

# MENTOS: A statistical thermodynamic approach for modeling metabolism

Jeremy Zucker<sup>1</sup>, nkkchem<sup>2</sup>, and william.cannon<sup>2</sup>

<sup>1</sup>Pacific Northwest National Laboratory

<sup>2</sup>Affiliation not available

January 24, 2018

## 1 Introduction

One of the key challenges in systems biology is to characterize the energetic state of the cell. However, unlike transcriptomics and proteomics data sets, almost all metabolomics data sets are sparse. Part of the reason for this sparsity is that the chemical properties of small molecules vary widely, so no single instrument can measure every type of metabolite. Sparsity is also due to the fact that biologically relevant metabolite concentrations may span several orders of magnitude, which is greater than what typical instrument sensitivities can handle. Therefore, new computational and theoretical methods are needed that can interpolate unmeasured or unmeasurable metabolite levels within a metabolic network using the least biased assumptions that are consistent with our current knowledge.

Constraint-based methods such as flux balance analysis (FBA) have been successfully used to predict the steady-state metabolic fluxes at genome-scale, but net flux predictions by themselves provide no information about the metabolite concentrations. Conversely, kinetic models can be used to predict steady-state metabolite concentrations, but cannot generally be applied at the genome scale due to the challenges involved in measuring rate constants in a high-throughput manner. Thermodynamic-based methods have been used to constrain the reaction direction of metabolic fluxes or to remove thermodynamically infeasible cycles. More recently, an optimization method has been developed that can predict metabolite concentrations by hypothesizing that a metabolic tug of war (?) exists between osmotic pressure to keep metabolite concentrations low, and the need to maintain thermodynamic non-equilibrium in order utilize enzymes efficiently.

An alternative approach for modeling metabolism is to predict energetic states using non-equilibrium thermodynamics. In this approach, we showed that the likelihood ratio of the forward and backward reaction rates are proportional to the thermodynamic driving force (?). In a recent study of the coupled reaction theorem, we demonstrated that both thermodynamic and kinetic optimal states are obtained at the maximum rate of entropy production (?).

Building on this body of work, we present Maximum ENtropy production rate Of the Stoichiometric matrix (MENTOS): a statistical thermodynamic optimization method for predicting the metabolite concentrations and fluxes that extract energy from the environment as quickly and efficiently as possible, given the constraints of the network. MENTOS was developed to address two major goals. First, the ability to predict reasonable values for steady state metabolite concentrations can be used as a foundation for developing thermokinetic models that can predict non-steady-state dynamics in the absence of rate constants. Second, MENTOS provides a quantitative measure by which we can test the hypothesis that natural systems have

evolved to move to the most probable state through the fastest path by minimizing energetic costs. In the theory section below, we develop the basic concepts of statistical thermodynamic that are necessary to formalize this hypothesis, and demonstrate the application of this hypothesis to a simple toy model. We will then apply the hypothesis to the central carbon metabolism of *E. coli* and compare the metabolite predictions with actual measurements.

## 2 Theory/Methods

In this section, we will introduce the basic concepts of statistical thermodynamics that are required for the formulation of the MENTOS approach.

### 2.1

### 2.2 Derivation of mass action rate constants from the steady state net flux and concentrations

Consider a simple bimolecular reversible reaction in which  $\gamma_A$  moles of  $A$  and  $\gamma_B$  moles of  $B$  are transformed into  $\gamma_C$  moles of  $C$  and  $\gamma_D$  moles of  $D$ . This reaction proceeds at forward reaction rate  $r_+$  and reverse reaction rate  $r_-$  :



According to the law of mass action, both the forward and reverse rates are dependent on the concentration of their corresponding reactant species:

$$r_+ = k_+ [A]^{\gamma_A} [B]^{\gamma_B} \quad (2)$$

$$r_- = k_- [C]^{\gamma_C} [D]^{\gamma_D} \quad (3)$$

$$(4)$$

Where  $k_+$  and  $k_-$  are the rate constants of the forward and the reverse reaction, respectively. To solve for the two unknown rate constants, we need two equations. The first equation comes from the definition of the net rate  $\Delta r$ :

$$\Delta r = r_+ - r_- \quad (5)$$

$$= k_+ [A]^{\gamma_A} [B]^{\gamma_B} - k_- [C]^{\gamma_C} [D]^{\gamma_D} \quad (6)$$

$$(7)$$

To find the second equation, we introduce the concept of the thermodynamic driving force, or reaction affinity  $\mathcal{A}$ , which is defined in terms of the forward and reverse rates:

$$\frac{\mathcal{A}}{RT} = \log \frac{r_+}{r_-} \quad (8)$$

Where  $R$  is the gas constant, and  $T$  is the temperature. Combining equations (2) and (8), we have

$$\frac{\mathcal{A}}{RT} = \log \frac{r_+}{r_-} \quad (9)$$

$$= \log \left( \frac{k_+}{k_-} \frac{[A]^{\gamma_A} [B]^{\gamma_B}}{[C]^{\gamma_C} [D]^{\gamma_D}} \right) \quad (10)$$

$$= \log \left( \frac{k_+}{k_-} \right) + \log \left( \frac{[A]^{\gamma_A} [B]^{\gamma_B}}{[C]^{\gamma_C} [D]^{\gamma_D}} \right) \quad (11)$$

$$= \log \left( \frac{k_+}{k_-} \right) - \log \left( \frac{[C]^{\gamma_C} [D]^{\gamma_D}}{[A]^{\gamma_A} [B]^{\gamma_B}} \right) \quad (12)$$

$$= \log \left( \frac{k_+}{k_-} \right) - \log Q_r \quad (13)$$

$$(14)$$

where the reaction quotient  $Q_r = \frac{[C]^{\gamma_C} [D]^{\gamma_D}}{[A]^{\gamma_A} [B]^{\gamma_B}}$ . The thermodynamic driving force for equation (1) can also be defined in terms of the molar free energy of the species involved, usually referred to as a chemical potential ( $\mu$ ):

$$\frac{\mathcal{A}}{RT} = \frac{1}{RT} (\gamma_A \mu_A + \gamma_B \mu_B - \gamma_C \mu_C - \gamma_D \mu_D) \quad (15)$$

The chemical potential of each metabolite is defined in terms of the metabolite concentrations and the standard chemical potentials ( $\mu^\circ$ )

$$\mu_A = \mu_A^\circ + RT \log[A] \quad (16)$$

$$\mu_B = \mu_B^\circ + RT \log[B] \quad (17)$$

$$\mu_C = \mu_C^\circ + RT \log[C] \quad (18)$$

$$\mu_D = \mu_D^\circ + RT \log[D] \quad (19)$$

$$(20)$$

Substituting equation (16) into equation (15), we have

$$\frac{\mathcal{A}}{RT} = \frac{1}{RT} (\gamma_A \mu_A + \gamma_B \mu_B - \gamma_C \mu_C - \gamma_D \mu_D) \quad (21)$$

$$= \frac{1}{RT} (\gamma_A (\mu_A^\circ + RT \log[A]) + \gamma_B (\mu_B^\circ + RT \log[B]) - \gamma_C (\mu_C^\circ + RT \log[C]) - \gamma_D (\mu_D^\circ + RT \log[D])) \quad (22)$$

$$= \frac{1}{RT} (\gamma_A \mu_A^\circ + \gamma_B \mu_B^\circ - \gamma_C \mu_C^\circ - \gamma_D \mu_D^\circ) + (\gamma_A \log[A] + \gamma_B \log[B] - \gamma_C \log[C] - \gamma_D \log[D]) \quad (23)$$

$$= \log e^{\frac{1}{RT} (\gamma_A \mu_A^\circ + \gamma_B \mu_B^\circ - \gamma_C \mu_C^\circ - \gamma_D \mu_D^\circ)} - \log \left( \frac{[C]^{\gamma_C} [D]^{\gamma_D}}{[A]^{\gamma_A} [B]^{\gamma_B}} \right) \quad (24)$$

$$= \log K^\circ - \log Q_r \quad (25)$$

$$(26)$$

where the standard equilibrium constant  $K^\circ = e^{\frac{1}{RT}(\gamma_C \mu_C^\circ + \gamma_D \mu_D^\circ - \gamma_A \mu_A^\circ - \gamma_B \mu_B^\circ)}$  and the reaction quotient  $Q_r = \frac{[C]^{\gamma_C} [D]^{\gamma_D}}{[A]^{\gamma_A} [B]^{\gamma_B}}$ . Comparing the last line of equation (9) with the last line of equation (21), we see that the ratio of unknown forward and reverse rate constants is the standard equilibrium constant, which can be calculated from the standard chemical potentials using empirical methods such as the group contribution method and the component contribution method (?) (?) (?), or from nonempirical quantum chemistry methods such as density functional theory, post Hartree-Fock and molecular dynamics (?).

$$\frac{k_+}{k_-} = K^\circ \quad (27)$$

$$= e^{\frac{1}{RT}(\gamma_C \mu_C^\circ + \gamma_D \mu_D^\circ - \gamma_A \mu_A^\circ - \gamma_B \mu_B^\circ)} \quad (28)$$

$$(29)$$

Therefore, given the net flux  $v_{net}$ , the steady-state concentrations  $[A]_{SS}, [B]_{SS}, [C]_{SS}, [D]_{SS}$ , the temperature  $T$  and the standard chemical potentials  $\mu_A^\circ, \mu_B^\circ, \mu_C^\circ, \mu_D^\circ$ , we can now solve for the forward and reverse rate constants  $k_+$  and  $k_-$  by combining the last line of equation (5) with the last line of equation (27),

$$v_{net} = k_+[A]_{SS}^{\gamma_A}[B]_{SS}^{\gamma_B} - k_-[C]_{SS}^{\gamma_C}[D]_{SS}^{\gamma_D} \quad (30)$$

$$k_+ = k_- e^{\frac{1}{RT}(\gamma_C \mu_C^\circ + \gamma_D \mu_D^\circ - \gamma_A \mu_A^\circ - \gamma_B \mu_B^\circ)} \quad (31)$$

$$(32)$$

### 2.3 Relationship between net likelihoods and net flux

There is an important relationship between the net rate  $\Delta r = r_+ - r_-$  and the net thermodynamic force. Recall the thermodynamic force in the forward direction is related to the likelihood ratio of the forward and backward rates  $e^{\mathcal{A}/RT} = r_+/r_- = L_+$ , and the thermodynamic force in the backward direction is related to the likelihood ratio of the backwards and forward rates  $e^{-\mathcal{A}/RT} = r_-/r_+ = L_-$

Then the net rate  $\Delta r$  can be defined in terms of the net likelihood ratio  $\Delta L = L_+ - L_-$

$$\Delta L = L_+ - L_- \quad (33)$$

$$= \frac{r_+}{r_-} - \frac{r_-}{r_+} \quad (34)$$

$$= \frac{r_+ r_+}{r_+ r_-} - \frac{r_- r_-}{r_- r_+} \quad (35)$$

$$= \frac{r_+^2 - r_-^2}{r_+ r_-} \quad (36)$$

$$= \frac{(r_+ + r_-)(r_+ - r_-)}{r_+ r_-} \quad (37)$$

$$= \frac{r_+ + r_-}{r_+ r_-} \Delta r \quad (38)$$

$$\frac{r_+ r_-}{r_+ + r_-} \Delta L = \Delta r \quad (39)$$

$$\mu_r \Delta L = \Delta r \quad (40)$$

$$(41)$$

We call the quantity  $\mu_r = r_+ r_- / (r_+ + r_-)$  the reduced rate, which is analogous to the reduced mass when computing relative force in Newtonian or Lagrangian mechanics. Note that

$$\mu_r \leq r_+ \quad (42)$$

$$\mu_r \leq r_- \quad (43)$$

$$\mu_r = \frac{r_+}{2} = \frac{r_-}{2} \text{ if } r_+ = r_- \quad (44)$$

$$\mu_r \approx r_- \text{ if } r_+ \gg r_- \quad (45)$$

$$\mu_r \approx r_+ \text{ if } r_- \gg r_+ \quad (46)$$

$$(47)$$

## 2.4 Solving for the metabolite concentrations that maximize entropy production

Here, we formalize the concept of extracting energy from the environment as quickly and efficiently as possible. Assume a metabolic network that can convert an initial substrate (energy source) to biomass. To extract energy from the environment as efficiently as possible, the total free energy change for the macroscopic reaction should be distributed as evenly as possible across each microscopic reaction in the pathway in order to prevent bottlenecks that require an exponential amount of enzyme to generate biomass.

For a pathway, we define the probability that a reaction  $i$  will fire in the forward direction  $\mathcal{P}_{+i}$  by normalizing its forward likelihood ratio by the sum of all forward and backward likelihood ratios in the pathway. Similarly, we define the probability that reaction  $i$  will fire in the backward direction  $\mathcal{P}_{-i}$  by normalizing its backward likelihood ratio by the sum of all forward and backward likelihood ratios in the metabolic network

$$\mathcal{P}_{+i} = \frac{r_{+i}/r_{-i}}{\sum_j^n (r_{+j}/r_{-j} + r_{-j}/r_{+j})} = \frac{e^{\mathcal{A}_i/RT}}{\sum_j^n (e^{\mathcal{A}_j/RT} + e^{-\mathcal{A}_j/RT})} \quad (48)$$

$$\mathcal{P}_{-i} = \frac{r_{-i}/r_{+i}}{\sum_j^n (r_{+j}/r_{-j} + r_{-j}/r_{+j})} = \frac{e^{-\mathcal{A}_i/RT}}{\sum_j^n (e^{\mathcal{A}_j/RT} + e^{-\mathcal{A}_j/RT})} \quad (49)$$

$$(50)$$

The entropy of the reaction, or entropy production,  $\Delta S$ , is a measure of the evenness of the reaction probability distribution.

$$\Delta S = - \sum_j^n \mathcal{P}_{+j} \log \mathcal{P}_{+j} - \sum_j^n \mathcal{P}_{-j} \log \mathcal{P}_{-j} \quad (51)$$

Microscopic entropy production rate is the entropy of each reaction, weighted by its rate

$$\dot{S}_{micro} = - \sum_j^n \mathcal{P}_{+j} \log \mathcal{P}_{+j} r_{+j} - \sum_j^n \mathcal{P}_{-j} \log \mathcal{P}_{-j} r_{-j} \quad (52)$$

Macroscopic entropy production rate is the entropy of each reaction, weighted by the overall growth rate:

$$\dot{S}_{macro} = - \sum_j^n \mathcal{P}_{+j} \log \mathcal{P}_{+j} v_{growth} - \sum_j^n \mathcal{P}_{-j} \log \mathcal{P}_{-j} v_{growth} \quad (53)$$

## Application of MENTOS to the ABC model

Now that we have defined entropy production, we can solve for the metabolite concentrations that maximize entropy production.

maximize	$-\sum_j (\mathcal{P}_{+j} \log(\mathcal{P}_{+j}) + \mathcal{P}_{-j} \log(\mathcal{P}_{-j}))$	Maximize Entropy Production
subject to	$S \cdot (\vec{\mathbf{L}}_+ - \vec{\mathbf{L}}_-) = 0$	Steady-state Net Likelihood Constraint
	$\log \vec{\mathbf{L}}_+ + \frac{1}{RT} S^T \cdot \vec{\mu}^0 + S^T \cdot \log \vec{\mathbf{c}} = 0$	Thermodynamic Constraint
	$\vec{\mathbf{L}}_+ \geq 0$	Positive forward likelihood Constraint
	$\vec{\mathbf{L}}_- \geq 0$	Positive backward likelihood Constraint
	$\log [\mathbf{A}_{ext}] = \log [\mathbf{E}_{ext}] = \log c_{upper}$	Input boundary Constraint
	$\log [\mathbf{E}_{ext}] = \log [\mathbf{D}_{ext}] = \log c_{lower}$	Output boundary Constraint

Where

- $\vec{\mathbf{L}}_+, \vec{\mathbf{L}}_-$  are decision variables representing the likelihood ratios of the forward and backward rates
- $L_{+i} = r_{+i}/r_{-i}$  for each reaction  $i$
- $L_{-i} = r_{-i}/r_{+i}$  for each reaction  $i$ .
- $\mathcal{P}_{+i}, \mathcal{P}_{-i}$  are the normalized likelihood ratios of the forward and backward rates
- $\mathcal{P}_{+i} = \frac{r_{+i}}{r_{-i}} \left( \sum_j \frac{r_{+j}}{r_{-j}} + \frac{r_{-j}}{r_{+j}} \right)^{-1}$
- $\mathcal{P}_{-i} = \frac{r_{-i}}{r_{+i}} \left( \sum_j \frac{r_{+j}}{r_{-j}} + \frac{r_{-j}}{r_{+j}} \right)^{-1}$
- $\log \vec{\mathbf{c}}$  is a decision variable representing the log concentrations of each metabolite
- $S$  is the  $m \times n$  stoichiometric matrix of representing  $m$  metabolites and  $n$  reactions of the model
- $\vec{\mu}^0$  is the vector of standard chemical potentials

## 2.5 Material view

## 2.6 Energy view

## 2.7 Entropy production view

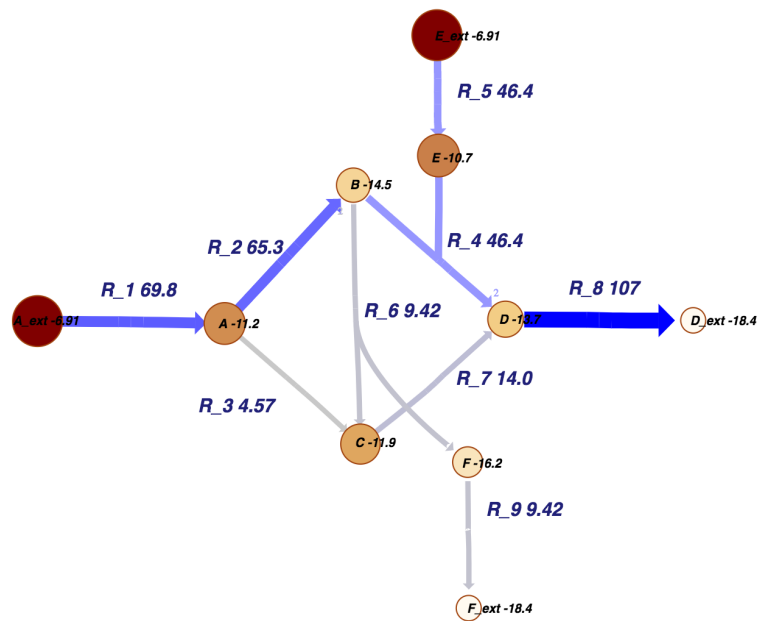


Figure 1: MENTOS prediction of the net difference between the likelihood ratio of forward to backward rates and the likelihood ratio of backward to forward rates (edges) and log concentrations (nodes). Net likelihoods satisfy the steady-state constraint, and are proportional to the fluxes that are predicted to be associated with these metabolite concentrations

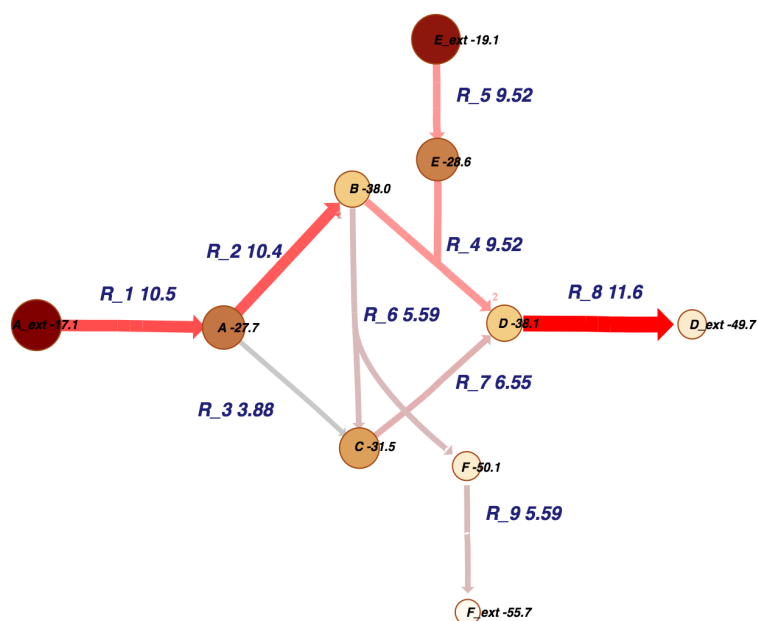


Figure 2: MENTOS prediction of the thermodynamic force (edges) and chemical potential (nodes). The thermodynamic force is the difference between reactant and product chemical potentials, weighted by their respective stoichiometries, and is equal to the negative change in gibbs free energy of the reaction.



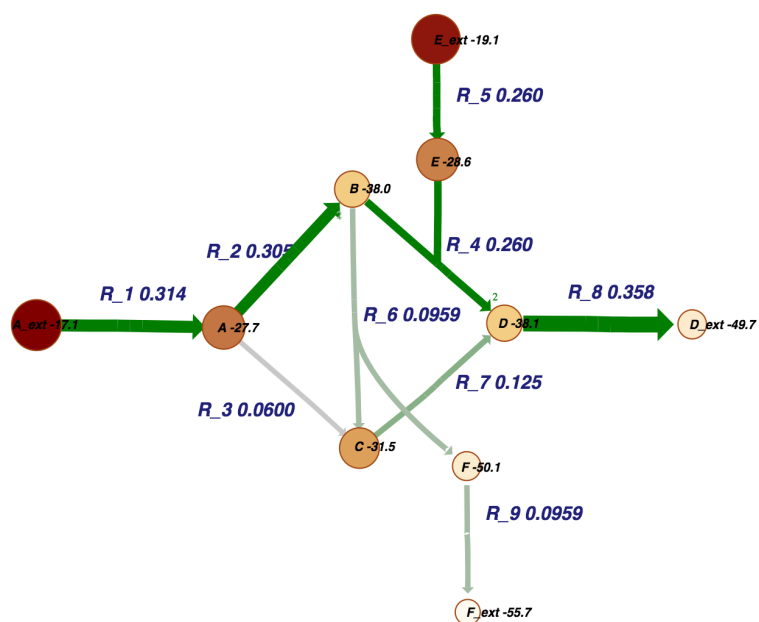


Figure 3: MENTOS prediction of the reaction entropy (edges) and chemical potential (nodes). The reaction entropy is the entropy of the normalized likelihood ratios.