of oxygen by hemoglobin do not require enzyme catalysis, but the equilibrium extents of binding reactions and the influence of other ligands are determined by thermodynamics. With knowledge of pKs, a measurement of the apparent equilibrium constant K' of a biochemical reaction at a single pH and ionic strength makes it

possible to calculate K' as a function of pH and ionic strength. When enthalpies of reaction are available and enthalpies of acid dissociation are known, a measurement of K' at a single temperature, pH, and ionic strength makes it possible to calculate K' and other transformed thermodynamic properties as functions of temperature, pH, and ionic strength.

The apparent equilibrium constants of about 500 enzyme-catalyzed reactions have been determined under various sets of conditions, and enthalpy changes have been measured calorimetrically for some of these reactions or can be calculated from the effect of temperature on the apparent equilibrium constant. In principle these data can be used to calculate standard Gibbs energies of formation and standard enthalpies of formation for the species of about 1000 reactants. The current number of known species matrices is 199. For 94 of these reactants, the $\Delta_f H°$ of all species are known. Further analysis of existing experimental data and new measurements will make it possible to extend the current database BasicBiochemData3. The most efficient way to store thermodynamic information on enzyme-catalyzed reactions is to store data on species because then apparent equilibrium constants and other transformed thermodynamic properties of reactants and reactions can be calculated for specified conditions. Such a database makes it possible to calculate apparent equilibrium constants and transformed thermodynamic properties for many more reactions than it takes to make the data table. An even larger data set can be based on analogies because of similarities in the underlying chemistry of some reactions. Such a database can be used to calculate apparent equilibrium constants that are too large to measure directly. The number of apparent equilibrium constants that can be calculated from a database increases exponentially with the number of reactants in the data base. The calculation of species properties from different enzyme-catalyzed reactions reveals inconsistencies between different equilibrium and calorimetric experiments.

Apparent equilibrium constants cannot be determined experimentally on reactions that go nearly completion. Calorimetric measurements of enthalpies of reaction do not have this problem, Proteins may be reactants in enzyme-catalyzed reactions. When apparent equilibrium constants can be measured on reactions involving proteins, the thermodynamic properties of the reaction site in the protein can be calculated.

It is assumed that the reader has had some introduction to thermodynamics at the level of an undergraduate course in physical chemistry. My previous book "Thermodynamics of Biochemical Reactions," Wiley, Hoboken, NJ (2003) provides a more complete introduction to the structure of thermodynamics and its relation to statistical mechanics. This successor book is needed because more recent research has clarified the structure of biochemical thermodynamics and opened up new possibilities for learning about the flow of energy in living things. Three aspects of these calculations are as follows:

1. Experimental data on enzyme-catalyzed reactions are in the form of apparent equilibrium constants K', heats of reaction, and pKs (and corresponding heats of dissociation), but the most efficient way to store the thermodynamic properties of biochemical reactions is by means of small matrices that give standard Gibbs energies of formation, standard enthalpies of formation, charge numbers, and numbers of hydrogen atoms in each species of a reactant. The bridge between treating enzyme-catalyzed reactions in terms of species and in terms of reactants, like ATP, which is a sum of species, is provided by the Legendre transform $G' = G - n_c(H)\mu(H^+)$, where G' is the transformed Gibbs energy of the system, G is the Gibbs energy of the system, $n_c(H)$ is the amount of the hydrogen component in the system (total amount of hydrogen atoms), and μ (H^+) is the specified chemical potential of hydrogen ions, which is determined by the pH. The standard transformed Gibbs energy of a biochemical reaction is given by $\Delta_r G'° = -RT\ln K'$. The dependence of thermodynamic properties on ionic strength can be calculated using the extended Debye-Hückel equation, which involves a temperature-dependent parameter. When a reactant consists of pseudoisomers that are at equilibrium at a specified pH, isomer group thermodynamics has to be used to calculate the standard transformed Gibbs energy of formation $\Delta_f G'°$ of the reactant. This process leads to functions of temperature, pH, and ionic strength that are too complicated to be written out by hand, but *Mathematica* can be used to derive these functions and to calculate the standard transformed enthalpy of formation, standard transformed entropy of formation, average number of hydrogen ions bound, and other thermodynamic properties by taking partial derivatives.

2. Going from the experimental thermodynamic properties K' and transformed enthalpies of reaction to properties of species involves the concept of the inverse Legendre transform $(G = G' + n_c(H)\mu(H^+))$. Computer programs can be written to go from the experimental properties directly to the standard Gibbs energies of formation and standard enthalpies of formation of the species involved in a reactant. These programs are more complicated than the programs using properties of species to derive the standard transformed thermodynamic properties of reactants.

3. Equilibrium compositions of systems of chemical reactions or systems of enzyme-catalyzed reactions can only be calculated by iterative methods, like the Newton-Raphson method, and so computer programs are required. These computer programs involve matrix operations for going back and forth between conservation matrices and stoichiometric number matrices. A more global view of biochemical equilibria can be obtained by specifying steady-state concentrations of coenzymes. These are referred to as calculations at the third level to distinguish them from the first level (chemical thermodynamic calculations in terms of species) and the second level (biochemical thermodynamic calculations at specified pH in terms of reactants).

In *Mathematica* reactants need to named with words starting with lower case letters because words starting with capital letters refer to operations. Also the names of reactants need to be as short as convenient and cannot involve spaces, subscripts, superscripts, hyphens, dots or other symbols that are *Mathematica* operations. Therefore, ATP is referred to as atp both in the text and in computer programs. Most of these abbreviated names will be recognized immediately, but a glossary of names is provided in the Appendix.

The Appendix contains a copy of the *Mathematica* notebook BasicBiochemData3.nb, Tables of Transformed Thermodynamic Properties, the Glossary of Names of Reactants, the Glossary of Symbols for Thermodynamic Properties, a List of *Mathematica* programs, and Sources of Biochemical Thermodynamic Information on the Web. The *Mathematica* package BasicBiochemData3.m , which is also available at

http : // library.wolfram.com / infocenter / MathSource / 5704

contains all of the species data at 298.15 K and zero ionic strength. It also contains functions of pH and ionic strength for the standard transformed Gibbs energies of formation of 199 reactants at 298.15 K; these functions are named atp, adp,... The functions are also given for the average number of hydrogen atoms in the reactant at 298.15 K as functions of the pH and ionic strength; these functions are named atpNH, adpNH,... Since $\Delta_f H°$ values are known for all species of 94 reactants at 298.15 K and zero ionic strength, the functions of temperature, pH, and ionic strength are given for these 94 reactants for the following transformed thermodynamic properties: $\Delta_f G'°$ (named atpGT, adpGT,...), $\Delta_f H'°$ (named atpHT, adpHT,...), $\Delta_f S'°$ (named atpST, adpST,...), and $\overline{N}_H$ (named atpNHT, adpNHT,...).

Since functions of pH and ionic strength for $\Delta_f G'°$ and $\overline{N}_H$ are known for 199 reactants at 298.15 K, $\Delta_r G'°$ and $\Delta_r N_H$ are calculated in Chapter 12 for 229 enzyme-catalyzed reactions as functions of pH and ionic strength. Since $\Delta_f G°$ and $\Delta_f H°$ are known for all the species of 94 reactants, functions of temperature, pH, and ionic strength that yield $\Delta_f G'°$, $\Delta_f H'°$, $\Delta_f S'°$, and $\overline{N}_H$ for 90 enzyme-catalyzed reactions are given in Chapter 13.

It is not necessary to be a programmer in order to use the programs and procedures illustrated in this book. Names of reactants, temperatures, pHs, and ionic strengths are readily changed in using the various programs.

The CD at the back of the book contains the whole book in *Mathematica*. It can be downloaded into a personal computer with *Mathematica* installed, but it can be read on a computer with *MathReader*, which is freely available from Wolfram Research, Inc. (100 Trade Center Drive, Champaign, IL 61820-7237, and www.wolfram.com). A chapter can be downloaded into a personal computer as a notebook. The following chapters do not require that BasicBiochemData3 be loaded: Chapters 2, 3, 5, 6, and 14. Chapters 1, 4, 7, 8, 9, 10, 11, 12, 13, and 15 need BasicBiochemData3 to be loaded by typing <<BiochemThermo`BasicBiochemData3` (see Use of *Mathematica*).

Robert A. Alberty
Cambridge, Massachusetts

Use of Mathematica

Even if you are not familiar with *Mathematica* (Wolfram Research, Inc. 100 Trade Center Drive, Champaign, IL 61820-7237 and www.wolfram.com), you should be able to read this book. The concepts and calculations in biochemical thermodynamics are explained in words in the textual parts of the book. And the results of calculations are discussed in words. When *Mathematica* is used to make tables and figures, explanatory titles are given. Names of biochemical reactants in *Mathematica* call up mathematical functions of temperature, pH, and ionic strength or mathematical functions of pH and ionic strength at $T = 298.15$ K (25 °C). Since mames of mathematical functions in *Mathematica* have to start with lowercase letters and cannot contain dots, dashes, or spaces, short names like atp are used for reactants. There is a complete list of these short names and the corresponding scientific names in the Appendix 3.

BasicBiochemData3 contains functions of pH and ionic strength that give standard transformed thermodynamic properties of 199 reactants for which values of $\Delta_f G°$ are known for all species that have significant concentrations in the pH 5 to pH 9 range. Using ATP as an example, these two functions are as follows:

atp: This function of pH and ionic strength yields the standard transformed Gibbs energy of formation of ATP at 298.15 K.
atpNH: This function of pH and ionic strength yields the average number of hydrogen atoms in ATP at 298.15 K.

For 94 of these reactants $\Delta_f G°$ and $\Delta_f H°$ are known for all species that have significant concentrations in the pH 5 to pH 9 range. Using ATP as an example, these four functions are as follows:

atpGT: This function of T, pH, and ionic strength yields the standard transformed Gibbs energy of formation of ATP.
atpHT: This function of T, pH, and ionic strength yields the standard transformed enthalpy of formation of ATP.
atpST: This function of T, pH, and ionic strength yields the standard transformed entropy of formation of ATP.
atpNHT: This function of T, pH, and ionic strength yields the average number of hydrogen atoms in ATP.

Since *Mathematica* is a high level language that uses words like Integrate and abbreviations like D for differentiate, you can see what mathematical operations are involved in a program. Everything that is involved in making these calculations is shown. When a calculation is made, a semicolon is often put at the end of the input so that the result is not shown immediately. When calculations are performed in your computer, the semicolon can be deleted to see what the result is, but semicolons are used in the book to save space.

The CD contains the whole book in *Mathematica*. When you put the CD in your computer, you will see a list of chapters and appendices. If *Mathematica* is installed in your computer, you can click on a chapter or appendix, and it will come up on the screen. If it is Chapter 2, 3, 5, 6, or 14, all the calculations can be run by using Kernel/Evaluation/Evaluate Notebook. This will take a few seconds or a couple of minutes depending on the chapter. When you load a new chapter, you can also run a cell at a time in the order they are in the chapter. The values of arguments in programs can be changed to make calculations at different temperatures, pHs, and ionic strengths. Each chapter should be opened in a fresh workspace.

If you want to run programs in Chapters 1, 4, 7, 8, 9, 10, 11, 12, 13, or 15, BasicBiochemData3 is needed (see Appendix 1). BasicBiochemData3.nb and BasicBiochemData3.m are also available at

`http : // library.wolfram.com / infocenter / MathSource / 5704`

Note that there is a BasicBiochemicalData3.m file in the CD that is not in the printed book. When the notebook BasicBiochemData3.nb was made, a package version BasicBiochemData3.m was made automatically. The package consists of the *Mathematica* input without the text. The package BasicBiochemData3.m needs to be installed in your personal computer as described in the following *Instructions for the use of the package BasicBiochemData3.m*. When it is installed, it possible to load it into a workspace by use of <<BiochemThermo`BasicBiochemData3`. The value of a mathematical function can be calculated using ReplaceAll (/.x->). For example, the standard transformed Gibbs energy of formation of ATP in kJ mol^{-1} at

298.15 K, pH 7, and ionic strength 0.25 M can be calculated by typing atp/.pH->7/.is->.25. This value can also be obtained by typing atpGT/.t->298.15/.pH->7/.is->.25.

In calculations on enzyme-catalyzed reactions, one of the ways a reaction equation can be entered is in the form atp+h2o+de==adp+pi. Note that hydrogen ions are never shown in a reaction equation at specified pH. Other programs may require that the reaction equation be written in the form ec3x6x1x3=adp+pi-(atp+h2o), where the name of the reaction is the EC (Enzyme Nomenclature) number with decimal points replaced by x's. It is especially simple to change the ranges of independent variables in tables and figures.

When a change is made in a chapter, the chapter can be saved in your computer, but the version in the CD cannot be changed. If you do not want to save the whole chapter, you can copy a calculation and paste it into a new notebook, where it can be saved where you want it in your computer. When you make a new notebook it needs to contain the programs that are used.

Mathematica provides for several styles of programing: functional programming, procedural programing, rule-based programming, and recursion. Almost all of the programs in this book are examples of functional programming. This is a style that is quite distinct from what is available in traditional computer languages. A functional program is a mathematical function and the inputs to the program are the arguments of the function. When the program is run, the function is applied to its arguments. The way of writing a functional program looks like a mathematical equation. For example, a function of x and y is written as functionname[x_,y_]:=body, where x and y are arguments. The body of the function can be a single expression or a series of expressions. The := is referred to as a delayed assignment. When the program is typed in, nothing is returned. But when the name of the function is typed in with values for the arguments, the program returns the result of the calculation. For example,

```
In[260]:=
        square[x_] := x^2

In[261]:=
        square[5]

Out[261]=
        25
```

Instructions for the use of the package BasicBiochemData3.m

When the 199 small matrices of species data or the 774 functions are needed, the command
<<BiochemThermo`BasicBiochemData3`
is used to make this information available. It is necessary for the user of this book to put BasicBiochemData3.m
in their computer so that
<<BiochemThermo`BasicBiochemData3`
will work. To load this package properly, it is first necessary to create a folder named BiochemThermo and put
BasicBiochemData3.m into this folder. Then it is necessary to find out where to put this folder as follows:
Open a *Mathematica* session and evaluate $UserBaseDirectory. On my Mac computer, this yields
/Users/robertalberty/Library/*Mathematica*. In this *Mathematica* file, you will find Applications. Put the folder
BiochemThermo (with BasicBiochemData3.m in it) into this Applications directory. Now this package becomes
available whenever you load it using
<<BiochemThermo`BasicBiochemData3`

When this package is loaded, all the species data on 199 reactants and the 774 functions under this package will be available for calculations. Each chapter should be opened in a fresh workspace. A whole chapter can be run by use of Kernel/Evaluation/Evaluate Notebook. The properties for adenosine triphosphate are named atpsp, atp, atpNH, atpGT, atpHT, atpST, and atpNHT. These functions are all protected; that is, none of them can be changed without unprotecting them. These thermodynamic values are based on the usual conventions of chemical thermodynamic tables that $\Delta_f G° = \Delta_f H° = 0$ for elements in defined reference states and for H^+ ($a=1$). Additional canventions are that $\Delta_f G° = \Delta_f H° = 0$ for $glutathione_{ox}^{2-}$, NAD_{ox}^{-1}, $NADP_{ox}^{3-}$, $retinal^0$, $thioredoxin_{ox}^0$, and $ubiquinone_{ox}^0$.

Sources of Biochemical Thermodynamic Information on the Web

In the lists of references in the chapters, some have URLs (Uniform Resource Locator). These URLs in the CD that contains this book are active in the sense that if you click on them in a computer connected with the Web, the data source will come up on your screen. These URLs are all given in one place in Appendix 6, which includes a short description of their content.

A number of books have been written to help people get started in *Mathematica*: Three of these are
B. F. Torrence and E. A. Torrence, The Student's Introduction to *Mathematica*, Cambridge University Press, 1999.
P. Wellin,R. Gaylord, and S. Kumin, An Introduction to Programming with *Mathematica*, Cambridge University Press, 2005.
R. J. Silbey, R. A. Alberty, and M. Bawendi, Solutions Manual to Accompany Physical Chemistry, Wiley, Hoboken, NJ, 2005.

Biochemical Thermodynamics:
Applications of Mathematica