

## Research Article

# Optimal Fluxes, Reaction Replaceability, and Response to Enzymopathies in the Human Red Blood Cell

A. De Martino,<sup>1,2</sup> D. Granata,<sup>2</sup> E. Marinari,<sup>2</sup> C. Martelli,<sup>3</sup> and V. Van Kerrebroeck<sup>2</sup>

<sup>1</sup> CNR Institute for Physico-Chemical Processes (IPCF), Rome Sapienza Unit, Roma, Italy

<sup>2</sup> Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro 2, 00185 Roma, Italy

<sup>3</sup> Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, CT 06520, USA

Correspondence should be addressed to A. De Martino, andrea.demartino@roma1.infn.it

Received 30 November 2009; Accepted 14 May 2010

Academic Editor: Jamey Young

Copyright © 2010 A. De Martino et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Characterizing the capabilities, key dependencies, and response to perturbations of genome-scale metabolic networks is a basic problem with important applications. A key question concerns the identification of the potentially most harmful reaction knockouts. The integration of combinatorial methods with sampling techniques to explore the space of viable flux states may provide crucial insights on this issue. We assess the replaceability of every metabolic conversion in the human red blood cell by enumerating the alternative paths from substrate to product, obtaining a complete map of the potential damage of single enzymopathies. Sampling the space of optimal steady state fluxes in the healthy and in the mutated cell reveals both correlations and complementarity between topologic and dynamical aspects.

## 1. Introduction

Understanding metabolic activity from the underlying genotype is one of the most addressed problems in computational biology. Of particular interest is the issue of the identification of the reactions that are indispensable for an organism to survive, grow or perform a specific function in a given growth medium or, conversely, of the potentially most harmful knock-outs or enzymopathies. Several experimental protocols are able to assess the essentiality of gene products (and hence of the corresponding metabolic reactions), ranging from individual knock-outs to transposon mutagenesis and RNA interference [1–5]. Computational approaches on the other hand might provide important clues on the system-level organization by investigating genome-scale network reconstructions.

The functional modularity of metabolic networks suggests that topological aspects may provide a key to identify a class of essential pathways [6, 7]. However the metabolic genotype only constitutes the frame on the top of which the dynamic phenotype is built. The essentiality of a metabolic pathway will in general depend on both structural considerations based on the network reconstruction from

genomic information, and on the “model of metabolism” defined on it, for example, on the corresponding steady state fluxes. In *E.coli*, phenotypical essentiality of metabolic genes has been associated with a reduced allowed variability of the corresponding fluxes, suggesting that dynamically stiff reactions may constitute an evolutionarily robust backbone of metabolism conserved over different species [8].

Here we attempt a more thorough integration of topological and dynamical views to obtain a more comprehensive insight into a metabolic network’s organization, efficiency, and ability to respond to perturbations. We will first associate the essentiality of a reaction with a measure of its topological replaceability by enumerating the alternative paths from substrate to product along the network edges, with the rationale that from a purely structural viewpoint more replaceable reactions are less likely to be crucial nodes of the network. Then we will validate and compare the essentiality map thus obtained with the metabolic phenotype resulting from the definition of a general constraint-based model for metabolic flux prediction. We shall see that dynamical and structural measures of essentiality may offer complementary views of a reaction network’s robustness.

We carry out our analysis on the metabolic network of the human red blood cell (hRBC), one of the most studied complexes in systems biology, from the earliest mathematical models of single biochemical pathways [9, 10] to the currently available genome-scale reconstructions [11]. The reason for this choice lies essentially in its limited size. On the one hand, it allows to compute reaction replaceabilities *exactly* by a suitable modification of Johnson’s algorithm for counting loops in a directed graph [12]. On the other, it allows for the efficient application of various sampling methods to the space of viable flux states [8, 13]. The latter is a vital ingredient to address many important properties of erythrocytes. Indeed for some organisms under certain conditions it is reasonable to assume that the metabolic activity is aimed at maximizing a subset of the metabolic reactions (or a function of them) associated with a certain biological function. In such cases the relevant flux configuration can be computed by standard optimization algorithms. For example, *E. coli*’s metabolism has been shown to maximize biomass production under evolutionary pressure [14], but after a genetic knockout it responds with a minimum rearrangement of fluxes [15]. While the production of the cofactors ensuring the maintenance of osmotic balance and the release of oxygen may be argued to be their metabolic goal, erythrocytes do not generically allow for such a simplification. Information-rich directions in flux space must be retrieved by coupling the underlying constraints on fluxes with other types of analyses. Much understanding has indeed been obtained from the uniform sampling of feasible states [13, 16, 17] and by functional studies, like the computation of extreme pathways [18], of metabolic regulatory structures [19, 20] and of metabolic pools [21]. These aspects combined make hRBCs a key benchmark for both theories of metabolism and computational tools.

It is worth noting that the detailed structural information we derive (i.e., the full map of alternative paths for each substrate/product pair) cannot be retrieved by other methods. Unluckily, computation times still prevent scaling the approach we employ up to networks larger than a few hundred nodes. More refined algorithms are currently being developed to overcome this limitation.

## 2. Approach

**2.1. Structural Analysis.** Given a reaction network, we want to compute, for any pair of metabolites  $a$  and  $b$  that are, respectively, substrate and product in a reaction  $i$  (this situation will be indicated by  $a \xrightarrow{i} b$ ), the number  $\mathcal{N}_{a \rightarrow b}^{(i)}(\ell)$  of alternative pathways, excluding reaction  $i$ , of length  $\ell$  allowing for the conversion  $a \rightarrow b$ , see Figure 1. The rationale is that a reaction performing a metabolite conversion  $a \xrightarrow{i} b$  for which  $\mathcal{N}_{a \rightarrow b}^{(i)}(\ell)$  (or, more properly,  $\sum_{\ell} \mathcal{N}_{a \rightarrow b}^{(i)}(\ell)$ ) is large will be more easily substituted, in case of an enzymopathy or a knockout, than one for which the above quantity is small.

Finding paths connecting two points of a directed network is a long-studied problem in computer science. The focus is usually on locating the shortest paths or the fastest

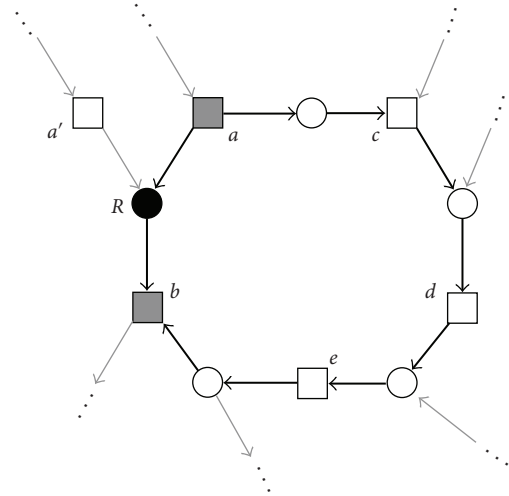


FIGURE 1: Bipartite graph representation of a reaction network, with circles (resp. squares) denoting reactions (resp. metabolites). Here, reaction  $R$  uses metabolites  $a$  and  $a'$  as substrates to produce metabolite  $b$ . If  $R$  is removed, the conversion of  $a$  to  $b$  is still permitted by the alternative pathway  $a \rightarrow c \rightarrow d \rightarrow e \rightarrow b$ . When  $R$  is fictitiously reversed, this chain forms a directed loop of length 5 reactions, formed by  $R$  and by a path passing through  $\ell = 4$  other reactions.

way to find any path. Enumerating *all* the distinct paths between two vertices is however a less confronted issue. In our case it is crucial to avoid overcounting, for example, due to self-intersecting paths. Therefore we shall resort to an exhaustive algorithm. We will identify the substitutive paths using the following trick: for each pair  $(a, b)$  of metabolites such that  $a \xrightarrow{i} b$ , revert  $i$  fictitiously. This results in a new graph where an auxiliary edge  $b \xrightarrow{i} a$  replaced the edge  $a \xrightarrow{i} b$ , see again Figure 1. Counting the number of alternative reaction chains producing  $b$  from  $a$  then comes down to computing the number of directed cycles, that is, non self-intersecting directed closed paths along the edges of the new graph, passing through the fictitious edge  $b \xrightarrow{i} a$ . Thanks to the limited size of the hRBC network it is possible to solve this enumeration problem exactly via Johnson’s algorithm [12], briefly described in the following section.  $\mathcal{N}_{a \rightarrow b}^{(i)}(\ell)$  can now be trivially inferred. For simplicity,  $\ell$  will denote here the number of reactions in the alternative pathway ( $\ell = 4$  in Figure 1).

**2.2. Flux Analysis.** The space of viable fluxes will be defined through a constraint-based approach which relies on more general assumptions than flux-balance analysis (FBA, [22]). FBA is the standard method to model steady-state reaction networks where mass balance constraints are imposed to every metabolite. For a reaction network with  $N$  reactions and  $M$  metabolites, let us denote by  $\mathbf{A}$  and  $\mathbf{B}$ , respectively, the  $M \times N$  matrices of output and input stoichiometric coefficients. The stoichiometric matrix is given by  $\mathbf{S} = \mathbf{A} - \mathbf{B}$ . Letting  $\boldsymbol{\nu} = (\nu_i)_{i=1}^N$  denote a vector of fluxes (with properly chosen bounds  $\nu_i^{\min} \leq \nu_i \leq \nu_i^{\max}$ ), the concentrations

$\mathbf{c} = (c^a)_{a=1}^M$  of metabolites vary in time according to  $\dot{\mathbf{c}} = \mathbf{S}\boldsymbol{\nu} - \mathbf{u}$ , where  $\mathbf{u} = (u^a)_{a=1}^M$  stands for the net cellular uptake of metabolite  $a$  ( $u^a > 0$  if  $a$  is a global output of metabolism,  $u^a < 0$  if  $a$  is consumed by the organism,  $u^a = 0$  if  $a$  is mass-balanced). Assuming a steady state, the concentrations are constant in time (i.e.,  $\dot{\mathbf{c}} = 0$ ) and vectors  $\boldsymbol{\nu}$  satisfying  $\mathbf{S}\boldsymbol{\nu} = \mathbf{u}$ , or

$$(\mathbf{A} - \mathbf{B})\boldsymbol{\nu} = \mathbf{u}, \quad (1)$$

represent flux configurations ensuring that each metabolite meets its production or consumption constraints at fixed concentrations. As  $N$  is typically larger than  $M$ , the system is underdetermined and feasible flux states form a convex set of dimension  $N - \text{rank}(\mathbf{S})$  embedded in the  $N$ -dimensional space of fluxes. In absence of a selection criterion that allows to pick one solution out of this set (as e.g., a maximum biomass principle), a uniform sampling of the solution space should be carried out. When  $N$  is sufficiently small (as for hRBCs), this can be achieved effectively, albeit at a considerable computational cost, by Monte Carlo methods [13, 16] or by message-passing procedures [17].

Here we will consider a different but related flux scheme based on Von Neumann's (VN) model of reaction networks [8]. In the VN framework, one fixes the environment through a small set of intakes on nutrients and defines a self-consistent flux problem where the network chooses, given a target growth rate, how much of the nutrients to use and which metabolites are globally produced. Mass balance then emerges as a property of the solutions for some metabolites.

The equations describing the VN model have been studied by statistical mechanics methods in [23, 24]. For an intuitive derivation, note that the quantities  $\mathbf{A}\boldsymbol{\nu}$  and  $\mathbf{B}\boldsymbol{\nu}$  represent, respectively, the total output and the total input of each metabolite for a given flux vector  $\boldsymbol{\nu}$ . Then a flux vector such that  $\mathbf{A}\boldsymbol{\nu} \geq \rho\mathbf{B}\boldsymbol{\nu}$ , with some constant  $\rho > 0$ , describes a network state where metabolites are being produced at a rate at least equal to  $\rho$ , since for each of them the total output is at least  $\rho$  times the total input. It is simple to see that as  $\rho$  increases the volume of such flux vectors shrinks continuously (for  $\rho = 0$  every flux vector is a solution). In particular, there exists a value  $\rho^*$  of  $\rho$ , representing the maximum metabolic production rate compatible with the stoichiometric constraints, above which no suitable flux vectors exist. The presence of conserved metabolic pools [25] implies  $\rho^* = 1$  [26], so that in metabolic networks optimal steady state fluxes correspond to the solutions of

$$(\mathbf{A} - \mathbf{B})\boldsymbol{\nu} \geq \mathbf{0}. \quad (2)$$

The solutions of (2) do not coincide with those of (1) even for  $\mathbf{u} = \mathbf{0}$ . Interestingly, a finite volume of (optimal) flux states turns out to satisfy the above constraints [8]. This trait is at odds with both the behavior of the solutions of (2) for a random reaction network (where a single solution survives at  $\rho^*$  [24]) and with the optimization that is usually coupled to FBA (where typically a single flux state maximizes the objective function), and points to the robustness of metabolic phenotypes. For *E.coli*, in particular, the solutions

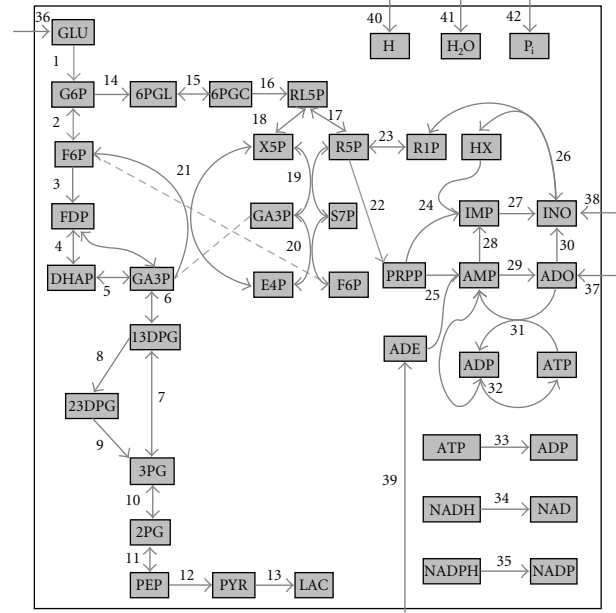


FIGURE 2: Scheme of the hRBC metabolic network used in our analysis. Squares correspond to metabolites, numbers to reactions (see Table 1).

of (2) have been shown to reproduce both the large-scale organization of fluxes and the individual measured rates. In addition, fluxes with the smallest solution-to-solution fluctuations, representing the most susceptible parts of the network, turn out to be strongly correlated with *E.coli*'s phenomenologically essential genes [8]. The main technical advantage in using the VN scheme lies in the fact that its solution space can be sampled uniformly at very modest computational costs even for genome-scale models. The algorithm allowing for this, which has been recently applied to sample *E.coli*'s solution space [8], is detailed in the following section. Its running times for hRBCs are negligible.

### 3. Methods

**3.1. Reconstructed Network.** We consider the hRBC metabolic network studied in [16], a map of which is shown in Figure 2; Table 1 lists reactions and the corresponding abbreviations. The network comprises three main pathways, namely, glycolysis (reactions 1–13), the pentose phosphate (PP) pathway (14–21) and the adenosine metabolism, with a total of  $M = 39$  metabolites linked by  $N = 59$  reactions: 49 internal reactions (34 of which come from the splitting of 17 reversible processes), 3 auxiliary fluxes to maintain the osmotic equilibrium and the redox state of the cell (ATPase, NADHase, NADPHase) and 7 uptake reactions to guarantee the intake of the necessary nutrients (GLU, ADE, ADO, INO), and of the cytosol elements ( $\text{H}_2\text{O}$ , H,  $\text{P}_i$ ). The forward and backward parts of reversible reactions are treated separately throughout this study, both in the structural and in the flux analysis.

TABLE 1: List of reactions considered in this work, including the corresponding number in the map of Figure 2, the abbreviation and the process. The 7 uptake fluxes, numbered 36 to 42, are as shown in Figure 2.

No.	Abbreviation	Chemical reaction
1	HK	$\text{GLU} + \text{ATP} \rightarrow \text{G6P} + \text{ADP} + \text{H}$
2	PGI	$\text{G6P} \leftrightarrow \text{F6P}$
3	PFK	$\text{F6P} + \text{ATP} \rightarrow \text{FDP} + \text{ADP} + \text{H}$
4	ALD	$\text{FDP} \leftrightarrow \text{GA3P} + \text{DHAP}$
5	TPI	$\text{DHAP} \leftrightarrow \text{GA3P}$
6	GAPDH	$\text{GA3P} + \text{NAD} + \text{P}_i \leftrightarrow \text{13DPG} + \text{NADH} + \text{H}$
7	PGK	$\text{13DPG} + \text{ADP} \leftrightarrow \text{3PG} + \text{ATP}$
8	DPGM	$\text{13DPG} \rightarrow \text{23DPG} + \text{H}$
9	DPGase	$\text{23DPG} + \text{H}_2\text{O} \rightarrow \text{3PG} + \text{P}_i$
10	PGM	$\text{3PG} \leftrightarrow \text{2PG}$
11	EN	$\text{2PG} \leftrightarrow \text{PEP} + \text{H}_2\text{O}$
12	PK	$\text{PEP} + \text{ADP} + \text{H} \rightarrow \text{PYR} + \text{ATP}$
13	LDH	$\text{PYR} + \text{NADH} + \text{H} \leftrightarrow \text{LAC} + \text{NAD}$
14	G6PDH	$\text{G6P} + \text{NADP} \rightarrow \text{6PGL} + \text{NADPH} + \text{H}$
15	PGL	$\text{6PGL} + \text{H}_2\text{O} \leftrightarrow \text{6PGC} + \text{H}$
16	PDGH	$\text{6PGC} + \text{NADP} \rightarrow \text{RL5P} + \text{NADPH} + \text{CO}_2$
17	RPI	$\text{RL5P} \leftrightarrow \text{R5P}$
18	XPI	$\text{RL5P} \leftrightarrow \text{X5P}$
19	TKI	$\text{X5P} + \text{R5P} \leftrightarrow \text{S7P} + \text{GA3P}$
20	TA	$\text{GA3P} + \text{S7P} \leftrightarrow \text{E4P} + \text{F6P}$
21	TKII	$\text{X5P} + \text{E4P} \leftrightarrow \text{F6P} + \text{GA3P}$
22	PRPPsyn	$\text{R5P} + \text{ATP} \rightarrow \text{PRPP} + \text{AMP}$
23	PRM	$\text{R1P} \leftrightarrow \text{R5P}$
24	HGPRT	$\text{PRPP} + \text{HX} + \text{H}_2\text{O} \rightarrow \text{IMP} + 2\text{P}_i$
25	AdPRT	$\text{PRPP} + \text{ADE} + \text{H}_2\text{O} \rightarrow \text{AMP} + 2\text{P}_i$
26	PNPase	$\text{INO} + \text{P}_i \leftrightarrow \text{HX} + \text{R1P}$
27	IMPase	$\text{IMP} + \text{H}_2\text{O} \rightarrow \text{INO} + \text{P}_i + \text{H}$
28	AMPDA	$\text{AMP} + \text{H}_2\text{O} \rightarrow \text{IMP} + \text{NH}_3$
29	AMPase	$\text{AMP} + \text{H}_2\text{O} \rightarrow \text{ADO} + \text{P}_i + \text{H}$
30	ADA	$\text{ADO} + \text{H}_2\text{O} \rightarrow \text{INO} + \text{NH}_3$
31	AK	$\text{ADO} + \text{ATP} \rightarrow \text{ADP} + \text{AMP}$
32	ApK	$2\text{ADP} \leftrightarrow \text{ATP} + \text{AMP}$
33	ATPase	$\text{ATP} \rightarrow \text{ADP} + \text{P}_i$
34	NADHase	$\text{NADH} \rightarrow \text{NAD} + \text{H}$
35	NADPHase	$\text{NADPH} \rightarrow \text{NADP} + \text{H}$

3.2. *Structural Analysis.* Structural vulnerabilities are identified by analyzing the loop structure of a modified metabolic reaction network, created from the original one by inverting—in turn—the direction of the single reaction for which we want to compute the replaceability, as explained in Figure 1. The fastest known exact algorithm (for the worst case scenario) of this cycle enumeration problem for a directed graph was introduced by Johnson [12]. We shall now shortly describe its key ideas, referring to [12] for a pseudocode.

Given a directed graph with  $n$  vertices and  $e$  edges, the algorithm is designed to build non self-intersecting paths from a root vertex  $r$  to itself, loading them onto stacks. The

main ingredients allowing for an optimal exploration of the graph are (a) a smart choice of the root vertex, and (b) an efficient method to avoid duplicating cycles and repeating searches on the same portions of the graph. To achieve this, vertices are initially ordered in a lexicographic sequence, and the algorithm only selects as roots those nodes that are the “least” vertex (in the initial ordering) of at least one cycle. The algorithm described in [27] guarantees to find such vertices in  $O(n + e)$  operations. Moreover, to avoid self-intersections, each time a node is loaded onto a stack it is also given a “blocked” status. It was proven by Johnson that if a vertex  $v$  stays blocked as long as every path from  $v$  to the root vertex  $r$  intersects the current path at a vertex

other than  $r$ , the algorithm outputs all cycles exactly once. By sufficiently delaying the unblocking of each of these vertices and by keeping track of the portions of the graph that have been searched holding the current stack, the maximum time that can elapse between two consecutive cycle outputs can be reduced to  $O(n + e)$ . The same holds for the time window before the first cycle is delivered and for the one after the output of the last cycle. Hence, the total time needed to list the, say,  $c$  cycles of the graph is  $O((n + e)(c + 1))$ . In our case, each fictitious reaction reversal generates a new graph, so that computing the complete substitutability map for a network of  $N$  reactions requires a time of the order  $O(N(n + e)(c + 1))$ . For practical reasons, we perform this analysis on the bipartite metabolic network (as in Figure 1) rather than the reduced network of Figure 2. This implies that in our case  $n = N + M$ .

One can in principle consider different measures of replaceability of a metabolic conversion  $a \xrightarrow{i} b$ . The quantity  $\mathcal{R}_{a \rightarrow b}^{(i)} = \sum_{\ell} \mathcal{N}_{a \rightarrow b}^{(i)}(\ell)$ , counting the total number of paths alternative to  $i$  from  $a$  to  $b$  of any length, is perhaps the most obvious option. Taking into account the fact that, typically, longer detours can be less convenient than shorter ones from an energetic viewpoint one could instead consider  $\ell$ -weighted functions like  $\mathcal{W}_{a \rightarrow b}^{(i)} = \sum_{\ell} \exp(-\ell) \mathcal{N}_{a \rightarrow b}^{(i)}(\ell)$ , with the caveat that shorter pathways might require more ATP than longer ones.  $\mathcal{R}$ -based and  $\mathcal{W}$ -based rankings of metabolic conversions are rather different. They are fully available from <http://chimera.roma1.infn.it/SYSBIO>. To focus on the basics, here we limit ourselves to identifying three key reaction groups that are *independent* of the replaceability measure used:

- (a) the group of reactions such that each substrate-product pair involved in them can be substituted (this is putatively the part of the network that is most robust to enzymopathies);
- (b) the group of reactions that cannot be substituted, corresponding to the most harmful enzymopathies;
- (c) the group of reversible reactions that are only replaceable in one direction, corresponding to the situation in which a conversion  $a \leftrightarrow b$  can only be substituted in one direction in case of a knockout.

All essentiality maps we show relate to this classification. Note that, for topological reasons, intakes are not replaceable.

**3.3. Flux Analysis.** Optimal flux vectors, that is, solutions of (2), are computed by the algorithm introduced in [24] based on [28]. The idea is to modify fluxes iteratively until all inequalities in (2) are satisfied. Specifically, for a fixed  $0 \leq \rho < \rho^*$  (with  $\rho^* = 1$  in our case) define  $\Xi(\rho) = \mathbf{A} - \rho \mathbf{B}$  and let  $\xi^a(\rho)$  denote the rows of  $\Xi(\rho)$ , for  $a \in \{1, \dots, M\}$ . Let also, for each iteration step  $t$ ,  $\mathbf{v}(t)$  be the flux vector at step  $t$  and

$$m(t) = \arg \min_a \xi^a(\rho) \cdot \mathbf{v}(t), \quad (3)$$

At each  $t$ , the algorithm runs as follows. If  $\xi^{m(t)}(\rho) \cdot \mathbf{v}(t) < 0$ , update fluxes according to

$$v_i(t+1) = \max\{0, v_i(t) + \xi_i^{m(t)}(\rho)\}, \quad (4)$$

and iterate in  $t$ . Else, if  $\xi^{m(t)}(\rho) \cdot \mathbf{v}(t) \geq 0$  stop, that is,  $\mathbf{v}(t)$  is a solution.

Convergence to a solution is rigorously ensured for all  $0 \leq \rho < \rho^*$ , and  $\rho^*$  can be approximated with the desired resolution by iterating the above process for increasing values of  $\rho$  [24]. To guarantee that solutions are well defined one can either resort to setting fixed upper bounds on  $v_i$ 's or, as we do, impose a linear constraint of the form  $\sum_i v_i(t) = N$  on the solutions (this is equivalent to singling out one flux as the reference unit for the other fluxes). It is convenient to initialize the algorithm with a random vector  $\mathbf{v}(0)$ . Different initial points generate trajectories to different solutions at  $\rho^*$  and the sampling of the solution space thus obtained turns out to be uniform [8].

Contrary to FBA, the solution space of VN's model is generically not a polytope. Indeed much useful information can be retrieved from its shape. As a means to characterize it we employ the average overlap between different optimal flux vectors, defined as follows. Let  $\mathbf{v}_\alpha$  and  $\mathbf{v}_\beta$  denote two distinct solution vectors of (2) and, for each flux  $i$ , let

$$q_{\alpha\beta}(i) = \frac{2v_{i\alpha}v_{i\beta}}{v_{i\alpha}^2 + v_{i\beta}^2}. \quad (5)$$

This quantity, called the ‘‘overlap’’ between solutions  $\alpha$  and  $\beta$ , equals 1 if flux  $i$  takes on the same value in solutions  $\alpha$  and  $\beta$  and decreases as the values differ more and more. Averaging  $q_{\alpha\beta}(i)$  over different pairs of solutions provides a measure of the allowed variability of flux  $i$  (smaller variability corresponds to larger average overlap), complementary to the standard deviation of the resulting flux distribution. The complexity of the solution space can then be roughly understood by distinguishing narrower directions with larger overlap or less variable fluxes from broader ones. It is reasonable to think that a cell will be more sensitive to perturbations (e.g., knockouts) of fluxes with larger overlap. Analyzing the susceptibility of the solution space to perturbations along the directions identified by different fluxes then allows to extract a list of the potentially more deleterious perturbations, in analogy with previous work on *E.coli* [8].

**3.4. Response to Enzymopathies.** In order to test the hRBC network against enzymopathies, we can focus on two types of perturbations. One can first employ a structural criterion: the knockout of a metabolic conversion  $a \rightarrow b$  that is less easily ‘‘substituted’’ is more likely to be deleterious for the cell than the knockout of a highly replaceable conversion. As said above, we concentrate here on a coarse-grained view of replaceability based on classifying reactions into the groups (a), (b), and (c) defined above, with groups (b) and (c) containing potentially essential reactions. The second criterion is based on fluxes: fluxes with smaller allowed variability (i.e., larger overlap) in the healthy cell are more likely to be essential links of the network than fluxes whose

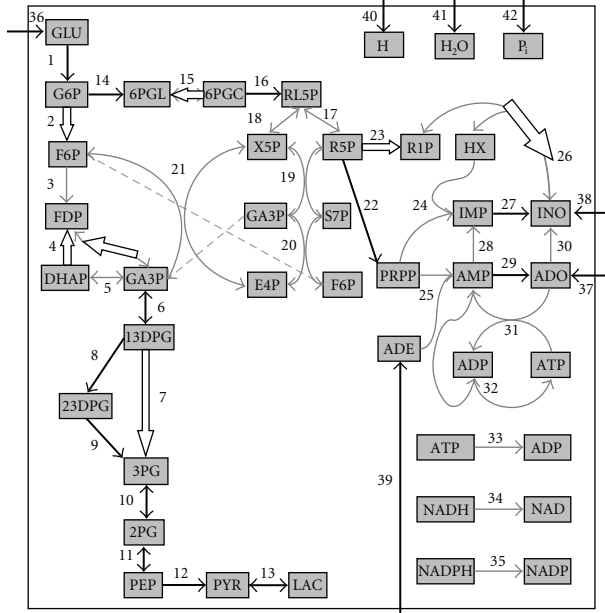


FIGURE 3: hRBC's structural reaction replaceability map. Thick black arrows denote unreplaceable reactions (group (b) above); thick white-filled arrows denote reversible reactions that can be replaced only in the direction indicated by the arrow (group (c)); all other reactions are fully replaceable (group (a)).

value can be changed over a larger range without losing optimality.

As is to be expected, the essentiality maps produced in these ways have a large degree of similarity, and reactions in the group (b) discussed above coincide with the physiologically most critical parts of hRBC's metabolism. The simplest way to simulate an enzymopathy on flux  $i$  is to constrain its value below a certain upper bound  $\bar{v}_i$ . Deficiencies can be partial, that is, of a smaller degree, the closer  $\bar{v}_i$  is to the upper limit of the allowed range in the healthy cell, or total if  $\bar{v}_i = 0$ . Such constraints cause in principle a modification of the solution space along the direction  $i$  which in turn cascades on the entire volume, modifying the optimal states of the metabolic network.

## 4. Results

**4.1. Structural Analysis.** The substitutability map derived from the loop analysis is displayed in Figure 3. (For the sake of simplicity we exclude the highly replaceable currency exchange fluxes from this discussion.) The most replaceable core of the network lies in the PP pathway (reactions 17–21), which constitutes the main source of NADPH, the key metabolite that in erythrocytes limits the accumulation of peroxides protecting the cell from hemolysis. The high reliability coming with replaceability partly explains the reason why this group of reactions plays a central role not just as an auxiliary pathway for glycolysis, see the following analysis of fluxes. Unreplaceable reactions are instead lined up along glycolysis (numbers 1, 6, 8–13), in

the bridge between glycolysis and the PP-pathway (14 and 16) or in auxiliary modules (22, 27, 29; the ADE  $\rightarrow$  AMP conversion in 25 is also not replaceable being directly linked to the ADE uptake). The physiologically most deleterious knockouts (HK, PK, and G6PDH) all belong to this group. For instance, deficiency in the level of G6PDH is the basis of different types of hemolytic anemias, including favism, and is also linked to malaria resistance [29]. Finally, there is a group of reversible reactions (numbers 2, 4, 7, 15, 23, 26) that can be replaced only in one direction. Note however that the last three of these could still be replaced in case of an enzymopathy if a proper medium is selected. For instance, if reaction 15 is removed, it could be substituted by an alternative chain of reactions provided 6PGC is externally supplied. This is instead not possible for reaction 4 and possibly 23 (depending on the directionality of reaction 26), as a knockout in these cases would necessarily result in a net production of FDP and R1P.

**4.2. Flux Analysis.** The flux distribution corresponding to optimal states in the healthy and enzyme deficient hRBC are displayed reaction by reaction in Figure 4, obtained by sampling 10000 solutions of (2), while a pictorial representation of the optimal flux states is given in Figure 5. For the healthy cell (black line in Figure 4 and top left panel in Figure 5) the large flux backbone is formed by the second part of glycolysis (crucial for ATP, NADH, and 23DPG production) and the PP pathway (NADPH production). The latter gives a substantial contribution to the former, not just as salvage way. The adenosine metabolism shows instead lower flux values. In addition to GLU, which is the fundamental substrate for hRBCs, the INO uptake plays an important role as an alternative way to the PP pathway. It is worth stressing that these solutions imply a net production of 23DPG, the crucial regulator for oxygen release, which is obtained without any imposed constraint. This picture is strongly reminiscent of the first eigenpathway obtained by extreme pathways analysis in [19], though the thermodynamic constraints and production requirements used in [19], including one on 23DPG, are more strict than the self-consistent analysis presented here. Comparing the distributions with FBA studies on the same system [16], one notices instead a general rearrangement of fluxes in the network apart from glycolysis. A close inspection reveals that such a rearrangement is mostly quantitative, as preferred reaction directions are generically preserved, the noteworthy exception being the RPI flux, that in the VN solution strengthens the PP pathway with respect to the FBA solution. This scenario is not surprising in view of the basic difference between FBA and the VN approach. It should be kept in mind however that the a priori constraints on flux variability are quite more strict in FBA than they are in the VN model, and the flux distribution appear to be particularly sensitive to the assumed upper and lower bounds for the fluxes.

In Figure 6 we report the overlap map of the hRBC. Comparing this with Figure 3 one sees that the large overlap backbone (signaling dynamically stiff fluxes) coincides to

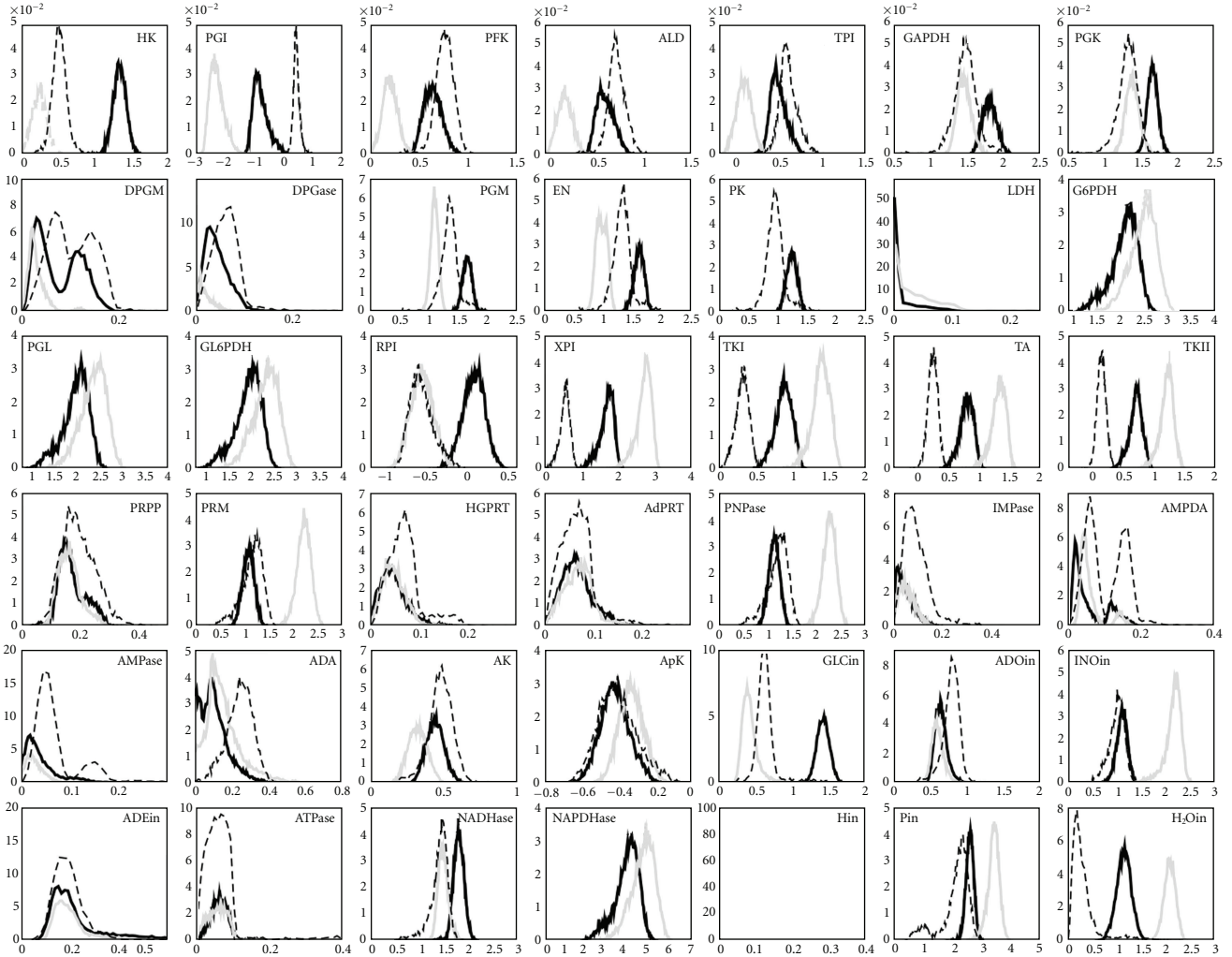


FIGURE 4: Distributions of differential fluxes (measured for each reaction by the difference between the direct and the reverse flux, when possible) for the healthy cell (thick black lines) and for the hRBC with knockout PK (thick gray lines) e G6PDH (dashed lines).

a large degree with the structurally most vulnerable parts of the network. Note that the overlap of reactions 2, 4, 15 and 26 is larger in the direction that cannot be replaced, further pointing to a higher susceptibility, and that currency reactions (31–35) belong to the most constrained part of the network. Revealingly, however, topological and dynamical characterizations prove to be complementary in some cases. This is seen, for example, from reaction 3, which is flux-constrained but also highly replaceable, so that the damage due to removal is limited even in presence of a small allowed dynamical range. (A similar picture holds for reaction 23.)

To conclude, we remind that in our framework uptake fluxes are optimized variables not fixed by boundary conditions. In the optimal state five of the uptakes have a limited allowed variability, implying rather severe constraints on the cell's environment.

**4.3. Response to Enzymopathies.** We have simulated the most studied enzymopathies by constraining the flux of the corresponding reaction. Generically speaking, the hRBC

metabolism displays a large resilience against partial perturbations. Indeed, we have observed appreciable differences in relevant cellular functions compared to the nondeficient case only under full enzyme deficiencies, as also observed in [11] within a standard FBA optimization approach. Even under the most serious enzyme deficiencies the network appears to be able to maintain the production of ATP, NADH and NADPH almost constant, see also [30]. We focus here on PK and G6PDH deletions. As shown in Figure 4, the alterations in the flux distributions are not particularly striking and indeed we do not observe global flux rearrangements on the network's scale. The G6PDH enzymopathy appears to only cause local changes, confirming the structural predictions, the overlap calculations and also in agreement with clinical observations [31]. The response to PK knockout is instead more marked. The synoptic analysis of Figure 5 shows that in general the response to the perturbation consisted in a drop of the GLU uptake, and in a reduction of the glycolytic flux, while the Rapoport-Leubering shunt (reactions 8-9) for the production of 23DPG remains

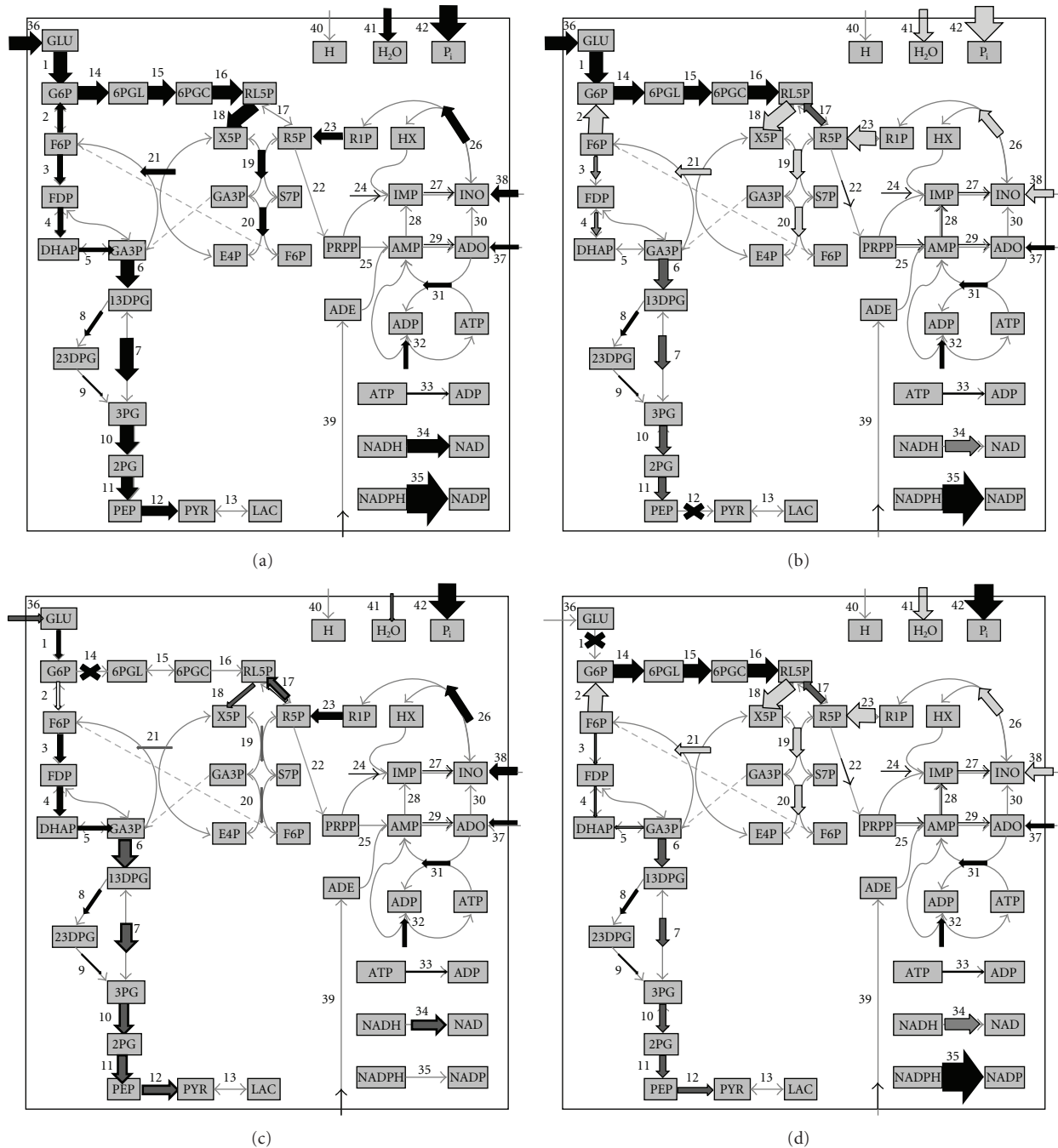


FIGURE 5: Maps of fluxes in the healthy hRBC (a) and flux rerouting in three full enzymopathies: PK (b), G6PDH (c) and HK (d). The thick arrows are dark grey if the reaction rates diminish with respect to the healthy cell, light grey if they are higher, black if they remains the same. The widths of the arrows are proportional to the mean value of the corresponding distributions.

particularly stable, as does the adenosine metabolism. For the glycolytic deficiencies PK and HK we further observe an increase of the INO uptake to sustain the PP pathway and allow for the second part of glycolysis, and with it the production of ATP and NADH, to take place. Detailed flux configurations corresponding to the next most severe enzymopathies (HK, EN, PGK and PGM) are available from <http://chimera.roma1.infn.it/SYSBIO>.

## 5. Final Remarks

In this work we compare two robustness measures for biochemical networks, one based on structural properties (the reaction replaceability), the other based on dynamical stiffness (the overlaps). The former can be exactly assessed by enumerating the alternative paths joining substrates and products of a given reaction in a network. The latter depends



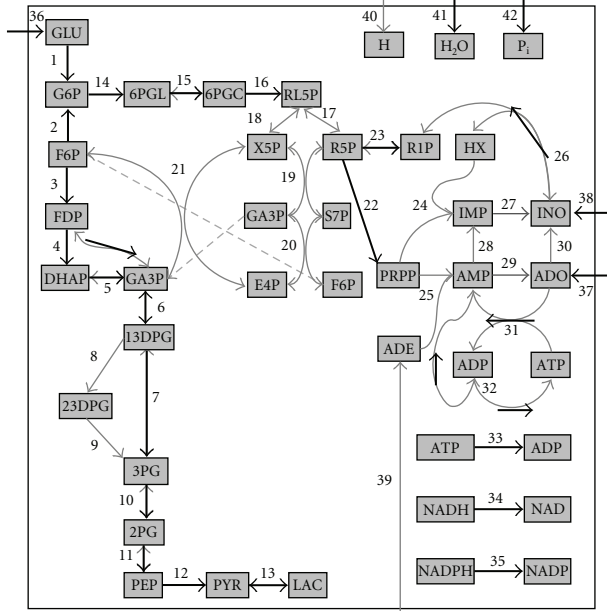


FIGURE 6: Large overlap backbone of hRBC: thick black arrows mark reactions with an overlap larger than 0.9 in an optimal healthy cell.

on both the network topology and the model defined on it. Within VN's frame, we found that unreplaceable reactions mostly correspond to processes with a smaller allowed flux variability. In such directions, reaction removals as well as constraints on the fluxes are expected to be generically harmful. Reactions with limited (but non zero) replaceability tend to have instead smaller overlap, so that while the reaction is difficult to substitute still its flux can be largely adjusted. In an evolutionary perspective [32], the former pathways appear as "frozen", and perturbations at these nodes will require large-scale flux rearrangement, while a mutation affecting the latter group may be neutral and could be preserved across generations. Interestingly, some reactions have both a large overlap and a large replaceability. These, albeit structurally robust, are dynamically constrained and should be considered as essential pathways of the metabolism as well. Integrating dynamical and structural characterizations may thus provide a rather complete picture of the emerging network robustness. The fact that topological and dynamical essentiality may not coincide could also prove to be important in view of the present challenges to understand the dynamical basis of topological modularity [33].

Extended flux state sampling was achieved here by an algorithm that is easily scalable to larger networks, see [8]. Exhaustive structural analysis instead was made possible by the small size of hRBC's metabolic network, which has served here as a model system to test basic concepts and algorithms. The use of the same procedure on a larger network, such as *E. coli*'s, is likely to be prevented by CPU time growth. However, message-passing algorithms, designed specifically to solve combinatorial optimization or counting problems—albeit approximately—on graphical models, may be a suitable replacement [34].

## Acknowledgment

The authors would like to express their pleasure to thank S. Jain for pointing them to [7] and D. Segrè for important comments and suggestions. This work was supported by the IIT (Italian Institute of Technology) through the Seed Project DREAM.

## References

- [1] K. Kobayashi, S. D. Ehrlich, A. Albertini et al., "Essential *Bacillus subtilis* genes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 8, pp. 4678–4683, 2003.
- [2] C. M. Sasseti, D. H. Boyd, and E. J. Rubin, "Comprehensive identification of conditionally essential genes in mycobacteria," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 22, pp. 12712–12717, 2001.
- [3] B. J. Akerley, E. J. Rubin, V. L. Novick, K. Amaya, N. Judson, and J. J. Mekalanos, "A genome-scale analysis for identification of genes required for growth or survival of *Haemophilus influenzae*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 2, pp. 966–971, 2002.
- [4] S. Y. Gerdes, M. D. Scholle, J. W. Campbell et al., "Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655," *Journal of Bacteriology*, vol. 185, no. 19, pp. 5673–5684, 2003.
- [5] R. S. Kamath, A. G. Fraser, Y. Dong et al., "Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi," *Nature*, vol. 421, no. 6920, pp. 231–237, 2003.
- [6] R. Mahadevan and B. Ø. Palsson, "Properties of metabolic networks: structure versus function," *Biophysical Journal*, vol. 88, no. 1, pp. L07–L09, 2005.
- [7] A. Samal, S. Singh, V. Giri, S. Krishna, N. Raghuram, and S. Jain, "Low degree metabolites explain essential reactions and enhance modularity in biological networks," *BMC Bioinformatics*, vol. 7, article 118, 2006.
- [8] C. Martelli, A. De Martino, E. Marinari, M. Marsili, and I. P. Castillo, "Identifying essential genes in *Escherichia coli* from a metabolic optimization principle," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 8, pp. 2607–2611, 2009.
- [9] T. A. Rapoport, R. Heinrich, and S. M. Rapoport, "The regulatory principles of glycolysis in erythrocytes in vivo and in vitro. A minimal comprehensive model describing steady states, quasi-steady states and time-dependent processes," *Biochemical Journal*, vol. 154, no. 2, pp. 449–469, 1976.
- [10] H.-G. Holzhutter, G. Jacobasch, and A. Bisdorff, "Mathematical modelling of metabolic pathways affected by an enzyme deficiency. A mathematical model of glycolysis in normal and pyruvate-kinase-deficient red blood cells," *European Journal of Biochemistry*, vol. 149, no. 1, pp. 101–111, 1985.
- [11] N. Jamshidi and B. Ø. Palsson, "Systems biology of the human red blood cell," *Blood Cells, Molecules, and Diseases*, vol. 36, no. 2, pp. 239–247, 2006.
- [12] D. B. Johnson, "Finding all the elementary circuits of a directed graph," *SIAM Journal on Computing*, vol. 4, p. 77, 1975.
- [13] S. J. Wiback, I. Famili, H. J. Greenberg, and B. Ø. Palsson, "Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space," *Journal of Theoretical Biology*, vol. 228, no. 4, pp. 437–447, 2004.

- [14] R. U. Ibarra, J. S. Edwards, and B. Ø. Palsson, "Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth," *Nature*, vol. 420, no. 6912, pp. 186–189, 2002.
- [15] D. Segrè, D. Vitkup, and G. M. Church, "Analysis of optimality in natural and perturbed metabolic networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 23, pp. 15112–15117, 2002.
- [16] N. D. Price, J. Schellenberger, and B. Ø. Palsson, "Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies," *Biophysical Journal*, vol. 87, no. 4, pp. 2172–2186, 2004.
- [17] A. Braunstein, R. Mulet, and A. Pagnani, "Estimating the size of the solution space of metabolic networks," *BMC Bioinformatics*, vol. 9, article 240, 2008.
- [18] S. J. Wiback and B. Ø. Palsson, "Extreme pathway analysis of human red blood cell metabolism," *Biophysical Journal*, vol. 83, no. 2, pp. 808–818, 2002.
- [19] N. D. Price, J. L. Reed, J. A. Papin, S. J. Wiback, and B. Ø. Palsson, "Network-based analysis of metabolic regulation in the human red blood cell," *Journal of Theoretical Biology*, vol. 225, no. 2, pp. 185–194, 2003.
- [20] C. L. Barrett, N. D. Price, and B. Ø. Palsson, "Network-level analysis of metabolic regulation in the human red blood cell using random sampling and singular value decomposition," *BMC Bioinformatics*, vol. 7, article 132, 2006.
- [21] K. J. Kauffman, J. D. Pajeroski, N. Jamshidi, B. Ø. Palsson, and J. S. Edwards, "Description and analysis of metabolic connectivity and dynamics in the human red blood cell," *Biophysical Journal*, vol. 83, no. 2, pp. 646–662, 2002.
- [22] K. J. Kauffman, P. Prakash, and J. S. Edwards, "Advances in flux balance analysis," *Current Opinion in Biotechnology*, vol. 14, no. 5, pp. 491–496, 2003.
- [23] A. De Martino and M. Marsili, "Typical properties of optimal growth in the von Neumann expanding model for large random economies," *Journal of Statistical Mechanics*, no. 9, pp. 19–27, 2005.
- [24] A. De Martino, C. Martelli, R. Monasson, and I. Pérez Castillo, "Von Neumann's expanding model on random graphs," *Journal of Statistical Mechanics*, no. 5, Article ID P05012, 2007.
- [25] I. Famili and B. Ø. Palsson, "The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools," *Biophysical Journal*, vol. 85, no. 1, pp. 16–26, 2003.
- [26] A. De Martino, C. Martelli, and F. A. Massucci, "On the role of conserved moieties in shaping the robustness and production capabilities of reaction networks," *Europhysics Letters*, vol. 85, no. 3, Article ID 38007, 2009.
- [27] R. Tarjan, "Depth-first search and linear graph algorithms," *SIAM Journal on Computing*, vol. 1, no. 2, pp. 146–160, 1972.
- [28] W. Krauth and M. Mezard, "Learning algorithms with optimal stability in neural networks," *Journal of Physics*, vol. 20, no. 11, pp. L745–L752, 1987.
- [29] K. B. Storey, Ed., *Functional Metabolism*, Wiley-Liss, Nashville, Tenn, USA, 2004.
- [30] S. Durmuş Tekir, T. Çakir, and K. Ö. Ülgen, "Analysis of enzymopathies in the human red blood cells by constraint-based stoichiometric modeling approaches," *Computational Biology and Chemistry*, vol. 30, no. 5, pp. 327–338, 2006.
- [31] G. Jacobasch, "Biochemical and genetic basis of red cell enzyme deficiencies," *Baillière's Clinical Haematology*, vol. 13, no. 1, pp. 1–20, 2000.
- [32] D. Vitkup, P. Kharchenko, and A. Wagner, "Influence of metabolic network structure and function on enzyme evolution," *Genome Biology*, vol. 7, no. 5, article R39, 2006.
- [33] R. P. Alexander, P. M. Kim, T. Emonet, and M. B. Gerstein, "Understanding modularity in molecular networks requires dynamics," *Science signaling*, vol. 2, no. 81, p. pe44, 2009.
- [34] V. Van Kerrebroeck and E. Marinari, "Ranking vertices or edges of a network by loops: a new approach," *Physical Review Letters*, vol. 101, no. 9, Article ID 098701, 2008.