

# Planning optimal measurements of isotopomer distributions for estimation of metabolic fluxes \*

Ari Rantanen<sup>1†</sup>, Taneli Mielikäinen<sup>1</sup>, Juho Rousu<sup>1</sup>, Hannu Maaheimo<sup>2</sup>, Esko Ukkonen<sup>1</sup>

<sup>1</sup>Department of Computer Science, P.O. Box 68 (Gustaf Hällströmin katu 2b), 00014 University of Helsinki, Finland

<sup>2</sup>NMR Laboratory, VTT Technical Research Centre of Finland, P.O. Box 65, 00014 Helsinki, Finland

## ABSTRACT

**Motivation:** Flux estimation using isotopomer information of metabolites is currently the most reliable method to obtain quantitative estimates of the activity of metabolic pathways. However, the development of isotopomer measurement techniques for intermediate metabolites is a demanding task. Careful planning of isotopomer measurements is thus needed to maximize the available flux information while minimizing the experimental effort.

**Results:** In this paper we study the question of finding the smallest subset of metabolites to measure that ensure the same level of isotopomer information as the measurement of every metabolite in the metabolic network. We study the computational complexity of this optimization problem in the case of the so-called positional enrichment data, give methods for obtaining exact and fast approximate solutions, and evaluate empirically the efficacy of the proposed methods by analyzing a metabolic network that models the central carbon metabolism of *S. cerevisiae*.

**Contact:** ajrantan@cs.helsinki.fi

## 1 INTRODUCTION

The goal of metabolic flux analysis is to discover the steady state conversion velocities of metabolites to each other through chemical reactions catalyzed by the enzymes of an organism. Information about reaction rates, or fluxes, represents the final outcome of cellular regulation and thus constitutes an important aspect of the phenotype of a cell in different conditions [18]. Flux information can be harnessed in many different applications ranging from pathway optimization in metabolic engineering [26] and from characterization of the physiology of an organism [12] to more efficient drug design for human diseases such as cancer [5].

The most accurate information about the fluxes can be obtained by conducting isotopomer tracer experiments where the cell is fed with a mixture of natural and <sup>13</sup>C-labeled external substrates. The fate of the <sup>13</sup>C atoms can be observed by measuring the resulting NMR [29] or mass spectrum [6, 34] of metabolic products and intermediates. From the measurements one obtains information about the fluxes of the alternative pathways producing a metabolite. This methodology has been successfully applied in numerous cases to explicitly solve

key fluxes of specific metabolic networks in various experimental conditions of interest [16, 26, 27].

A popular general framework for estimating the flux distribution of an arbitrary metabolic network is based on non-linear optimization procedure where candidate values for the fluxes are generated successively and the isotopomer distributions of metabolites corresponding candidate fluxes are computed until the isotopomer distributions fit the measured data well enough [9, 21, 23, 32]. In each step of such an iteration a nonlinear equation system stating isotopomer distributions as the function of the fluxes has to be solved.

Recently Rousu *et al.* [20] proposed a direct flux estimation method that first propagates the measured information on isotopomer distributions over the metabolic network and then augments the basic linear stoichiometric equation system with generalized isotopomer balance equations. Rantanen *et al.* [19] improved the propagation of isotopomer information by introducing a partition of metabolite fragments to sets having equal isotopomer distributions in all steady states. The direct flux estimation method can be seen as a member of another flux estimation framework originating from METAFoR (metabolic flux ratio) analysis [7, 25, 27, 29].

Both flux estimation frameworks have their strengths and weaknesses. The iterative methods require often fewer measured metabolites than METAFoR methods (see Section 2) to produce an estimate of a flux distribution. On the other hand, due to the nonlinearity of the optimization task, it is hard to guarantee that the iterative method converges to the global, unique optimum. If measured data does not fully determine the flux distribution, iterative methods may only output points from a solution space, not the set of all feasible solutions. By performing an a priori identifiability analysis [11, 30] and analysing the sensitivity of a point solution [17] it is possible to examine the uniqueness of the solution. In the method of [20] the equation system bounding the flux distribution is linear. Thus the equation system is easily solvable in a fully determined case. Also, if the equation system is underdetermined the linear subspace containing all solutions can be explicitly stated and well known techniques can be applied to solve as many fluxes as possible [13] as well as to obtain upper and lower bounds to the rest of the fluxes. The general sensitivity of a solution to measurement errors can be easily estimated by inspecting the condition number a linear equation system. Furthermore, as solving and manipulation of a linear equation system can be done efficiently, computationally intensive

\*Preliminary version of this article appeared in the proceedings of German Conference on Bioinformatics 2005. Lecture Notes in Informatics Vol. P-71 (2005), pp. 177 – 191.

†to whom correspondence should be addressed

but conceptually simple Monte Carlo techniques can be applied to estimate the sensitivities of individual fluxes.

Linear flux estimation techniques thus have attractive properties compared to nonlinear, iterative techniques. However, the need for more measurement data should not be overlooked. Developing measurement techniques for intermediate metabolites and conducting the measurements using the technique are nontrivial, laborious and costly processes. This experimental burden could be eased by concentrating the measurements to non-redundant subsets of metabolites that alone give as much information about the fluxes as measuring every metabolite in the network would give. In this paper we formalize this problem and derive algorithms for finding such optimal subsets of metabolites to measure for a flux estimation method by Rousu *et al.* [20]. The methods are facilitated by the recent flow analysis method of [19], that enables us to discover redundancies within sets of measurements of metabolites. Practically interesting variations of the problem analyzed are discussed in Section 4.5.

The structure of the paper is the following. Section 2 gives a short introduction to a linear method for flux estimation using isotopomer information. Section 3 introduces the concept of fragment equivalence. In Section 4 we motivate the problem of selecting an optimal set of metabolites to measure, define the measurement optimization problem at hand, study its computational complexity and give exact and fast approximate algorithms for solving it. Section 5 presents the results of experiments conducted with a metabolic model of central carbon metabolism of *Saccharomyces cerevisiae*. A summary of the related work is given in Section 6, together with discussion on possible future research directions.

## 2 A LINEAR METHOD FOR METABOLIC FLUX ESTIMATION USING <sup>13</sup>C ISOTOPIC TRACERS

A metabolic network  $G = (\mathcal{M}, \mathcal{R})$  is composed of a set  $\mathcal{M} = \{M_1, \dots, M_m\}$  of metabolites and a set  $\mathcal{R} = \{\rho_1, \dots, \rho_n\}$  of reactions that perform their interconversions.

For the purposes of <sup>13</sup>C isotopic tracing, only carbon atoms are of interest. Thus, we represent a  $k$ -carbon metabolite as a set of carbon locations  $M = \{c_1, \dots, c_k\}$ . Fragments of metabolites are simply subsets  $F = \{f_1, \dots, f_h\} \subseteq M$  of the metabolite. A fragment  $F$  of  $M$  is denoted as  $M|F$ . By a slight abuse of notation, we denote by  $F$  and  $M$  also the corresponding physical pools of molecules that have the required molecular structure, respectively.

Isotopomers—different isotopic versions of the molecule  $M = \{c_1, \dots, c_k\}$ —are represented by binary sequences  $b = (b_1, \dots, b_k) \in \{0, 1\}^k$  where  $b_i = 0$  denotes a <sup>12</sup>C and  $b_i = 1$  denotes a <sup>13</sup>C in location  $c_i$ . Molecules that belong to the  $b$ -isotopomer of  $M$  are denoted by  $M(b)$ . Isotopomers of metabolite fragments  $M|F$  are defined in an analogous manner: a molecule belongs to the  $F(b)$ -isotopomer of  $M$ , denoted  $M|F(b_1, \dots, b_h)$ , if it has a <sup>13</sup>C atom in all locations  $f_j$  that have  $b_j = 1$ , and <sup>12</sup>C in other locations of  $F$ .

The isotopomer distribution  $D(M)$  of metabolite  $M$  gives the relative abundances  $0 \leq P_M(b) \leq 1$  of each isotopomer  $M(b)$  in the pool of  $M$  such that

$$\sum_{b \in \{0,1\}^{|M|}} P_M(b) = 1.$$

Isotopomer distribution  $D(M|F)$  of fragment  $M|F$  is defined analogously.

Reactions are pairs  $\rho_j = (\alpha_j, \lambda_j)$  where  $\alpha_j = (\alpha_{1j}, \dots, \alpha_{mj}) \in \mathbb{Z}^m$  is a vector of stoichiometric coefficients—denoting how many molecules of each kind are consumed and produced in a single reaction event—and  $\lambda_j$  is a carbon mapping describing the transition of carbon atoms in  $\rho_j$ . Bidirectional reactions are modeled as a pair of reactions. (For simplicity of presentation, we assume in this paper that the reactions have simple stoichiometries  $\alpha_{ij} \in \{-1, 0, 1\}$  and that the carbon mappings  $\lambda_j$  are bijections. Both of these restrictions, however, can be lifted without great difficulty.) Metabolites  $M_i$  with  $\alpha_{ij} < 0$  are called reactants and with  $\alpha_{ij} > 0$  are called products of  $\rho_j$ . A pathway from metabolite fragments  $\{F_1, \dots, F_p\}$  to fragment  $F'$  is a sequence of reactions that define a bijective (composite) mapping from all carbons of  $\{F_1, \dots, F_p\}$  to all carbons of  $F'$ .

For notational convenience it will be useful to distinguish between the subpools of a metabolite produced by different reactions and pathways. Therefore, we denote by  $M_{ij}$ ,  $j > 0$ , the subpool of  $M_i$  produced ( $\alpha_{ij} > 0$ ) or consumed ( $\alpha_{ij} < 0$ ) by reaction  $\rho_j$ . As metabolite molecules produced by different reactions are mixed in a metabolite pool it is not possible to directly measure isotopomer distributions of inflow subpools  $\{M_{ij} \mid \alpha_{ij} > 0\}$ . However, the isotopomer distributions of inflow subpools can be (in some cases) derived from the isotopomer distributions of the reactants of reactions  $\rho_j$  (see Section 3). All flux estimation methods that are based on tracing isotopomers, in some form or another, rely on this fact. Note further that the subpools related to the outflows of the junction are the same, that is  $M_i = M_{ij}$  for all  $j : \alpha_j < 0$ .

By  $M_{i0}$  we denote the subpool of  $M_i$  that is related to the external inflow ( $\beta_i < 0$ ) or external outflow ( $\beta_i > 0$ ) of  $M_i$ . We call the sources of external inflows external substrates. We denote by  $I_i = \{M_{ij} \mid \alpha_{ij} < 0\}$  the inflow and by  $O_i = \{M_{ij} \mid \alpha_{ij} > 0\}$  the outflow subpools of  $M_i$ . Subpools of fragments are defined analogously.

In flux estimation, the quantities of interest are the velocities  $v_j \geq 0$  of the reactions  $\rho_j$ , giving the number of reaction events of  $\rho_j$  per time unit. If the velocities  $v_j$  of reactions  $\rho_j \in \mathcal{R}$  and the sizes of metabolite pools stay constant over time, we say that the metabolic network is in a steady state. In such a state the metabolite balance

$$\sum_{j=1}^n \alpha_{ij} v_j = \beta_i \quad (1)$$

holds for each metabolite  $M_i$ , which tells that the rate of production and consumption of the intermediate metabolite is the same. Here  $\beta_i$  is the measured external inflow ( $\beta_i < 0$ ) or external outflow ( $\beta_i > 0$ ) of metabolite  $M_i$ . If in addition to the velocities the isotopomer distributions of metabolites also remain constant, the system is in an isotopomeric steady state. In such a state, the rate of production and consumption of each metabolite  $M_i$  satisfies the isotopomer balance

$$\sum_{j=1}^n \alpha_{ij} v_j P_{M_{ij}}(b) = \beta_i P_{M_{i0}}(b) \quad (2)$$

for any  $b \in \{0, 1\}^{|M_i|}$ .

The isotopomer distributions of the outflow subpools  $M_{ij}$  are always identical to the distribution of the whole metabolite pool  $M_i$  as we assume that reactions uniformly sample their reactant

pools. If, however, the pathways leading to a *junction metabolite*—a metabolite with more than one producer—manipulate the carbons of the metabolite differently, then the isotopomer distributions of the inflows ( $\alpha_{ij} > 0$ ) often have differences. Furthermore, (2) together with (1) gives linearly independent equations that constrain the fluxes, ideally so that a single (correct) flux distribution can be pinpointed. The stoichiometric linear system (1) alone can be underdetermined, hence the additional equations (2) are used.

Applying (2) suffers from two difficulties [20]:

1. The (mass spectrometric and NMR) measurements do not in general come in the form of fully determined isotopomer distributions, but as a set of linear *isotopomer constraints*

$$d_{ih} = \sum_b s_{bih} P_{M_i}(b), \quad (3)$$

for the metabolites  $M_i$ , where the coefficients  $s_{bih} \in \mathbb{R}$  depend on the measurement technique and the metabolite. For example, the constraints for a metabolite given by a mass spectrometric measurement depend on the fragmentation pattern of a metabolite in the tandem mass spectrometer and how many of the produced fragments have sufficiently high frequency to exceed the detection limit. NMR technology gives more direct access to relative frequencies of some isotopomers. In general, however, some isotopomers cannot be uncovered and the sensitivity is lower than that of a mass spectrometer.

Because of these practical hindrances, instead of (2), we will have to resort to a weaker form of balances

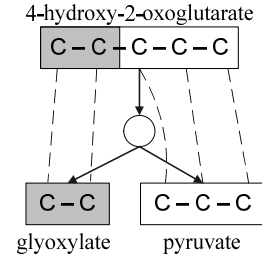
$$\sum_{j=1}^n \alpha_{ij} v_j d_{ih} = \beta_i d_{i0h}, \quad (4)$$

(for all metabolites  $M_i$ ) that at worst only contain the metabolite balance equations (1) and in the best case, meet (2).

2. With current technologies, not all metabolites can be measured, so not all isotopomer frequencies are available. This calls for methods that can be used to derive isotopomer frequencies or their combinations from measurements made for other metabolites in the network. In [20], a general methodology was proposed, where measurements of the form of (3) can be propagated in between two junction metabolites to obtain the constraints of the form of (4) for the junction metabolite with as many linearly independent equations as possible. The method relies on the fact that in individual reactions—and in general pathways with no junction metabolites—one can compute isotopomer constraints to products from isotopomer constraints (3) of reactants, and vice versa, by using vector space operations and the carbon maps.

### 3 FRAGMENT EQUIVALENCE

In [19], a flow analysis method was developed that extends the scope of propagation to make it possible to propagate information through the junction metabolites. The idea is to find fragments  $M|F$  and  $M_i|F'$  that are *equivalent* (denoted by  $M|F \equiv M_i|F'$ ) in the sense that in all isotopomeric steady states their isotopomer distributions are the same, no matter what kind of labelings are used for the external substrates. Intuitively, the source fragment  $M|F$  and the target fragment  $M_i|F'$  are equivalent if all pathways producing  $F'$  from



**Fig. 1.** An example of a metabolic reaction. In 4-hydroxy-2-oxoglutarate glyoxylate-lyase reaction a 4-hydroxy-2-oxoglutarate ( $C_5H_6O_6$ ) molecule is split into pyruvate ( $C_3H_4O_3$ ) and glyoxylate ( $C_2H_2O_3$ ) molecules. Carbon maps are shown with dashed lines. The gray fragment of 4-hydroxy-2-oxoglutarate is equivalent to glyoxylate and the white fragment is equivalent to pyruvate.

the fragments of external substrates go through  $F$ , in all pathways from  $F$  to  $F'$  carbons of  $F$  stay intact and all pathways have the same carbon maps between  $F$  and  $F'$ . We assume that if a metabolic reaction transfers a fragment from a reactant to a product, then the carbons of the fragment stay intact in the reaction. Now, because  $F$  is always a precursor of  $F'$ , carbons of  $F$  always travel intact to  $F'$  and the fragments  $F$  are similarly oriented when reaching  $F'$  via every pathway (carbon maps are the same), the isotopomer distribution of  $F'$  is equal to the one of  $F$  in every steady state and with every labeling of external substrates.

The basic example of such equivalence is that of reactant and product fragments of a single reaction, which easily follows from the assumed intactness of fragments in a single reaction (see Figure 1):

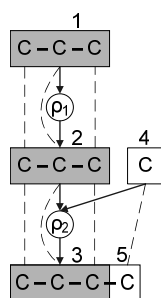
**LEMMA 1.** Let  $\rho_j = (\alpha_j, \lambda_j)$  be a reaction with a reactant  $M$  and a product  $M_i$ . If fragment  $M|F$  satisfies  $\lambda_j(F) \subset M_{ij}$ , then  $M|F \equiv M_{ij}|\lambda_j(F)$ .

If  $M_i$  is not a junction metabolite, that is, it has only one inflow, then  $M_{ij} = M_i$  in Lemma 1, and we have  $M|F \equiv M_i|\lambda_j(F)$ . By the transitivity of the equivalence relation  $\equiv$ , the result can be applied to an *unbranched pathway* i.e., to a pathway that does not contain two reactions producing the same metabolite. Here, intactness of the fragment is ensured by requiring that the image of the fragment belongs to a single metabolite in each intermediate step of the pathway (see Figure 2):

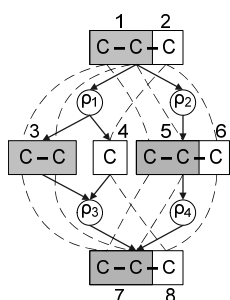
**LEMMA 2.** Let  $M, M_{i_h}, 1 \leq h \leq r$ , be the metabolites and  $\rho_{j_h}, 1 \leq h \leq r$ , the reactions of a pathway. Let  $M$  be a reactant of  $\rho_{j_1}$  and let  $\rho_{j_h}$  be the sole producer of  $M_{i_h}$ , and denote by  $\Lambda^h = \lambda_{j_h} \circ \dots \circ \lambda_{j_1}$  the composite carbon mapping of the pathway  $(\rho_{j_1}, \dots, \rho_{j_h})$ . If for some fragment  $F$  of  $M$ ,  $\Lambda^h(F) \subseteq M_{i_h}$  for each  $1 \leq h \leq r$ , then  $M|F \equiv M_{i_r}|\Lambda^r(F)$ .

For a source fragment  $M|F$  and a target fragment  $M_i|F'$  having several pathways in between them, it is further required that  $F$  must always be a precursor of  $F'$  and that the composite carbon mappings of the pathways are the same (see Figure 3).

**LEMMA 3.** Let  $R_1, \dots, R_p$  be the set of unbranched pathways connecting  $M$  and  $M_i$  with associated composite carbon mappings  $\Lambda^1, \dots, \Lambda^p$  and let  $M_{ik}$  denote the subpool of  $M_i$  produced by  $R_k$ . For fragment  $F = (f_1, \dots, f_h)$  of  $M$ , if  $M|F \equiv M_{ik}|\Lambda^k(F)$  for  $1 \leq k \leq p$ ,  $\Lambda^1(f_t) = \dots = \Lambda^p(f_t)$  for each  $1 \leq t \leq h$  and every



**Fig. 2.** An example of equivalence between fragments in an unbranched pathway. Every metabolite in the pathway has only one producer and fragment 1 travels intact to fragment 2 which travels intact to fragment 3, so  $1 \equiv 2 \equiv 3$ . Also,  $4 \equiv 5$ .



**Fig. 3.** An example of equivalence between fragments in a branched pathway. Fragment 7 is produced by two unbranched pathways ( $\rho_1, \rho_3$ ) and ( $\rho_2, \rho_4$ ) from precursor fragment 1. Composite carbon mappings from 1 to 7 are the same in both pathways, so  $1 \equiv 3 \equiv 5 \equiv 7$ . Also,  $2 \equiv 4 \equiv 6 \equiv 8$ .

pathway producing  $F'$  from some fragments of external substrates contains  $F$  then  $M|F \equiv M_i|F'$ .

The equivalence relation defined with above lemmas is symmetric and transitive. As every fragment is equivalent to itself, the equivalence partitions a metabolic network to *equivalence classes* of fragments with equivalent isotopomer distributions. The equivalence classes of fragments can be efficiently (in polynomial time) found by constructing the fragment flow graph of a metabolic network and applying dominator tree analysis [14] to it [19]. Briefly, in a dominator tree constructed from the fragment flow graph a fragment  $F'$  is a descendant of  $F$  if and only if all pathways from the fragments of external substrates to  $F'$  contain  $F$  and carbons of  $F$  always travel intact and similarly oriented to  $F'$ .

For flux estimation, these equivalences serve in several roles: first, the fragment isotopomer distributions of the subpools of a junction can differ only if the fragment is not equivalent to any fragment in its reactants. It is only those fragment isotopomers that can potentially induce linearly independent constraints to the fluxes. This gives us possibilities to remove redundant equations from the flux system to be solved. Second, we can use interchangeably any isotopomer measurement of two equivalent fragments, thus enabling us to form balance equations in junctions where adjacent metabolites are poorly or not at all measured. As by Lemma 3 the equivalence classes may extend beyond junctions, the technique improves the propagation methods described in [20].

In this paper, these equivalences are utilized in yet another way. Namely, they turn out to be powerful tools for selecting a small subset of metabolites to be measured so that the fluxes of a metabolic network can still be discovered from the measurements.

## 4 MEASUREMENT OPTIMIZATION IN THE POSITIONAL ENRICHMENT CASE

Using the linear system consisting of equations (1) and (4) for solving the fluxes  $\rho_j$  is not straightforward. There are several problems. First of all, the system should be of full rank to give a point solution instead of just some linear constraints to the fluxes. Full rank means that there should be sufficiently many linearly independent equations. In the extreme case it is possible that full rank is not achievable at all (for example, if a one-carbon metabolite has a large number of producers).

The independence of equations depends on the actual values of the isotopomer abundances which in turn depend intricately on the isotopomer abundances of the external substrates and on the actual fluxes in the steady state under consideration. Hence the quality of the equation system obtained depends on the measurements performed as well as on the  $^{13}\text{C}$  labeling patterns used for external substrates.

On the other hand, the NMR and MS methods producing isotopomer data are not trivial either. In (tandem) mass spectrometer it is necessary to separate metabolites in the cell extract and develop for each metabolite or metabolite group a specific experimental protocol. In NMR, spectral overlap, large differences in concentration and available spectral edition techniques make fractional positional labeling of some metabolites more accessible than others. Therefore there is a need to design an optimized measurement strategy that minimizes the experimental effort. Here the equivalence concept of Section 3 can be utilized: measuring more than one representative of an equivalence class does not add new information; the distribution observed for one member of the class can also be used in association with the others to write equations (4).

In the rest of the paper we will consider the measurement optimization problem in a special case (positional enrichment) satisfying rather strong but experimentally justified conditions. Even in this case the optimization problem turns out to be computationally hard but still tractable in practice. Solutions to the problem given below can be used to guide an iterative experiment planning process towards a small set of metabolites to measure. Our computational techniques can also be generalized for less constrained situations, for example to fractional enrichments of larger fragments [4, 10] or to the mass isotopomers of whole metabolites [31]. Hence the present exercise is of wider interest.

### 4.1 Optimization problems

In the so-called *positional enrichment measurements* of  $^{13}\text{C}$ , one observes the  $^{13}\text{C}$  isotopomer distribution of an individual carbon  $c_h$  of a metabolite  $M$  which simply is the relative abundance of  $^{13}\text{C}$  in  $c_h$ . Positional enrichment data can be obtained by NMR [8, 16, 22, 24, 25, 33] or tandem mass spectrometry. Also GC-MS with different derivatisation techniques can provide such data for many metabolites [28]. Thus the availability of positional enrichment data for the carbons of some metabolites in our network is a reasonable – and computationally convenient – first approximation in the early stages of an experiment process when more detailed information

about the measurable constraints to the isotopomer distributions is not yet available.

Formally, we assume that one  $^{13}\text{C}$  positional enrichment measurement gives for a metabolite  $M = \{c_1, \dots, c_k\}$  the  $^{13}\text{C}$  labeling frequencies  $P_{M|c_h}(1)$  and  $P_{M|c_h}(0)$  for each carbon  $c_h$  of  $M$ . Note that

$$P_{M|c_h}(1) = \sum_{b \in \{0,1\}^{|M|}: b_h=1} P_M(b),$$

i.e.,  $P_{M|c_h}(1)$  is the full isotopomer distribution of  $M$  marginalized to  $b_h = 1$ .

With the positional enrichment data available for all metabolites  $M$ , we can infer by Lemma 1 the positional enrichment frequencies  $P_{M_{i,j}|c_h}(1)$  for all subpools  $M_{i,j}$  of all junction metabolites  $M_i$ . Then the generalized isotopomer balances (4) get the form

$$\sum_{j=1}^n \alpha_{ij} v_j P_{M_{i,j}|c_h}(1) = \beta_i P_{M_{i0}|c_h}(1). \quad (5)$$

Note here that the corresponding equation for  $b_h = 0$  is linearly dependent of equations (1) and (5) as  $P_{M|c_h}(0) = 1 - P_{M|c_h}(1)$ .

Based on the positional enrichment data, we can thus write  $|M_i|$  new equations, in addition to the mass balance (1), for each junction metabolite  $M_i$ . This is the strongest system of equations we can hope to obtain from positional enrichment measurements. (Note, however, that there is no guarantee that this system would allow solving all the fluxes: the system may be underdetermined in some junctions and overdetermined in some others.)

Now it should be clear that because of the equivalence of carbons of different metabolites, measuring all metabolites may sometimes be redundant. Some subset would already allow us to write the full set of  $|M_i|$  equations (5) for each junction metabolite  $M_i$ . This leads to the following optimization problem:

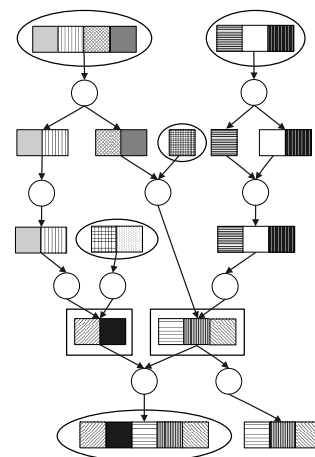
**Problem 1** (Positional enrichment measurement (PEM) minimization problem). Given a metabolic network  $G = (\mathcal{M}, \mathcal{R})$ , find a smallest set of metabolites to measure for positional enrichment data such that, noting the equivalences of the carbons, it is possible to write a full set of  $|M_i|$  equations (5) for each junction metabolite  $M_i$  of the network  $G$ .

Figure 4 gives an example of an optimal solution to the PEM minimization problem in a small metabolic network consisting of 13 metabolites and ten reactions. There exist two junction metabolites consisting of three and two carbons in the example (enclosed by boxes). In order to write the full set of five balance equations (5) for these carbons, the positional enrichment data is required for the junction carbons and for all the carbons of the reactants of reactions that produce junction metabolites. In the example this data can be derived, with the help of equivalences, from five metabolites (enclosed by ovals). So instead of measuring all 13 metabolites it is sufficient to measure only five of them.

## 4.2 Solution by set covering techniques

By its combinatorial nature, the PEM minimization is a variant of the well-known minimum set cover problem:

**Problem 2** (Minimum set cover [3]). Given a collection  $\mathcal{C}$  of subsets of a finite set  $S$ , find a smallest subcollection  $\mathcal{C}'$  of  $\mathcal{C}$  such that  $\bigcup \mathcal{C}' = S$ .



**Fig. 4.** An optimal solution to the PEM minimization problem in a small example network. Positional enrichment data is required for all carbons of the two junction metabolites (enclosed by boxes) and for all reactants of reactions producing the junction metabolites. Carbon positions having the same colouring pattern are equivalent. An optimal solution to the problem consists of metabolites circled with ovals.

To see the connection between PEM and set cover, observe that an equation (5) can be written as soon as we know  $P_{M|c_h}(1)$  for all subpools  $M_{i,j}$  of  $M_i$ . This is the case when we have measured for each  $M_{i,j}$  some carbon  $c_t$  of some metabolite  $M_r$  such that  $M_r|c_t \equiv M_{i,j}|c_h$ . We now say that measuring  $M_r$  covers a subpool carbon  $M_{i,j}|c_h$  of a junction  $M_i$  if  $M_{i,j}|c_h \equiv M_r|c_t$  for some carbon  $c_t$  of  $M_r$ . The metabolites  $M_r$  define in this way the collection of subsets of the set cover instance. The basic set of the instance consists of all one-carbon subpools of the junction metabolites. Then we get the following technical formulation of PEM minimization:

**LEMMA 4.** A set  $\mathcal{M}' \subseteq \mathcal{M}$  of metabolites is a solution of the PEM minimization if  $\mathcal{M}'$  is a smallest set such that it covers all subpool carbons  $M_{i,j}|c_h$  of all junction metabolites  $M_i$ . Furthermore, each set  $\mathcal{K} \subseteq \mathcal{M}$  covers certain subpool carbons.

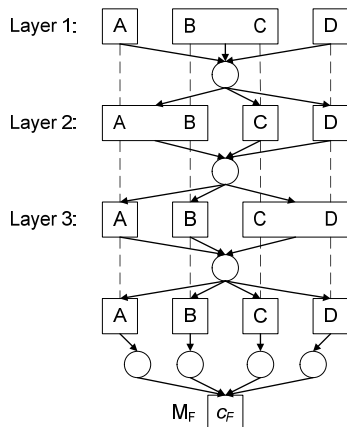
The NP-hardness and the inapproximability of our problem can be shown by a polynomial-time reduction from the set cover:

**THEOREM 1.** PEM minimization is NP-hard.

**PROOF.** Let  $(\mathcal{C}, S)$  be an instance of set cover.

We will describe a polynomial-time construction of a metabolic network  $G(\mathcal{C}, S)$  that satisfies the following claim: the PEM minimization of  $G(\mathcal{C}, S)$  has solution of size  $k + 1$  if and only if  $(\mathcal{C}, S)$  has a cover of size  $k$ .

Network  $G(\mathcal{C}, S)$  will have only one junction metabolite, denoted  $M_F$ . Metabolite  $M_F$  has only one carbon  $c_F$  which is produced along  $|S|$  separate paths induced by the carbon maps. Each path corresponds to an element of  $S$ . The paths go through metabolites that correspond to the sets in the collection  $\mathcal{C}$ . The PEM minimization is then equivalent to finding the smallest number of metabolites that intersect all paths. This is equivalent to finding the smallest subcollection of  $\mathcal{C}$  that covers entire  $S$ . (For an illustration, see Figure 5.)



**Fig. 5.** Reduction of an example instance ( $\mathcal{C} = \{\{B, C\}, \{A, B\}, \{C, D\}\}$ ,  $S = \{A, B, C, D\}$ ) of Set Cover to an instance of PEM minimization. For clarity, we denote each carbon by the corresponding element of  $S$ .

More technically, let  $S = \{s_1, \dots, s_n\}$  and  $\mathcal{C} = \{C_1, \dots, C_m\}$ . The metabolites of the network  $G(\mathcal{C}, S)$  form  $m + 2$  layers. Layers  $1, \dots, m$  represent sets  $C_1, \dots, C_m$  while layer  $m + 1$  is an extra technical layer, and layer  $m + 2$  is the final layer containing only the junction metabolite  $M_F$ . The metabolites on layer  $i$ ,  $i < m + 2$ , consist of carbons  $c_{ij}$ ,  $1 \leq j \leq n$ , that correspond to the elements of  $s_j$  of  $S$  such that  $c_{ij}$  corresponds to  $s_j$ . The carbons are grouped into metabolites such that  $\{c_{ij} : s_j \in C_i\}$  is a metabolite representing  $C_i$ , and the remaining carbons  $c_{ij}, s_j \notin C_i$ , each form a one-carbon metabolite. The carbons  $c_{m+1,j}$  of layer  $m + 1$  form  $n$  separate one-carbon metabolites. The reaction  $\rho_i$ ,  $1 \leq i \leq m$ , of  $G(\mathcal{C}, S)$  transforms the metabolites of layer  $i$  into metabolites of layer  $i + 1$ , the associated carbon map taking carbon  $c_{ij}$  to  $c_{i+1,j}$  for  $1 \leq j \leq n$ . Finally, reaction  $\rho'_j$ ,  $1 \leq j \leq n$ , transforms the metabolite of carbon  $c_{m+1,j}$  to metabolite  $M_F$  (which has carbon  $c_F$  only).

Obviously, the inflow subpool of  $M_F$  from reaction  $\rho'_j$  and all one-carbon fragments  $c_{ij}$ ,  $1 \leq i \leq m + 1$ , are equivalent (and correspond to  $s_j$ ). Hence to solve PEM for  $G(\mathcal{C}, S)$  we need to find metabolites that cover all these equivalence classes as well as the outflow from  $M_F$  (Lemma 4). Noting that the smallest such covering set of metabolites can always be selected from metabolites representing  $\mathcal{C}$  (plus metabolite  $M_F$  which must always be included), the smallest PEM solutions of  $G(\mathcal{C}, S)$  are in one-to-one correspondence, the former being larger by one. Hence,  $G(\mathcal{C}, S)$  satisfies the claim.  $\square$

In the reduction given in the proof of Theorem 1 there is a bijective mapping between the elements of the set  $S$  and the equivalence classes. Let  $|\mathcal{E}_G|$  be the number of equivalence classes in the metabolic network  $G$ . As the minimum set cover problem is known to be hard to approximate within a factor  $c' \log |S|$  for some  $c' > 0$  (see [3]), we get the following inapproximability result:

**COROLLARY 1.** *The PEM minimization problem is hard to approximate within a factor  $c \log |\mathcal{E}_G|$  for some  $c > 0$ , unless  $P = NP$ .*

On the other hand, the PEM minimization problem can be reduced to the minimum set cover problem in polynomial time, as

pointed out by Lemma 4. It is known that the well-known Greedy Set Cover algorithm is a factor  $1 + \ln |S|$  approximation algorithm for Problem 2, see [3]. Hence, we get the following result:

**COROLLARY 2.** *The greedy set cover algorithm finds for the PEM minimization problem a solution which is within a factor of  $1 + \ln |\mathcal{E}_G|$  from the optimum.*

The number of equivalence classes in the metabolic network can be bounded above by e.g. the number of carbons.

### 4.3 Solution by integer linear programming

Most metabolic network models used in flux estimation contain only a few dozens of metabolites and reactions. Thus, methods for obtaining the optimal metabolite sets might be of interest, even if they have exponential worst-case time complexity. One versatile approach is to model the problem as a mixed integer linear program (MILP), i.e., as a minimization of some linear objective function in a polytope, with the additional constraint that certain variables in the optimal solutions are integral (see [15] for further details).

The objective function to minimize is the sum of costs of the metabolites that provide the maximum positional enrichment information (see Problem 1). Let  $m_1, \dots, m_{|\mathcal{M}|}$  be indicator variables whether or not the metabolite  $M_i \in \mathcal{M}$  is measured. Let  $w_i$  be the cost of measuring the metabolite  $M_i$ . Thus, the objective function to be minimized is

$$\min_{m_1, \dots, m_{|\mathcal{M}|}} \sum_{i=1}^{|\mathcal{M}|} w_i m_i$$

with the constraints  $m_M \in \{0, 1\}$  for each metabolite  $M \in \mathcal{M}$ . The other constraints are as follows.

The requirement of Problem 1 that we have to write a balance equation for every carbon  $c \in M_i$  can sometimes lead to sets of measured metabolites that are too laborious to measure in practice. In that case we can try to use MILP to find less expensive solutions by requiring for each junction  $M_i$  only  $k_i \leq |M_i|$  balance equations. The cost of this relaxation is that we might lose some independent balances constraining the fluxes in (5).

Let  $x_{i,c}$  be the indicator variable that the balance equation can be written for a carbon  $c \in M_i$ . Then the constraints can be stated as

$$\sum_{c \in M_i} x_{i,c} - k_i \geq 0.$$

The balance equation can be written for a carbon  $c \in M_i$  if for all corresponding carbons  $c_{ij}$  in the inflow subpools  $M_{ij}$  and for carbon  $c$  there is a measured metabolite  $M'$  with equivalent carbon  $c'$ . Let  $R_{i,c}$  be the set of reactions with the carbon  $c$  as a product. Let  $E_{i,c,j}$  be the set of indices  $p$  of metabolites that have a carbon equivalent to the inflow subpool carbon  $c_{ij}$  produced by the reaction  $\rho_j$ , and let  $E_{i,c}$  be the set of indices  $p$  of metabolites that have a carbon equivalent to  $c$ . Now

$$\sum_{p \in E_{i,c,j}} m_p - x_{i,c} \geq 0$$

and

$$\sum_{p \in E_{i,c}} m_p - x_{i,c} \geq 0$$

must hold.

Combining these constraints we obtain the following mixed integer linear program (we assume  $k_i = 0$  for each non-junction metabolite  $M_i$ ):

$$\begin{aligned}
 \min_{m_1, \dots, m_{|\mathcal{M}|}} & \sum_{i=1}^{|\mathcal{M}|} w_i m_i \\
 \text{s.t.} & \sum_{c \in M_i} x_{i,c} - k_i \geq 0 & \forall M_i \in \mathcal{M} \\
 & \sum_{p \in E_{i,c,j}} m_p - x_{i,c} \geq 0 & \forall \rho_j \in R_{i,c}, c \in M_i \in \mathcal{M} \\
 & \sum_{p \in E_{i,c}} m_p - x_{i,c} \geq 0 & \forall c \in M_i \in \mathcal{M} \\
 & m_i \in \{0, 1\} & \forall M_i \in \mathcal{M} \\
 & x_{i,c} \in \{0, 1\} & \forall c \in M_i \in \mathcal{M}
 \end{aligned}$$

The number of variables in the program is

$$|\mathcal{M}| + \sum_{M_i \in \mathcal{M}} |M_i|$$

and the number of inequalities is

$$|\mathcal{M}| + \sum_{M_i \in \mathcal{M}} \sum_{c \in M_i} (|R_{i,c}| + 1).$$

If the number of integer variables is too high, the integer linear program can be relaxed to a linear program, i.e., the requirement of the variables being binary can be relaxed to the requirement that the variables have their values in the unit interval  $[0, 1]$ . In fact, the constraints  $x_{i,c} \in \{0, 1\}$  can be relaxed to  $x_{i,c} \in [0, 1]$  without affecting the solution since the cost of the solution is minimized when all variables  $x_{i,c}$  are either 0 or 1; the values of  $x_{i,c}$  can be replaced by  $\lceil x_{i,c} \rceil$  without violating the constraints or increasing the cost of the solution. If the constraints  $m_i \in \{0, 1\}$  are relaxed, then the obtained solutions are not necessarily integral, but the solution can be transformed to (possibly suboptimal) integral solution by randomized rounding techniques. For example, the following procedure can be applied:

1. Find the optimal solution for the linear program. If the values of all variables  $m_i$  are integral, output the solution and halt.
2. Choose one variable  $m_i$  with non-integral value randomly with probability proportional to its value and replace the occurrences of the variables  $m_i$  in the linear program by the constant 1 and go to step 1.

#### 4.4 Full Isotopomer data

It is possible to modify the above techniques of PEM minimization for other types of measurement data. At the other end of the spectrum is the full isotopomer data, i.e., the distributions  $P_M(b)$  for full metabolites  $M$ . Similar set cover and MILP algorithms can be used in this case as well.

Note that, in the PEM case, the equivalence classes are the largest possible (because they correspond to the smallest possible, one-carbon, fragments) and hence their number is the smallest possible which leads to a small number of measurements. For the full metabolite isotopomer data the sets are smaller and hence a larger

number of measurements may be necessary. However, full metabolite measurements give more data per measurement and therefore allow writing more equations (4).

#### 4.5 Practical extensions

The relevant cases in which positional enrichment data is available only for some but not all carbons of the metabolites or more than one measurement per equivalence class is required can be handled with straightforward modifications to the techniques given above. In the first case, the unmeasurable carbons of metabolite  $M$  are disregarded when the cover of  $M$  is computed. For example, this way a situation where tandem MS produces positional enrichments only for some carbons of a metabolite can be modelled. In the latter case, we require more than one carbon to be measured from an equivalence class before it is considered as covered. By setting high costs to metabolites overlapping in the NMR spectrum one can test whether the separation of these metabolites is necessary or it is possible to derive the same isotopomer information from some other metabolites that are easier to measure. Our techniques can also be used to find out how many equations (4) per each junction can be written, given a set of metabolites whose positional enrichments are measurable.

Above we study the problem of selecting small sets of metabolite to measure that make possible to write as many generalized balance equations (4) as the measurement of every metabolite would allow. However, due to the dependencies in the basic equation system (1), some equations (4) might be redundant. By applying standard techniques of linear algebra to the equation system (1) it is possible to find subsets of fluxes whose values alone would suffice to fix the whole flux distribution [11]. Such sets are called *free fluxes*. Now, if the number of metabolites to measure proposed by the above methods is still too high, we can enumerate different minimal sets of free fluxes and for each such set require that the equations (4) are written only to junction carbons that are produced by some free flux. In the best case, that is, with good labeling of external reactants and small measurement errors, equations bounding only free fluxes are enough to solve all the fluxes in the network.

## 5 COMPUTATIONAL EXPERIMENTS

We tested our method for finding fragment equivalence sets with the model of central carbon metabolism of *Saccharomyces cerevisiae* containing glycolysis, pentose phosphate pathway and citric acid cycle. Carbon mappings were provided by the ARM project [2]. The network consisted of 31 metabolites and 35 reactions of which 15 were bidirectional. Bidirectional reactions were modeled as two unidirectional reactions. The rank of the linear equation system (1) for the 50 fluxes of the model network was 30. Cofactor metabolites not accepting or donating carbon atoms were excluded from the analysis. The only carbon source of the network was glucose. There were two external products in the model. Twenty metabolites were produced by more than one reaction and thus formed junctions. Visualizations of the metabolic network used in the experiments are available at [http://www.cs.helsinki.fi/group/sysfys/images/bioinformatics\\_metabolic\\_net.pdf](http://www.cs.helsinki.fi/group/sysfys/images/bioinformatics_metabolic_net.pdf).

In this demonstration of the computational methods given above we assumed that either positional enrichment or full isotopomer information was available for every metabolite in the network. We

also assumed that the effort needed to measure the isotopomer information is equal for every metabolite. In the real experimental design situations these parameters can be easily set to correspond the task at hand (see Section 4.5). Two alternative forms of balance equation systems were analyzed: (a) a balance equation should be written for every carbon of a junction metabolite (there were 45 such carbons in an example model) or (b) a balance equation should be written only for (*the number of producing reactions - 1*) of equations for each junction (this would ask for writing 21 equations in the example). The equations were not generated for carbons known to be uninformative by fragment equivalence analysis [19]. The minimum sets of metabolites to measure were found by Greedy Set Cover algorithm and MILP programs with guaranteed optimal solution. (In the case (b) only MILP technique is applicable.) MILP programs were solved by using the publicly available MILP solver `lp_solve 5.5` package ([http://groups.yahoo.com/group/lp\\_solve/](http://groups.yahoo.com/group/lp_solve/)).

The results of the tests are summarized in Table 1. Solutions to the MILP programs were computed instantaneously with a PC with 2,4 GHz Pentium 4 processor, except for the fifth problem of Table 1 that took 25 seconds to finish. According to our tests it is enough to measure 16 metabolites in the model to be able to construct as many balance equations as the measurement of every metabolite would allow. If (*the number of producing reactions - 1*) equations for each junction were constructed, 14 metabolites have to be measured. The actual number of metabolites to measure is sensitive to changes in the metabolic network. In a simpler model of central carbon metabolism of *Saccharomyces cerevisiae* that had more one-directional reactions it was enough to measure only one fourth of the metabolites.

The fragment equivalence analysis (see Section 3) revealed five junction metabolites whose isotopomer distributions are uninformative for flux estimation. This implies that not all the fluxes of the model can be solved by utilizing only isotopomer information, regardless of the estimation method, the quality measurement data and input substrate labeling. The sets of metabolites to measure contained pentose phosphates ribose 5P, xylulose 5P and ribulose 5P that are hard to measure with current techniques. Unfortunately their information is needed to write as many generalized balance equations as the measurement of every metabolite would allow.

In our experiments Greedy Set Cover algorithm found equally good solutions as the MILP solver. Optimal sets for the positional enrichment information were found equal to the optimal sets for full isotopomer data indicating certain robustness against assumptions about the measurement data. There existed many optimal solutions consisting mostly of the same key metabolites in all the test cases.

## 6 DISCUSSION AND CONCLUSION

To the authors' best knowledge the problem of selecting an optimal set of metabolites to measure for flux estimation has not been systematically studied in the context of direct flux estimation. However, the selection of metabolites to measure is related to the studies of structural flux identifiability analysis that examines if a set of measurements suffices to determine all the fluxes of metabolic network in the context of iterative flux estimation. In this area van Winden *et al.* [30] give analytical methods to study the information content of positional enrichment and the so-called cumomer data [33] before the actual measurements are carried out. They

**Table 1.** Sizes of the minimum sets of metabolites whose isotopomer information is required to measure in different settings.

measurement	target	algorithm	# of metabolites
full	(a)	greedy	16
full	(a)	MILP	16
pos	(a)	greedy	16
pos	(a)	MILP	16
pos	(b)	MILP	14

First column (measurement) indicates whether full isotopomer distribution (full) or positional enrichment information (pos) was assumed to be available, second column shows whether equation system (a) or (b) was constructed. Third column indicates whether greedy algorithm or MILP program was used to obtain the solution. Finally, the last column gives the sizes of minimum sets of metabolites out of 31 whose isotopomer information is required.

also present methods for checking the structural identifiability in the case of mass isotopomer or NMR multiplet data. Isermann and Wiechert [11] give analytical and numerical methods for structural identifiability problem in the case of full isotopomer information. For the case of arbitrary measurement information they give an algorithm that checks whether an iterative flux fitting algorithm will converge to a (possibly suboptimal) point solution. They also study information content of single metabolites.

The selection of the optimal set of metabolites to measure covers only one half of the experimental design of isotopomer tracing experiments. The other half is the selection of the labeling mixture of external substrates. Together with the actual flux distribution the labeling of external substrates induce the isotopomer distributions of every metabolite in the network thus having a strong effect on the linear independence of isotopomer balance equations. The selection of optimal labeling in the context of iterative flux estimation methods is studied by Möllney *et al.* [17] and by Araúzo-Bravo and Shimizu [1].

An interesting future work is to utilize the free flux approach in our context as well as to look for methods that combine the contribution of this article with techniques that propose optimal labelings of substrates in the direct flux estimation context of [20]. Moreover, by applying branch & bound algorithm and isotopomer constraint manipulation techniques introduced in [20] it is possible to extend the metabolite selection to the case where we know what kind of constraints can be measured to the isotopomer distribution of each metabolite. From an experimental point of view it would also be useful to develop an experimental design method that tries to falsify the given model of a metabolic network by proposing a cost effective set of measurements of metabolite fragments whose isotopomer distributions should be equal if the model were correct.

## Acknowledgement

The authors thank the reviewers of this paper for many useful comments on the manuscript, Esa Pitkänen and Paula Jouhten for fruitful discussions, and Marina Kurtén for correcting the language. This work was supported by grant 203668 from the Academy of Finland (SYSBIO program). Juho Rousu was also supported by

the EU/Marie Curie Individual Fellowship grant (HPMF-CT-2002-02110) and Taneli Mielikäinen by APrIL II (FP6-508861).

## REFERENCES

- [1] M. Araúzo-Bravo and K. Shimizu. An improved method for statistical analysis of metabolic flux analysis using isotopomer mapping matrices with analytical expressions. *Journal of Biotechnology*, 105:117–133, 2003.
- [2] Automatic reconstruction of metabolism. <http://www.metabolome.jp/index.html>.
- [3] G. Ausiello, P. Crescenzi, Kann V., A. Marchetti-Spaccamela, and M. Protasi. *Complexity and Approximation: Combinatorial Optimization Problems and Their Approximability Properties*. Springer, 1999.
- [4] L. Blank and U. Sauer. Uca cycle activity in *saccharomyces cerevisiae* is a function of the environmentally determined specific growth and glucose uptake rates. *Microbiology*, 150:1085–1093, 2004.
- [5] L. Boros, N. Serkova, M. Cascante, and W-N. Lee. Use of metabolic pathway flux information in targeted cancer drug design. *Drug Discovery Today: Therapeutic Strategies*, 1(4):435–443, 2004.
- [6] B. Christensen and J. Nielsen. Isotopomer analysis using GC-MS. *Metabolic Engineering*, 1:E8–E16, 1999.
- [7] E. Fisher, Zamboni N., and U. Sauer. High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived  $^{13}\text{C}$  constraints. *Analytical Biochemistry*, 325:308–316, 2004.
- [8] B. Follstad and G. Stephanopoulos. Effect of reversible reactions on isotope label redistribution analysis of the pentose phosphate pathway. *European Journal of Biochemistry*, 252:360–371, 1998.
- [9] S. Ghosh, T. Zhu, I.E. Grossmann, M.M. Ataa, and M.M. Domach. Closing the loop between feasible flux scenario identification for construct evaluation and resolution of realized fluxes via nmr. *Journal of Bacteriology*, 183:1441–1451, 2001.
- [10] A.K. Gombert, M. Moreira dos Santos, B. Christensen, and J. Nielsen. Network identification and flux quantification in the central metabolism of *saccharomyces cerevisiae* under different conditions of glucose repression. *Computers and Chemical Engineering*, 29:459–466, 2005.
- [11] N. Isermann and W. Wiechert. Metabolic isotopomer labeling systems. part ii: structural identifiability analysis. *Mathematical Biosciences*, 183:175–214, 2003.
- [12] J. Kelleher. Flux estimation using isotopic tracers: Common ground for metabolic physiology and metabolic engineering. *Metabolic Engineering*, 3:100–110, 2001.
- [13] S. Klamt and S. Schuster. Calculating as many fluxes as possible in underdetermined metabolic networks. *Molecular Biology Reports*, 29:243–, 2002.
- [14] T. Lengauer and R. Tarjan. A fast algorithm for finding dominators in a flowgraph. *ACM Transactions on Programming Languages and Systems*, 1:121–141, 1979.
- [15] A. Martin. General mixed integer programming: Computational issues for branch-and-cut algorithms. In Michael Jünger and Denis Naddef, editors, *Computational Combinatorial Optimization: Optimal and Provably Near-Optimal Solutions*, volume 2241 of *Lecture Notes in Computer Science*, pages 1–25. Springer, 2001.
- [16] A. Marx, A. de Graaf, W. Wiechert, L. Eggeling, and H. Sahm. Determination of the fluxes in the central metabolism of *corynebacterium glutamicum* by nuclear magnetic resonance spectroscopy combined with metabolite balancing. *Biotechnology and Bioengineering*, 49:111–129, 1996.
- [17] M. Möllney, W. Wiechert, D. Kownatzki, and A. de Graaf. Bidirectional reaction steps in metabolic networks IV: Optimal design of isotopomer labeling systems. *Biotechnology and Bioengineering*, 66:86–103, 1999.
- [18] J. Nielsen. It is all about metabolic fluxes. *Journal of Bacteriology*, 185, 2003.
- [19] A. Rantanen, J. Rousu, E. Pitkänen, H. Maaheimo, and E. Ukkonen. Flow analysis of metabolite fragments for flux estimation. In *Proceedings of the 3rd International Workshop on Computational Methods in Systems Biology, CMSB 2005, Edinburgh*, pages 242–255, 2005.
- [20] J. Rousu, A. Rantanen, H. Maaheimo, E. Pitkänen, K. Saarela, and E. Ukkonen. A method for estimating metabolic fluxes from incomplete isotopomer information. In *Computational Methods in Systems Biology, Proceedings of the First International Workshop, CMSB 2003*, volume 2602 of *Lecture Notes in Computer Science*, pages 88–103, 2003.
- [21] K. Schmidt, M. Carlsen, J. Nielsen, and J. Viladsen. Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices. *Biotechnology and Bioengineering*, 55:831–840, 1997.
- [22] K. Schmidt, L. Nørregaard, B. Pedersen, A. Meissner, J. Duus, J. Nielsen, and J. Villadsen. Quantification of intracellular metabolic fluxes from fractional enrichment and  $^{13}\text{C}$ - $^{13}\text{C}$  coupling constraints on the isotopomer distribution in labeled biomass components. *Metabolic Engineering*, 1:166–179, 1999.
- [23] V. Selivanov, J. Puigjaner, A. Sillero, J. Centelles, A. Ramos-Montoya, P. Lee, and M. Cascante. An optimized algorithm for flux estimation from isotopomer distribution in glucose metabolites. *Bioinformatics*, 20:3387–3397, 2004.
- [24] J. Shen, K. Petersen, K. Behar, P. Brown, T. Nixon, G. Mason, O. Petroff, G. Shulman, R. Shulman, and D. Rothman. Determination of the rate of the glutamate/glutamine cycle in the human brain by in vivo  $^{13}\text{C}$  NMR. *Proceedings of the National Academy of Sciences of the United States of America*, 96(14):8235–8240, 1999.
- [25] A. Sola, H. Maaheimo, K. Ylönen, P. Ferrer, and T. Szyperki. Amino acid biosynthesis and metabolic flux profiling of *pichia pastoris*. *FEBS Journal*, 271:2462–2470, 2004.
- [26] G. Stephanopoulos, A. Aristidou, and J. Nielsen. *Metabolic engineering: Principles and Methodologies*. Academic Press, 1998.
- [27] T. Szyperki. Biosynthetically directed fractional  $^{13}\text{C}$ -labelling of proteinogenic amino acids. *European Journal of Biochemistry*, 232:433–448, 1995.
- [28] T. Szyperki.  $^{13}\text{C}$ -nmr, ms and metabolic flux balancing in biotechnology research. *Quarterly Reviews of Biophysics*, 31:41–106, 1998.
- [29] T. Szyperki, R. Glaser, M. Hochuli, J. Fiaux, U. Sauer, J. Bailey, and K. Wütrich. Bioreaction network topology and metabolic flux ratio analysis by biosynthetic fractional  $^{13}\text{C}$  labeling and two-dimensional NMR spectrometry. *Metabolic Engineering*, 1:189–197, 1999.
- [30] W. van Winden, J. Heijnen, P. Verheijen, and J. Grievink. A priori analysis of metabolic flux identifiability from (13)c-labeling data. *Biotechnology and Bioengineering*, 74:505–516, 2001.
- [31] W. van Winden, J. van Dam, C. Ras, R. Kleijn, J. Vinke, W. Gulik, and J. Heijnen. Metabolic-flux analysis of *saccharomyces cerevisiae* cen.pk113-7d based on mass isotopomer measurements of  $^{13}\text{C}$ -labeled primary metabolites. *FEMS Yeast Research*, 5:559–568, 2005.
- [32] W. Wiechert, M. Möllney, S. Petersen, and A. de Graaf. A universal framework for  $^{13}\text{C}$  metabolic flux analysis. *Metabolic Engineering*, 3:265–283, 2001.
- [33] W. Wiechert, C. Siefke, A. de Graaf, and A. Marx. Bidirectional reaction steps in metabolic networks: II. flux estimation and statistical analysis. *Biotechnology and Bioengineering*, 55:118–134, 1997.
- [34] C. Wittmann and E. Heinzle. Mass spectrometry for metabolic flux analysis. *Biotechnology and Bioengineering*, 62:739–750, 1999.