Duality, Thermodynamics, and the Linear Programming Problem in Constraint-Based Models of Metabolism

Patrick B. Warren and Janette L. Jones

Unilever R&D Port Sunlight, Bebington, Wirral, CH63 3JW, United Kingdom

(Received 9 February 2007; published 7 September 2007)

It is shown that the dual to the linear programming problem that arises in constraint-based models of metabolism can be given a thermodynamic interpretation in which the shadow prices are chemical potential analogues, and the objective is to minimize free energy consumption given a free energy drain corresponding to growth. The interpretation is distinct from conventional nonequilibrium thermodynamics, although it does satisfy a minimum entropy production principle. It can be used to motivate extensions of constraint-based modeling, for example, to microbial ecosystems.

DOI: 10.1103/PhysRevLett.99.108101 PACS numbers: 87.16.−b, 05.70.Ln, 82.60.−s

In biology, the metabolism of an organism provides energy and raw materials for maintenance and growth. As such, an interesting and important question concerns the application of thermodynamics to metabolic reaction networks [1–4]. For example, Prigogine and Wiame suggested a long time ago that an organism's metabolism might be governed by a minimum entropy production (MEP) principle [5]. From the physical point of view, a metabolic reaction network is an excellent example of a system in a nonequilibrium steady state, since one can usually assume that the metabolite concentrations are unchanging after a short transient relaxation period. The appropriate generalization of thermodynamics and statistical mechanics to nonequilibrium steady states is a large field [6], which continues to attract attention to this present day [7]. In this Letter, we show that a novel thermodynamic interpretation can be given to the dual linear programming problem which arises in constraint-based models of metabolism. The resulting interpretation is rigorously defined and uniquely determined by the mathematics. It is closely analogous to, but distinctly different from, conventional nonequilibrium thermodynamics. We also show that it satisfies an MEP principle similar to that proposed by Prigogine and Wiame.

Constraint-based modeling (CBM) of metabolic networks has been pioneered by Palsson and co-workers [8]. In a typical application, described in more detail below, the steady-state assumption is combined with a target function to make a linear optimization or linear programming (LP) problem. The LP variables are the fluxes through the various reactions that comprise the network, and the LP constraints arise from basic considerations of stoichiometry and from the reversibility or otherwise of the reactions. The LP objective function is biologically motivated, for example, a "growth" reaction is commonly inserted, and the target is to maximize flux through this reaction to correspond to maximal growth rate. CBM has been applied to microorganisms from all three domains of life [9–11] and has been remarkably successful in predicting phenotypic behavior [12–14].

Mathematically, every LP problem has a unique dual [15]. It was in determining the dual to the CBM LP problem that we noticed a striking analogy to nonequilibrium thermodynamics. Let us start therefore with a general discussion of LP duality, before specializing to the case of CBM. We recall that the basic or primal LP problem, in standard form, is to maximize an objective function $z = \sum_{\alpha=1}^{n} a_\alpha x_\alpha$ given $\sum_{\alpha=1}^{n} A_{i\alpha} x_\alpha = b_i$ $(i = 1 \ldots m,\ m < n)$, where the $x_\alpha \geq 0$ are variables, the a_α are coefficients, $A_{i\alpha}$ is a matrix, and the b_i are constants (we use Greek and Roman indices to emphasize that different components live in different vector spaces). The dual problem is then to minimize an objective function $w = \sum_{i=1}^{m} \pi_i b_i$ subject to $\sum_{i=1}^{m} \pi_i A_{i\alpha} \geq a_\alpha$, with no restriction on the sign of the dual variables π_i. The LP strong duality theorem guarantees that $\max z = \min w$, provided both problems have optimal solutions. In addition, at optimality, "complementary slackness" (CS) conditions hold. To formulate these, first define the "slack" in the inequalities in the dual problem to be $y_\alpha = \sum_{i=1}^{m} \pi_i A_{i\alpha} - a_\alpha$. The CS conditions state that the inequalities $x_\alpha y_\alpha \geq 0$ are saturated (i.e., $= 0$) at optimality, and only at optimality.

In many applications of LP, the dual problem can be given an economic interpretation, which has led to the dual variables being generically known as "shadow prices". We note that shadow prices can be obtained directly from the solution to the primal problem [15], so the dual problem need never be explicitly formulated. This may be the reason why the remarkably simple structure of the dual problem in CBM has not been described before. The use of shadow prices in CBM was pioneered by Varma and Palsson to assess efficiencies in a model of the central metabolism of *E. coli* [16].

Now let us turn to the LP problem in CBM. We start with the set of chemical rate equations that describe the metabolic reaction network, $dc_i/dt = \sum_\alpha S_{i\alpha} v_\alpha$, where the c_i

are metabolite concentrations, the v_α are reaction velocities or fluxes, and $S_{i\alpha}$ is a stoichiometry matrix giving the number of moles of the ith metabolite involved in the αth reaction. Making the steady-state assumption, the chemical rate equations reduce to a set of flux-balance conditions $\sum_\alpha S_{i\alpha} v_\alpha = 0$. At this point, in CBM, attention shifts from the metabolite concentrations to the reaction fluxes. From this point of view, the flux-balance conditions become a set of linear constraints on the v_α. In addition, one usually imposes the "thermodynamic" constraint that $v_\alpha \geq 0$ if a reaction is irreversible.

In modern approaches [8], the reactions in the network are elementally and charge balanced. To make the network "do" something, two kinds of imbalanced reactions are typically added. The first, as mentioned already, is a growth reaction. This reaction drains the end points of metabolism in the appropriate ratios and represents the combined effect of the biochemistry subsequent to metabolism. The flux through the growth reaction (the growth rate) will be labeled v_{gr}. The second type of imbalanced reaction is an "exchange" reaction, which represents the exchange of an extracellular metabolite with the environment (the model additionally includes transporter reactions which allow extracellular metabolites to enter and leave the intracellular environment). The exchange reactions enable the uptake of food substrates, trace minerals, dissolved gases, and vitamins, and the discharge of metabolic waste products. By convention, a positive (negative) flux through an exchange reaction represents the discharge (uptake) of the corresponding metabolite. Exchange reactions may be reversible, or irreversible if discharge only is possible. A special case arises when one wishes to represent *limited* availability, for example, of a food substrate. In this case the exchange flux is allowed to become negative to a limited extent, thus $v_\alpha \geq -v_\alpha^{\min}$ with $v_\alpha^{\min} > 0$ representing a "cap" on the (negative) reaction flux. The value of $v_\alpha^{\min}$ is typically empirically determined to agree with experimentally measured uptake rates. Figure 1 shows schematically how the exchange reactions and the growth reaction

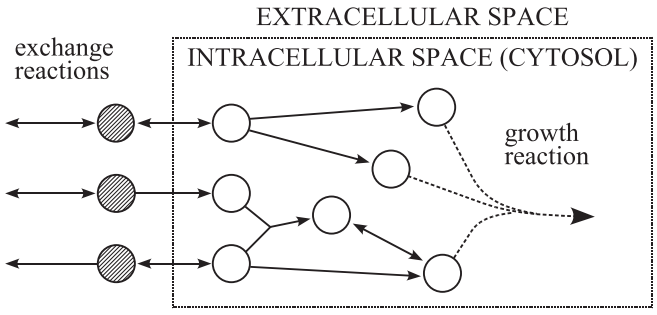

FIG. 1. Schematic metabolic network for a prokaryote like *E. coli*, showing intracellular metabolites (open circles), extracellular metabolites (hatched circles), internal and exchange reactions (arrows), and a growth reaction (dashed arrow).

are connected into the rest of the metabolic network. For high accuracy work, an ATP (adenosine tri-phosphate) → ADP (adenosine di-phosphate) maintenance reaction with a specified flux is sometimes included in the model [11]. We omit this here although it can easily be accommodated with a small extension to the formalism.

The LP problem in CBM is then to find values for the fluxes v_α which maximize v_{gr} subject to the above constraints. This is summarized in Table I. Usually it is the limited availability of substrates through the exchange reactions that prevents the problem being unbounded. Technically the LP problem is not quite in the standard form but it does not take much to make it so.

We formulate the dual to the above (primal) LP problem following the textbook approach described above. After some straightforward simplifications, the following picture emerges. Each metabolite has an associated shadow price π_i which is unrestricted in sign. Each reaction has an associated constraint, of the form $\sum_i \pi_i S_{i\alpha} = 0$ for reversible and unlimited exchange reactions, $\sum_i \pi_i S_{i\alpha} \geq 0$ for irreversible and limited exchange reactions, and $\sum_i \pi_i S_{i\alpha} \geq 1$ for the growth reaction. The objective function is $w = \sum_{i\alpha} \pi_i S_{i\alpha} v_\alpha^{\min}$, where the sum is over the limited exchange reactions only. The LP problem is to find values for the shadow prices π_i which *minimize* this objective function subject to the constraints. Note that, although the $v_\alpha^{\min}$ appear in the dual objective function w, these are numerical constants common to both the primal and dual problems. The actual fluxes v_α do not feature in the dual problem.

We now show that the dual problem admits a striking thermodynamic interpretation. The motivation is the standard expression for the free energy change in a chemical reaction, or reaction affinity, $A_\alpha = \sum_i \mu_i S_{i\alpha}$, where the μ_i are chemical potentials [6]. The similarity between this expression and the rules for formulating the dual LP problem above makes it natural to interpret the shadow prices as chemical potential analogues. To aid the interpretation, we rescale the dual problem by a factor $-B^* < 0$, set $\mu_i = -\pi_i B^*$, and write $B_\alpha = \sum_i \mu_i S_{i\alpha}$ as the analogue of reaction affinity. The resulting thermodynamic formulation of this rescaled dual LP problem is summarized in Table I. We have introduced B_α to distinguish our interpretation from conventional nonequilibrium thermodynamics; in general $B_\alpha \neq A_\alpha$, as explained in more detail below.

Let us discuss the thermodynamic interpretation in a bit more depth. We see that the constraints assert that $B_\alpha = 0$ for a reversible reaction, and $B_\alpha \leq 0$ for an irreversible reaction. These are precisely in accord with equilibrium chemical thermodynamics. In addition, we interpret the fact that $B_\alpha \leq -B^* < 0$ for the growth reaction to mean that a minimum free energy drain equal to B^* is required for growth. The magnitude of B^* sets the overall energy scale and can be arbitrarily chosen. Finally, in the rescaled dual LP problem the objective is to minimize

TABLE I. The primal and dual linear programming problems in constraint-based models of metabolism. In the dual objective function, the sum is over the limited exchange reactions only. The complementary slackness (CS) inequalities are saturated (i.e., $= 0$) at optimality, where also $\max z = \min w$. The shadow prices for the primal problem are given by $-\mu_i/B^*$.

	Primal	Dual	CS Inequalities
Variables	fluxes, v_α	chemical potentials, μ_i	
Flux balance or thermodynamics	$\sum_\alpha S_{i\alpha} v_\alpha = 0$	$B_\alpha = \sum_i \mu_i S_{i\alpha}$	
Reversible reactions	v_α unlimited	$B_\alpha = 0$	
Irreversible reactions	$v_\alpha \geq 0$	$B_\alpha \leq 0$	$v_\alpha B_\alpha \leq 0$
Growth reaction	$v_{\rm gr} \geq 0$	$B_\alpha \leq -B^* < 0$	$v_{\rm gr}(B_\alpha + B^*) \leq 0$
Limited exchange reactions	$v_\alpha \geq -v_\alpha^{\rm min}$	$B_\alpha \leq 0$	$(v_\alpha + v_\alpha^{\rm min})B_\alpha \leq 0$
Objective function	$z = v_{\rm gr}$	$w = \sum_\alpha v_\alpha^{\rm min} \lvert B_\alpha \rvert / B^*$	

$wB^* = \sum_\alpha v_\alpha^{\rm min} |B_\alpha|$, in other words, a weighted sum of the free energy consumption associated with the limited exchange reactions. At optimality, one has $\max z = \min w$, hence the growth rate is easily calculated from the solution to the thermodynamic LP problem using $v_{\rm gr} = \sum_\alpha v_\alpha^{\rm min} |B_\alpha| / B^*$.

We now show that the formalism satisfies an MEP principle. To derive this, we consider the internal entropy production due to the chemical transformations, $T\dot{S} = -\sum_\alpha v_\alpha B_\alpha$. The sum excludes the exchange reactions since the flux-balance condition implies the total entropy production $\sum_\alpha v_\alpha B_\alpha \equiv 0$ when the sum is over *all* reactions. It is straightforward to show that $z \leq T\dot{S}/B^* \leq w$. Thus, at optimality, the entropy production is "pinched" between the two objective functions. Alternatively, for a fixed growth rate $v_{\rm gr}$, one has $v_{\rm gr} B^* \leq T\dot{S}$. Since the minimum value is attained at the combined solution of the primal and dual problems, this gives the desired MEP principle.

What does the dual solution look like for a constraint-based model of metabolism? To give an example, we computed the chemical potential analogues for a genome-scale model of the metabolism of *E. coli* growing aerobically on a glucose "minimal medium", for which uptake of extracellular glucose is the limiting exchange reaction [17]. Lack of space precludes a detailed discussion of the results, but we find that the vast majority of chemical potentials are positive and there is a broad distribution over several decades of magnitude. An interesting observation is that the chemical potentials increase with increasing molecular complexity. This is shown in Fig. 2, using molecular weight (discounting metal ions) as a stand-in for molecular complexity. This correlation arises because the chemical potential of a complex molecule is given approximately by the sum of the chemical potentials of its constituent parts. This in turn follows from the CS conditions which imply $B_\alpha = \sum_i \mu_i S_{i\alpha} = 0$ for all reactions with a flux $v_\alpha \neq 0$ (see further discussion below).

Let us discuss our findings in a wider context. Our results show that the dual to the CBM LP problem has a thermodynamic interpretation in which the dual variables are analogous to chemical potentials. The rules to formulate this thermodynamic LP problem are summarized in Table I. In principle, we could strip away the CBM "scaffolding" and let the thermodynamic LP problem stand on its own, since LP duality guarantees this is equivalent to solving the original (primal) LP problem. Such a viewpoint motivates a number of interesting questions.

First, a technical point arises since the primal LP problem is often *degenerate*, in the sense that alternative optimal flux distributions exist [18]. This reflects the fact that multiple pathways may exist in the metabolism. But this is not a serious problem, for the dual problem will be similarly degenerate but the strong duality theorem and the CS conditions still hold, allowing one to move from a solution of the dual problem to a solution of the primal problem, and *vice versa*.

A more serious discussion point concerns the relationship to conventional nonequilibrium thermodynamics. For the reversible reactions, $B_\alpha = 0$ is a constraint. For the irreversible reactions, the CS conditions (Table I) show that $B_\alpha = 0$ if there is a flux ($v_\alpha > 0$) through a reaction, and

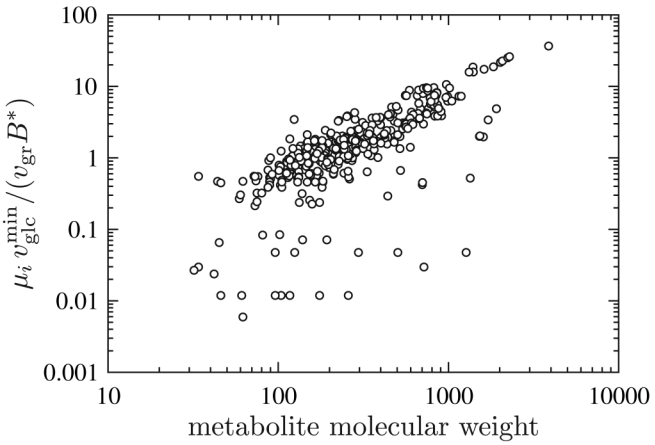

FIG. 2. Chemical potential analogues (shadow prices) plotted as a function of metabolite molecular weight for the intracellular metabolites with $\mu_i > 0$ in a genome-scale constraint-based model of *E. coli* [17]. A normalizing factor is included to make the μ_i dimensionless.

$B_\alpha < 0$ only if there is no flux ($v_\alpha = 0$) through a reaction (except for the growth reaction where we expect $v_{gr} > 0$ and hence $B_\alpha = -B^*$). This presents a sharp contrast to conventional nonequilibrium thermodynamics where a flux through a reaction ($v_\alpha > 0$) is associated with a (negative) affinity driving force $A_\alpha < 0$. This clearly demonstrates that $B_\alpha \neq A_\alpha$, and the thermodynamics described in Table I is not simply the same as conventional nonequilibrium thermodynamics. We must therefore regard the rules described in Table I as describing a *novel* but *tightly constrained* thermodynamics for the CBM class of problems, derived from the (unique) dual to the primal LP problem. Whether the close analogy to *equilibrium* chemical thermodynamics (and the unexpected appearance of an MEP principle) is indicative of deeper principles or not remains a problem for future investigation. It would, for example, be an interesting exercise to compare the *E. coli* shadow prices with what is known about the thermodynamic metabolic state of this organism [2]. We should emphasize that our MEP principle is couched in terms of B_α and not A_α, and therefore our results do not constitute a proof of the original proposition of Prigogine and Wiame [5].

Another interesting remark is that the choice in the primal LP problem to maximize the flux through a growth reaction seems to be ''pure biology''. Experiments demonstrate that this works well under controlled conditions [12,13], and it can be supported by examining population dynamics for continuous culture growth in a chemostat [19]. Other choices could and perhaps should be made in different circumstances [20–22]. In terms of the thermodynamic LP problem, this biologically motivated component is translated into the existence of a growth reaction with a minimal free energy drain B^*. We could turn this observation to our advantage, to suggest extensions to the CBM approach which are perhaps not obvious in the primal LP problem. Consider, for example, metabolism in a microbial *ecosystem*, comprising multiple species which share a pool of common extracellular metabolites. The obvious generalization of the thermodynamic LP problem is to include a growth reaction with a minimal free energy drain B^* for each organism and to seek to minimize the free energy consumption *of the ecosystem* through the exchange reactions of the extracellular metabolites. Of course LP duality means there is a corresponding primal model (in this case the primal objective function becomes a weighted sum of growth rates [23]). Further exploration of this we leave to future work.

We thank J. D. Trawick and S. J. Wiback of Genomatica, Inc., for useful correspondence, M. E. Cates, W. C. K. Poon, and P. R. ten Wolde for a critical reading of the manuscript, and B. Ø. Palsson and his group for helpful discussions and for generously providing the *E. coli* model.

[1] D. A. Beard, S. Liang, and H. Qian, Biophys. J. **83**, 79 (2002); D. A. Beard *et al.*, J. Theor. Biol. **228**, 327 (2004); H. Qian and D. A. Beard, Biophys. Chem. **114**, 213 (2005).
[2] C. S. Henry *et al.*, Biophys. J. **90**, 1453 (2006).
[3] U. von Stockar *et al.*, J. Biotechnol. **121**, 517 (2006).
[4] A. Kümmel, S. Panke, and M. Heinemann, BMC Bioinformatics **7**, 512 (2006).
[5] I. Prigogine and J. M. Wiame, Experientia **2**, 451 (1946); see also P. J. Stoward, Nature (London) **194**, 977 (1962); I. W. Richardson, Biophys. J. **9**, 265 (1969).
[6] S. R. de Groot and P. Mazur, *Non-Equilibrium Thermodynamics* (Dover, New York, 1984).
[7] M. R. Evans *et al.*, Phys. Rev. Lett. **80**, 425 (1998); T. Hatano and S. I. Sasa, Phys. Rev. Lett. **86**, 3463 (2001); R. M. L. Evans, Phys. Rev. Lett. **92**, 150601 (2004).
[8] B. Ø. Palsson, *Systems Biology: Properties of Reconstructed Networks* (Cambridge University Press, Cambridge, England, 2006).
[9] J. L. Reed *et al.*, Genome Biol. **4**, R54 (2003) (*E. coli*).
[10] N. C. Duarte, M. J. Herrgård, and B. Ø. Palsson, Genome Res. doi:10.1101/gr.2250904 (2004) (*S. cerevisiae*).
[11] A. M. Feist *et al.*, Molecular Systems Biol. doi:10.1038/msb4100046 (2006) (archeal methanogen *M. barkeri*).
[12] R. U. Ibarra, J. S. Edwards, and B. Ø. Palsson, Nature (London) **420**, 186 (2002).
[13] C. Zhang *et al.*, Biochem. Eng. J. **25**, 99 (2005).
[14] J. D. Trawick and C. H. Schilling, Biochemical Pharmacology **71**, 1026 (2006).
[15] G. B. Dantzig, *Linear Programming and Extensions* (Princeton University Press, Princeton, NJ, 1963).
[16] A. Varma and B. Ø. Palsson, J. Theor. Biol. **165**, 477 (1993); **165**, 503 (1993).
[17] The model is based on ''iJR904'' described in Ref. [9]. We set $v_{glc}^{min} = 10$ mmol/[(g dry weight) hr] for which we find a growth rate $v_{gr} = 0.957$/hr.
[18] R. Mahadevan and C. H. Schilling, Metabolic Eng. **5**, 264 (2003).
[19] S. J. Pirt, *Principles of Microbe and Cell Cultivation* (Blackwell Scientific, Oxford, 1975).
[20] D. Segre, D. Vitkup, and G. M. Church, Proc. Natl. Acad. Sci. U.S.A. **99**, 15 112 (2002).
[21] T. D. Vo, H. J. Greenberg, and B. Ø. Palsson, J. Biol. Chem. **279**, 39 532 (2004).
[22] S. J. Wiback and B. Ø. Palsson, Biophys. J. **83**, 808 (2002).
[23] A weighted sum of growth rates has been used very recently for modeling a syntrophic microbial coculture, although the motivation was purely empirical; see S. Stolyar *et al.*, Mol. Syst. Biol. doi:10.1038/msb4100131 (2007).