Elucidation and Structural Analysis of Conserved Pools for Genome-Scale Metabolic Reconstructions

Evgeni V. Nikolaev, Anthony P. Burgard, and Costas D. Maranas
Department of Chemical Engineering, The Pennsylvania State University, University Park, Pennsylvania 16802

ABSTRACT In this article, we introduce metabolite concentration coupling analysis (MCCA) to study conservation relationships for metabolite concentrations in genome-scale metabolic networks. The analysis allows the global identification of subsets of metabolites whose concentrations are always coupled within common conserved pools. Also, the minimal conserved pool identification (MCPI) procedure is developed for elucidating conserved pools for targeted metabolites without computing the entire basis conservation relationships. The approaches are demonstrated on genome-scale metabolic reconstructions of *Helicobacter pylori*, *Escherichia coli*, and *Saccharomyces cerevisiae*. Despite significant differences in the size and complexity of the examined organism's models, we find that the concentrations of nearly all metabolites are coupled within a relatively small number of subsets. These correspond to the overall exchange of carbon molecules into and out of the networks, interconversion of energy and redox cofactors, and the transfer of nitrogen, sulfur, phosphate, coenzyme A, and acyl carrier protein moieties among metabolites. The presence of large conserved pools can be viewed as global biophysical barriers protecting cellular systems from stresses, maintaining coordinated interconversions between key metabolites, and providing an additional mode of global metabolic regulation. The developed approaches thus provide novel and versatile tools for elucidating coupling relationships between metabolite concentrations with implications in biotechnological and medical applications.

INTRODUCTION

An organism's metabolic network stoichiometry plays an important role in the organization and regulation of cellular metabolism by establishing structural barriers and limits on reaction fluxes and metabolite concentrations (Cornish-Bowden and Hofmeyr, 2002; Famili and Palsson, 2003; Almaas et al., 2004; Burgard et al., 2004). Specifically, conservation relationships for metabolite concentrations are important biophysical barriers, selected through evolution, to protect cellular organisms from stresses (e.g., osmotic) and provide global metabolic regulation (Bakker et al., 1999, 2000). The conservation relationships are linear combinations of metabolite concentrations that do not change over time and are solely determined by the organism's stoichiometry and uptake/secretion transport conditions. A simple example of conservation relationships is the ATP, ADP, and AMP cofactors pool in energy metabolism that preserves the total adenylate moiety. An increase in the concentration of any cofactor will be immediately compensated by a drain in the other cofactors implying that none of these concentrations can be changed arbitrarily. Thus, constraining energy equivalents within a common conserved pool will automatically coordinate energy-dependent and energy-replenishing processes by shifting the adenylate charge between charged and uncharged forms (Atkinson, 1968; Reich and Selkov, 1981; Gottschalk, 1986).

Conservation relationships and conserved pools have proven very important in understanding fundamental design principles of living systems. In a recent study of *Trypanosoma brucei*, the unicellular eukaryotic parasite that causes the African sleeping sickness, the role that moiety conserved pools, formed within glycosomes, play in the protection of trypanosomes from toxic accumulation of intracellular metabolites and regulation of the organism's glycolysis have been revealed and highlighted (Bakker et al., 1999, 2000). Consequently, conservation relationships may cause biotechnological or medical interventions to fail if not well understood. This is because, the aim of metabolic engineering or medical treatments is often to increase or decrease concentrations of certain biochemicals inside the cell (Wiemer et al., 1995; Eisenthal and Cornish-Bowden, 1998). If the concentration changes are limited by conservation relationships, the concentration may reach a steady state long before triggering the desired metabolic effect. Therefore, detailed information on conserved pools can help identify unattainable concentration changes. In this respect, a rational design approach to engineer organisms to desired metabolite concentrations and fluxes has been recently suggested and its limitations due to moiety-conserved cycles were discussed (Kholodenko et al., 1994, 1998, 2000).

In the last decade, several theoretic approaches and elegant concepts, based on the utilization of convex analysis (Rockafellar, 1970), have been developed to systematically identify and analyze biologically meaningful conservation relationships (Schuster and Höfer, 1991), maximal conserved moieties (Schuster and Hilgetag, 1995), and extreme pools (Famili and Palsson, 2003). Although they provide a solid theoretic foundation for the systemic analysis of metabolically meaningful pools, convex analysis approaches do not scale well to genome-scale metabolic networks due to

Submitted March 25, 2004, and accepted for publication September 21, 2004.

Address reprint requests to Costas D. Maranas, Tel.: 814-863-9958; Fax: 814-865-7846; E-mail: costas@psu.edu.

0006-3495/05/01/37/13 $2.00

doi: 10.1529/biophysj.104.043489

the combinatorial explosion (Chvátal, 2000). Traditional methods of linear algebra used to find conservation relationships (Reder, 1988; Cornish-Bowden and Hofmeyr, 2002) are based on identifying linear dependences between mass balances or, equivalently, rows in the corresponding stoichiometric matrices. Although they are very efficient for moderately sized networks, the methods encounter significant computational difficulties when applied to genome-scale stoichiometric matrices due to computational round-off errors (Golub and van Loan, 1996). Furthermore, existing convex analysis and linear algebra approaches do not allow for the computation of conservation relationships for targeted metabolites before the entire basis set of conservation relationships is found. However, in many practical situations, only certain key metabolic pools are of special interest or importance. For example, these may be pools of adenylate moieties or targeted amino acids. Given these computational and practical limitations, new approaches for the elucidation and structural analysis of conserved pools for genome-scale metabolic reconstructions of complex microorganisms are needed. To this end, we introduce an optimization-based framework to elucidate and analyze conservation relationships for metabolite concentrations in the context of genome-scale reaction networks. The framework is comprised of metabolite concentration coupling analysis (MCCA) and the minimal conserved pool identification (MCPI) procedure. MCCA draws from and complements flux coupling analysis (FCA) developed earlier by Burgard et al. (2004). Both FCA and MCCA circumvent computationally prohibitive convex and linear algebra analyses of a genome-scale stoichiometric models by requiring instead the solution of a series of linear programs.

In analogy to enzymes subsets (Pfeiffer et al., 1999; Schilling et al., 2002), which can be identified using FCA (Burgard et al., 2004), MCCA allows for the efficient identification of subsets of metabolites that have to be simultaneously present in all common conserved pools (see Fig. 1). The analysis of metabolite subsets is important for both a greater understanding of fundamental design principles of cellular metabolism, and applications in biotechnology and medicine. Specifically, in the context of biotechnological applications, lowering the concentrations of secondary metabolites coupled with targeted amino acids within common subsets can increase concentrations of the desired amino acids. This can be achieved, for example, by knocking out appropriate genes with the goal of optimally redirecting major flux routes (Burgard et al., 2003; Pharkya et al., 2003). In the context of medical applications, finding metabolite subsets including a few external metabolites can help suggest new antiparasitic or antibiotic drug targets. Suppressing the transport of such metabolites could lead to osmotic destabilization and ultimate death of the pathogenic organism (Cornish-Bowden and Hofmeyr, 2002).

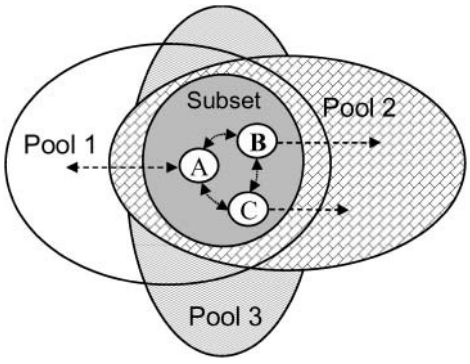


FIGURE 1 A metabolite subset that comprises metabolites A, B, and C simultaneously present in conserved pools 1, 2, and 3.

In the next section, we provide a detailed discussion of conservation relationships for metabolite concentrations in the context of genome-scale metabolic reconstructions. The MCCA and MCPI procedures are introduced in the Methods section. In the Results section, metabolite subsets and conservation relationships are identified in genome-scale models of the three organisms of increasing size and complexity: 1), *Helicobacter pylori*, 2), *Escherichia coli*, and 3), *Saccharomyces cerevisiae*.

CONSERVATION RELATIONSHIPS

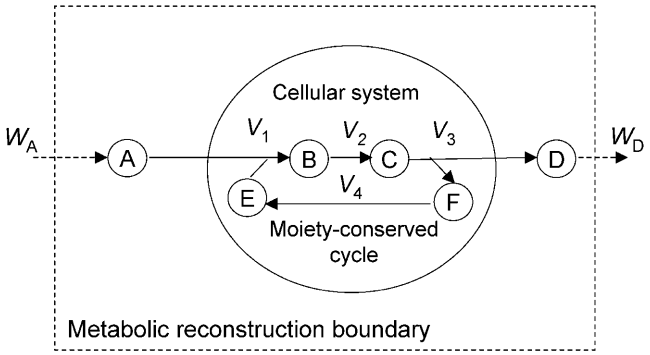
Complete metabolic reconstructions of cellular systems encompass both internal and external metabolites (see Fig. 2) and thus the corresponding stoichiometric matrices describe *closed* metabolic systems (Famili and Palsson, 2003). However, cellular systems are often open and some of the conservation relationships, inferred from ''closed'' metabolic reconstructions, can disappear when external metabolites are exposed to exchange fluxes (e.g., W_A and W_D depicted in Fig. 2). Therefore, a careful interpretation of metabolite pools obtained from closed metabolic reconstructions is needed. In this section, we first consider closed metabolic systems that provide the maximal number of conserved pools and then extend the analysis to open metabolic systems. The definitions and concepts introduced will be highlighted with an example addressing conserved pools in glycolysis.

Closed systems

The dynamic mass balances for a closed metabolic system can be described with a set of ordinary differential equations

$$\frac{dC_i}{dt} = \sum_{j=1}^{M} S_{ij} \cdot v_j \quad \forall i \in \mathcal{N}. \tag{1}$$

Here C_i is the concentration of metabolite i, S_{ij} is the stoichiometric coefficient of metabolite i in reaction j, and v_j is the rate of reaction j. Sets $\mathcal{N} = \{1, \ldots, N\}$ and $\mathcal{M} = \{1, \ldots, M\}$ correspond to metabolites and reactions,



Reactions:	Mass balance equations:
V_1: A+E → B	$dA/dt = W_A - V_1$
V_2: B → C	$dB/dt = V_1 - V_2$
V_3: C → F + D	$dC/dt = V_2 - V_3$
V_4: F → E	$dD/dt = V_3 - W_D$
	$dE/dt = V_4 - V_1$
	$dF/dt = V_3 - V_4$

FIGURE 2 Example representation of a cellular system in the context of complete genome-scale metabolic reconstructions. A complete metabolic reconstruction encompasses both internal (i.e., B, C, E, and F) and external (i.e., A and D) metabolites and thus the corresponding stoichiometric matrix describes a closed system. Here V_1 and V_3 are transport reaction fluxes across the cellular system's boundary, V_2 and V_4 are intracellular reaction fluxes, and W_A and W_D are external exchange fluxes.

respectively. All closed systems satisfy the total mass conservation over all metabolites

$$\sum_{i=1}^{N} \mu_i \cdot C_i(t) \equiv \text{const}, \tag{2}$$

where coefficient μ_i is the molar mass or molecular weight of metabolite i (Heinrich and Schuster, 1996). The conservation relationship Eq. 2 corresponds to the most encompassing metabolite pool. Additional conservation relationships for metabolite concentrations are represented in the general form

$$\sum_{i=1}^{N} \beta_i \cdot C_i(t) \equiv \text{const}. \tag{3}$$

A linear combination of concentrations $\sum_{i=1}^{N} \beta_i \cdot C_i(t)$ does not change over time t if and only if coefficients β_i satisfy the following system of linear equations (see Appendix A for proof):

$$\sum_{i=1}^{N} \beta_i \cdot S_{ij} = 0 \quad \forall j \in \mathcal{M}. \tag{4}$$

This means that all coefficients β_i of any arbitrary conservation relationship Eq. 3 must fulfill condition Eq. 4. Conversely, any linear combination $\sum_{i=1}^{N} \beta_i \cdot C_i(t)$ with coefficients β_i satisfying Eq. 4 is time invariant. Thus, irrespectively of specific (nonlinear) dependencies of reaction rates v_j on kinetic parameters and concentrations, conservation relationships for metabolite concentrations in closed systems are solely determined by the system's stoichiometry.

In closed systems, a complete set of linearly independent conservation relationships Eq. 3 can be constructed with nonnegative coefficients β_i (Heinrich and Schuster, 1996). Following Heinrich and Schuster (1996), we refer to any nonnegative vector β that satisfies Eq. 4 as a conservation vector and its elements β_i as conservation coefficients. A set of metabolites that enter a conservation relationship with positive β_i is referred to as a conserved metabolic pool. A simple example of this is the dissociation reaction, $H_2O = 2 \cdot H^+ + O^{2-}$, which yields conservation relationships $2 \cdot [H_2O] + [H^+] = \text{const}_1$ for the conserved proton moiety and $[H_2O] + [O^{2-}] = \text{const}_2$ for the conserved oxygen moiety. The corresponding conservation vectors are $\beta_1 = (2, 1, 0)$ and $\beta_2 = (1, 0, 1)$, respectively, where $\beta = (\beta_{H_2O}, \beta_H, \beta_O)$. Note that conservation coefficients can be alternatively interpreted as the number of moiety units present in reacting species (e.g., for β_1 coefficient 2 corresponds to two proton moiety units present in the H_2O molecule). A detailed classification and interpretation of conservation relationships and the corresponding conserved pools can be found in Heinrich and Schuster (1996) and Famili and Palsson (2003).

Open systems

Certain conserved pools persist for open cellular systems whereas others may disappear if certain conditions are not fulfilled. A simple example of this is a moiety-conserved cycle encompassing metabolites B, C, E, and F (see Fig. 2). Concentrations of these metabolites always satisfy the conservation relationship $P_1 = [B] + [C] + [E] + [F] = \text{const}_1$. In contrast, by summing the mass balances for metabolites A, B, C, and D (see Fig. 2), one obtains $dP_2/dt = W_A - W_D$, where $P_2 = [A] + [B] + [C] + [D]$. Thus unlike P_1, the global substrate-product pool $P_2 = \text{const}_2$ only if the supply of the cellular system with substrate A and the removal of product D are mass balanced (i.e., $W_A - W_D = 0$). Importantly, for the interpretation of metabolite pools, the time hierarchy of metabolism and transport processes should also be taken into account. Because of the wide separation of time constants for different cellular processes, conserved pools can exist within certain timescales and disappear within the other timescales (Heinrich and Schuster, 1996). In many applications and modeling studies, external concentrations are clamped or vary extremely slowly in comparison with intracellular processes (Heinrich and Schuster, 1996). Therefore the corresponding conservation relationships encompassing such external concentrations can be automatically fulfilled. Thus, closed metabolic reconstructions provide the most encompassing lists of conserved pools maintained within open cellular systems under certain transport influx/outflux conditions and timescales.

Glycolysis example

Glycolysis provides a surprisingly diversified set of metabolically meaningful extreme pools depicted in Fig. 3. Recall that extreme pools are defined as the convex basis vectors that lie at the edge of the high-dimensional cone generated by nonnegative solutions to Eq. 4 and any conserved pool can be obtained as a nonnegative linear combination of extreme pools (Famili and Palsson, 2003). Given the small size of the pathway, the eight extreme pools P1–P8 are computed using the Fourier-Motzkin-Chernikova algorithm (Chernikov, 1968) of convex analysis (Rockafellar, 1970). Similar techniques and approaches have been developed and used to analyze metabolically meaningful pools in moderately sized reaction networks (Heinrich and Schuster, 1996; Famili and Palsson, 2003). Following the classification scheme introduced by Famili and Palsson (2003), extreme pools P1–P8 can be interpreted as follows. Pool P1 represents the total NAD moiety, whereas P2 represents the total adenylate moiety. Pool P3 represents the total carbon moiety. Pool P4 represents the total phosphate moiety in the form of both free inorganic phosphate and an organic phosphate moiety. Pool P5 represents an oxygen moiety, formed within reactions 7–10, where oxygen is incorporated by glyceraldehyde-3-phosphate dehydrogenase through the phosphate group (i.e., $O - [PO_3]$) and is released by enolase to form H_2O. Pool P6 represents the oxidized state of metabolites, whereas P7 counts redox equivalents. Pool P8 represents the potential of each carbon-containing compound to release the high-energy ATP molecule.

METHODS

Although the identification and analysis of conserved pools in small metabolic systems such as glycolysis is straightforward, the extension of existing approaches to genome-scale metabolic systems becomes intractable due to the combinatorial explosion of convex bases (Chvátal, 2000), and round-off errors leading to serious difficulties in the computation of linearly independent solutions to Eq. 4 that span the left null space of stoichiometric matrix **S**, generated by all solutions to Eq. 4 (Golub and van Loan, 1996). Here we introduce MCCA and the MCPI procedure to elucidate and analyze conserved pools for genome-scale cellular systems.

Metabolite concentration coupling analysis

Following the approach utilized by Burgard et al. (2004) for the calculation of flux ratios, a similar sequence of linear programs (Eq. 5) is used to compute the minimal R_{min} and maximal R_{max} values of conservation ratios $\beta_{i''}/\beta_{i'}$ (i.e., $R_{min} = \min \beta_{i''}/\beta_{i'}$ and $R_{max} = \max \beta_{i''}/\beta_{i'}$) to investigate coupling characteristics of every pair of metabolites.

Here variables $\tilde{\beta}_i$ are the original conservation coefficients β_i normalized by $\beta_{i'}$ (i.e., $\tilde{\beta}_i = \beta_i/\beta_{i'}$). Once the minimal and maximal values of the ratio $\beta_{i''}/\beta_{i'}$ are computed, the minimal and maximal values of the reverse ratio $\beta_{i'}/\beta_{i''}$ are automatically known. Thus, at most only $N \cdot (N-1)/2$ ratios need to be computed to find the minimal and maximal values of conservation ratios for every pair of metabolites. Note that despite the fact that alternative optimal solutions to Eq. 5 are often present, formulation Eq. 5 is guaranteed to obtain unique optimal values for the conservation ratios.

The analysis of the minimal R_{min} and maximal R_{max} values of ratio $\beta_{i''}/\beta_{i'}$ for every pair of metabolites i' and i'' results in the following classification of metabolite couplings within conserved pools (see Fig. 4):

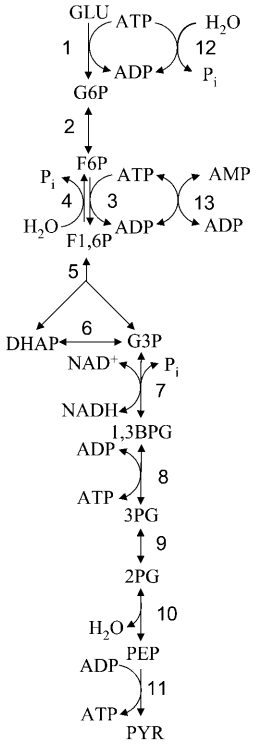
1. Full coupling ($\beta_{i'} \Leftrightarrow \beta_{i''}$), if a nonzero value of $\beta_{i'}$ implies a fixed value of $\beta_{i''}$ and vice versa.
2. Partial coupling ($\beta_{i'} \leftrightarrow \beta_{i''}$), if a nonzero $\beta_{i'}$ implies a nonzero, yet variable, value for $\beta_{i''}$ and vice versa.
3. Directional coupling ($\beta_{i'} \rightarrow \beta_{i''}$), if a nonzero $\beta_{i'}$ implies a nonzero $\beta_{i''}$ but not reverse.
4. Uncoupled metabolites, if any arbitrary value of $\beta_{i'}$ does not impact arbitrary value of $\beta_{i''}$.

The first case ($\beta_{i'} \Leftrightarrow \beta_{i''}$) occurs when $R_{min} = R_{max} = c$ with a nonzero finite constant c. In this case, ratio $\beta_{i''}/\beta_{i'}$ has the same fixed value c (i.e., $\beta_{i''}/\beta_{i'} \equiv c$) for all conserved pools where metabolites i' and i'' are present. The second case ($\beta_{i'} \leftrightarrow \beta_{i''}$) occurs when $R_{min} = c_1$ and $R_{max} = c_2$, where c_1 and c_2 are different nonzero finite constants (i.e., $0 < c_1 < c_2 < \infty$). In this case, ratio $\beta_{i''}/\beta_{i'}$ varies within the range $[c_1, c_2]$, depending on the pool where metabolites i' and i'' are present. The third case ($\beta_{i'} \rightarrow \beta_{i''}$) occurs when R_{min} is zero and R_{max} is equal to some finite constant c_2 or when R_{min} is equal to a finite constant c_1 and R_{max} is unbounded. Metabolites are uncoupled when the ratio $\beta_{i''}/\beta_{i'}$ can take on any arbitrary nonnegative value. Fully and partially coupled metabolites are always present together within common conserved pools and we refer to such sets of coupled metabolites as metabolite subsets (see Fig. 5). Directional coupling captures the one-way type of connectivity between metabolites, where the presence of a given metabolite in a pool implies the simultaneous presence of another metabolite in the same pool and not necessary reverse (see Fig. 5). Note that there may be encountered metabolite i absent from all conserved pools (i.e., metabolite D in Fig. 5). Such metabolites correspond to identically zero conservation coefficients β_i (i.e., $\beta_i \equiv 0$) and this case will be considered separately below in detail.

Application of MCCA to glycolysis (see Fig. 3) yields three fully coupled (see Table 1) and nine directionally coupled metabolite subsets (see Table 2). Within each of the tables, the first metabolite corresponds to a reference metabolite i' for which constraint $\beta_{i'} = 1$ is set in Eq. 5. For example, for the second fully coupled metabolite subset in Table 1, DHAP is a reference metabolite. The corresponding computed conservation ratios are $\beta_{F1,6P}/\beta_{DHAP} = 2$ and $\beta_{G3P}/\beta_{DHAP} = 1$. For the analysis of directional coupling in Table 2, only the reference metabolites from Table 1 are listed because partially and fully coupled metabolites have identical directional coupling relationships.

The first fully coupled subset (2PG, 3PG) (see Table 1) is formed by phosphoglycerate mutase (see Fig. 3). The second subset (F1,6P, DHAP, G3P) is formed by aldolase and triose phosphate isomerase. Lastly, the third

$$\left\{\begin{array}{lll} \text{minimize} & R_{min} = \tilde{\beta}_{i''} & (\text{or maximize } R_{max} = \tilde{\beta}_{i''}) \\ \text{subject to} & \tilde{\beta}_{i'} = 1 & \\ & \sum_{i=1}^{N} \tilde{\beta}_i \cdot S_{ij} = 0 & \forall j \in \mathcal{M} \\ & \tilde{\beta}_i \geq 0 & \forall i \in \mathcal{N} \end{array}\right\} \forall (i'', i') \in \mathcal{N} \quad \text{and} \quad i'' > i'. \tag{5}$$



(P1) Total NAD moiety: $[NAD^+] + [NADH]$

(P2) Total adenylate moiety: $[ATP] + [ADP] + [AMP]$

(P3) Total carbon moiety: $2[GLU] + 2[G6P] + 2[F6P] + 2[F1,6P] + [DHAP] + [G3P] + [1,3BPG] + [3PG] + [2PG] + [PEP] + [PYR]$

(P4) Total phosphate moiety: $[G6P] + [F6P] + 2[F1,6P] + [DHAP] + [G3P] + 2[1,3BPG] + [3PG] + [2PG] + [PEP] + 2[ATP] + [ADP] + [P_i]$

(P5) Total oxygen moiety: $[1,3BPG] + [3PG] + [2PG] + [P_i] + [H_2O]$

(P6) Oxidized state of metabolites: $[1,3BPG] + [3PG] + [2PG] + [PEP] + [PYR] + [NAD^+]$

(P7) Reduced state of metabolites: $2[GLU] + 2[G6P] + 2[F6P] + 2[F1,6P] + [DHAP] + [G3P] + [NADH]$

(P8) High energy potential release: $2[GLU] + [G6P] + [F6P] + [3PG] + [2PG] + [PYR] + [ADP] + 2[AMP] + [H_2O]$

FIGURE 3 Extreme pools identified for glycolysis. Reactions are: 1), hexokinase, 2), phosphoglucoseisomerase, 3), phosphofructokinase, 4), fructose 1,6-bisphosphatase, 5), aldolase, 6), triose phosphate isomerase, 7), glyceraldehyde 3-phosphate dehydrogenase, 8), phosphoglycerate kinase, 9), phosphoglycerate mutase, 10), enolase, 11), pyruvate kinase, 12), ATPase, and 13), adenylate kinase. Metabolites are: glucose (Glu), glucose 6-phosphate (G6P), fructose 6-phosphate (F6P), fructose 1,6-bisphosphate (F1,6P), dihydroxyacetone phosphate (DHAP), glyceraldehyde 3-phosphate (G3P), 1,3-bisphosphoglycerate (1,3BPG), 3-phosphoglycerate (3PG), 2-phosphoglycerate (2PG), phosphoenolpyruvate (PEP), pyruvate (Pyr), and inorganic phosphate (P_i).

subset (G6P, F6P) is formed by phosphoglucoseisomerase. The directional metabolite couplings are shown in the metabolite coupling graph comprised of three disjoint subgraphs (see Fig. 6). The first subgraph corresponds to biochemical transformations of hexoses into trioses in the upper part of glycolysis. The second subgraph corresponds to further transformations of trioses at the lower part of glycolysis. The third subgraph corresponds to the couplings between adenylate moieties. This subgraph schema may be better understood when compared with the schematic representation of couplings between metabolites and the corresponding common conserved pools shown in Figs. 3 and 5. Consider, for example, the fully coupled subset (2PG, 3PG) (subset 1 from Table 1) implied by the molecule H_2O (subset 6 from Table 2). Using the notation of Fig. 5, one then can define $A = 2PG$, $B = 3PG$, and $C = H_2O$. All the three metabolites belong to the same common pools 1 and 2 (see Fig. 5), which correspond to pools P5 and P8, respectively (see Fig. 3). Pool 3, which does not include metabolite C (see Fig. 5), corresponds to pools P3, P4, or P6 (see Fig. 3), which do not include H_2O.

Interestingly, the species NAD^+ and NADH are completely uncoupled within glycolysis. This is because each of the extreme pools P6 and P7 (see

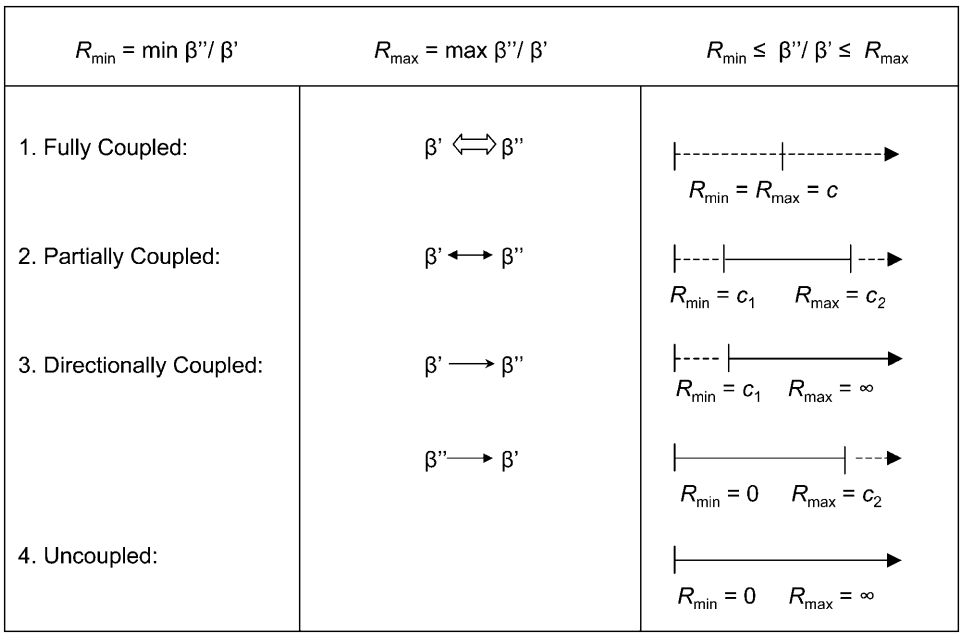


FIGURE 4 The various potential types of coupling between two metabolites are related to their conservation coefficient ratio limits R_{min} and R_{max} as shown. Two metabolites are: 1), fully coupled if the presence of one metabolite within conserved pools implies the presence of the other within the same conserved pools and vice versa, 2), partially coupled if they are always present within common conserved pools but their conservation coefficients are not fixed, and 3), directionally coupled if the presence of one metabolite within conserved pools implies the other and not vice versa.

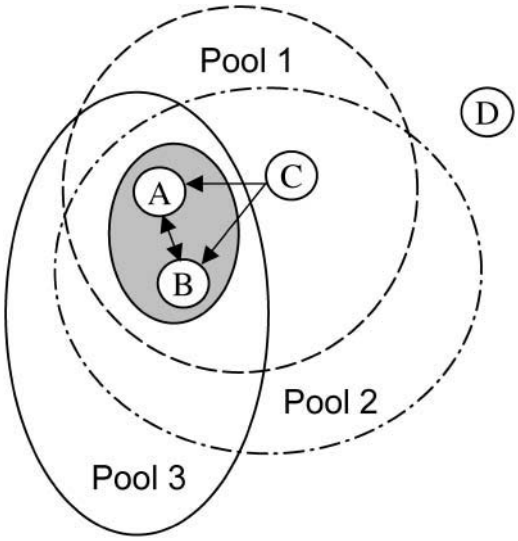


FIGURE 5 A schematic representation of couplings between metabolites. Metabolites A and B are fully or partially coupled, and are simultaneously present in all common conserved pools (i.e., pools 1, 2, and 3). These metabolites form a metabolite subset. Metabolite C is directionally coupled with A and B in the sense that the presence of C in conserved pools forces A and B to be present in the same conserved pools and not vice versa (i.e., C is absent from pool 3, which encompasses A and B). Metabolite D is absent from all conserved pools and its concentration is not constrained by these pools.

Fig. 3) encompasses only one of them (i.e., NAD^+ belongs to pool P6 whereas NADH to pool P7) and their nonnegative linear combinations with pool P1 lead to arbitrary values of ratio β_{NADH}/β_{NAD}. Consider, for example, conserved pools P_a and P_b, where $P_a = a{\cdot}P1 + P6$ and $P_b = b{\cdot}P1 + P7$. For P_a one can infer the ratio $\beta_{NADH}/\beta_{NAD} = a/(a + 1)$ and for P_b, the ratio $\beta_{NADH}/\beta_{NAD} = (b + 1)/b$. Given arbitrary nonnegative a and b, one can conclude that ratio β_{NADH}/β_{NAD} can take on any nonnegative value from within the range $0 \le \beta_{NADH}/\beta_{NAD} < \infty$.

Minimal conserved pools identification

Although it provides complete information regarding metabolites simultaneously present in all common conserved pools, MCCA does not allow for the elucidation of particular conserved pools. However, in some applications, only certain targeted conserved pools are of a special interest. For example, these can be pools of targeted amino acids. In all such cases, it is desirable to elucidate only those conserved pools that encompass targeted metabolites, without computation of entire sets of all conservation relationships. To identify minimal conserved pools, encompassing targeted metabolites, the following mixed integer linear program (MILP) is solved to compute conservation relationships Eq. 3 with the minimal number of non-zero coefficients β_i (i.e., with the minimal number of metabolites present in the corresponding pool),

$$\begin{aligned} \text{minimize} \quad & Z = \sum_{i=1}^{N} y_i \\ \text{subject to} \quad & \beta_{i'} \ge \varepsilon \\ & \sum_{i=1}^{N} \beta_i \cdot S_{ij} = 0 \quad \forall j \in \mathcal{M} \\ & \sum_{i=1}^{N} \beta_i = 1 \\ & 0 \le \beta_i \le y_i, y_i \in \{0,1\}\ \forall i \in \mathcal{N}. \end{aligned} \tag{6}$$

The first constraint $\beta_{i'} \ge \varepsilon$ in Eq. 6 corresponds to a targeted metabolite i' for which a minimal conserved pool is sought, where ε is an arbitrarily small number. The second constraint is the definition of coefficients β_i satisfying Eq. 4. The third constraint in Eq. 6 scales all coefficients β_i to sum to one. Finally, inequalities $0 \le \beta_i \le y_i$ ensure that each conservation coefficient β_i is set to 0 when metabolite i is absent from the minimal pool (i.e., $y_i = 0$). The optimal value of the objective function Z is the minimal number of metabolites present in a conserved pool encompassing metabolite i'. Although alternative optimal solutions to Eq. 6 are often present, the optimal value Z is always unique.

Alternative pools for targeted metabolites can be found by incorporating Eq. 6 within an iterative procedure using integer cuts to exclude previously identified solutions. For example, pools P5 and P8 (see Fig. 3) for the targeted molecule H_2O can be calculated without considering the other pools. By setting $\beta_{H_2O} \ge \varepsilon$ in Eq. 6, the minimal conserved pool P5 can be identified. The next minimal conserved pool P8 for H_2O can be found by appending a simple inequality

$$y_{1,3BPG} + y_{3PG} + y_{2PG} + y_{P_i} + y_{H_2O} \le 4, \tag{7}$$

called an integer cut, to Eq. 6. The integer cut described by Eq. 7 will exclude pool P5 from future iterations because all five binary variables used in Eq. 7 cannot simultaneously take on unit values and, hence, at least one of the corresponding β_i will be forced to zero. Note that the first constraint

TABLE 1 Subsets of fully coupled metabolites in glycolysis

No.	Subset	Conservation ratio	Visual interpretation
1	2PG	Reference	2PG ⇔ 3PG
	3PG	1	
2	DHAP	Reference	
	F1,6P	2	DHAP ⇔ F1,6P ⇔ G3P
	G3P	1	
3	F6P	Reference	F6P ⇔ G6P
	G6P	1	

TABLE 2 Subsets of directionally coupled metabolites in glycolysis

No.	Subset	Range of conservation ratios	Visual interpretation
1	1,3BPG	Reference	1,3BPG → 2PG
	2PG	[0.5, ∞)	
2	AMP	Reference	AMP → ADP
	ADP	[0.5, ∞)	
3	ATP	Reference	ATP → ADP
	ADP	[0.5, ∞)	
4	DHAP	Reference	DHAP → F6P
	F6P	[1, ∞)	
5	Glu	Reference	Glu → F6P
	F6P	[0.5, ∞)	
6	H2O	Reference	H_2O → 2PG
	2PG	[1, ∞)	
7	PEP	Reference	
	1,3BPG	[1, ∞)	1,3BPG ← PEP → 2PG
	2PG	[1, ∞)	
8	PYR	Reference	PYR → 2PG
	2PG	[1, ∞)	
9	Pi	Reference	
	1,3BPG	[1, ∞)	1,3BPG ← P_i → 2PG
	2PG	[1, ∞)	

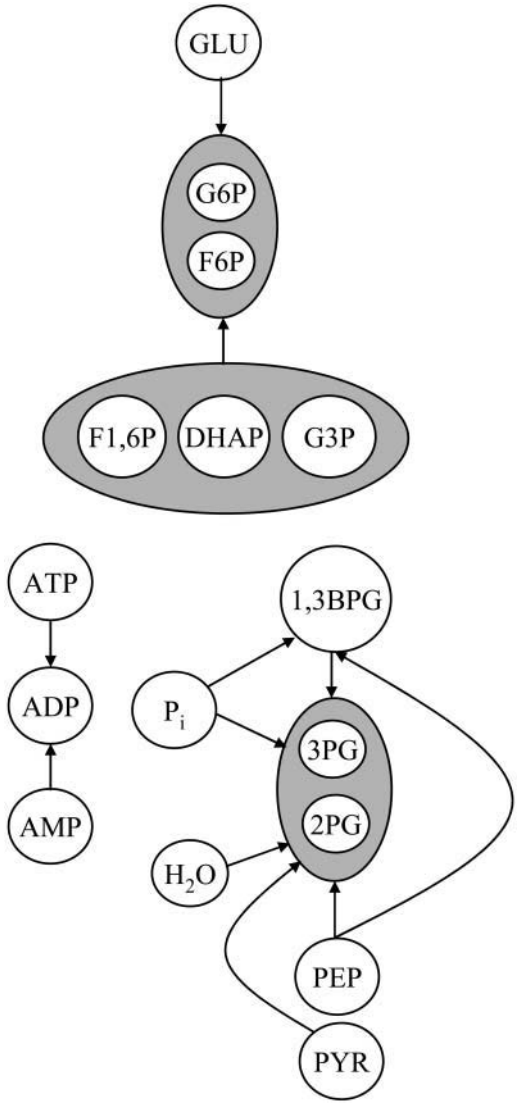


FIGURE 6 Metabolic concentration coupling in glycolysis. Glycolysis admits three metabolite subsets, (G6P, F6P), (F1,6P, DHAP, G3P), and (3PG, 2PG). Directional coupling within glycolysis means that, for example, all conserved pools where PEP is present will always include both metabolite 1,3BPG and subset (3PG, 2PG).

$\beta_{H_2O} \geq \varepsilon$ in Eq. 6 will always enforce $y_{H_2O} = 1$. By analogy, pool P8 can be excluded from future iterations by introducing the integer cut

$$y_{GLU} + y_{G6P} + y_{F6P} + y_{3PG} + y_{2PG} + y_{PYR} + y_{ADP} + y_{AMP} + y_{H_2O} \leq 8. \quad (8)$$

After augmenting Eq. 6 with the above integer cuts, the procedure will terminate because all other pools encompassing H_2O must include nonnegative linear combinations of the excluded extreme pools P5 and P8 (i.e., the formulation will become infeasible).

Unconserved metabolites

While computing conservation ratios using Eq. 5, identically zero conservation coefficients $\beta_{i'}$ (i.e., $\beta_{i'} \equiv 0$) may be encountered. Whereas identically zero fluxes correspond to blocked or unused reactions (Burgard et al., 2004), an identically zero $\beta_{i'}$ implies that metabolite i' is absent from all conserved pools (i.e., it is no longer conserved). Because in closed metabolic systems every metabolite must be present in at least one conserved pool (e.g., Eq. 2) (Heinrich and Schuster, 1996), absence of metabolite i' from all conserved pools signifies incompleteness of the closed metabolic reconstruction as the corresponding stoichiometric matrix, encompassing both internal and external metabolites, must represent a closed system (Famili and Palsson, 2003).

Even though the balanceability of complex metabolic reconstructions can be ensured by checking the atomic balance for every particular reaction, even curated databases can have errors and omissions. In this respect, any additional quick check, like identification of identically zero conservation coefficients, may provide valuable insights into the origin of inaccuracy. Unconserved metabolites absent from all conserved pools can be identified as those accepting zero maximal values for the corresponding conservation coefficients $\beta_{i'}$ (i.e., max $\beta_{i'} = 0$) found from the following linear programs (LPs)

$$\left\{ \begin{array}{lll} \text{maximize} & Z_{i'} = \beta_{i'} & \\ \text{subject to} & \sum_{i=1}^{N} \beta_i \cdot S_{ij} = 0 & \forall j \in \mathcal{M} \\ & \beta_i \geq 0 & \forall i \in \mathcal{N} \end{array} \right\} \forall i' \in \mathcal{N}. \quad (9)$$

Here each Eq. 9 is solved only once for every metabolite i'. If the maximal value $Z_{i'} = \beta_{i'}^{max}$ is zero, then metabolite i' is absent from all conserved pools.

To illustrate, consider first a ''modified'' version of glycolysis with the molecule H_2O omitted from the ATPase reaction only (see Fig. 3). In this case, pools P_5 and P_8 disappear, resulting in the absence of H_2O from all conserved pools and, therefore, signifying the incompleteness of the pathway with respect to the balanceability of the proton and oxygen molecules. Importantly, H_2O is not a dead-end metabolite in the modified glycolysis as it is a substrate for fructose 1,6-bisphosphatase and is a product of enolase (see Fig. 3). Alternatively, if the molecule ATP is removed from pyruvate kinase (see Fig. 3), it is still present in conserved pools despite the fact that there is an unbalanced reaction in the network. Thus the absence of unconserved metabolites does not guarantee that a closed network is balanced and, therefore, this condition can be viewed only as a necessary and not sufficient condition for the balanceability of closed networks. Nevertheless, elucidation of unconserved metabolites, if present, can be used as a fast and final check of the balanceability of complete (closed) metabolic reconstructions.

Computational implementation

The computational implementation for metabolite concentration coupling analysis is similar to the flux coupling finder (FCF) procedure developed by Burgard et al. (2004). The MCCA computational procedure substantially reduces the number of conservation ratios to be calculated by utilizing the transitive nature of partial and full coupling (i.e., if A↔B and B↔C then A↔C) to skip redundant calculations. The conservation ratios for MCCA are calculated using Lindo API whereas the MCPI procedure is implemented in the GAMS modeling environment using CPLEX 7.0. Computational requirements are in the order of minutes for genome-scale models involving as many as 1173 reactions and 811 metabolites on an Intel Pentium IV, 2.4-GHz, 512 MB RAM computer.

RESULTS

In this section, we investigate metabolite subsets and minimal pools for four genome-scale models: *H. pylori* 26695 (Schilling et al., 2002), *E. coli* K-12 iJE660a (Edwards and Palsson, 2000), *E. coli* K-12 iJR904 (Reed

et al., 2003), and *S. cerevisiae* (Foster et al., 2003). Despite their different organization and complexity, the analysis of these genome-scale metabolic reconstructions reveals the following key classes of metabolite subsets: 1), carbon moiety input/output; 2), amino acids; 3), organic phosphate; 4), energy and redox cofactors; 5), sulfur; 6), coenzyme A (CoA); 7), acyl carrier protein (ACP); and 8), various small subsets. Below we will first describe in detail the metabolite subsets for the *E. coli* K-12 iJR904 model and then compare these results with the metabolite subsets identified for the other models.

Genome-scale analysis of metabolite subsets

The *E. coli* K-12 iJR904 model comprises 931 unique reactions and 626 unique metabolites, organized in 30 metabolic pathways (Reed et al., 2003). In our analysis, we consider internal and external metabolites such as H_2O and $H_2O_{(ext)}$ separately. Also, an artificial biomass forming reaction is added to the model and biomass is considered as an individual metabolite. The extended model is thus comprised of 932 reactions and 762 metabolites, where 618 metabolites are present in the cytosol and 144 metabolites are external. All metabolites are found to be present in conserved pools (i.e., the concentration of every metabolite is constrained by conservation relationships).

Application of MCCA led to the identification of the following metabolite subsets:

1. Carbon moiety input/output. Each metabolite of this subset contains carbon and the subset comprises metabolites involved in biotransformations present in 28 of the 30 pathways in the *E. coli* K-12 iJR904 model (see Table 3). We find that 190 metabolites or 24.9% of the 762 metabolites are fully or partially coupled in the subset. The coupled metabolites are present in the central metabolic pathways, biosynthetic pathways, and transport pathways, and are absent only from the nitrogen and putative pathways. Examples of metabolites present in the subset are acetate, 2-oxoglutarate, citrate, formate, fructose, glucose, maltose, oxaloacetate, pyruvate, ubiquinone, ribulose, ribose, succinate, and trehalose.
2. Amino acids. This metabolite subset includes 164 metabolites or 21.5% of all metabolites. The coupled metabolites are present in 21 pathways (see Table 3). This subset includes 18 amino acids where only cysteine and methionine are absent and instead are present together in the sulfur subset discussed later. Some of

TABLE 3 Metabolic pathways spanned by metabolite subsets

Pathway	Carbon influx/outflux	Amino acids	Organic phosphate	Energy and redox cofactors	Sulfur	CoA	ACP
1. Alanine and asparate metabolism	X	X		X			
2. Alternate carbon metabolism	X	X	X	X		X	
3. Anaplerotic reactions	X		X	X		X	
4. Arginine and proline metabolism	X	X		X	X	X	
5. Cell envelope biosynthesis	X	X	X	X		X	X
6. Citrate cycle (TCA)	X			X		X	
7. Cofactor and prothetic group biosynthesis	X	X	X	X	X	X	X
8. Cysteine metabolism	X	X		X	X	X	
9. Folate metabolism	X	X		X			
10. Glutamate metabolism	X	X		X			
11. Glycine and serine metabolism	X	X	X	X		X	
12. Glycolysis/gluconeogenesis	X		X	X		X	
13. Glyoxylate metabolism	X		X	X			
14. Histidine metabolism	X	X	X	X			
15. Membrane lipid metabolism	X	X	X	X		X	X
16. Methionine metabolism	X	X		X	X	X	
17. Methylglyoxal metabolism	X		X		X		
18. Nitrogen	X	X					
19. Nucleotide salvage pathway		X	X	X			
20. Oxidative phosphorylation	X		X	X			
21. Pentose phosphate cycle	X		X	X			
22. Purine and pyrimidine biosynthesis	X	X	X	X			
23. Putative		X		X			
24. Putative transporters	X	X		X			
25. Pyruvate metabolism	X		X	X		X	
26. Threonine and lysine metabolism	X	X		X		X	
27. Transport, extracellular	X	X	X	X	X		
28. Tyrosine, trytophan, and phenylalanine metabolism	X	X	X	X			
29. Unassigned	X			X	X		
30. Valine, leucine, and isoleucine metabolism	X	X		X		X	

the other metabolites present in the subset include 10-formyltetrahydrofolate, *N*-acetyl-L-glutamate, adenosine, cytosine, and glycine. The fact that almost all of the amino acids are simultaneously present within common conserved pools hints at the important role that conserved pools play in the biosynthesis and interconversion of amino acids. From a practical viewpoint, the existence of a single amino acids subset may complicate attempts to arbitrarily increase the concentrations of desired amino acids. Therefore, only limited increases in the concentrations of targeted amino acids can be achieved at the expense of decreasing the concentrations of other amino acids and metabolites coupled in the corresponding conserved pools.

3. Organic phosphate. This subset includes phosphate in the form of a phosphate moiety bound to 75 intermediates (e.g., ribulose-5-phosphate, phosphoenolpyruvate). We find that the organic phosphate subset spans more than half of all the pathways present in the model (see Table 3). This indicates the importance of the corresponding conserved pools for the automatic coordination of the pathways through the transfer of the phosphate moiety between metabolites present in these pools.
4. Energy and redox cofactors. This metabolite subset comprises 147 metabolites or 19.2% of the total 762 metabolites. We find that the subset includes all energy and redox cofactors such as ATP, ADP, AMP, IMP, CTP, GTP, UTP, NAD^+, NADH, $NADP^+$, NADPH, FAD, and $FADH_2$. The well-known fact that the energy and redox cofactors, often referred to as ''currency metabolites'', are ubiquitous in cellular metabolism is supported by the fact that all such currency metabolites are coupled within one large subset spanning 28 pathways. Interestingly, in textbooks, research articles and monographs, examples of the adenylate and NAD^+ conserved moiety pools are usually introduced as two completely independent pools. Although this is true for metabolic subsystems such as glucolysis (see Fig. 3), this genome-scale study reveals that both energy and redox cofactors are present together in conserved pools (i.e., the energy cofactors imply the redox cofactors to be present in the same pools and vice versa). This is because the cofactor and prosthetic group biosynthesis pathway and other biosynthetic pathways functioning in complex organisms couple energy and redox cofactors with one another. For example, the NAD-diphosphatase reaction (EC-3.6.1.22),

$$H_2O + NAD^+ \rightarrow AMP + 2H^+ + NMN,$$

couples NAD^+ and AMP. Another example is the FMN adenylyltransferase reaction (EC-2.7.7.2),

$$ATP + FMN + H^+ \rightarrow FAD + PPi,$$

which couples ATP and FAD. In contrast, in energy metabolism cofactors are transformed in pairs as ATP and ADP, NAD^+ and NADH, or FAD and FADH, resulting in independent pools of the corresponding conserved moieties. Thus, biosynthetic pathways additionally coordinate energy- and redox-dependent metabolic states and processes by coupling the energy and redox cofactors within common conserved pools.
5. Sulfur. This metabolite subset comprises only 16 metabolites coupled to the amino acids, cysteine and methionine. The reason that these two amino acids are decoupled from the other 18 amino acids is that cysteine and methionine carry a sulfur-containing R-group. This subset spans only a very small number of the *E. coli* metabolic pathways (see Table 3).
6. Coenzyme A. This small subset includes coenzyme A. The importance of this small subset that comprises only 16 metabolites is alluded from the fact that it spans 14 out of the 30 pathways present in the model. In comparison, the organic phosphate subset, which comprises almost five times more coupled metabolites, spans only 18 pathways. Interestingly, the presence of the artificially created biomass compound in the subset reveals the fact that changes in concentrations of biomass and CoA groups, including acetyl-CoA, are automatically coordinated.
7. Acyl carrier protein. This metabolite subset represents coupled metabolites carrying acyl carrier protein (e.g., octadecenoyl-ACP, palmitoyl-ACP). The subset includes only 1.4% of all metabolites and is formed within the three biosynthetic pathways, cell envelope biosynthesis, cofactor and prosthetic group biosynthesis, and membrane lipid metabolism.
8. Small metabolite subsets. In addition, there are 21 small metabolite subsets that comprise two metabolites, seven subsets that comprise three metabolites, and one subset that comprises seven metabolites. Most of the two-metabolite subsets couple internal and external concentrations for metabolites CO_2, Fe_2, K, NH_4, NO_2, NO_3, SO_4, taurine, and thiamine. For example, internal CO_2 and external CO_2^{ext} metabolites are always simultaneously present within all common conserved pools. Thus, if CO_2 is present in a pool then CO_2^{ext} is also present in the same pool and vice versa (i.e., a nonzero value of β_{CO_2} implies a nonzero value of $\beta_{CO_2^{ext}}$ and vice versa). An example of a metabolite subset that comprises three internal metabolites is the subset (Heme O, protoheme, siroheme) formed within cofactor and prosthetic group biosynthesis. There are also four metabolite subsets comprising internal and external metabolites that correspond to the proton (e.g., H^+, H_2, and $H^+_{(ext)}$), oxygen, and inorganic phosphate moieties.

Directional coupling and minimal conserved pools

Fig. 7 illustrates directional couplings between metabolite subsets identified for the *E. coli* K-12 iJR904 model. The

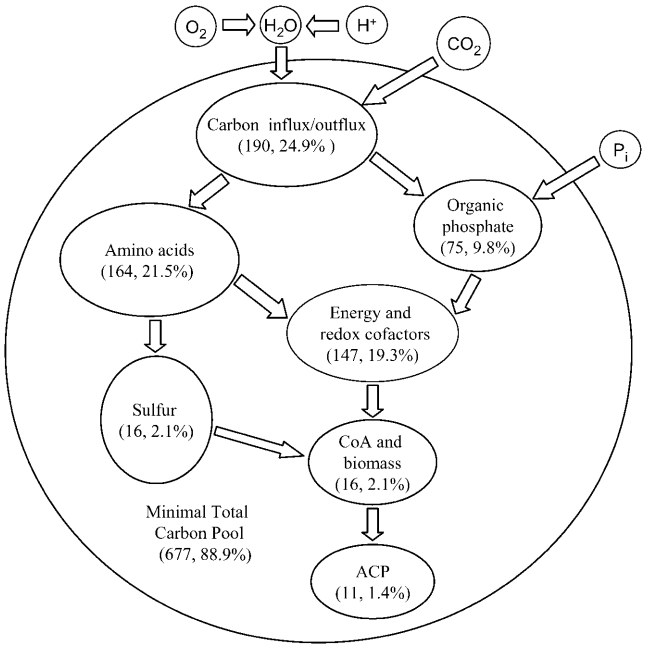


FIGURE 7 Directional coupling between metabolite subsets for the genome-scale *E. coli* (iJR904) model.

directional couplings for the other models are similar. We find, for example, that if metabolites from the CoA subset are present in any conserved pool, then all the metabolites from the ACP subset must also be present in the same pool. Therefore, conserved pools with CoA groups will always include at least $16 + 11 = 27$ coupled metabolites from the two directionally coupled subsets. Similarly, conserved pools encompassing the carbon moiety will always include at least 619 metabolites from all the metabolite subsets directionally coupled with the carbon moiety subset. This implies that conserved ''carbon'' moiety pools always include at least 81% of the 762 metabolites present in the *E. coli* model. It is also interesting to investigate conserved pools encompassing small molecules such as H^+, O_2, H_2O, CO_2, and inorganic phosphate P_i. Fig. 7 shows that most of *E. coli*'s metabolites are present in large common pools shared with these small molecules. Directional couplings between metabolite subsets imply large minimal pools for most metabolites. The MCPI procedure indicated that 692 or 91% of all the *E. coli*'s metabolites are present in the minimal pool for H_2O. The minimal pool for the total carbon moiety comprises 677 metabolites.

Organization of conserved pools within central metabolism for *E. coli*

The coupling interactions between metabolite concentrations in *E. coli* central metabolism are depicted in Fig. 8. We find that almost all metabolites in central metabolism are coupled within four large metabolite subsets, the carbon moiety input/output, organic phosphate, energy and redox cofactors, and coenzyme A subsets. Specifically, the concentrations of all primary metabolites in the pentose phosphate pathway (PPP) and the concentrations of almost all primary metabolites in glycolysis are coupled within common conserved pools by the transfer of organic phosphate groups. We find that TCA cycle metabolites are present in both carbon moiety input/output and CoA metabolite subsets. This is in full agreement with the two important functions of the TCA cycle: 1), coupling pyruvate, produced in glycolysis, to the aerobic production of ATP; and 2), providing initial building blocks for the biomass synthesis (Mathews et al., 2000; Cornish-Bowden and Hofmeyr, 2002). It is important to stress the complementary role that flux and metabolite concentration couplings play in central metabolism. For example, the three major pathways (i.e., glycolysis, PPP, and TCA cycle) are largely uncoupled from each other in the sense that no glycolysis steady-state flux is capable of forcing flux through the TCA cycle or PPP based on stoichiometry alone (Burgard et al., 2004). In contrast, concentration changes in glycolysis affect the PPP and TCA cycle by introducing coordinated changes in the concentrations of metabolites that belong to the same subset, yet, are present in different pathways.

Comparative analysis of metabolite subsets for different models

The *E. coli* K-12 (iJE660a GSM) model comprises 629 unique reactions and 538 metabolites, organized in 82 pathways. In this model 438 metabolites are present in cytosol and 100 metabolites are external. A set of five metabolites absent from all conserved pools was identified, H_2O_2, O_2, $H_{2(ext)}$, H^+_{ext}, and $O_{2(ext)}$ implying that they are unbalanced. This is likely due to the fact that the water and proton molecules were omitted from the *E. coli* (iJE660a) model, although present in the detailed *E. coli* (iJR904) model. Despite the differences between the *E. coli* K-12 iJE660a and iJR904 models, their metabolite subsets are similar. We find that the CoA, ACP, and sulfur subsets are lumped into one subset that also includes biomass (see Table 4). Thus, the most up to date *E. coli* iJR904 model more accurately accounts for the grouping of different conserved moieties within specific conserved pools. Although the compositions of most metabolite subsets are very similar, the compositions and sizes of the carbon moiety input/output subsets for these two models are substantially different. This is because the more recent model includes 224 new metabolites, absent from the earlier model, and almost half of the added metabolites are coupled within the carbon moiety input/output subset.

The *H. pylori* 26695 model includes 385 unique reactions and 402 metabolites, organized in seven pathways. In this model, 338 metabolites are internal and 64 are external. The same five metabolites are absent from all conserved pools as

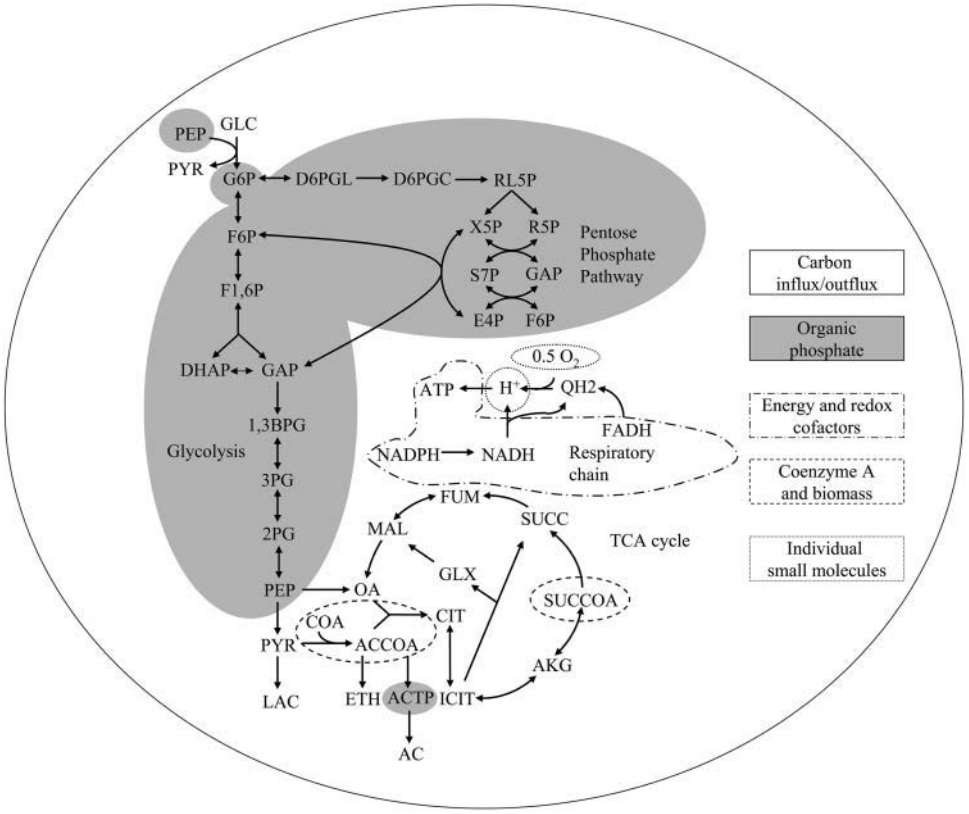


FIGURE 8 Conserved pools spanning central metabolism of the genome-scale *E. coli* (iJR904) model. All metabolites within the four major central metabolism pathways (i.e., glycolysis, TCA cycle, PPP, and respiratory chain) belong to the four metabolite subsets, carbon influx/outflux, organic phosphate, energy and redox cofactors, and CoA subsets.

in the *E. coli* K-12 (iJE660a GSM) model. The metabolite subsets of the *H. pylori* model are also similar to metabolite subsets considered earlier. However, one of the most important differences between the *H. pylori* and the *E. coli* models is that in the *H. pylori* model there are eight amino acids (i.e., cysteine, glycine, histidine, isoleucine, leucine, methionine, phenylalanine, and valine) that are decoupled from all internal metabolite subsets. Each one of these eight amino acids is coupled with its external analog (e.g., valine and valine$_{(ext)}$).

The *S. cerevisiae* model includes 1173 reactions and 811 metabolites, organized in 73 pathways. After removal of all redundant isoenzyme reactions one obtains 841 distinct reactions. In the model there are 692 internal and 118 external metabolites. We find that 360 metabolites or 44.2% of all the metabolites in the model are absent from all conserved pools (i.e., unbalanced). Interestingly, key central carbon metabolites such as pyruvate and phosphoenolpyruvate are also absent from all conserved pools, whereas all cofactors are present in conserved pools. This hints at the fact that there is at least one reaction in the model that is likely to be unbalanced with respect to carbon atoms.

DISCUSSION

In this article, we introduced: i), metabolite concentration coupling analysis for identifying subsets of metabolites, which are simultaneously present in common conserved pools, and ii), the minimal conserved pool identification procedure for elucidation of targeted metabolite pools. These approaches remain tractable for large metabolic networks that comprise hundreds or even thousands of reactions and metabolites, inferred from genome-scale metabolic reconstructions. MCCA draws from and complements flux coupling analysis and the flux coupling finder procedure introduced earlier by Burgard et al. (2004) for the identification of blocked and coupled reactions in genome-scale metabolic networks. Although certain conserved pools elucidated from closed metabolic reconstructions may disappear depending on uptake/secretion transport conditions, the analyses provide complete sets of metabolite subsets and conserved pools that can be formed within the cellular systems under a set of specified conditions. For example, *E. coli*, a free living bacterium, can synthesize all necessary cellular structures from several sugar substrates (Neidhardt et al., 1990). When *E. coli* is fed on a minimal medium, it neither consumes amino acids, absent from the medium, nor secretes the amino acids needed for its optimal growth. Thus, under certain growth conditions, the conserved pools encompassing the amino acids subset can be viewed as ''internal'' conserved pools, which do not exchange mass with the environment. In contrast, *H. pylori*, a human parasite, requires the uptake of certain amino acids for growth. Accordingly, the eight essential amino acids (i.e.,

TABLE 4 Distribution of metabolites across metabolite subsets

Subset classification	*H. pylori*	*E. coli* (iJE660a)	*E. coli* (iJR904)
	Number of coupled metabolites		
Carbon influx and outflux	44 (12)	107 (33)	190 (55)
Amino acids	101 (25)	158 (45)	164 (48)
Organic phosphate	41 (0)	57 (0)	75 (3)
Energy and redox cofactors	91 (0)	116 (2)	147 (1)
Sulfur	–	–	16 (2)
CoA	–	–	11 (0)
ACP	–	–	16 (1)
CoA/Sulfur	19 (0)	33 (3)	–

Numbers enclosed in parentheses show the number of external metabolites present in the subsets.

cysteine, glycine, histidine, isoleucine, leucine, methionine, phenylalanine, and valine) are absent from all internal metabolite subsets in the *H. pylori* model. Instead they are coupled with their external equivalents indicative of the need to uptake and inability to synthesize them.

After applying MCCA to the genome-scale stoichiometric models of *H. pylori* and *E. coli*, the following key classes of metabolite subsets were identified, 1), the carbon influx/outflux subset, 2), the amino acids subset, 3), the organic phosphate subset, 4), the energy and redox cofactors subset, 5), the sulfur subset, 6), the coenzyme A subset, 7), the acyl carrier protein subset, and 8), various small subsets. These subsets encompass ~80% of all metabolites present in the models. It was also determined that many metabolite subsets are directionally coupled to one another (see Fig. 7). For example, all metabolite subsets are simultaneously present in conserved pools encompassing the carbon influx/outflux subset. For the *E. coli* K-12 iJR904 model, the MCPI procedure applied to the largest carbon influx/outflux metabolite subset revealed that the minimal conserved pools forming the subset are comprised of almost 90% of all metabolites. Given that glucose can serve as the sole carbon source for *E. coli* (Neidhardt et al., 1990), all carbon atoms in *E. coli* metabolites must be reachable from glucose (Arita, 2004). This is supported by the fact that all the key metabolite subsets are directionally coupled with the carbon influx/outflux metabolite subset and thus the corresponding large metabolite pools can be regarded as a ''carbon resource'' for the bacterium. Although the steady-state flux organization of genome-scale metabolic networks is characterized by and is a direct consequence of the networks' scale-free architectures (Almaas et al., 2004; Burgard et al., 2004), in contrast almost all metabolites are present within a handful of key metabolite subsets. Conserved pools provide an upper limit on the attainable metabolite concentrations in the cell and, so, can be viewed as global biophysical barriers protecting cellular systems from large concentration changes (Bakker et al., 2000). However, the role that this limit can play in regulation of cellular systems in vivo is not still well understood. This is because allosteric regulation and other factors can keep metabolite concentrations far below the maximal upper limit.

Examination of metabolite subsets and conserved pools provides not only fundamental insights into the genome-scale organization of concentration interactions within metabolic pathways but may provide insights in designing strategies for efficient therapeutic interventions, guiding metabolic engineering efforts, and aiding metabolic reconstructions of complex organisms. In both biotechnological and therapeutic interventions, fluxes through appropriate pathways and/or metabolite concentrations are changed to trigger desired effects. Recently a strategy to kill pathogenic organisms by blocking transport of byproducts has been suggested (Wiemer et al., 1995; Cornish-Bowden and Hofmeyr, 2002). The inhibition of all transporters will lead to the accumulation of byproducts to toxic amounts and, as a result, to cell death. This strategy appears to be feasible for organisms having a few transported metabolites. However, if there are many external metabolites simultaneously present in common conserved pools such as in *E. coli* and *H. pylori*, then the inhibition of only a few transporters will be insufficient to trigger the accumulation of byproducts inside the cell. This is because byproducts can be converted into the other biochemicals present in common pools and then released from the cell. At the same time, large conserved pools can help advance biotechnological applications. Specifically, the presence of secondary metabolites in common pools shared with targeted amino acids provides many degrees of freedom that can be used to increase amino acids production by reducing concentrations of appropriate secondary metabolites.

APPENDIX A

Conservation relationships theorem

In this Appendix, we show that all conservation relationships for metabolite concentrations in a closed metabolic system are solely determined by the system's stoichiometry.

Theorem

Consider an arbitrary closed metabolic system that comprises N metabolites and M reactions. Any linear combination of the system's metabolite concentrations, $\sum_{i=1}^{N} \beta_i \cdot C_i$, is time invariant if and only if the coefficients β_i introduce linear dependencies between rows of the system's stoichiometric matrix $\mathbf{S}$ (i.e., satisfy linear equations $\sum_{i=1}^{N} \beta_i \cdot S_{ij} = 0, j = 1, \ldots, M$).

Proof
(Necessary condition)

First, we prove that coefficients β_i of the conservation relationship $\sum_{i=1}^{N} \beta_i \cdot C_i = \text{const}$ (P) satisfy equations $\sum_{i=1}^{N} \beta_i \cdot S_{ij} = 0$, $j = 1,\ldots, M$ (E). Suppose that the conservation relationship (P) holds for some coeffici-

ents β_i. Differentiation of (P) with respect to time t leads to the following chain of equalities

$$\frac{d}{dt}\left(\sum_{i=1}^{N}\beta_i \cdot C_i\right) = \sum_{i=1}^{N}\beta_i \cdot \frac{dC_i}{dt} = \sum_{i=1}^{N}\beta_i \cdot \left(\sum_{j=1}^{M} S_{ij} \cdot v_j\right) = \sum_{j=1}^{M}\left(\sum_{i=1}^{N}\beta_i \cdot S_{ij}\right) \cdot v_j \equiv 0. \quad (10)$$

The last identity in Eq. 10 can be rewritten as $\sum_{j=1}^{M} \tilde{S}_j \cdot v_j \equiv 0$ with $\tilde{S}_j = \sum_{i=1}^{N} \beta_i \cdot S_{ij}$. Because the identity holds for any arbitrary values of v_j, the coefficients $\tilde{S}_j$ must be identically zero. Therefore, equations (E) are fulfilled.

Sufficient condition

We need to prove that any solution of (E) corresponds to a conservation relationship (P). Given coefficients β_i satisfying (E), differentiation of (P) with respect to t leads to

$$\frac{d}{dt}P = \frac{d}{dt}\left(\sum_{i=1}^{N}\beta_i \cdot C_i\right) = \sum_{j=1}^{M}\left(\sum_{i=1}^{N}\beta_i \cdot S_{ij}\right) \cdot v_j = 0.$$

The proof follows.

We gratefully acknowledge support for this research from the Department of Energy and National Science Foundation BES 0120277.

REFERENCES

Almaas, L., B. Kovacs, T. Vicsek, Z. N. Oltval, and A.-L. Barabasi. 2004. Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature*. 427:839–843.

Arita, M. 2004. The metabolic world of *Escherichia coli* is not small. *Proc. Natl. Acad. Sci. USA*. 101:1543–1547.

Atkinson, D. E. 1968. The energy charge of the adenylate pool as a regulatory parameter. Interactions with feedback modifiers. *Biochemistry*. 7:4030–4034.

Bakker, B. M., F. I. C. Mensonides, B. Teusink, P. van Hoek, P. A. M. Michels, and H. V. Westerhoff. 2000. Compartmentation protects trypanosomes from the dangerous design of glycolysis. *Proc. Natl. Acad. Sci. USA*. 88:3268–3271.

Bakker, B. M., P. A. M. Michels, F. R. Opperdoes, and H. V. Westerhoff. 1999. What controls glycolysis in bloodstream form *Trypanosoma brucei*? *J. Biol. Chem.* 274:14551–14559.

Burgard, A. P., E. V. Nikolaev, C. H. Schilling, and C. D. Maranas. 2004. Flux coupling analysis of genome scale metabolic network reconstructions. *Genome Res.* 14:301–312.

Burgard, A. P., P. Pharkya, and C. D. Maranas. 2003. OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng*. 84:647–657.

Chernikov, S. N. 1968. Linear Inequalities. Nauka, Moscow [in Russian]. (German translation. 1971. Lineare Ungleichungen. Deutscher Verlag der Wissenschaften, Berlin).

Chvátal, V. 2000. Linear Programming. W. H. Freeman and Company, New York.

Cornish-Bowden, A., and J. H. Hofmeyr. 2002. The role of stoichiometric analysis in studies of metabolism: an example. *J. Theor. Biol.* 216:179–191.

Edwards, J. S., and B. O. Palsson. 2000. The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. USA*. 97:5528–5533.

Eisenthal, R., and A. Cornish-Bowden. 1998. Prospects for antiparasitic drugs. *J. Biol. Chem.* 273:5500–5505.

Famili, I., and B. O. Palsson. 2003. The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools. *Biophys. J*. 85:16–26.

Foster, J., I. Famili, P. C. Fu, B. O. Palsson, and J. Nielsen. 2003. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* 13:244–253.

Golub, G. H., and C. F. van Loan. 1996. Matrix Computations. Johns Hopkins University Press, Baltimore, MD.

Gottschalk, G. 1986. Bacterial Metabolism, 2nd Ed. Springer, New York.

Heinrich, R., and S. Schuster. 1996. The Regulation of Cellular Systems. Chapman & Hall, New York.

Kholodenko, B. N., M. Cascante, J. B. Hoek, H. V. Westerhoff, and J. Schwaber. 1998. Metabolic design: how to engineer a living cell to desired metabolite concentrations and fluxes. *Biotechnol. Bioeng*. 59: 239–247.

Kholodenko, B. N., H. M. Sauro, and H. V. Westerhoff. 1994. Control by enzymes, coenzymes and conserved moieties: a generalization of the connectivity theorem of metabolic control analysis. *Eur. J. Biochem.* 225: 179–186.

Kholodenko, B. N., H. V. Westerhoff, J. Schwaber, and M. Cascante. 2000. Engineering a living cell to desired metabolite concentrations and fluxes: pathways with multifunctional enzymes. *Metab. Eng*. 2:1–13.

Mathews, C. K., E. E. van Holde, and K. G. Ahern. 2000. Biochemistry. Addison Wesley Longman, New York.

Neidhardt, F. C., J. L. Ingraham, and M. Schaechter. 1990. Physiology of the Bacterial Cell. A Molecular Approach. Sinauer Associates, Sunderland, MA.

Pfeiffer, T., I. Sanchez-Valdenebro, H. C. Nuno, F. Montero, and S. Schuster. 1999. METATOOL: for studying metabolic networks. *Bioinformatics*. 15:251–257.

Pharkya, P., A. P. Burgard, and C. D. Maranas. 2003. Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. *Biotechnol. Bioeng*. 84:887–899.

Reder, C. 1988. Metabolic control theory: a structural approach. *J. Theor. Biol*. 135:175–201.

Reed, J. L., T. D. Vo, C. H. Schilling, and B. O. Palsson. 2003. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.* 4:R54.

Reich, J. G., and E. E. Selkov. 1981. Energy Metabolism of the Cell: A Theoretic Treatise. Academic Press, London, UK.

Rockafellar, R. T. 1970. Convex Analysis. Princeton University Press, Princeton, NJ.

Schilling, C. H., M. W. Covert, I. Famili, G. M. Church, J. S. Edwards, and B. O. Palsson. 2002. Genome-scale metabolic model of *Helicobacter pylori* 26695. *J. Bacteriol*. 184:4582–4593.

Schuster, S., and C. Hilgetag. 1995. What information about the conserved-moiety structure of chemical reaction systems can be derived from their stoichiometry? *J. Phys. Chem.* 99:8017–8023.

Schuster, S., and T. Höfer. 1991. Determining all extreme semi-positive conservation relations in chemical reaction systems: a test criterion for conservativity. *J. Chem. Soc. Faraday Trans.* 87: 2561–2566.

Wiemer, E. A. C., P. A. M. Michels, and F. R. Opperdoes. 1995. The inhibition of pyruvate transport across the plasma membrane of the bloodstream form of *Trypanosoma brucei* and its metabolic implications. *Biochem. J*. 312:479–484.