Available online at www.sciencedirect.com
SCIENCE DIRECT®
Metabolic Engineering 5 (2003) 264–276
METABOLIC ENGINEERING
http://www.elsevier.com/locate/ymben

The effects of alternate optimal solutions in constraint-based genome-scale metabolic models ☆

R. Mahadevan* and C.H. Schilling

Genomatica, Inc., Bioprocessing Division, 5405 Morehouse Drive, Suite 210, San Diego, CA, 92121, USA

Received 12 June 2003; received in revised form 19 September 2003; accepted 25 September 2003

Abstract

Genome-scale constraint-based models of several organisms have now been constructed and are being used for model driven research. A key issue that may arise in the use of such models is the existence of alternate optimal solutions wherein the same maximal objective (e.g., growth rate) can be achieved through different flux distributions. Herein, we investigate the effects that alternate optimal solutions may have on the predicted range of flux values calculated using currently practiced linear (LP) and quadratic programming (QP) methods. An efficient LP-based strategy is described to calculate the range of flux variability that can be present in order to achieve optimal as well as suboptimal objective states. Sample results are provided for growth predictions of *E. coli* using glucose, acetate, and lactate as carbon substrates. These results demonstrate the extent of flux variability to be highly dependent on environmental conditions and network composition. In addition we examined the impact of alternate optima for growth under gene knockout conditions as calculated using QP-based methods. It was observed that calculations using QP-based methods can show significant variation in growth rate if the flux variability among alternate optima is high. The underlying biological significance and general source of such flux variability is further investigated through the identification of redundancies in the network (equivalent reaction sets) that lead to alternate solutions. Collectively, these results illustrate the variability inherent in metabolic flux distributions and the possible implications of this heterogeneity for constraint-based modeling approaches. These methods also provide an efficient and robust method to calculate the range of flux distributions that can be derived from quantitative fermentation data.

Keywords: In silico model; *Escherichia coli*; Alternate optima; Spectrum analysis; Flux variability

1. Introduction

Development of high throughput technologies for probing various biological processes in an organism at the gene, protein, and metabolite levels has resulted in the generation of large amounts of information. This influx of information has motivated the development of several quantitative approaches to analyze biological systems and interpret the large-scale data sets. The importance of such computational approaches for enhancing the understanding of biological systems and for the generation of experimentally verifiably hypotheses has been recognized (Endy and Brent, 2001; Kitano, 2002). One of the well studied areas is the analysis of metabolic networks (Reich and Selkov, 1981; Fell, 1996; Heinrich and Schuster, 1996; Stephanopoulos et al., 1998; Varner and Ramkrishna, 1999; Fell, 1996). Among the quantitative approaches that exist for the analysis of metabolic networks, constraint-based modeling has attracted attention due to its ability to analyze genome-scale metabolic networks while using very few model parameters (Bailey, 2001; Palsson, 2000).

Constraint-based modeling involves the application of a series of constraints arising from the consideration of stoichiometry, thermodynamics, flux capacity, and regulatory restraints under which reactions operate in a metabolic network (Varma and Palsson, 1994; Price et al., 2003; Edwards et al., 2002; Bonarius et al., 1997). These constraints serve to limit the range of attainable flux distributions or metabolic phenotypes that can be achieved by an organism. Applying mass balancing

☆ Supplementary data associated with this article can be found, in the online version

* Corresponding author.
E-mail address: rmahadevan@genomatica.com (R. Mahadevan).

1096-7176/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymben.2003.09.002

approaches to a metabolic network leads to a convex mathematical representation of linear equations and inequalities that define an underdetermined system. One approach that is commonly used to explore the metabolic capabilities defined by these constraints is linear programming (LP). In this approach suitable metabolic objectives are posed for the system to maximize or minimize, yielding optimal flux distributions. Typical examples of objective functions include maximization of ATP synthesis, minimization of substrate utilization, and maximization of growth rate.

Linear programming problems with growth rate maximization are commonly used to investigate the metabolic network of microorganisms (Edwards and Palsson, 2000; Edwards and Palsson, 1999; Schilling et al., 2002). Recently, the ability of the constraint-based models to predict the growth and by-product secretion patterns on various substrates for *E. coli* has been experimentally validated (Edwards et al., 2001), along with the ability to predict the outcome of adaptive evolution of suboptimal laboratory strains (Ibarra et al., 2002). In addition to the use of LP and other related approaches based on convexity principles (Schilling et al., 1999), the constraint-based modeling framework has been extended to include energy balance constraints (Beard et al., 2002) and dynamic rate of change of flux constraints (Mahadevan et al., 2002). Recent advances in the constraint-based approach have seen alternative optimization procedures developed to enable the exploration of an extended range of metabolic function including the use of mixed integer linear programming (MILP) (Burgard et al., 2001; Lee et al., 2000; Hatzimanikatis et al., 1996). Recently a QP-based approach (termed minimization of metabolic adjustment—MOMA) was developed to examine the growth characteristics of mutant strains, in which specific genes have been deleted (Segre et al., 2002).

In every approach that is used to explore the solution space defined by the variety of metabolic constraints a flux distribution is determined that describes the flow of mass and energy through the set of reactions comprising the network. While these flux distributions are at times unique, in the majority of cases they are non-unique particularly with reference to how the network can achieve various objectives. For any given optimal flux distribution there may exist alternate optimal solutions that characterize a region in the flux space that generates the same objective value (e.g., growth rate) via different flux distribution patterns. These represent multiple solutions that are feasible based on the LP formulation and could represent biologically meaningful solutions. The source of these alternate optima and non-unique solutions lies in the biological design of metabolism and the inherent redundancies built into the reaction network. Very few publications with the exception of Lee et al. (Phalakornkule et al., 2001; Lee et al., 2000) have looked at the characterization and significance of alternate optimal flux distributions in metabolic systems.

In this manuscript we address the topic of redundancy and flux variability in metabolic systems. We focus on the impact of such redundancies on current LP and QP solution methods used within the constraint-based approach to modeling. We analyze the effect of the alternate optimal solutions on the predictions of the mutant growth rates made using the QP-based analysis, where a reference flux is chosen based on a previous LP solution. In addition we characterize the range and variability of flux values under conditions (growth on glucose, lactate, and acetate) where alternate optimal solutions exist using a genome-scale model of *E. coli* as our example system, and assess the underlying biological significance. Furthermore we present an algorithm for the identification of redundancies in metabolic networks that give rise to the existence of alternate optimal solutions. These efforts demonstrate the biological source of variability inherent in metabolic flux distributions and the modeling considerations that must be made due to such functional heterogeneity.

2. Constraint-based modeling

2.1. Linear programming based analysis

A metabolic network is constrained by the imposition of stoichiometric constraints that correspond to the mass balance around each metabolite. When the metabolic network has a steady state distribution of fluxes the constraints are described by a system of linear equations as shown below:

$$S \cdot v = 0, \tag{1}$$

where S is the $m \times n$ stoichiometric matrix of all the reactions in the metabolic network, m is the number of metabolites, n is the number of fluxes (reaction rates), and v is the flux vector of the metabolic network. In addition to these linear constraints there exist thermodynamic constraints on reactions that may restrict the directional flow of the reaction along with capacity constraints that provide potential upper limits to the flux levels of the reactions. A typical objective function used for the simulation of microbial genome-scale metabolic networks has been the growth rate as represented by a reaction draining biomass components. The LP problem that is solved is summarized below:

$$\begin{aligned} &\text{Max } f^T v, \\ &\text{s.t. } S.v = 0, \\ &\quad 0 \leqslant v \leqslant v_{\max}, \end{aligned} \tag{2}$$

where f is the objective function vector and v_{max} is the vector containing the maximum capacities of the fluxes. Maximum measured uptake rates are used to constrain the exchange fluxes.

2.2. Quadratic programming based analysis

In addition to LP it is possible to use QP-based approaches to examine the ability of the metabolic network to minimize the extent of a perturbation to the metabolic network that may result from a change in environmental and/or underlying genetic conditions. The rationale behind this approach is that the mutant strain, in the absence of sufficient evolutionary pressure, will exhibit sub-optimal growth and that the flux distribution in the mutant strain will be "closest" to the wild type flux distribution (in terms of the Euclidean distance metric) in the flux space. In other words, the organism will attempt to absorb the genetic loss with as little disturbance to its metabolic flux distribution as possible. The mathematical formulation for this approach results in a quadratic programming problem that is summarized below:

$$\begin{aligned} &\text{Min} \quad (v-w)^T(v-w), \\ &\text{s.t.} \quad S.v = 0, \\ &\qquad 0 \leqslant v \leqslant v_{max} \\ &\qquad v_d = 0, \end{aligned} \tag{3}$$

where w is the base/wild type flux distribution (typically determined via LP-based analysis) and v_d is the vector of constraints on the flux through the reactions that are associated with the deleted gene(s). It has been suggested that this QP-based approach predicts the initial growth characteristics of mutant strains more accurately than using LP-based approaches (Segre et al., 2002).

2.3. Alternate optimal solutions

An issue that may be encountered in any constraint-based optimization analysis is the possible existence of alternate optimal solutions. These alternate optima represent situations wherein a different set of reactions or flux distributions can be used by the system to reach the exact same quantitative objective value. If alternate optima exist then the solution to any LP problem will not be unique. The impact of this situation on the biological conclusions drawn from a simulated flux distribution could range from being negligible to being highly significant. Therefore, awareness of the possible existence of these alternate optimal solutions is important for any constraint-based analysis, although the impact will be variable. We next explore the impact of alternate optimal solutions on flux distributions calculated in a genome-scale metabolic model of *E. coli*. In particular we focus on examining the flux variability that results due to alternate optimal solutions, and explore the impact of such variability on both LP and QP constraint-based analyses.

3. Calculation of flux variability due to alternate optima

LP problems can have multiple solutions that have the *exact* same optimal value for the objective function and satisfy all of the constraints. In order to investigate the effects of these alternate optima, approaches must be developed to calculate the multiplicity of solutions. Previously, a study of the multiple flux distributions that satisfy the stoichiometric and capacity constraints and have the same objective function has been presented in Lee et al. (2000). In that study, all the alternate optimal solutions were identified for a metabolic network of 33 reactions and 30 metabolites. An MILP formulation with at least 60 variables was used to determine all the alternate optimal solutions. Recent studies using the MILP approach for the genome scale model of *E. coli* have indicated that the number of optimal solutions could be large (on the order of thousands of solutions) and the MILP approach could be computationally intractable for certain conditions (Reed et al., 2003, Personal Communication). In general, an MILP algorithm can be computationally expensive and intractable for a genome-scale network due to the exponential number of extreme points that might exist if there are several redundant pathways in the metabolic network. As a result we have developed an alternative strategy to investigate the issue of alternate optimal solutions and their biological significance.

The focus of the current study is the characterization of alternate solutions within the *E. coli* genome-scale network consisting of 953 fluxes and 536 metabolites (Edwards and Palsson, 2000). In this LP-based approach the emphasis is on determining the maximum and minimum values of all the fluxes that will satisfy the constraints and allow for the same optimal objective value. It must be noted that this approach does not identify all possible alternate optimal solutions, but rather the range of flux variability that is possible within any given solution.

The approach begins with determining the base/wild-type value of the objective function by solving the LP problem outlined in Eq. (2). From this solution the range of variability that can exist in each flux in the network due to alternate optimal solutions can be calculated through a series of LP problems wherein the value of the original objective is fixed and each reaction in the network is maximized and subsequently minimized to determine the feasible range of flux values for each reaction. Similar analysis has been used to identify bounds on the fluxes to further constrain the flux space for the identification of minimal reaction sets (Burgard

et al., 2001). The mathematical formulation of this approach is described below:

Case 1:

$$\begin{aligned} &\text{Max } v_i \\ &\text{s.t. } S.v = 0 \\ &\quad f^T v = Z_{\text{obj}} \\ &\quad 0 \leqslant v \leqslant v_{\max} \quad \text{for } i = 1 \ldots n, \end{aligned} \tag{4a}$$

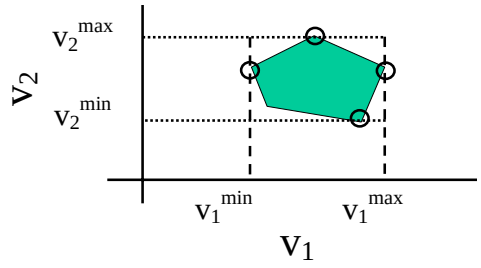

Fig. 1. A schematic describing the optimal solution space. The shaded area represents the region where the objective function can take on the same maximum value, and the circles represent the solutions that can be identified using the approach proposed in the paper. Note that not all the optimal solutions can be identified. However, the maximum and minimum values of the flux that allow the optimal objective function value can be determined.

Case 2:

$$\begin{aligned} &\text{Min } v_i, \\ &\text{s.t. } S.v = 0, \\ &\quad f^T v = Z_{\text{obj}}, \\ &\quad 0 \leqslant v \leqslant v_{\max} \quad \text{for } i = 1 \ldots n, \end{aligned} \tag{4b}$$

where Z_{obj} is the value of the objective function calculated previously from Eq. (2), and n is the number of fluxes. The solution of the $2n$ LP problems outlined in Eq. (4) determines the upper and lower bounds of every reaction flux that will result in the same value for the original objective function. Once again we note that this approach provides only the bounds on all solutions, as opposed to providing all the possible alternate optimal solutions (Fig. 1). All the algorithms were formulated in MATLAB (Natick, MA) and the LP and QP problems were solved using the LINDO API interface (Lindo Systems Inc. IL).

This approach is illustrated for the example network shown in Fig. 2a and b (Schilling et al., 2000). For this case, a sample flux distribution that maximizes the growth rate in the presence of 4 units of A resulting in one unit of biomass, is assumed to be the reference flux distribution for the analysis of flux variability among alternate optima. Note that for this sample network, there are two futile cycles (Fig. 2c) that would be

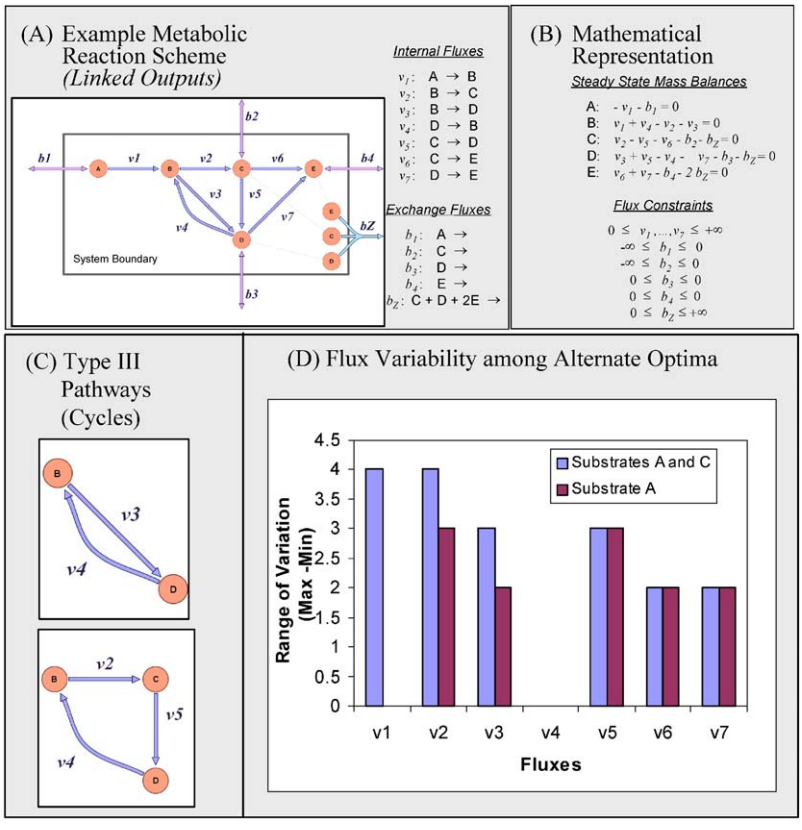

Fig. 2. The example metabolic network. (a) The metabolic network and the associated reactions. (b) Mathematical formulation of the metabolic network. (c) Type III pathways or cycles as defined in Schilling et al. (2000). (d) Flux variability among alternate optima for the example network.

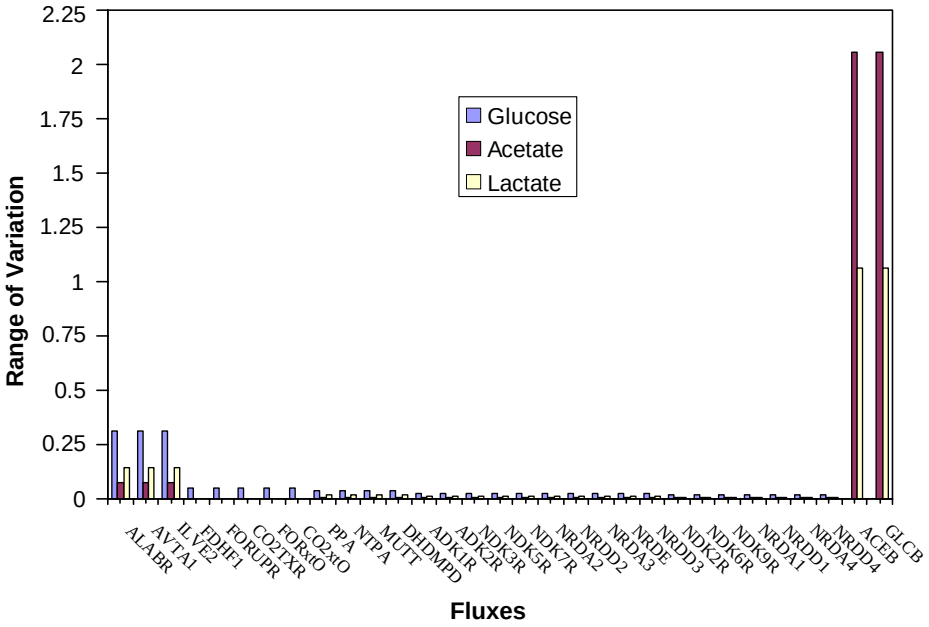

Fig. 3. The absolute range of variation (maximum–minimum) of all the fluxes that show variation in the alternate optima for growth on glucose, acetate and lactate. The rest of the fluxes have the same value for all the alternate optima. The reaction names are consistent with the description in (Edwards and Palsson, 2000).

infeasible from thermodynamic considerations (Beard et al., 2002; Price et al., 2002a). These cycles are eliminated from the network by setting the flux through v_4 in the network to zero. The presence of these cycles can cause the objective function for certain fluxes (v_4) to have an infinite range of variation.

To illustrate the basic concepts of the approach described in Eq. (4) we analyze the flux variability for this example under different environments (i.e., when A alone is present and when both A and C are present) with the results are depicted in Fig. 2d. For this example, when both A and C are allowed to be taken up (for biomass flux of one unit), all the fluxes except v_4 can vary and the range of the variation for each flux can be calculated. It can be seen that fluxes v_6 and v_7 show less variation relative to the fluxes v_1 and v_2. When only A is allowed to be taken up (biomass flux of one unit), the flux v_1 does not vary and the range of flux variation for the fluxes v_2, v_3, is reduced relative to the previous case, where as the remaining fluxes show the same variability.

As further demonstration we have applied this algorithm to separately analyze the flux variation for aerobic growth on glucose, lactate, and acetate in the *E. coli* model. The maximum allowed uptake rates for glucose and oxygen chosen for the analysis are 10 and 15 mmol/gdw h^{-1}, respectively. The maximum and minimum values for every flux are calculated using Eq. (4). The futile cycles (Type III pathways) are identified using the extreme pathway algorithm applied to a reaction network that only contains the fluxes that have infinite values when maximized/minimized using the algorithm of Eq. (4). Isozymes are eliminated from the model by inspection to ensure that the maximum and minimum values for all the fluxes are finite. It is observed that for growth on glucose, 29 of the 953 (3.0%) reactions show flux variations (greater than 1e–5 mmol/gdw h^{-1}) in the alternate optima and the remaining reactions have the same flux value for all the alternate optima. We also considered growth on acetate and lactate and calculate 19 (2.0%) and 28 (2.9%) fluxes that vary, respectively. Fig. 3 depicts the variation in the fluxes for these reactions for growth on glucose, lactate and acetate (uptake rates, 10, 10, and 10 mmol/gdw h^{-1}, respectively). It can be seen that the set of fluxes that vary among the three substrates are different, for example, the reactions corresponding to the glyoxalate shunt show variation only for growth on acetate and lactate, and do not vary for growth on glucose.

The fluxes that had different values in the alternate flux distributions can be readily identified for further investigation to generate equivalent reaction sets (redundant pathways) from these fluxes. From the analysis of the example network shown in Fig. 2a, it is clear that the redundancies within the subset of fluxes that had a non-zero range of variation results in the alternate optimal solutions.

4. Effect of flux variability on QP-based analysis

The existence of alternate optimal solutions will have an effect on QP-based analyses. In QP-based

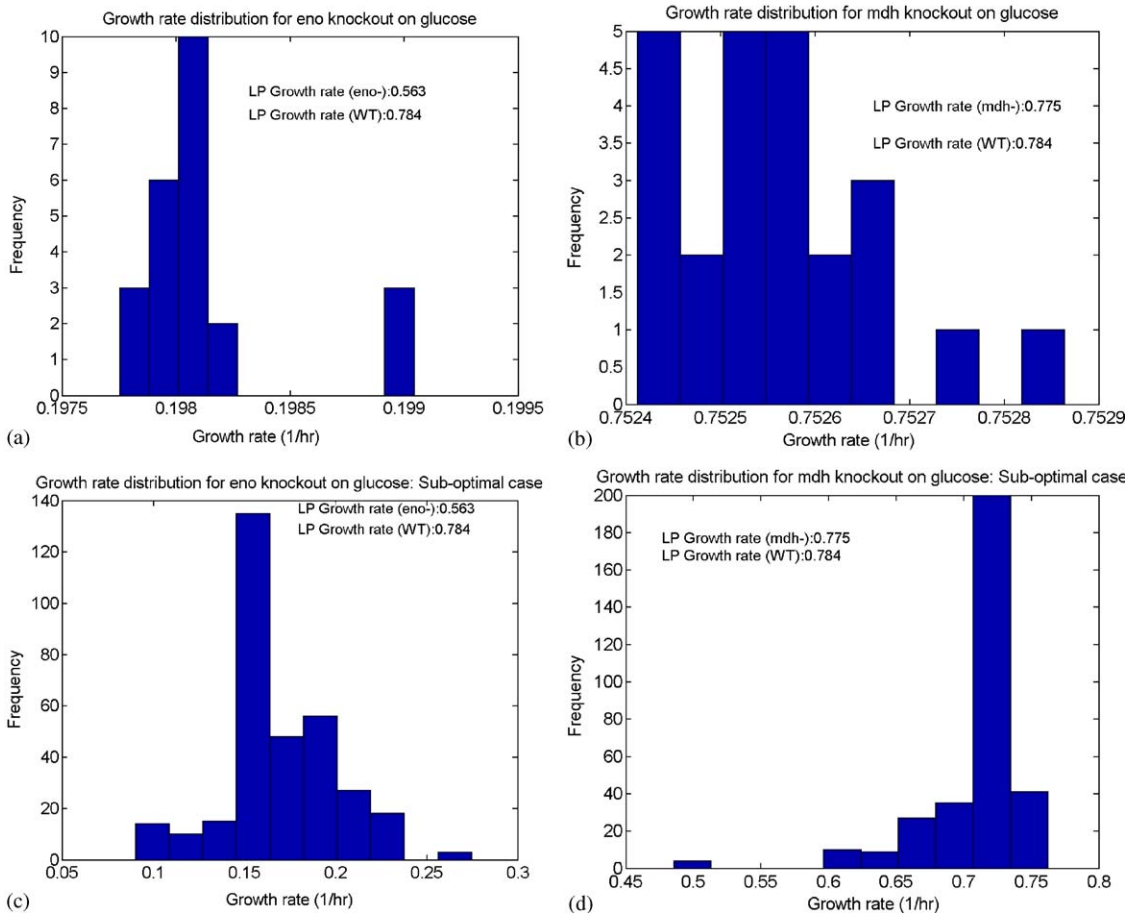

Fig. 4. Histogram of the knockout growth rates calculated using the QP-based analysis for growth on glucose with (a) the optimal flux distributions (the alternate optimal solutions) as the reference for the *eno* knockout; (b) the optimal flux distributions (the alternate optimal solutions) as the reference for the *mdh* knockout; (c) with the suboptimal flux distributions as the reference for the *eno* knockout; (d) with the suboptimal flux distributions as the reference for the *mdh* knockout. The variation in the mutant growth rates is greater for the case where the suboptimal flux distributions are used as the reference due to increased flux variability among the suboptimal solutions.

approaches a reference flux distribution is selected that is often derived from an LP optimization problem. The selection of this reference flux distribution can be complicated by the presence of alternate optima as discussed in the previous section. Thus while the flux distribution calculated in the QP-based approach is unique, the reference state may not be unique leading to a different result based on the selection of the reference flux distribution. Below we perform an initial investigation of the effects of the non-uniqueness of the reference solution on the location of the projection.

The first step in assessing the impact of alternate optima on QP-based approaches is the identification of distinct alternate optima from the set of optimal flux distributions with maximum flux variability identified previously. This subset corresponds to flux distributions that contain reaction fluxes at their maximum/minimum value and have the optimal value of the objective function and are distinct as some of the $2n$ solutions identified in Eq. (4) can have the same optimal flux distribution. Twenty-four distinct alternate optima were identified for the growth on glucose for *E. coli*. It is important to note that these distinct solutions are a subset of all the possible alternate optima. These 24 distinct flux distributions were used as the reference flux distributions in the calculations using the QP-based approach. The growth rates for specific single gene knockouts (*eno* or *mdh*) were calculated for all the distinct optima, and a histogram of the frequency of the growth rates on glucose was determined for both cases. This histogram is depicted in Fig. 4a and b. It is seen that the effect of the variation of wild type flux distribution has little effect on the predicted mutant growth rate on glucose, as the variation in growth rate ranges from 0.1978 to 0.1990 h^{-1} (0.65% change) for the *eno* knockout and the variation in growth rate ranges from 0.7524 to 0.7529 h^{-1} (0.06% change) for the *mdh* knockout.

5. Effect of sub-optimal flux distributions on LP & QP-based analysis

To this point, flux variation has been characterized under the assumption that the optimal solution is attained by the metabolic network. However, metabolic systems may not necessarily operate at the full optimal state, but may be operating in near optimal or sub-optimal modes. Unlike optimization of engineering processes where the tolerances in the value of the objective function are small, biological systems can be expected to have tolerances that are larger and thus the objective values can have a greater range of variability around the optimal solution. Hence, it is important to consider situations where the metabolic network can operate in a suboptimal fashion. Sub-optimal solutions can be calculated that constitute flux distributions wherein the value of the objective function is marginally below the optimal value. There are several instances where sub-optimal solutions are of interest, for example, in cases where the cell has not been evolved on a specific substrate (Ibarra et al., 2002; Fong et al., 2003). Furthermore, the characterization of near optimal or suboptimal solutions is important for comparisons with experimental flux data as well as online fermentation measurements of uptake/secretion rates wherein measurement error must be considered.

For investigating the effect of sub-optimal solutions, the optimality condition is relaxed for calculating the maximum and minimum values of all the fluxes. The variation in the flux distribution for sub-optimal growth was also investigated as a part of the current study. The objective was constrained to be above 95% of the optimal growth rate (Eq. (5)) and the solutions that had the maximum and minimum values for all the fluxes were identified for growth on glucose. The formulation for this case is shown below:

Case 1:

$$\begin{aligned} &\text{Max } v_i \\ &\text{s.t. } S.v = 0 \\ &\quad f^T v \geqslant 0.95\ Z_{\text{obj}} \\ &\quad 0 \leqslant v \leqslant v_{\max} \quad \text{for } i = 1 \ldots n, \end{aligned} \tag{5a}$$

Case 2:

$$\begin{aligned} &\text{Min } v_i, \\ &\text{s.t. } S.v = 0, \\ &\quad f^T v \geqslant 0.95\ Z_{\text{obj}}, \\ &\quad 0 \leqslant v \leqslant v_{\max} \quad \text{for } i = 1 \ldots n. \end{aligned} \tag{5b}$$

This analysis resulted in the identification of 473 (49.63% of the total) fluxes that varied in the case of glucose and 470 (49.32%) fluxes in the case of lactate and 468 (49.11% of the total) fluxes in the case of acetate. Thus, the number of fluxes that show variation for the sub-optimal cases increases several fold ($\sim$17 fold) compared to the optimal case for glucose, lactate, and acetate. The set of distinct solutions (326 solutions) is identified from all the solutions resulting from Eq. (5) and QP-based analysis is used to calculate the growth rate with the distinct sub-optimal solutions as the reference flux distribution for the *eno* and *mdh* knockout as presented in Segre et al. (2002). A similar less formal analysis evaluating the effect of the suboptimal solutions on the QP-based analysis is presented in the supplementary information in Segre et al. (2002). However, the range of the predicted mutant growth rate in that preliminary analysis was found to be less significant. The results of the distribution of the mutant growth rates on glucose for the distinct suboptimal solutions are shown in Fig. 4c and d. These results indicate that there is considerable variability in the flux distributions that have a suboptimal objective within 5% of the wild type value. This variability significantly affects the growth rates for mutant phenotypes based on the suboptimal flux distributions as seen from Fig. 4c and d. Similar results were obtained for the *pykA*, *pykF* (pyruvate kinase) and the *ppc* (phosphoenolpyruvate carboxylase) mutants when growth of the mutants on glucose were compared for the optimal and the suboptimal cases (results are shown in the supplementary information at the following website, http://www.genomatica.com/papers/AltOpt/SuppI.pdf).

The analysis of the flux variation in alternate optima for the case of growth on alternate substrates such as acetate and lactate is depicted in Fig. 3. The effect of flux variation in the alternate optima on the QP-based analysis was investigated in the same fashion as described for the growth on glucose. The histogram depicting the variation in the predicted growth rates on lactate for the *sucA* and *mdh* mutants due to alternate optima is shown in Fig. 5a and b. This analysis was also carried with the suboptimal solutions as the reference. The histogram of the mutant growth rates with the suboptimal solutions as the reference is shown in Fig. 5c and d. It can be seen that for both the substrates, the flux variability of the sub-optimal case leads to a very broad distribution of the mutant growth rates.

6. Identification of equivalent pathways

Often times there are many different routes through a metabolic network that can provide the same net stoichiometric conversion of substrates into products (Price et al., 2002b). The existence of alternate optimal solutions motivates the identification of these different routes that may exist in a network. These different pathways can be termed as equivalent pathways and are a condition-independent or structural property of the network. The complete set of these equivalent pathways

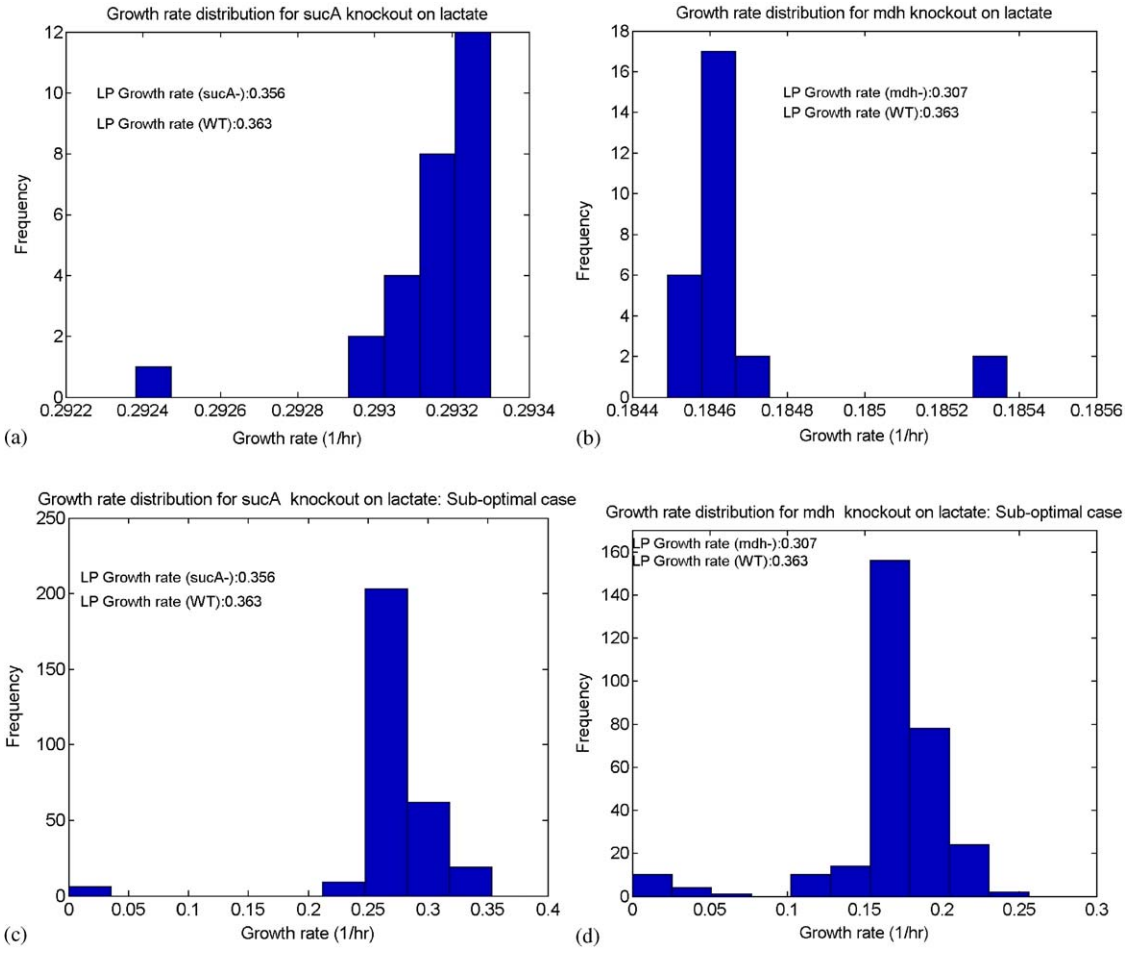

Fig. 5. Histogram of the knockout growth rates calculated using the QP-based analysis for growth on lactate with (a) the optimal flux distributions (the alternate optimal solutions) as the reference for the *sucA* knockout; (b) the optimal flux distributions (the alternate optimal solutions) as the reference for the *mdh* knockout; (c) with the suboptimal flux distributions as the reference for the *sucA* knockout; (d) with the suboptimal flux distributions as the reference for the *mdh* knockout. The variation in the mutant growth rates is greater for the case where the suboptimal flux distributions are used as the reference due to increased flux variability among the suboptimal solutions.

can be found using a direct extension of the extreme pathway algorithm presented in Schilling et al. (2000). In the current implementation presented herein, the exchange fluxes are eliminated from the network and internal fluxes that represent the negative of all the original internal fluxes are introduced essentially creating an inverse copy of the metabolic network. This expands the number of reactions in the network by introducing an inverse set of internal fluxes, which perform the exact same reaction as the original set of internal fluxes, but in the opposite direction. The pathway algorithm discussed in Schilling et al (2000) to calculate the extreme pathways is then applied to the network to calculate internal pathways that satisfy the stoichiometric and thermodynamic constraint of each of the fluxes in the expanded reaction network. This set of pathways contains a number of pathway vectors, which can be removed in post-processing to yield the complete set of equivalency relationships. Pathways utilizing an original internal flux and its inverse, and pathways that utilize fluxes from only the original set or only the inverse set are removed. The pathways that remain utilize different internal fluxes from both sets.

To demonstrate the process, the equivalent reaction sets for the network in Fig. 2 are calculated. An inverse network is added to the existing network and the algorithm is applied to generate pathway vectors. The processing of these pathway vectors is best illustrated in Table 1, where the network in Fig. 2 is analyzed for redundancies. The analysis results in the comprehensive discovery of three equivalent reaction sets. These sets are illustrated in Fig. 6 for the example network.

We next employ the approach described above to calculate the equivalent reaction sets in the fluxes that varied among the alternate optima identified earlier for growth on glucose (Fig. 3). The metabolic network containing only the fluxes that vary among the alternate optima for growth on glucose alone is considered for the identification of the equivalent reaction sets. The

Table 1

Pathway number	Original set of internal fluxes						Inverse set of internal fluxes						Equivalent reaction sets
	v_1	v_2	v_3	v_5	v_6	v_7	d_1	d_2	d_3	d_5	d_6	d_7	
P_1	0	1	0	1	0	0	0	0	1	0	0	0	$v_2+v_5 \sim v_3$
P_2	0	0	1	0	0	0	0	1	0	1	0	0	
P_3	0	0	0	0	1	0	0	0	0	1	0	1	$v_7+v_5 \sim v_6$
P_4	0	0	0	1	0	1	0	0	0	0	1	0	
P_5	0	0	1	0	0	1	0	1	0	0	1	0	$v_2+v_6 \sim v_3+v_7$
P_6	0	1	0	0	1	0	0	0	1	0	0	1	
P_7	1	0	0	0	0	0	1	0	0	0	0	0	Identical reaction used from both sets of internal fluxes
P_8	0	1	0	0	0	0	0	1	0	0	0	0	
P_9	0	0	1	0	0	0	0	0	1	0	0	0	
P_{10}	0	0	0	1	0	0	0	0	0	1	0	0	
P_{11}	0	0	0	0	1	0	0	0	0	0	1	0	
P_{12}	0	0	0	0	0	1	0	0	0	0	0	1	

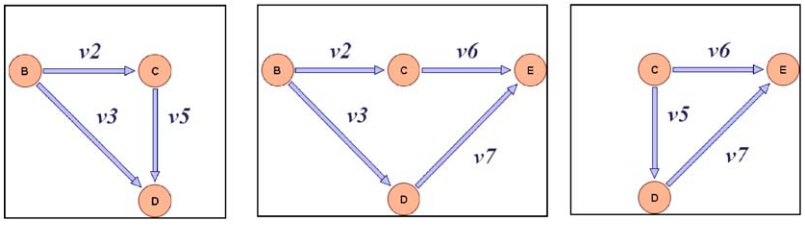

Fig. 6. The equivalent reaction sets identified for the example metabolic network described in Fig. 2.

equivalent reaction sets identified in this way are specific to the growth condition, namely growth on glucose, as the flux of these reactions show variation under this condition. Note that the equivalent reaction sets calculated are a subset of all the possible equivalent reaction sets that are based on the network stoichiometry alone and therefore, the overall equivalent reaction sets are not condition dependent. However, the recruitment of the equivalent reaction sets to the optimal solution will be condition dependent and hence, the active equivalent reaction sets can be different in each condition. These equivalent reaction sets represent the different optimal pathways that will enable different flux distributions to have the same objective function. Fig. 7 shows the various equivalent reaction sets that were identified using the analysis procedures described in the previous section. Fig. 8 (visualized using Simpheny™, Genomatica Inc., San Diego, CA) shows the various ways in which the equivalent reaction set in Fig. 7a can be realized.

Several of the equivalent reaction sets arise due to the cycle of reactions that are catalyzed by ribonucleoside-diphosphate reductase (RNDR) (that reduces nucleoside diphosphate to deoxyribonucleoside-diphosphate) and nucleoside diphosphate kinase (that transfers phosphate groups among the nucleoside phosphates). In all the eight pairs of equivalent reaction sets shown in Fig. 8, RNDR is used in both the sets, although the substrates (nucleoside diphosphates) are different. However, all the RNDR reactions are catalyzed by the gene products of the *nrdA*, *nrdB* genes, which have been shown to be essential for growth under aerobic conditions (Jordan et al., 1996). Thus, for this case, the proposed equivalent reaction sets cannot be vali-

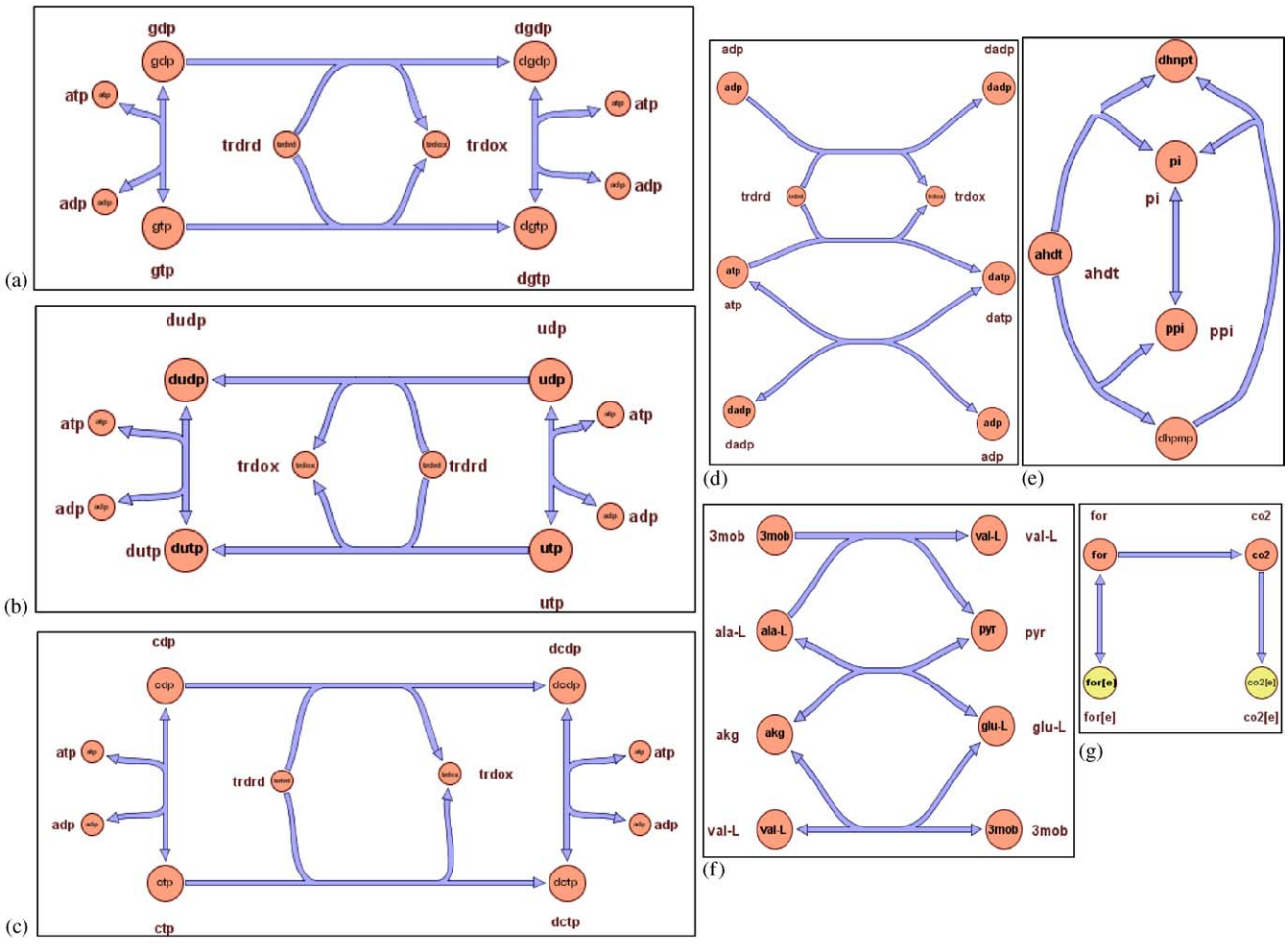

Fig. 7. The reaction systems containing the equivalent reaction sets in the genome-scale in silico model for growth on glucose. (a–d) The equivalent reaction sets in the nucleotide metabolism. (e) Diphosphate phosphohydrolase. (f) Transamination reactions among pyruvate, alphaketoglutarate, valine and oxoisovalerate. (g) Formate dehydrogenase reactions (note that this equivalent reaction set does not represent redundancy and that the protons and water produced/consumed are not incorporated in the model reactions. This set implies that formate and CO_2 are equivalent sinks). Metabolites: trdrd: Reduced thioredoxin, trdox: Oxidized thioredoxin, ahdt: 2-Amino-4-hydroxy-6-(erythro-1-2-3-trihydroxypropyl) dihydropteridine, dhnpt: Dihydroneopterin, dhpmp: Dihydroneopterin phosphate, 3mob: Oxoisovalerate, ala: Alanine, glu: Glutamate, akg: Alpha ketoglutarate.

dated in a straightforward fashion through the characterization of mutant phenotypes. Other equivalent reaction sets include transamination reactions among glutamate, pyruvate, alphaketoglutarate, valine, and 2-oxoisovalerate and the formate dehydrogenase that converts formate to carbon dioxide. The equivalent reaction set involving the transamination reactions (Fig. 7f) could explain the observations that the *avtA* mutation is silent (loss of oxoisovalerate valine transaminase) and that the *ilvE-avtA*$^+$ (loss of oxoisovalerate alphaketglutarate transaminase) mutants are prototrophic for valine (Berg et al., 1983; Umbarger, 1996). The other explanation could be that the transaminase is promiscuous and could catalyze the transamination of multiple substrates. Thus, the equivalent reaction sets provide insights into the redundancy of metabolic networks. These redundant equivalent reaction sets directly results in the variation in the fluxes seen in Fig. 3 for growth on glucose and lead to the presence of the alternate optima explored in the previous sections.

7. Discussion

Biological systems often contain redundancies that contribute significantly to their robustness. Herein, the redundancies in the metabolic network that lead to alternate optimal flux distributions and hence contribute to robustness in the context of optimal growth and flux variability are analyzed. An approach to characterize and study the variability of the fluxes in alternate optimal solutions is presented. This approach for the characterization of the alternate optima is demonstrated using the genome-scale metabolic model of *E. coli*. The alternate optimal solutions for growth of *E. coli* on glucose, acetate, and lactate have been identified using

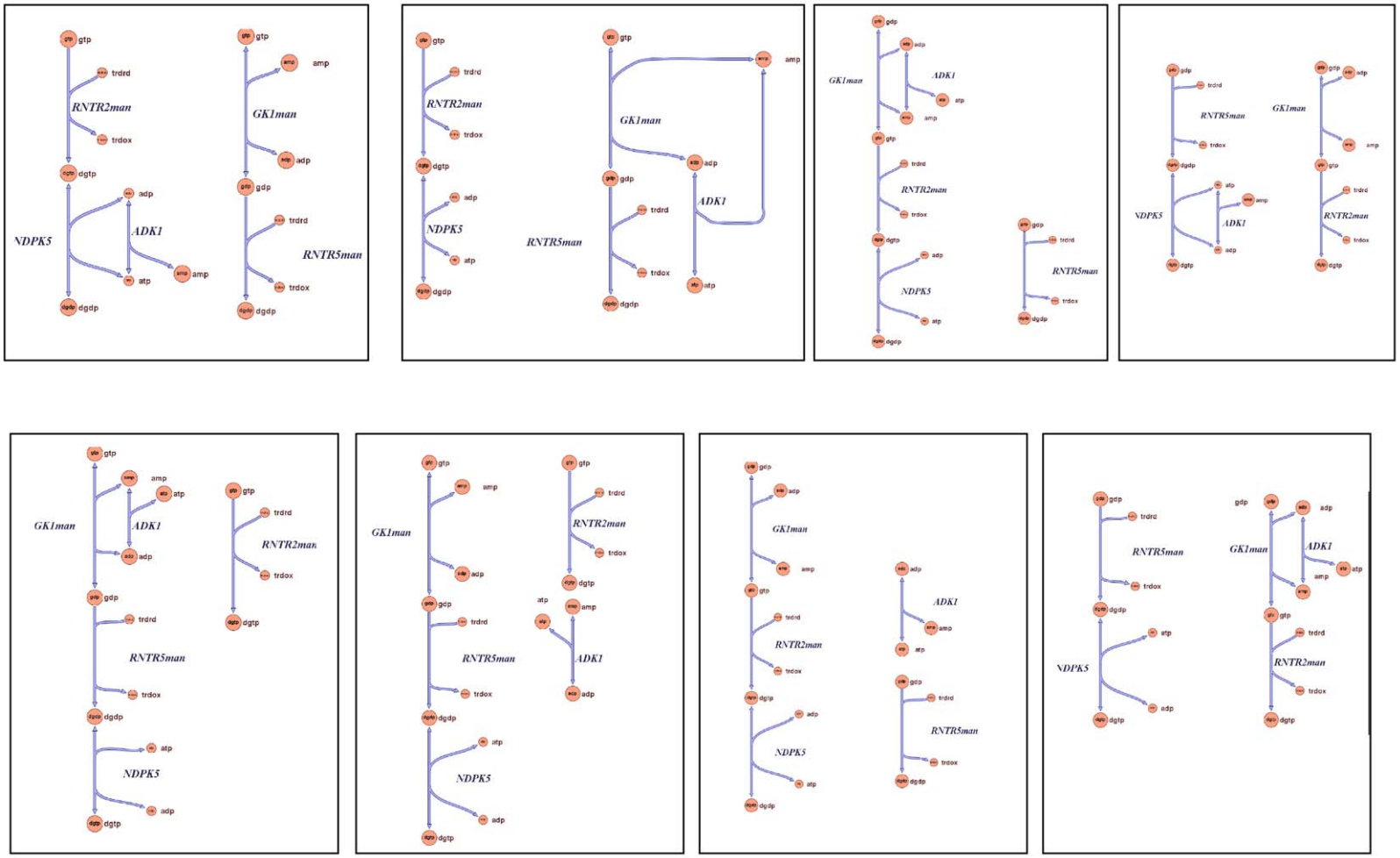

Fig. 8. All the possible equivalent reaction sets visualized using Simpheny^{TM} (Genomatica, San Diego, CA) for the reaction system in Fig. 7a.

the proposed approach. The results presented indicate that the number of pathways with optimal biomass yield and hence the number of fluxes that vary is dependent on the substrate that is metabolized.

The redundancy in the network that leads to the alternate optimal solutions for growth on glucose has been characterized by determining the equivalent reaction sets. These alternate solutions represent different flux distributions that can be active and result in optimal growth based on the metabolic network. This redundancy in the network could be used to explain the observations presented in Fong et al. (2003), where two strains evolved on lactate reach the same optimal endpoint, but through different evolutionary paths and the observed differences in flux distributions in *E. coli* strains with an inactive phosphotransferase system reported by (Flores et al., 2002). It is conceivable that different equivalent reaction sets are operational in these two strains that lead to differences in the flux distributions in the initial suboptimal phase (as evident in the flux variability analysis for the suboptimal case) and the evolutionary paths that result in the same optimal endpoint.

The effect of these alternate optima on the growth rate predictions of mutant strains (using the MOMA QP-based approach) has been investigated. It is observed that the variability in the flux distributions of alternate optimal solutions can affect the predictions of the QP-based analysis significantly, particularly when considering slightly sub-optimal solutions as the reference. Therefore, the choice of the reference distribution for the QP-based analysis has to be determined after comparison of the alternate optimal flux distribution with experimental data (e.g., analysis of gene expression data to identify active equivalent reaction sets) and after the consideration of established regulatory interactions to eliminate a subset of the alternate optimal solutions.

In the discussion presented above, it was assumed that there exists a unique flux distribution that could be determined through the use of additional constraints (gene expression, regulatory rules, kinetics etc.). However, it is possible that multiple flux distributions could be viable in reality. In such a case, the analysis presented here contains interesting implications for population dynamics of bacterial cultures. A bacterial culture could contain a population of heterogeneous individuals that grow at the same maximal rate with underlying differences in the flux distribution due to the equivalent reaction sets. The flux distributions active in these individuals could correspond to different alternate optimal solutions obtained in the LP-based analysis. These differences in the flux distribution could lead to significant differences in the phenotype when subjected to a perturbation, e.g., a gene deletion. For example, in the case of *mdh* mutant for growth on lactate, there would be a distribution of growth rates as predicted by the QP-based analysis (Fig. 5d), and the mutation would be lethal for certain individuals in the population.

The analysis of the alternate optimal solutions has generated insights into the redundancies that arise due to the equivalent reaction sets and has provided a physiological interpretation for the presence of alternate optimal solutions. It must be noted that all possible alternate optimal solutions have not been evaluated in this study. The optimal solutions, where the fluxes are present at their maximum and minimum value are investigated in this paper. The information regarding the fluxes that do not vary in any of the alternate optima can be used to reduce the solution space in determining all possible alternate optimal solutions. Thus, preliminary analysis using the approach presented in this paper can also be used to design more efficient algorithms for obtaining all the optimal solutions.

The fluxes that varied in the alternate optima were used to identify equivalent reaction sets for the growth of *E. coli* on glucose. The use of any of the equivalent reactions does not change the objective function and thus alternate optimal solutions can be realized through the combination of the different equivalent reaction sets that are valid for the specific growth conditions. The analysis of the in silico model for growth on glucose reveals the presence of seven such sets. These equivalent reaction sets are based on the characteristics of the metabolic network, namely, the stoichiometry and thermodynamics (that determines reversibility). However, not all the equivalent reaction sets are involved in the alternate optima for growth on a specific substrate. Therefore, the reaction sets and the number of alternate optimal solutions are determined by the growth conditions. For the analysis presented in the paper, it is assumed that all equivalent reaction sets are available for the cell to obtain the optimal solution. However, a subset of these sets may not be available due to regulatory and other inhibitory effects that are not considered in the analysis. The inclusion of these regulatory effects could be the subject of future investigation.

The analysis of the flux variability on the QP-based analysis for determining mutant growth rate reveals the importance of obtaining the physiologically relevant (wild type) flux distribution for cases where there is significant variability in the optimal flux distributions. Thus, experimental techniques to probe the flux distribution, gene expression data, proteomic, and metabolomic data to identify the physiologically relevant flux distribution are critical to distinguish between the optimal flux distributions.

In summary, we have presented an approach to investigate alternate optimal solutions in genome-scale constraint-based models and assess the physiological significance of these alternate optima. It is seen that alternate optimal solutions have a biological basis as redundant metabolic pathways lead to the existence of

alternate optimal solutions. Furthermore, the choice of a particular alternate optimal solution as reference for subsequent analysis (QP-based) can significantly affect its predictions. This motivates the development of approaches for discerning the biologically relevant metabolic flux distribution based on additional experimental information such as expression data. Thus, the approach presented investigates the redundancy of the optimal solutions in the metabolic network and provides significant insights into the capabilities and the functioning of metabolic networks under different growth conditions.

Acknowledgments

Authors would like to acknowledge Sharon Wiback, Stephen Fong, Bernhard Palsson, Daniel Segre, Brian Gates and George Church for valuable comments on this manuscript. Financial support for this work provided by the Department of Energy (Grant No: DE-FG03-01-01ER25499/A000) is gratefully acknowledged.

References

Bailey, J.E., 2001. Complex biology with no parameters. Nat. Biotechnol. 19, 503–504.

Beard, D.A., Liang, S.C., Qian, H., 2002. Energy balance for analysis of complex metabolic networks. Biophys. J. 83, 79–86.

Berg, C.M., Whalen, W.A., Archambault, L.B., 1983. Role of alanine-valilne transaminase in Salmonella typhimurium and analysis of an avtA:: tn5 mutant. J Bacteriol. 155, 1009–1014.

Bonarius, H.P.J., Schmid, G., Tramper, J., 1997. Flux analysis of underdetermined metabolic networks: The quest for the missing constraints. Trends Biotechnol. 15, 308–314.

Burgard, A.P., Vaidyaraman, S., Maranas, C.D., 2001. Minimal reaction sets for *E. coli* metabolism under different growth requirements and uptake environments. Biotechnol. Prog. 17, 791–797.

Edwards, J.S., Palsson, B.O., 1999. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. J. Biol. Chem. 274, 17410–17416.

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

Edwards, J.S., Ibarra, R.U., Palsson, B.O., 2001. In silico predictions of *E. coli* metabolic capabilities are consistent with experimental data. Nat. Biotechnol. 19, 125–130.

Edwards, J.S., Covert, M., Palsson, B.O., 2002. Metabolic modeling of microbes: the flux balance approach. Environ. Microbiol. 4 (3), 133–140 (Ref Type: Generic).

Endy, D., Brent, R., 2001. Modelling cellular behaviour. Nature 409 (Suppl.), 391–395.

Fell, D.A., 1996. Understanding the Control of Metabolism. Portland Press, London.

Flores, S., Gosset, G., Flores, N., de Graaf, A.A., Bolivar, F., 2002. Analysis of carbon metabolism in *E. coli* strains with an inactive phosphotransferase system by (13)C labeling and NMR spectroscopy. Metab. Eng. 4, 124–137.

Fong, S.S., Marciniak, J.Y., Palsson, B., 2003. Description and Interpretation of Adaptive Evolution of *E. coli* K12 MG1655 Using a Genome-scale in silico Metabolic Model. Journal of Bacteriology, accepted for publication.

Hatzimanikatis, V., Floudas, C.A., Bailey, J.E., 1996. Optimization of regulatory architectures in metabolic reaction networks. Biotechnol. Bioeng. 52, 485–500.

Heinrich, R., Schuster, S., 1996. The Regulation of Cellular Systems. Kluwer Academic Publishers, Hingham, MA.

Ibarra, R.U., Edwards, J.S., Palsson, B.O., 2002. *E. coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature 420, 186–189.

Jordan, A., Aragall, E., Gibert, I., Barbe, J., 1996. Promoter identification and expression analysis of Salmonella typhimurium and *E. coli* nrdEF operons encoding one of two class I ribonucleotide reductases present in both bacteria. Mol. Microbiol. 19, 777–790.

Kitano, H., 2002. Computational systems biology. Nature 420, 206–210.

Lee, S., Phalakornkule, C., Domach, M.M., Grossmann, I.E., 2000. Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. Comput. Chem. Eng. 24, 711–716.

Mahadevan, R., Edwards, J.S., Doyle, F.J., 2002. Dynamic flux balance analysis of diauxic growth in *E. coli*. Biophys. J. 83, 1331–1340.

Palsson, B., 2000. The challenges of in silico biology. Nat. Biotechnol. 18, 1147–1150.

Phalakornkule, C., Lee, S., Zhu, T., Koepsel, R., Ataai, M.M., Grossmann, I.E., Domach, M.M., 2001. A MILP-based flux alternative generation and NMR experimental design strategy for metabolic engineering. Metab. Eng. 3, 124–137.

Price, N.D., Famili, I., Beard, D.A., Palsson, B.O., 2002a. Extreme pathways and Kirchhoff's second law. Biophys. J. 83, 2879–2882.

Price, N.D., Papin, J.A., Palsson, B.O., 2002b. Determination of redundancy and systems properties of the metabolic network of *Helicobacter pylori* using genome-scale extreme pathway analysis. Genome Res. 12, 760–769.

Price, N.D., Papin, J.A., Schilling, C.H., Palsson, B., 2003. Genome-scale microbial in silico models: the constraints-based approach. Trends Biotechnol. 21, 162–169.

Reed, J.L., Vo, T.D., Schilling, C.H., Palsson, B., 2003. *E. coli* IJR904: an expanded genome-scale model of *E coli* k-12. Genome Biol. 4, R54.1.

Reich, J.G., Selkov, E.E., 1981. Energy Metabolism of the Cell: A Theoretical Treatise. Acad. Press, New York, NY.

Schilling, C.H., Schuster, S., Palsson, B.O., Heinrich, R., 1999. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. Biotechnol. Prog. 15, 296–303.

Schilling, C.H., Letscher, D., Palsson, B.O., 2000. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. J. Theor. Biol. 203, 229–248.

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

Segre, D., Vitkup, D., Church, G.M., 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. USA 99, 15112–15117.

Stephanopoulos, G.N., Aristidou, A.A., Nielsen, J., 1998. Metabolic Engineering: Principles and Methodologies. Academic Press, San Diego, CA.

Umbarger, H.E., 1996. Biosynthesis of branched-chain amino acids. In: Neidhardt, F. (Ed.), *E. coli* and Salmonella: Cellular and Molecular Biology. ASM Press, Washington, DC.

Varma, A., Palsson, B., 1994. Metabolic flux balancing: basic concepts, scientific and practical use. Bio/Technology 12, 994–12998.

Varner, J., Ramkrishna, D., 1999. Metabolic engineering from a cybernetic perspective. 1. Theoretical preliminaries. Biotechnol. Prog. 15, 407–425.