

On the Computation of Minimal Cut Sets in Genome Scale Metabolic Networks

Marcin Imielinski and Calin Belta

Abstract—A cut set for an objective reaction in a metabolic network is a set of reactions whose knockout disables flux through that reaction at steady state. Cut sets represent a particular type of failure mode of a metabolic network and may correspond to novel drug targets. In this paper, we demonstrate how cut sets can be obtained from the computation of sub-Elementary Modes (sub-EM). The sub-EM's of a metabolic network are the Elementary Modes (EM) of a submatrix of the stoichiometry matrix formed by taking a subset of its rows. Sub-EM's emerge naturally in the intermediate steps of the standard tableau algorithm for computation of EM, and are thus obtainable for a network of any size. By employing properties of the feasible flux cone, we show how cut sets for a reaction can be constructed by enumerating minimal hitting sets for the sub-EM's containing that reaction. Though the resulting cut sets are not guaranteed to be minimal, they can be reduced to minimality via a second linear programming pruning step. We demonstrate the applicability of this approach to a recent genome scale metabolic model of *E.coli*.

I. INTRODUCTION

The metabolic network is the biochemical machinery with which a cell transforms a limited set of nutrients in its environment into the multitude of molecules required for growth and survival. It consists of hundred to thousands of small molecule species intricately linked by an even larger set of biochemical reactions. The expansive and highly connected nature of this important cellular system greatly limits the degree of insight that may be gained from the isolated study of a single component or module. The first step towards systems-level understanding of metabolism is the construction of a model that captures what is known regarding an organism's small molecule biochemistry and its underlying genetics. The advent of sequencing technology combined with general improvements in the organization of biological information [10], [16] has allowed the building of such *genome-scale metabolic models* for numerous microbial organisms, including *E. coli*, *S. cerevisiae*, *H. pylori*, and *S. aureus*, as well as the human red blood cell, hepatocyte, and mitochondrion [19], [21], [5], [6], [17], [18], [11], [2], [28], [25].

Current approaches to the study of genome-scale metabolic networks employ an analysis of feasible and optimal reaction fluxes through the network at steady state, subject to structural, thermodynamic, and flux capacity constraints [16], [18]. Structural constraints arise from the stoichiometry of the metabolic reactions, thermodynamic

constraints are imposed by the irreversibility of certain metabolic reactions, while flux capacity constraints can be derived from the availability of nutrients, the existence of a knockout, and biochemical data on the maximum throughput of enzymes. Finally, the steady state assumption follows from time-scale separation between rapid metabolic reactions and slower environmental and cellular regulatory changes. Given such constraints, the flux configuration through the network is limited to a feasible region, which can be checked for non-emptiness to characterize the production capacity of the network [8], [9], can be analyzed to find points that maximize biomass production [26], [27], [16], [1], [4] or minimize metabolic adjustment [24]. Alternatively, starting from the observation that these constraints limit the feasible set of fluxes to a polyhedral cone, one can try to compute its (unique) set of generators and thus identify the elementary functional units of a metabolic network [20]. The most prevalent notions of such units are elementary flux modes (EM) and extreme pathways (EP) [23], [19], [15], [14], [29].

A gene or reaction is considered *essential* for a given physiological function if its knockout abolishes that function. Relatively few metabolic genes or reactions can be considered essential for growth of an organism in rich media. However, a gene or reaction that may not be essential "alone" may be still be part of one or more "essential sets". Klamt and Gilles refer to such sets of reactions as *minimal cut sets*, since knockout of that set of reactions abolishes (i.e. "cuts") the physiological function, while knock out of any of its subsets preserves it [13]. Though execution of a linear program can validate whether a given reaction set is a cut set for an objective reaction, use of linear programming for the design of novel cut sets requires a brute force search through all reaction subsets. The latter approach is reasonable for determining essential reactions (i.e. single reaction cut sets) but scales poorly in the search for more complex (i.e. higher cardinality) cut sets. For example, a network of 1000 reactions yields 1000 potential single knockouts, however the set of potential double knockouts approaches 1 million, and the set of potential triple knockouts approaches 1 billion.

A more systematic approach for generating cut sets is the minimal cut set (MCS) algorithm of Klamt and Gilles [13], which employs the elementary modes (EM) of a metabolic network to determine minimal cut sets for a particular objective reaction. Underlying the MCS algorithm is the principle that a minimal cut set R for an objective reaction j is a minimal hitting set for all j -containing EM [12]. As a result, MCS can be enumerated for a network through the simple application of a minimal hitting set algorithm

M. Imielinski is with the University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA imielins@mail.med.upenn.edu

C. Belta is with the Center for Information and Systems Engineering, Boston University, Brookline, MA 02446, USA cbelta@bu.edu

to the collection of j -containing EM. The main limitation of this approach is the difficulty of calculating EM for large networks (i.e. larger than 200 reactions) given current computing resources and algorithms.

In this paper, we demonstrate how cut sets can be obtained from the computation of sub-EM's. The sub-EM's of a metabolic network are the elementary modes of a submatrix of the stoichiometry matrix formed by taking a subset of its rows. Alternatively stated, sub-EM's are flux configurations that place only a subset of species in the system at steady state. Sub-EM's naturally emerge in the intermediate steps of the tableau algorithm for EM computation [23], and are thus obtainable for a network of any size. By employing properties of the feasible flux cone, we show how cut sets for a reaction j can be constructed by enumerating minimal hitting sets for j -containing sub-EM's. Though the resulting cut sets are not guaranteed to be minimal, they can be reduced to minimality via a second linear programming based step. As we show, this method offers a practical approach for minimal cut set computation in genome-scale metabolic networks.

We demonstrate the applicability of this approach to a recent genome scale metabolic model *E. coli* iJR904 [18]. Our results reveal many complex (i.e. 2 or more reactions) minimal cut sets for the biomass reaction in each model. Knockouts of these reaction sets is predicted to be lethal by the *in silico* model, while knockouts of any of their subsets are predicted to be viable. These results lend insight into the function and vulnerabilities of several microbial metabolic networks. Furthermore, our results suggest important experiments for model validation and represent potential targets for drug design.

II. PROBLEM FORMULATION AND APPROACH

Notation In what follows, \mathbb{N} is the set of natural numbers, \mathbb{R} is the set of real numbers, \mathbb{R}_+^n is the set of all n -dimensional vectors with real and positive components, and $\mathbb{R}^{m \times n}$ is the set of all $m \times n$ matrices with real entries. Given $m, n \in \mathbb{N}$, we use the notation $M = \{1, \dots, m\}$ and $N = \{1, \dots, n\}$. For a set C , we use $|C|$ to denote its cardinality. If $A \in \mathbb{R}^{m \times n}$ and $U \subseteq M$, then A_U denotes the submatrix of A containing the rows with indices in the set U . Therefore, if $x \in \mathbb{R}^n$, $i \in N$, and $U \subset N$, then x_i and $x_U \in \mathbb{R}^{|U|}$ denote its i th component and the vector formed by taking components with indices in set U , respectively. The inequality $x \geq 0$ is interpreted componentwise, i.e., $x_i \geq 0$, $i = 1, \dots, n$, while the inequality $x > 0$ is interpreted as $x \geq 0$, $x \neq 0$. Each vector $x \in \mathbb{R}^n$ induces a *line*, $l = \{\alpha x \mid \alpha \in \mathbb{R}\}$ and a *ray*, $r = \{\alpha x \mid \alpha > 0\}$. We denote the set of nonzero indices of a line or ray $r \subset \mathbb{R}^n$ as $NZ(r) \subseteq N$. The notation $r^1 + r^2$, where r^1, r^2 are rays or lines, refers to the usual (Minkowski) set sum.

We represent a mass-balanced metabolic network of n chemical reactions involving m metabolites in a stoichiometry matrix $S \in \mathbb{R}^{m \times n}$. We assume that matrix S also incorporates stoichiometric information about all exchange reactions (uptake and secretion) and about the maintenance

and growth reactions. Each entry S_{ij} specifies the stoichiometric coefficient for metabolite i in reaction j , which is negative for substrates and positive for products. We represent the flux distribution through the reactions of the network by $v \in \mathbb{R}^n$, where a component v_j corresponds to the flux of reaction complex passing through reaction j . The concentrations of species in the system at time t are denoted by $x(t) \in \mathbb{R}_+^m$. Finally, thermodynamic constraints restrict a subset of reactions $T \subseteq N$ to be irreversible. Under these assumptions, the rate of change in time of species concentrations is given by:

$$\dot{x} = Sv, v_T \geq 0. \quad (1)$$

Metabolic reactions occur at a fast rate with respect to cell regulatory and environmental changes. When modeling at the slower time scale it is reasonable to apply the *quasi-steady state assumption*, under which we have:

$$Sv = 0, v_T \geq 0. \quad (2)$$

A (steady state) metabolic network with stoichiometry S and set of irreversible reactions T will be denoted for short by (S, T) . The set of all *feasible fluxes*

$$K = \{v \in \mathbb{R}^n \mid Sv = 0, v_T \geq 0\} \quad (3)$$

is a polyhedral cone in \mathbb{R}^n .

Definition 1: A set of reactions $C \subseteq N$ is a *cut set* for an objective reaction $j \in N$ in system (S, T) if

$$v_C = 0 \rightarrow v_j = 0, \forall v \in K. \quad (4)$$

A cut set C is minimal if no proper subset of C is a cut set.

In this paper we consider the following problem:

Problem 1: Compute minimal cut sets for a target reaction $j \in N$ in a metabolic network (S, T) .

A "brute force" approach commonly used to solve the latter problem involves using flux balance analysis (FBA) to test every possible reaction combination to determine whether it "cuts" the objective reaction j . To determine all minimal cut sets of cardinality one (essential reactions), one has to decide the non-emptiness of the set $\{v \mid v \in K, v_i = 0, v_j \neq 0\}$ for all $i \in N$, $i \neq j$, which can be achieved by solving linear programs. For example, if the solution to both linear programs $\max_v v_j$ and $\min_v v_j$ with constraint set $\{v \mid v \in K, v_i = 0\}$ is zero, then this is equivalent to i being an essential reaction for j (minimal cut set of cardinality 1). To determine all minimal cut sets of cardinality two, all pairwise combinations $\{i, k\} \in N$ ($i \neq k \neq j \neq i$) not containing essential reactions have to be considered. Again, $\{i, k\}$ is a minimal cut set for j if and only if the solution to both linear programs $\max_v v_j$ and $\min_v v_j$ with constraint set $\{v \mid v \in K, v_i = v_k = 0\}$ yields 0. The procedure continues for higher cardinality combinations.

Though applicable to the study of single and double knockouts, this "brute-force" approach fails to be useful for generation of higher order cut sets in genome-scale metabolic models, which generally have greater than 1000 reactions.

The solution we propose in this paper is a "rational" approach for discovering cut sets that employs computation of generators of a polyhedral cone and the subsequent application of a minimum hitting set algorithm. The idea is not new - this approach was proposed by Klamt and Gilles in [13], [12]. However, because of limitations due to complexity, their approach also cannot be applied to genome scale metabolic models. In this paper, we propose a two step procedure. In a first step, we determine cut sets based on the computation of generators for polyhedral cones over-approximating K . In the second, we use linear programming to prune these sets and make them minimal. We cannot guarantee that we compute all the minimal cut sets for j with our approach. However, as shown in Section V, our approach leads to the computation of a large set of non-trivial minimal cut sets at genome scale.

III. PRELIMINARIES

A. Computation of generators for polyhedral cones

A polyhedral cone is usually described as either the set sum of a finite number of rays (V-representation) or as the intersection of a finite number of half-spaces (H-representation) [22]. Such a polyhedral cone is called pointed. A ray in a collection of rays is called extreme if it does not belong to the set sum of the others. A collection of rays is called extreme if every ray in the collection is extreme. A pointed polyhedral cone can be written as the set sum of a unique set of rays in an extreme collection. These rays are called generators of the cone. An example of a pointed polyhedral cone is the intersection of the null space of a matrix with the positive orthant.

However, in general, a polyhedral cone $K \subset \mathbb{R}^n$ (such as the feasible cone of equation (3) can contain a subspace. In this case, a useful V-representation of K is as the set sum of a collection of rays and a collection of lines that span the subspace contained in K . We refer to this set of rays and lines as *generators* of K , and we denote them by $E(K)$. If the collection of rays is extreme and the number of lines is equal to the dimension of the subspace contained in K , then $E(K)$ constitutes a *minimal V-representation* of K .

The connection with the metabolic model in Section II is as follows: if $T = N$ (i.e., all the reactions were irreversible), then the feasible cone K from equation (3) would be pointed. Its generators are called *extreme pathways*, and can be computed using the algorithm described in [19]. However, in general $T \subset N$, which means that K can contain lines. The tableau algorithm described in [23] is the standard approach for computing generators of polyhedral cones associated with metabolic networks. In this case, the generators are called *elementary modes* (EM). This algorithm is a modification of the classical Gaussian elimination algorithm used to compute the null space of a matrix. Here we give a brief description of the algorithm, which is necessary to understand the rest of the paper.

The tableau algorithm takes as arguments the matrix $S \in \mathbb{R}^{m \times n}$ and the index set $T \subset N$, and returns a set of elementary modes $E(K)$ of K , corresponding to a collection

of rays and lines that generate K . If $T = N$, the output corresponds to the extreme rays of K .

The tableau algorithm proceeds in an iterative fashion by computing generators $E(K^i)$ for a series of cones $K^i \subset \mathbb{R}^n$, $i \in \{0, \dots, m\}$ given by:

$$K^i = \{v \mid S_{M_i}v = 0, v_T \geq 0, A_i \subseteq M, |M_i| = i\} \quad (5)$$

The algorithm is seeded with the initial cone $K^0 = \{v \in \mathbb{R}^n \mid v_T \geq 0\}$. The initial collection of generators $E(K^0)$ consists of rays pointing in the directions of Euclidean basis vectors $e^j \in \mathbb{R}^n, j \in T$ and lines corresponding to the remaining Euclidean basis vectors. At each iteration $i \in \{1, \dots, m\}$, the generators $E(K^i)$ are computed from the analysis of rays and lines in $E(K^{i-1})$ in three steps. In the first step, each ray / line in E^{i-1} is tested to determine whether it belongs to the hyperplane $S_i v = 0$. Rays and lines in $E(K^{i-1})$ that belong to this hyperplane are added to the collection $E(K^i)$. In the second step of iteration i , the remaining rays and lines in $E(K^{i-1})$ are paired to compute intersections of cone K^{i-1} with the hyperplane $S_i v = 0$. The final step of each iteration involves pruning of "decomposable" or dependent generators from $E(K^i)$. This step ensures minimality or near-minimality of the set $E(K^i)$ as a V-representation for the cone K^i .

Following iteration m , the tableau algorithm terminates, having computed a set of rays and lines (elementary modes) that generate the polyhedral cone $K = K^m$. The costliest part of the tableau algorithm is the pruning step applied at the end of each iteration. Furthermore, the memory requirements of this algorithm can be prohibitive, given the large size of generators for many polyhedral cones [7].

Performance-improving modifications to this algorithm include interlacing of the second and third steps of each iteration [23] and dynamically re-ordering the rows of S to minimize computation [3]. The latter modification applies a heuristic "local greedy" optimization strategy following each iteration i to choose a row from S that minimizes the number of generator pairs in $E(K^i)$ that must be combined in the second step of iteration $i + 1$ [3]. We apply both of these modifications in our implementation.

B. Minimal hitting set for a collection of lines and rays

In combinatorics, a *hitting set* of a collection of sets \mathcal{C} , each taken from a discrete universe of items U , is a set $H \subset U$ that intersects every set in \mathcal{C} . H is a *minimal hitting set* if none of its subsets are hitting sets of \mathcal{C} .

In this paper, we use a modification of this concept employed by Klamt to refer to hitting sets $H \subset N$ of a collection of elementary modes $E \subset \mathbb{R}^n$ [12].

Definition 2: A set $H \subset N$ is a hitting set for for a collection of elementary modes $E \subset \mathbb{R}^n$ if H intersects $NZ(r)$ for every $r \in E$. In addition, H is a *minimal hitting set* for E if none of its subsets are hitting sets.

Algorithm 1 provides a simple procedure for computing the set \mathcal{H} of all minimal hitting sets of cardinality k or less of a collection E of generating vectors (i.e. EM or sub-EM).

This algorithm differs from that proposed in [12] in that it iterates through the set of generators to construct minimal hitting sets rather than exhaustively considering all tuples of (reaction) indices. We find this algorithm to be much faster and computationally feasible for the calculation of higher-order minimal hitting sets.

Algorithm 1 $\mathcal{H} = \text{MinHit}(E, k)$

```

/* Input:  $E = \{r^1, \dots, r^{|E|}\}$ , a collection of lines and/or
rays in  $\mathbb{R}^n$  */
/* Output:  $\mathcal{H}$ , a collection of minimal hitting sets of  $E$ 
with cardinality  $k$  or less*/
/*  $\mathcal{H}$  is initialized with the empty set.*/
 $\mathcal{H} = \{\emptyset\}$ 
for all  $j \in 1, \dots, |E|$  do
  for all  $i$  for which  $r_i^j \neq 0$  do
    /* Expand sets in  $\mathcal{H}$  so that they "hit" nonzero */
    /* components of ray / line  $j$  of  $E$  */
    for all  $H \in \mathcal{H}$  for which  $i \notin H$  do
      replace  $H$  with  $H \cup \{i\}$  in  $\mathcal{H}$ 
    end for
  end for
  /* Prune duplicate, non-minimal, and high cardinality
sets from  $\mathcal{H}$  */
  for all  $H \in \mathcal{H}$  do
    if  $|H| > k$  OR there exists  $\hat{H} \in \mathcal{H}$  for which  $\hat{H} \subseteq H$ 
then
      remove  $H$  from  $\mathcal{H}$ 
    end if
  end for
end for

```

IV. COMPUTATION OF MINIMAL CUT SETS

Klamt and Gilles [13], [12] showed that a minimal hitting set $C \subset N$ for a collection of elementary modes E (see Definition 2) with positive entry in position j (such an elementary mode is said to contain reaction j) is a minimal cut set for reaction j in system (S, T) (see Definition 1).

This is easily shown to be true, since constraining $v_C = 0$ renders infeasible all flux configurations that lie in the set sum of j -containing EM of the feasible cone K . The remaining EM in E have a value of 0 in component j , constraining $v_j = 0$ for all remaining flux configurations when C is knocked out. Knocking out a subset of reactions C will leave one or more j -containing EM active, allowing flux configurations with nonzero flux in j to persist.

In this section, we demonstrate how application of Klamt and Gilles' cut set criteria to intermediate results of the tableau algorithm for EM computation (described in Section III-A) allows identification of cut sets for (S, T) .

At each iteration $i \in M$, the tableau algorithm determines a collection of generators $E(K^i) \subset \mathbb{R}^n$ for the cone K^i in equation (5). $E(K^i)$ corresponds to a collection of non-decomposable lines and rays that obeys the quasi steady state assumption for a subset $A_i \subseteq M$ of species in the system.

K^i can also be seen to correspond to the feasible flux cone of the system (S_{A_i}, T) with elementary modes $E(K^i)$. The rays and lines in $E(K^i)$ are therefore called *sub-elementary modes* (sub-EM).

By applying Klamt and Gilles' criteria to $E(K^i)$, a minimal hitting set R for the collection of j -containing sub-EM in $E(K^i)$ is a minimal cut set for j in the system (S_{A_i}, T) . This means that R is a minimal set of reactions satisfying the relation:

$$v_R = 0 \implies v_j = 0, \quad \forall v \in K^i \quad (6)$$

However, from Equations (3) and (5), it is clear that K^i over-approximates K , i.e.

$$K \subseteq K^i. \quad (7)$$

R thus also satisfies equation (4) and is a cut set for j in the full system (S, T) . As a result, the application of *MinHit* (Algorithm 1) to a collection of j -containing sub-EM taken from any iteration i of the tableau algorithm is guaranteed to yield valid cut sets for an objective reaction j .

Algorithm 2 $MCS = \text{BruteCut}(\mathcal{H}, j, S, T)$

```

/* Input:  $\mathcal{H}$ , a collection of cut sets for reaction  $j$  in
system  $(S, T)$  */
/* Output:  $MCS$ , a collection of minimal cut sets for  $j$ 
*/
 $MCS =$  empty collection
push sets in  $\mathcal{H}$  onto stack  $ToDo$ 
while  $ToDo$  is not empty do
   $H =$  pop set from  $ToDo$ 
  /* Run linear programs (i.e. flux balance analysis) to
find subsets of  $H$  that are also cut sets for  $j$  */
   $\mathcal{D} =$  collection of size  $|H| - 1$  subsets of  $H$  that are
cut sets for  $j$  in system  $(S, T)$ 
  if  $\mathcal{D}$  is empty then
    add  $H$  to collection  $MCS$ 
  else if  $\mathcal{D}$  contains one set then
    add sole set in  $\mathcal{D}$  to  $MCS$ 
  else
    push sets in  $\mathcal{D}$  onto stack  $ToDo$ 
  end if
end while

```

A collection of sub-EM arises from an incomplete analysis of the metabolic network, and thus only contain partial information about its dynamics. As a result, analysis of sub-EM leads to a sufficient but *not necessary* criterion for determining whether a reaction set C constitutes a cut set for reaction j . This results in two caveats regarding the "quality" of cut set obtained from the analysis of pre-elementary modes: 1) not all cut sets are guaranteed to be found and 2) cut sets that are found are not guaranteed to be minimal.

As may be intuitively expected, the quality of cut sets directly depends on the "quality" of sub-EM. The latter, in turn, depends on the iteration i of the tableau algorithm from

which the sub-EM are gathered, and the row order employed by the tableau algorithm (see Section III-A). A general rule of thumb for the quality of sub-EM is that later iterations yield "higher quality" sub-EM; in other words, "naive" sub-EM gathered from an early iteration of the algorithm will yield fewer cut sets that are farther from being minimal, while "mature" sub-EM gathered from a later iteration will yield larger numbers of cut sets that are closer to being minimal.

Since minimality of cut sets obtained from sub-EM is not guaranteed, we propose a brute-force post processing algorithm for 1) checking minimality of cut sets and 2) reducing them to their minimal subsets if they are not minimal. The procedure is iterative (Algorithm 2). For a given cut set H , subsets of cardinality $|H| - 1$ are tested to see if they are cut sets. If none is found, then the initial cut set H is minimal (see Definition 1). Otherwise, the new found cut sets are tested for minimality in the same way. To check whether a set $H \subset N$ is a cut set for reaction $j \in N$, we need to check the feasibility of the set $\{v \in K \mid v_H = 0, v_j \neq 0\}$, for which we can use linear programming (i.e. flux balance analysis) as described in Section II.

V. RESULTS

We use the computational framework developed in this paper to compute minimal cut sets for biomass production in the *E. coli* iJR904 genome scale model [18], which has 761 metabolites involved in 1835 chemical reactions (i.e., the stoichiometry matrix S in equation (3) is 761×1835). These reactions represent the inflow, outflow, and inter-conversion of small-molecule chemical species in an *E. coli* cell grown in a rich nutrient media. For each species in the model, there exists a "sink" reaction that represents its dilution during growth and consumption by macromolecular processes. Of the 761 species, 49 correspond to "biomass components" that are considered to be essential substrates for survival and growth [18]. In this model, knock out of biomass production (and thus growth and survival) corresponds to "cutting" flux through *at least one* of the sinks corresponding to an essential biomass component.

We computed sub-EM's using the tableau algorithm and applied a local greedy optimization strategy at each iteration to minimize computation (Section III-A). Sub-EM's were collected once progress in the algorithm reached memory limit or suffered very significant slow down. Cut sets for each biomass component sink j were computed by applying the *MinHit* (Algorithm 1) to the collection of j containing sub-EM. For computational ease and biological relevance we limited *MinHit* to generating cut sets of size 10 or below. Minimal cut sets for each j were computed by applying algorithm *BruteCut* (Algorithm 2) to the output of *MinHit*. To solve the linear programs involved in *BruteCut*, we used the semidefinite programming package SeDuMi (<http://sedumi.mcmaster.ca/>).

In a preliminary computation, we used linear programming to determine that there are 92 "essential" reactions whose

knockout disables production of at least one biomass component in rich media (i.e. single reaction cut sets).

Execution of the tableau algorithm to iteration 709 (of 761) on *E. coli* iJR904 in rich media yields 157207 sub-EM. 1375 complex MCS emerged from the application of *MinHit* and *BruteCut* to these results. MCS were successfully obtained for 34 to 49 of the biomass component sinks, which includes 17 components that are immune to attack by any of the 92 single reaction knockouts of biomass. These MCS range in cardinality between 2 and 10 and employ 223 reactions in the *E. coli* genome scale metabolic network (in contrast with the 92 essential reactions). These results show that though relatively few reactions are essential for biomass production in rich media, many more emerge as potential targets for knockout when included in the context of a multi-pronged attack.

For further analysis of MCS mechanisms, we used linear programming to determine the spectrum of reactions knocked out by each MCS. Despite the fact that most MCS (1235 of 1375) consist of 6 or more reactions, we find that the vast majority (1317 of 1375) of the corresponding knockouts appear to carry out a "surgical strike" on biomass production, disabling the producibility of only a single biomass component (e.g. L-threonine, lipopolysaccharide, GTP).

Reactions contribution to the 1375 MCS span 20 reaction subsystems (compared to only 10 subsystems spanned by the 92 single knockouts). Not surprisingly, a vast majority of the MCS discovered employ transporter reactions (subsystems "Transport, Extracellular" and "Putative Transporter"). Expression of MCS in terms of reaction subsystems yields 38 unique "cut set patterns", shown in Figure 1. The predominant two patterns (35 and 36, comprising 1160 of 1375 MCS) amongst the collection of cut sets consists of least one reaction from each of the following reaction subsystems: "Transporter, Extracellular" or "Putative Transporter", "Cell Envelope Biosynthesis", "Threonine and Lysine metabolism", and "Alternate Carbon Metabolism". Other subsystems significantly represented amongst the MCS are "Nucleotide Salvage Pathways" (78 minimal cut sets), "Puridine and Pyrimidine Biosynthesis" (35 cut sets), and "Arginine and Proline Metabolism" (31 minimal cut sets).

VI. CONCLUDING REMARKS

The results presented in this paper represent one of the first applications of a network-based pathway analysis method to a genome-scale metabolic model. So far, all applications of elementary mode or extreme pathway analysis have been used to analyze *subnetworks* of reactions or subsystem decompositions of genome-scale metabolic models. In contrast, this paper employs sub-EM obtained from the analysis of the *full* genome scale model. Though these sub-EM satisfy the quasi-steady state assumption for only a subset of species in the system, they take into account *all* reactions in the model. This study also illustrates the first genome-scale application of the "rational" approach to minimal cut set (MCS) computation outlined by Klamt and Gilles [13].

- [4] M. W. Covert, C. H. Schilling, and B. Palsson, "Regulation of gene expression in flux balance models of metabolism." *J Theor Biol*, vol. 213, no. 1, pp. 73–88, Nov 2001. [Online]. Available: <http://dx.doi.org/10.1006/jtbi.2001.2405>
- [5] I. Famili, J. Forster, J. Nielsen, and B. O. Palsson, "Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network." *Proc Natl Acad Sci U S A*, vol. 100, no. 23, pp. 13 134–9, Nov 2003. [Online]. Available: <http://dx.doi.org/10.1073/pnas.2235812100>
- [6] J. Forster, I. Famili, P. Fu, B. O. Palsson, and J. Nielsen, "Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network." *Genome Res*, vol. 13, no. 2, pp. 244–53, Feb 2003. [Online]. Available: <http://dx.doi.org/10.1101/gr.234503>
- [7] J. Gagneur and S. Klamt, "Computation of elementary modes: a unifying framework and the new binary approach." *BMC Bioinformatics*, vol. 5, no. 1, p. 175, Nov 2004. [Online]. Available: <http://dx.doi.org/10.1186/1471-2105-5-175>
- [8] M. Imielinski, C. Belta, A. Halasz, and H. Rubin, "Characterizing the essential metabolome through genome scale analysis of e.coli production capabilities," *Bioinformatics*, vol. 21, no. 9, pp. 2008–2016, 2005.
- [9] —, "Systematic analysis of conservation relations in e. coli genome-scale metabolic network reveals novel growth media," *Biophysical journal*, vol. 90, pp. 2659 – 2672, 2006.
- [10] P. D. Karp, C. A. Ouzounis, C. Moore-Kochlacs, L. Goldovsky, P. Kaipa, D. Ahn, S. Tsoka, N. Darzentas, V. Kunin, and N. Lpez-Bigas, "Expansion of the BioCyc collection of pathway/genome databases to 160 genomes." *Nucleic Acids Res*, vol. 33, no. 19, pp. 6083–6089, 2005. [Online]. Available: <http://dx.doi.org/10.1093/nar/gki892>
- [11] K. J. Kauffman, J. D. Pajeroski, N. Jamshidi, B. O. Palsson, and J. S. Edwards, "Description and analysis of metabolic connectivity and dynamics in the human red blood cell." *Biophys J*, vol. 83, no. 2, pp. 646–62, Aug 2002.
- [12] S. Klamt, "Generalized concept of minimal cut sets in biochemical networks." *Biosystems*, Nov 2005. [Online]. Available: <http://dx.doi.org/10.1016/j.biosystems.2005.04.009>
- [13] S. Klamt and E. D. Gilles, "Minimal cut sets in biochemical reaction networks." *Bioinformatics*, vol. 20, no. 2, pp. 226–34, Jan 2004.
- [14] S. Klamt and J. Stelling, "Two approaches for metabolic pathway analysis?" *Trends Biotechnol*, vol. 21, no. 2, pp. 64–9, Feb 2003.
- [15] J. A. Papin, J. Stelling, N. D. Price, S. Klamt, S. Schuster, and B. O. Palsson, "Comparison of network-based pathway analysis methods." *Trends Biotechnol*, vol. 22, no. 8, pp. 400–5, Aug 2004. [Online]. Available: <http://dx.doi.org/10.1016/j.tibtech.2004.06.010>
- [16] N. D. Price, J. L. Reed, and B. O. Palsson, "Genome-scale models of microbial cells: evaluating the consequences of constraints." *Nat Rev Microbiol*, vol. 2, no. 11, pp. 886–97, Nov 2004. [Online]. Available: <http://dx.doi.org/10.1038/nrmicro1023>
- [17] A. Raghunathan, N. D. Price, M. Y. Galperin, K. S. Makarova, S. Purvine, A. F. Picone, T. Cherny, T. Xie, T. J. Reilly, R. Munson, R. E. Tyler, B. J. Akerley, A. L. Smith, B. O. Palsson, and E. Kolker, "In Silico Metabolic Model and Protein Expression of Haemophilus influenzae Strain Rd KW20 in Rich Medium." *OMICS*, vol. 8, no. 1, pp. 25–41, 2004. [Online]. Available: <http://dx.doi.org/10.1089/153623104773547471>
- [18] J. L. Reed, T. D. Vo, C. H. Schilling, and B. O. Palsson, "An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR)." *Genome Biol*, vol. 4, no. 9, p. R54, 2003. [Online]. Available: <http://dx.doi.org/10.1186/gb-2003-4-9-r54>
- [19] C. H. Schilling and B. O. Palsson, "The underlying pathway structure of biochemical reaction networks." *Proc Natl Acad Sci U S A*, vol. 95, no. 8, pp. 4193–8, Apr 1998.
- [20] C. H. Schilling, S. Schuster, B. O. Palsson, and R. Heinrich, "Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era." *Biotechnol Prog*, vol. 15, no. 3, pp. 296–303, 1999. [Online]. Available: <http://dx.doi.org/10.1021/bp990048k>
- [21] C. H. Schilling, M. W. Covert, I. Famili, G. M. Church, J. S. Edwards, and B. O. Palsson, "Genome-scale metabolic model of Helicobacter pylori 26695." *J Bacteriol*, vol. 184, no. 16, pp. 4582–93, Aug 2002.
- [22] A. Schrijver, *Theory of Linear and Integer Programming*. Chichester: John Wiley, 1986.
- [23] S. Schuster, D. A. Fell, and T. Dandekar, "A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks." *Nat Biotechnol*, vol. 18, no. 3, pp. 326–332, Mar 2000. [Online]. Available: <http://dx.doi.org/10.1038/73786>
- [24] D. Segr, D. Vitkup, and G. M. Church, "Analysis of optimality in natural and perturbed metabolic networks." *Proc Natl Acad Sci U S A*, vol. 99, no. 23, pp. 15 112–15 117, Nov 2002. [Online]. Available: <http://dx.doi.org/10.1073/pnas.232349399>
- [25] I. Thiele, N. D. Price, T. D. Vo, and B. O. Palsson, "Candidate metabolic network states in human mitochondria. Impact of diabetes, ischemia, and diet." *J Biol Chem*, vol. 280, no. 12, pp. 11 683–95, Mar 2005. [Online]. Available: <http://dx.doi.org/10.1074/jbc.M409072200>
- [26] A. Varma, B. W. Boesch, and B. O. Palsson, "Stoichiometric interpretation of Escherichia coli glucose catabolism under various oxygenation rates." *Appl Environ Microbiol*, vol. 59, no. 8, pp. 2465–73, Aug 1993.
- [27] A. Varma and B. O. Palsson, "Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110." *Appl Environ Microbiol*, vol. 60, no. 10, pp. 3724–31, Oct 1994.
- [28] T. D. Vo, H. J. Greenberg, and B. O. Palsson, "Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data." *J Biol Chem*, vol. 279, no. 38, pp. 39 532–40, Sep 2004. [Online]. Available: <http://dx.doi.org/10.1074/jbc.M403782200>
- [29] C. Wagner and R. Urbanczik, "The Geometry of the Flux Cone of a Metabolic Network." *Biophys J*, Sep 2005. [Online]. Available: <http://dx.doi.org/10.1529/biophysj.104.055129>