von Bertalanffy 1.0 : a COBRA toolbox extension to establish a thermodynamically constrained metabolic model

September 11, 2010

R.M.T. Fleming[1]
Science Institute and Center for Systems Biology,
University of Iceland,
Reykjavik,
Iceland.

I. Thiele
Center for Systems Biology and
Faculty of Industrial Engineering, Mechanical Engineering
and Computer Science,
University of Iceland,
Reykjavik,
Iceland.

[1]Corresponding author. Address: Science Institute and Center for Systems Biology, University of Iceland. E-mail address: ronan.mt.fleming AT gmail.com Tel.:+354 618 6245

Abstract

Summary: In flux balance analysis of genome scale stoichiometric models of metabolism, the principal constraints are uptake or secretion rates, the steady state mass conservation assumption and reaction directionality. Here, we introduce an algorithmic pipeline for quantitative assignment of reaction directionality in multi-compartmental genome scale models based on an application of the second law of thermodynamics to each reaction. Given experimental or computationally estimated standard metabolite species Gibbs energy and metabolite concentrations, the algorithms bounds reaction Gibbs energy, which is transformed to *in vivo* pH, temperature, ionic strength and electrical potential. This cross platform MATLAB extension to the, COnstraint-Based Reconsruction and Analysis (COBRA) toolbox, is computationally efficient, extensively documented and open source.

Availability: http://opencobra.sourceforge.net

1 Introduction

Ludwig von Bertalanffy (1901-1972) sought a theory encompassing "universal principles applying to systems in general" [15] and emphasized the role of thermodynamics in understanding biological systems. In modeling the overall behavior of a genome scale stoichiometric model of metabolism [12], one of the key constraints is reaction directionality which is required to follow the second law of thermodynamics[6]. At constant temperature, pressure, pH and ionic strength, this requirement means that the stoichiometrically weighted sum of transformed Gibbs energy of net substrates must exceeds that of net products [1]. With experimentally derived (e.g. [13]) or computationally estimated standard reactant Gibbs energy [11], if one assumes a typically narrow physiological range of reactant concentrations, one may constrain reaction directionality on a genome scale [9, 7, 8].

The Constraint-Based Reconsruction and Analysis (COBRA) toolbox [3] is a growing suite of MATLAB functions for quantitative prediction of cellular metabolism. Here, we introduce an extension of the COBRA toolbox for quantitative assignment of stoichiometric model reaction directionality via integration of experimentally derived standard Gibbs energies [1], group contribution estimates of standard Gibbs energies [11] and reactant concentrations. Encoded in a single extensively documented pipeline is a thorough thermodynamic treatment of the necessary transformation of such data [8] to *in vivo* temperature as well as compartment specific pH, ionic strength and electrical potential for multi-compartment models.

The pipeline automatically generates extensive statistics and figures, focusing on quantitative reaction directions which disagree with reconstruction directions, manually assigned from literature [14]. The COBRA toolbox supports model exchange via the Systems Biology Markup Language [10] and libSBML [4], thereby the extension reported herein may be applied to an arbitrary mass and charge balanced metabolic reconstruction.

Data	Ref./Range
Metabolic reconstruction	[14]
Experimental $\Delta_i G^o$ & $\Delta_i H^o$	[2]
+/- Group contribution estimate of $\Delta_i G^o$	[11]
Metabolite structure-data file	[5]
Temperature	273 to 313 K
pH	5 to 9
Electrical potential	(mV) e.g. [16]
Ionic strength	0-0.35 Molar
Metabolomic data	

Table 1: Data required to establish a thermodynamically constrained stoichiometric model of metabolism. A group contribution estimate of $\Delta_i G^o$ is necessary if no experimental data is available.

2 Methods

We have previously documented the necessary biophysical chemistry theory whilst thermodynamically constraining a genome scale model of *E. coli* metabolism [8]. The necessary prerequisite data are indicated in Table 1. Alberty [2] has established thermodynamic tables of experimentally derived *metabolite species* standard Gibbs energy, $\Delta_i G^o$, and standard enthalpy $\Delta_i H^o$ for ~100 *reactants*, at $298.15K$, atmospheric pressure and zero ionic strength (See supplement 1). A reactant is a group of metabolite species differing only in state of protonation. We interface the COBRA toolbox extension with a web based implementation of a group contribution method tailored to organic molecules[11]. The estimated metabolite species standard Gibbs energy, $\Delta_i G^o$, for the most predominant protonated form of a metabolite species at a particular pH, is used to augment the experimental data.

Initiation of the pipeline (via the function setupThermoModel.m) is followed by algorithmic checks on the quality of the reconstruction, such as mass and charge balance, presence of chemical formulae, compartment identification, identification of reactions which exchange metabolite species with the environment, and feasibility of steady state reaction flux. Certain adjustments to reaction stoichiometry are made in order to maintain thermodynamic consistency. For example, CO_2 dissolved in water is represented thermodynamically as H_2CO_3 so each reaction involving CO_2must have H_2O added to the opposite side [1]

Experimentally derived standard transformed Gibbs energies, augmented by group contribution estimates, are assigned to each compartment specific reactant. Standard Gibbs energy of formation is usually reported at 298.15 K. We adjust this to a temperature between 273 and 313 K using the van't Hoff equation [1]. The effect of ionic strength on the activity of each metabolite species is taken into ac-

count with the extended Debye-Huckle equation [1]. Within a given compartment, Legendre transformations of metabolite species standard Gibbs energy of formation, for specified pH and electrical potential, allow a group of metabolite species to be thermodynamically represented as a single reactant. The hydrogen ion stoichiometric coefficient is adjusted to balance the number of hydrogen atoms in each reaction based on thermodynamic calculation of the average number of hydrogen ions bound by each reactant. In doing so, a compartment specific adjustment is made to recover H^+ concentration as from a measurement of H^+ *activity*. For reactions which transport between compartments at different pH, or electrical potential, thermodynamic consistency is maintained by adjustment to the standard Gibbs energy for the reaction, depending on the number of hydrogen atoms, or the charge, of the transported metabolite species.

The transformed reaction Gibbs energy, and thereby the directionality of each reaction, is assigned by combining standard transformed Gibbs energy with metabolomic data. Due to uncertainty in estimated standard Gibbs energy or lack of metabolomic data, typically many reactions seem quantitatively reversible considering the possible upper and lower bounds on transformed reaction Gibbs energy alone. However, the uncertainty associated with group contribution estimates of standard metabolite species Gibbs energy is normally distributed [11] so a calculation of the cumulative probability that a reaction is irreversible can be used to stratify such reactions by a measure of confidence that the reaction operates in a particular direction.

Allowing many reactions, irreversible *in vivo*, to be reversible *in silico,* under-constrains metabolism, therefore, in the pipeline, reactions that are irreversible with sufficient confidence, above a certain cutoff, are set to irreversible. The trade off between under- and over-constraining the feasible set of steady state solutions may be identified by decreasing the cutoff until a reaction direction essential for steady state flux is reversed. Ultimately, quantitative assignment of reaction directionality requires comparison with experimental literature. To aid this process, the pipeline generates a *directionality report*, for each quantitatively assigned reaction direction which conflicts with a reconstruction direction. This also includes details on the relative contributions of uncertainty in estimation or concentration for each metabolite involved. Moreover, to further aid comparison with literature, numerous figures are automatically generated which illustrate the stratification of directionality.

3 Implementation

The pipeline is implemented as approximately 14,000 lines of cross platform MATLAB (Mathworks Inc., Natick, Massachusetts) code supplemented by approximately 4,500 comment lines. With a typical personal computer, a thermodynamically constrained genome scale model of metabolism can be generated in 30 seconds. An installation of the COBRA toolbox is also necessary if it is desired to test for a feasible steady state before and after establishing a thermodynamically constrained model.

4 Conclusion

This pipeline facilitates the quantitative thermodynamic assignment of reaction directionality using experimentally derived standard metabolite species Gibbs energies[1] together with group contribution estimates [11]. The code is appropriate for reaction thermodynamics at biochemically relevant pH, ionic strength and electrical potential ranges. The stoichiometric metabolic reconstruction can be multicompartmental, but should pass certain quality controls [14], especially elemental and charge balance of each reaction.

Funding: This work was supported by the National Institute of Health, grant 5R01GM057089-11, and U.S. Department of Energy, Offices of Advanced Scientific Computing Research and the Biological and Environmental Research as part of the Scientific Discovery Through Advanced Computing program, grant DE-SC0002009.

Conflict of Interest: none.

References

[1] R. A. Alberty. *Thermodynamics of Biochemical Reactions*. Wiley-Interscience, Hoboken, NJ, 2003.

[2] R. A. Alberty. *Biochemical Thermodynamics: Applications of Mathematica*. Wiley-Interscience, Hoboken, NJ, 2006.

[3] S. A. Becker, A. M. Feist, M. L. Mo, G. Hannum, B. Ø. Palsson, and M. J. Herrgard. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nature protocols*, 2:727–738, 2007.

[4] Benjamin J Bornstein, Sarah M Keating, Akiya Jouraku, and Michael Hucka. Libsbml: an api library for sbml. *Bioinformatics*, 24(6):880–881, Mar 2008.

[5] A. Dalby, J.G. Nourse, W.D. Hounshell, A.K.I. Gushurst, D.L. Grier, B.A. Leland, and J. Laufer. Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. *Journal of Chemical Information and Computer Sciences*, 32(3):244–255, 1992.

[6] K. A. Dill and S. Bromberg. *Molecular driving forces: Statistical thermodynamics in Chemistry and Biology*. Garland Science, London, 2003.

[7] A. M. Feist, C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt,

V. Hatzimanikatis, and B. Ø. Palsson. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular systems biology*, 3(1):e121, 2007.

[8] R. M. T. Fleming, I. Thiele, and H. P. Nasheuer. Quantitative assignment of reaction directionality in constraint-based models of metabolism: application to *Escherichia coli*. *Biophysical chemistry*, 145(2-3):47–56, Dec 2009.

[9] C. S. Henry, M. D. Jankowski, L. J. Broadbelt, and V. Hatzimanikatis. Genome-scale thermodynamic analysis of *Escherichia coli* metabolism. *Biophysical journal*, 90(4):1453–1461, 2006.

[10] M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, and T. C. Hodgman. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics (Oxford, England)*, 19:524–31, 2003. 20030303 DCOM-20030926.

[11] M. D. Jankowski, C. S. Henry, L. J. Broadbelt, and V. Hatzimanikatis. Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophysical journal*, 95(3):1487–1499, Aug 2008.

[12] B. Ø. Palsson. *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press, Cambridge, 2006.

[13] Y.B. Tewari and R.N. Goldberg. Thermodynamics of the oxidation–reduction reaction {2 glutathionered (aq)+ NADPox (aq)= glutathioneox (aq)+ NADPred (aq)}. *The Journal of chemical thermodynamics*, 35(8):1361–1381, 2003.

[14] I. Thiele and B. Ø. Palsson. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature protocols*, 5:93–121, 2010.

[15] L. von Bertalanffy. *General System Theory: Foundations, Development, Applications*. G. Braziller, New York, 1973.

[16] B. Wan, C. Doumen, J. Duszynski, G. Salama, and K. F. LaNoue. A method of determining electrical potential gradient across mitochondrial membrane in perfused rat hearts. *Am J Physiol*, 265(2 Pt 2):H445–H452, Aug 1993.