Efficient Modeling of ^{13}C -Labeling Distributions in Microorganisms

Wouter A. van Winden^{1,2}, Peter J.T. Verheijen¹, Joseph, J. Heijnen²

¹ Process Systems Engineering, ² Kluwer Laboratory for Biotechnology,

Delft University of Technology, The Netherlands,

phone: +31-15-275307, fax: +31-15-2782355, e-mail: w.a.vanwinden@tnw.tudelft.nl

Abstract

Metabolic networks can be analysed using 2D [^{13}C , ^1H] COSY (NMR) measurements of ^{13}C -labeled metabolites. A framework is presented whereby the steady state reaction rates are deduced from conventional isotopomer balances. This model is reduced by removing redundant nodes and lumping equilibrium pools. Conversion of the balances to the recently introduced bondomer notation further reduces the complexity. When the reduction approaches are applied to the glycolysis and pentose phosphate pathway, the number of equations is reduced by a factor of three without loss of information.

1. Introduction

A central issue in metabolic engineering is the determination of steady state reaction rates (or: 'fluxes') in the metabolic reaction network in cells. During the past decade a ^{13}C -labeling based tracer technique has been developed for this purpose. ^{13}C -labeling experiments consist of cultivating a microorganism in a continuous bioreactor system fed with ^{13}C -enriched carbon feedstock (e.g. glucose), sampling ^{13}C -labeled cells from the reactor and measuring the ^{13}C -distribution of intracellular compounds by means of mass spectroscopy or nuclear magnetic resonance spectroscopy. Metabolic reaction rates are subsequently determined by fitting the labeling measurements with a model of the metabolic reaction network wherein the reaction rates are the free variables.

Figure 1: I: a small network showing carbon atom (circles) transitions. II: isotopomers of the compounds (white and black circles represent ^{12}C - and ^{13}C -isotopes respectively).

A mathematical model for this system traditionally consists of a set of balances for all isotopomers of each metabolite in the reaction network (Schmidt et al., 1997). Isotopomers are chemically identical molecules that only vary with respect to their ^{12}C and ^{13}C -distribution. The isotopomer concept is illustrated in Fig.1. The isotopomer balances of the eight isotopomers of compound D in Fig.1-I are given in Eq.1, where subscripts 0 and 1 represent ^{12}C -isotopes ^{13}C -isotopes, respectively and ' $\otimes$ ' denotes an elementwise multiplication of two vectors.

$$v_1 \cdot \begin{bmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} A_{00} \\ A_{01} \\ A_{10} \\ A_{11} \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} B_0 \\ B_1 \end{pmatrix} + v_2 \cdot \begin{pmatrix} C_{000} \\ C_{001} \\ C_{010} \\ C_{011} \\ C_{100} \\ C_{101} \\ C_{110} \\ C_{111} \end{pmatrix} = v_3 \cdot \begin{pmatrix} D_{000} \\ D_{001} \\ D_{010} \\ D_{011} \\ D_{100} \\ D_{101} \\ D_{110} \\ D_{111} \end{pmatrix} \quad (1)$$

The matrices here are called isotopomer mapping matrices (IMM), and the vectors are isotopomer distribution vectors (idv).

A compound that contains n carbon atoms has 2^n isotopomers. By consequence, the total set of all isotopomer balances for a metabolic network with tens of compounds that contain up to 7 carbon atoms each, consists of many hundreds of non-linear equations that have to be solved in each step of the iterative fitting procedure.

Here we present three model reduction approaches that lead to a substantial reduction of the size of the ^{13}C -labeling distribution models without any loss of information.

The model reduction is exemplified by the glycolysis and pentose phosphate pathway, which are two central reaction pathways in the carbon metabolism of most cells (see Fig.2). The full isotopomer model of this network consists of 618 isotopomer balances (64 for glucose 6-phosphate (g6p), fructose 6-phosphate (f6p), fructose 1,6-bisphosphate (fbp), 6-phosphoglucono- δ -lactone (6pgl) and 6-phosphogluconate (6pg); 32 for ribulose 5-phosphate (ru5p), xylulose 5-phosphate (x5p) and ribose 5-phosphate (r5p); 16 for erythrose 4-phosphate (e4p); 128 for sedoheptulose 7-phosphate (s7p); 8 for dihydroxyacetonephosphate (dhap), glyceraldehyde 3-phosphate (g3p), 1,3-bisphosphoglycerate (bpg), 3-phosphoglycerate (3pg), 2-phosphoglycerate (2pg), phosphoenolpyruvate (pep), and pyruvate (pyr); 2 for carbondioxide (cd)). Note that no isotopomer balances are made for glucose (glc) as its ^{13}C -labeling is chosen by the experimenter.

2. Model Reduction Approaches

2.1. Removal of linear and divergent nodes from the network

A first reduction of ^{13}C -labeling distribution models is based on the number of reactions entering and leaving metabolite pools (Van Winden et al., 2001). Three basic structures exist in metabolic networks: linear, divergent and convergent nodes. In their article

Figure 2: The glycolysis and pentose phosphate pathway. Double-headed arrows denote bidirectional reactions, 'f' and 'b' indicate the corresponding forward and backward reactions.

Wiechert *et al.* (1999) argue that pools with only one influx only yield so-called 'labeling redundancies' between isotopomer fractions and do not give information about fluxes. It can be shown that only the labeling distribution of metabolites at convergent nodes yield flux ratios. This is demonstrated for the metabolic network shown in Fig.2 by analyzing the three types of nodes. The flux balance and labeling balance of the linear node around metabolite 6pgl in Fig.2 and the combination of both balances is shown in Eq.2.

$$\left. \begin{aligned} v_{11} &= v_{12} \\ v_{11} \cdot \text{IMM}_{\text{g6p} \rightarrow \text{6pgl}} \cdot \text{g6p} &= v_{12} \cdot \text{6pgl} \end{aligned} \right\} \text{IMM}_{\text{g6p} \rightarrow \text{6pgl}} \cdot \text{g6p} = \text{6pgl} \quad (2)$$

$\text{IMM}_{\text{glc} \rightarrow \text{6pgl}}$ represents the isotopomer mapping matrix that describes which isotopomer of the reaction product 6pgl is formed from each isotopomer of the substrate g6p. The vectors **g6p** and **6pgl** contain the labeling information of the concerning metabolites. The combined balance in Eq.2 enables the calculation of the labeling distribution of 6pgl from that of g6p but does not give information about the fluxes. Eq.3 shows that a similar labeling redundancy follows from a divergent node (e.g. pep in Fig.2):

$$\left. \begin{aligned} v_{9f} &= v_{9b} + v_{10} \\ v_{9f} \cdot \text{IMM}_{\text{2pg} \rightarrow \text{pep}} \cdot \text{2pg} &= (v_{9b} + v_{10}) \cdot \text{pep} \end{aligned} \right\} \text{IMM}_{\text{2pg} \rightarrow \text{pep}} \cdot \text{2pg} = \text{pep} \quad (3)$$

A different set of relations results from combining the flux balance and the labeling balance of a convergent node (e.g. g6p in Fig.2):

$$\left. \begin{aligned} v_1 + v_{2b} &= v_{2f} + v_{11} \\ v_1 \cdot (\text{IMM}_{\text{glc} \rightarrow \text{g6p}} \cdot \text{glc}) + v_{2b} \cdot (\text{IMM}_{\text{f6p} \rightarrow \text{g6p}} \cdot \text{f6p}) &= (v_{2f} + v_{11}) \cdot \text{g6p} \end{aligned} \right\} \quad (4)$$

$$v_1 \cdot (\text{IMM}_{\text{glc} \rightarrow \text{g6p}} \cdot \text{glc} - \text{g6p}) + v_{2b} \cdot (\text{IMM}_{\text{f6p} \rightarrow \text{g6p}} \cdot \text{f6p} - \text{g6p}) = 0$$

By means of linear algebra Eq.4 can be rewritten to one single equation in which the ratio of the fluxes entering the g6p pool appears. The remaining equations are labeling redundancies. Eq.4 shows that only the ratio v_1/v_{2b} can be determined from this labeling balance and that one single element of vector **g6p** (or: one ^{13}C -labeling measurement) suffices to do so, provided that the labeling distributions of glc and f6p are known.

The above shows that all linear and divergent nodes can be removed from the network in Fig.2 without losing any ^{13}C -labeling information. Thus, metabolite pools fbp, pep, 6pgl, 6pg are removed. When pool pep is removed, 2pg becomes a single-influx pool and is removed as well. This latter procedure applies to 3pg and bpg as well.

2.2. Lumping equilibrium pools

A second possible reduction of metabolic networks is the lumping of metabolite pools of which the ^{13}C -labeling information may be considered instantaneously equilibrated by large exchange fluxes. This reduction is often applied to the hexose 6-phosphate pools (g6p and f6p) in the glycolysis and the pentose 5-phosphate pools (ru5p, x5p, r5p) in the pentose phosphate pathway (e.g. Schmidt et al., 1997; Follstad and Stephanopoulos, 1998). Lumping g6p and f6p and lumping ru5p, x5p and r5p further simplifies the metabolic reaction network of Fig.2 to the reduced version shown in Fig.3.

Figure 3: The reduced metabolic network.

2.3. Bondomer balancing

A measurement method that is often used to determine the ^{13}C -labeling of metabolites is 2D [^{13}C , ^1H] correlation nuclear magnetic resonance spectroscopy (COSY). This method yields relative intensities of multiplets in the NMR spectra that correspond to the relative amounts of groups of isotopomers in which the observed atom is ^{13}C -labeled and the adjacent carbon atoms in the carbon backbone are either ^{13}C -labeled or not (Szyperski, 1995). If the compound of which the labeling pattern is measured was synthesized by a microorganism growing on a mixture of uniformly ^{13}C -labeled and naturally labeled carbon substrate, then the relative intensities of NMR fine structures can be calculated from the fractions of molecules that stem from one or more substrate molecule(s) in the feed medium. This is done using so-called 'probability equations' (Szyperski, 1995) that require as input the fraction of uniformly ^{13}C -labeled medium substrate and the fraction of naturally ^{13}C -labeled carbon.

Molecules that stem from one or more medium substrate molecules and that are calculated using the probability equations are both chemically and physically identical. They only vary in the numbers and positions of C-C bonds that have remained intact since the medium substrate molecule entered the metabolism. These entities are hereby defined as 'bondomers' (Van Winden et al., 2002). Bondomers of a given compound are denoted by an abbreviation of the compound and a binary subscript. Whereas in isotopomer notation the binary subscript '0' denotes a ^{12}C -atom and a '1' a ^{13}C -atom (Schmidt et al., 1997), in bondomer notation '0' denotes a C-C bond that has been formed in one of the metabolic reactions and '1' denotes a C-C bond that was already present in the medium substrate molecule.

A linear or branched molecule that has a backbone of n carbon atoms has $n-1$ C-C bonds. Such a molecule has 2^n isotopomers and 2^{n-1} bondomers. The bondomer distribution of a molecule can be simulated in a way that is completely analogous to the simulation of isotopomers. Bondomer balancing is based on bondomer mapping matrices (BMMs) that indicate which bondomers of the reaction products are formed from the bondomers of the reaction substrates. Each column of a BMM corresponds with a bondomer of (one of) the reaction substrate(s), each row with a bondomer of (one of) the product(s). The bondomer distributions are given as bondomer distribution vectors (bdv). The bondomer balances of the four bondomers of compound D in Fig.1-I are given in Eq.5:

$$v_1 \cdot \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} A_0 \\ A_1 \end{pmatrix} + v_2 \cdot \begin{pmatrix} C_{00} \\ C_{01} \\ C_{10} \\ C_{11} \end{pmatrix} = v_3 \cdot \begin{pmatrix} D_{00} \\ D_{01} \\ D_{10} \\ D_{11} \end{pmatrix} \quad (5)$$

In a bondomer balance, the inflows and outflows of all the 2^{n-1} bondomers of a n -carbon compound are accounted for. The inflow terms in the balance are the products of the fluxes, BMMs and substrate-bdvs of the reactions that lead to the balanced metabolite. The outflow term is obtained by multiplying the bdv of the balanced metabolite by the sum of the fluxes of the reactions in which the metabolite itself serves as a substrate. In the inflow term corresponding to reaction v_1 in Eq.5, the bondomer distribution of B

does not appear for the simple reason that a one-carbon compound has no C-C bonds and therefore has only one ($=2^{1-1}$) bondomer fraction equalling 1. Eq.5 shows that bondomer modeling does not only lead to fewer, but also to simpler balances than isotopomer modeling (Eq.1).

As shown before, the isotopomer model of Fig.2 consisted of 618 isotopomer balances. Due to the removal and lumping of metabolite pools, the reduced model of Fig.3 only counts 258 isotopomer balances. As 2D [^{13}C , ^1H] COSY data can be calculated from bondomers, one need not simulate all isotopomers to fit these measured data with simulated data. Instead, one can apply a bondomer model, which counts half as few balances: only 129 (32 for hexose 6-phosphate; 16 for pentose 5-phosphate; 8 for erythrose 4-phosphate; 64 for sedoheptulose 7-phosphate; 4 for glyceraldehyde 3-phosphate, and pyruvate; 1 for carbondioxide). An even further reduction is possible by applying the 'cumomer' concept (Wiechert et al., 1999) to the bondomer balances, but this is outside the scope of this paper. See Van Winden et al. (2002) for details.

3. Conclusions

In this paper we discussed three methods that can be applied to reduce the size of mathematical models that give the ^{13}C -labeling distribution in the metabolites of a biochemical reaction network as a function of the reaction rates in that network. We illustrated the methods by applying them to the metabolic network that consists of the glycolysis and pentose phosphate pathway. For this case study, the model was reduced from the initial 618 balances to a mere 129 balances, without loss of information. This reduction shows that models can sometimes be trimmed by carefully checking that only the needed output is simulated and that all redundancies are removed from the model. This saves computational time, which is especially relevant in problems that require many repeated simulations, such as iterative parameter fitting procedures and optimization. Additionally, the achieved model reduction has the advantage that it makes the model amenable to a priori identifiability analysis. This allows better designing of ^{13}C -labeling experiments.

4. References

- Follstad, B.D. and Stephanopoulos, G., 1998, Eur. J. Biochem., 252.
- Schmidt, K., Carlsen, M., Nielsen, J. and Villadsen, J., 1997, Biotechnol. Bioeng., 55, 6.
- Szyperski, T., 1995, Eur. J. Biochem., 232.
- Van Winden, W.A., Heijnen, J.J., Verheijen, P.J.T., Grievink, J., 2001, Biotechnol. Bioeng., 74, 6
- Van Winden, W.A., Heijnen, J.J., Verheijen, P.J.T., 2002, Biotechnol. Bioeng., 80, 7
- Wiechert, W., Möllney, M., Isermann, N., Wurzel, M., De Graaf, A.A., 1999, Biotechnol. Bioeng., 66, 2.