Bayesian analysis of biochemical thermodynamics data

Teddy Groves

Introduction

The behaviour of networks of metabolic reactions is determined, among other factors, by the amount of energy each reaction releases or stores from its environment; that is, the reaction's Gibbs free energy change $\Delta_r G$. In order to make best use of metabolomic and fluxomic data about a biological system - for example to predict whether an intervention like over-expressing an enzyme would help or hinder an organism's metabolism - information about the system's thermodynamics is required.

While $\Delta_r G$ has been measured for some reactions, network modelling requires values for reactions where it has not yet been measured or for which direct measurement would be infeasible. These reactions' properties must be inferred from related measurements, together with biological and thermodynamic knowledge.

This paper proposes a novel approach to the task of inferring biochemical Gibbs free energy changes, framing it as a Bayesian statistical inference problem.

[EXPAND, REFER TO STRUCTURE OF PAPER]

Theoretical background

This section explains how biochemical reactions' Gibbs free energy changes are affected by the formation energies of their reactants, as well as experimental conditions and ion dissociation constants.

How Gibbs energies of reaction depend on formation energies of reactants

The formation energy of a compound is the amount of energy that is stored or released by the chemical reaction that creates it out of its constituent elements. For example, according to Wikipedia, liquid water has a formation energy of

-237.14 kilojoules per mole. This means that creating 1 mole of liquid water out of gaseous hydrogen and oxygen stores 237.14 kilojoules of energy from the surrounding environment.

The relationship between the condition-specific gibbs free energy change of a reaction $\Delta_r G'^o$ and the condition-specific formation energies $s_i \Delta_f G'^o$ and stoichiometric coefficients s of its reactants is as follows:

$$\Delta_r G^{o\prime} = \sum_{i \ is \ a \ reactant} s_i \Delta_f G_i^{o\prime}$$

Since stoichiometric coefficients are typically well-known, biochemical Gibbs energy changes are essentially determined by the formation energies of their reactants.

How formation energies of compounds depend on formation energies of microspecies

When dissolved in water, most biologically interesting compounds exist in several different forms called 'pseudoisomers' or 'microspecies', each with a different configuration of bindings to metal and hydrogen ions (the latter is sometimes referred to as a protonation state). The different microspecies have different thermodynamic characteristics and relative proportions respond differently to experimental conditions like temperature, ionic strength and concentration of metal and hydrogen ions. Consequently the different microspecies they need to be considered individually in order to capture the behaviour of the whole compound.

The relationship between a compound's condition-specific formation energy and those of its microspecies is as follows:[1]

$$\Delta_f G^{o\prime} = -RT \ln \sum_{i \ is \ a \ microspecies} \exp(-\frac{\Delta_f G_i^{o\prime}}{RT})$$

How formation energies of microspecies depend on conditions

The condition-specific formation energy of a microspecies has the following relationship with the conditions (i.e. the temperature T, ionic strength I, pH and pMg), the microspecies's standard-condition formation entropy $\Delta_f S^o$, the properties of the microspecies (i.e. its charge z, its number of protons nH and its number of magnesium ions nMg) and some temperature-specific quantities

[1] see Alberty (2003) section 4.5, 'thermodynamics of pseudoisomer groups at specified pH'

(the formation energy of magnesium $\Delta_f G^o_{Mg}(T)$ and the Debeye-Hükel number α):[2]

$$\begin{aligned}\Delta_f G^{o\prime} = &\ \Delta_f G^o \\ &- \ (T - 298.15) \cdot \Delta_f S^o \\ &+ \ nH \cdot RT \cdot \ln(10) \cdot pH \\ &- \ nM \cdot (\Delta_f G^o_{Mg}(T) - RT \cdot \ln(10) \cdot pMg) \\ &- \ RT \cdot \alpha \cdot (z^2 - nH) \cdot (\frac{\sqrt{I}}{1 + \sqrt{I}} - 0.3 \cdot I)\end{aligned}$$

How formation energies of microspecies can be inferred from dissociation constants

Standard-condition formation energies of microspecies are typically not all measured directly. Fortunately, the binding reactions that form diffent microspecies can be measured, making it possible to infer relative formation energies. The quantities measured are called dissociation constants. The dissociation constant for a binding reaction is the equilibrium ratio between the concentration of substrates to products, i.e. the unbound compound and ligand vs the bound compound. Negative log-scale acid dissociation constants are called PK_a and negative log-scale magnesium dissociation constants are called PK_{Mg}.

If a microspecies with minimum hydrogen ions has formation energy $\Delta_f G^o_0$, then a microspecies of the same compound with the same number of magnesium ions and n more hydrogen ions has formation energy

$$\Delta_f G^o_n = \Delta_f G^o_0 - \sum_{i=0}^{n} RT \cdot \log_{10}(PK_{ai})$$

Similarly for magnesium ions, if a microspecies with no magnesium ions has formation energy $\Delta_f G^o_0$, then a microspecies of the same compound with the same number of hydrogen ions and n more magnesium ions has formation energy

$$\Delta_f G^o_n = \Delta_f G^o_0 + n \cdot \Delta_f G^o_{Mg} - \sum_{i=0}^{n} RT \cdot \log_{10}(PK_{Mgi})$$

where $\Delta_f G^o_{Mg}$ is the standard condition formation energy of magnesium. This extra term appears because magnesium ions are bonded pairs of magnesium atoms and some energy is required to form the bond, whereas hydrogen ions have zero formation energy.

[2] see Du et al. (2018) supplementary material, equation 8

Summary

The theoretical machinery described above makes it possible to express the gibbs free energy change of a biochemical reactions in given conditions as a function of the conditions, the formation energies and entropies of its reactants and the dissociation constants of those reactants' microspecies. The procedure is as follows:

1. Find the relative condition-independent formation energies of all microspecies of all reactants using dissociation constants.
2. Find the condition-specific formation energy of each microspecies using its condition-independent formation energy and entropy, the conditions and supporting information like the Debeye-Hükel numbers.
3. Find the condition-specific formation energy of each reactant using those of its microspecies.
4. Find the condition specific gibbs free energy change of the reaction using the formation energies and stoichiometric coefficients of its reactants.

Proposed model

This paper proposes to treat the analysis of biochemical thermodynamics data as a Bayesian statistical inference problem. This means specifying unknown quantities, measurements, a measurement model specifying how the measurements provide information about the unknown quantities and a prior model representing the available pre-experimental information about the unknowns.

The unknown quantities in our model are the standard-condition compound formation energies and dissociation constants.

The measurements come in three categories: direct measurements of microspecies' formation energies, dissociation constant measurements and measurements of equilibrium constants. Another category that ought to be included is measurements of reduction potential change due to redox reactions; the current paper does not do this as no such reactions are involved in the available data.

The measurement model incorporates probabilistic assumptions about measurement error - i.e. how the measurements depend on the true values of the measured quantities - and deterministic assumptions about how the measured quantities depend on the unknown quantities.

For the probabilistic assumptions we use a regression model where the observed gibbs energy changes (derived from values of observed equilibrium constants), dissociation constants and formation energies are noisy draws from a normal distribution centered at the true values of these quantities, with the scale of the noise known in advance.

The deterministic component of the measurement model specifies the relationship between observed gibbs energy changes and partially latent compound formation energies and dissociation constants, using the theoretical relationships described above, with one deviation. Unlike in cite:duTemperatureDependent EstimationGibbs2018, we exclude the effect of temperature and formation entropy from the calculation of microspecies' condition-specific formation energies, and exclude ions other than Hydrogen and Magnesium. Both of these exclusions are for the sake of simplicity and introduce potential biases.

The prior model for standard-condition compound formation energies is a vector of independent normal distributions with known location and scale parameters derived from background knowledge.

The prior model for dissociation constants needs to take into account that for the same microspecies and ion kind (i.e. hydogen or magnesium), each successive binding reaction has a lower dissociation constant than the last. Our prior model enforces this constraint by representing information about the absolute value each compound and ion type's first dissociation constant with a normal distribution, and information about the differences between subsequent dissociation constants with lognormal distributions.

In mathematical formulation our model is as follows: [FORMULATION]

Difference from existing approaches

Existing approaches [REFER TO DU et al review paper]

Main difference: pipeline vs single model

Pros:

- uncertainty properly propagated
- easier to identify bad measurements and incorrect assumptions
- better predictions?

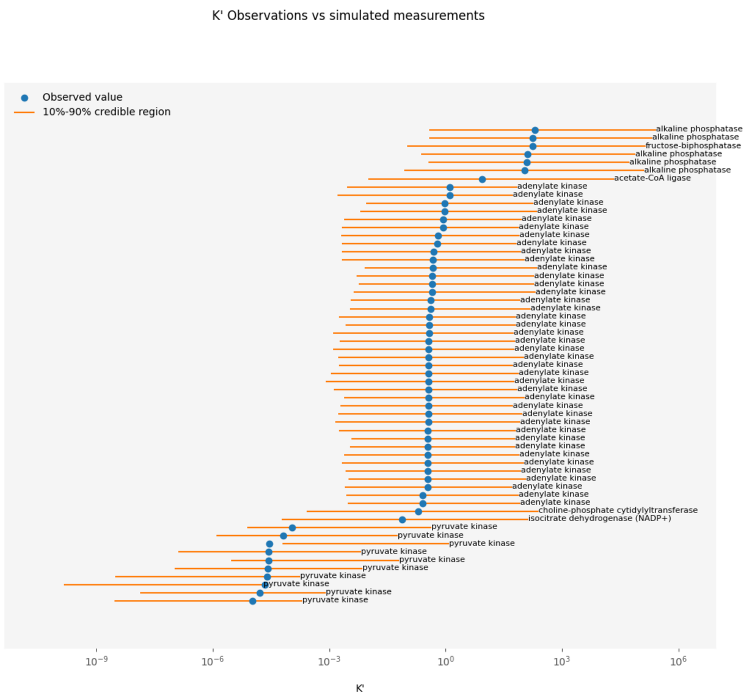
Cons:

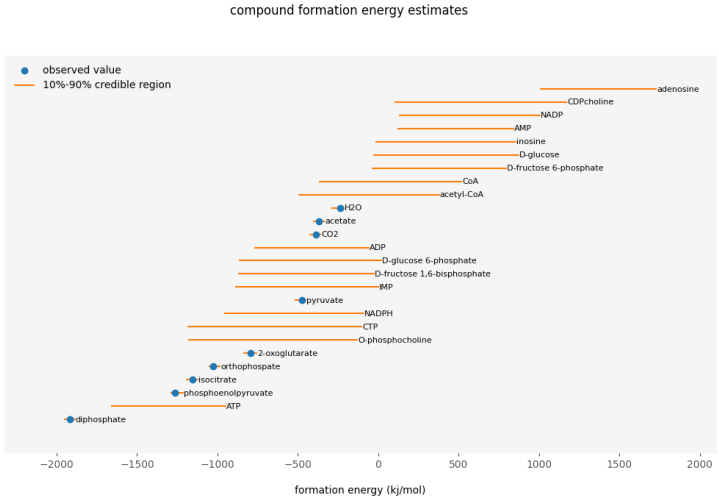
- no group contribution so much lower coverage
- slower to compute

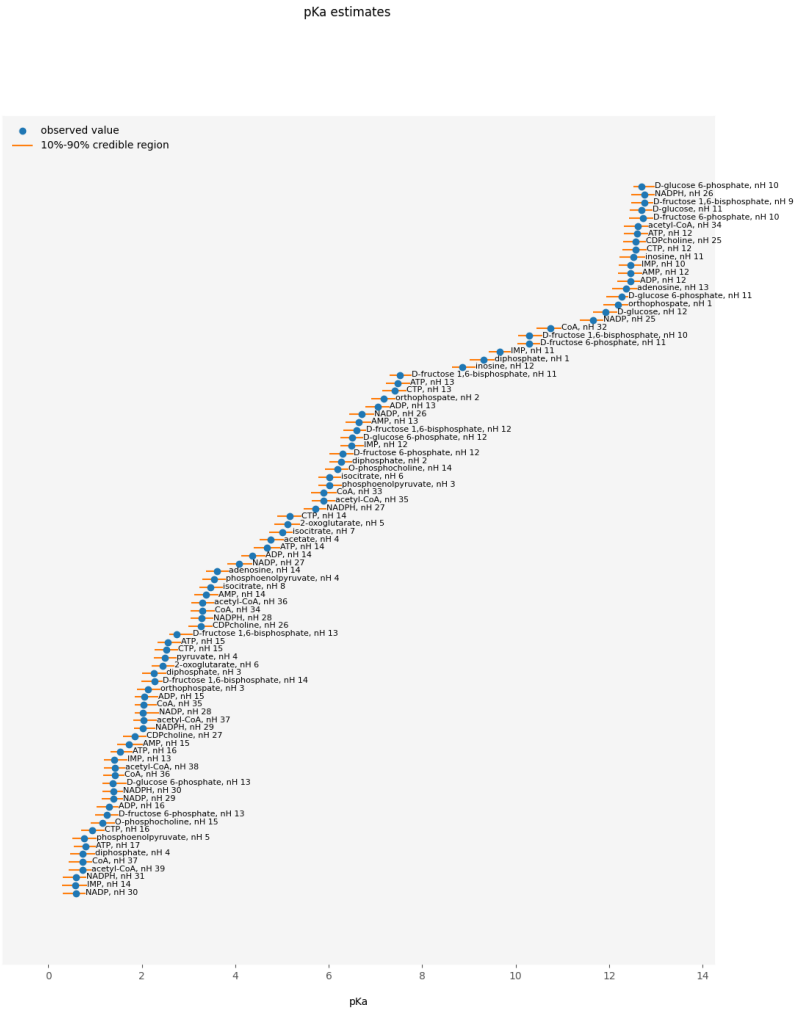
Results

Data

Model output




compound formation energy estimates
observed value
10%-90% credible region
adenosine
CDPcholine
NADP
AMP
inosine
D-glucose
D-fructose 6-phosphate
CoA
acetyl-CoA
H2O
acetate
CO2
ADP
D-glucose 6-phosphate
D-fructose 1,6-bisphosphate
IMP
pyruvate
NADPH
CTP
O-phosphocholine
2-oxoglutarate
orthophospate
isocitrate
phosphoenolpyruvate
ATP
diphosphate
−2000
−1500
−1000
−500
0
500
1000
1500
2000
formation energy (kj/mol)



Comparison with component contribution

Alberty, Robert A. 2003. *Thermodynamics of Biochemical Reactions*. Hoboken, N.J: Wiley-Interscience.

Du, Bin, Zhen Zhang, Sharon Grubner, James T. Yurkovich, Bernhard O. Palsson, and Daniel C. Zielinski. 2018. "Temperature-Dependent Estimation of Gibbs Energies Using an Updated Group-Contribution Method." *Biophysical Journal* 114 (11): 2691–2702. https://doi.org/10.1016/j.bpj.2018.04.030.