Description of the algorithm for computing elementary flux modes

by

Stefan Schuster[1], Thomas Dandekar[1,2], and David Fell[3]

[1] Department of Bioinformatics, Max Delbrück Centre for Molecular Medicine
D-13092 Berlin-Buch, Germany, e-mail: schuster@bp.biologie.hu-berlin.de
[2] Biocomputing and Structures Program, EMBL, D-69012 Heidelberg, Germany
[3] School of Biological and Molecular Sciences, Oxford Brookes University
Oxford OX3 0BP, U.K.

An elementary flux mode is a minimal set of enzymes that could operate at steady state, with all the irreversible reactions used in the appropriate direction (cf. Schuster et al., 2000). The elementary flux modes for biochemical reaction systems of any complexity can be detected by an algorithm to be outlined and illustrated in what follows. It is based on an algorithm for detecting the generating vectors of convex polyhedral cones given in Nožička *et al*. (1974). A more mathematical description as well as references to related methods have been given earlier (Schuster and Hilgetag, 1994; Schuster *et al*., 1996). In contrast to the algorithm proposed by Clarke (1981), reversible reactions need not be split into their forward and reverse steps. Even in the situation that all reactions are irreversible, the presented algorithm is faster than the method proposed by Clarke (1981) (for a comparison, see Schuster and Schuster, 1993). The algorithm has been implemented as computer programs in Smalltalk (program EMPATH, John Woods, Oxford), C (program METATOOL, Pfeiffer *et al.*, 1999) and MAPLE (program METAFLUX, Klaus Mauch, Stuttgart). The former two are available from ftp://bmshuxley.brookes.ac.uk/pub/mca/software/ibmpc, METATOOL also from http://www2.bioinf.mdc-berlin.de/metabolic/. The running time of the programs is less than one second for the system considered below on a usual PC or workstation. The programs start from a list of reaction equations and a declaration of reversible and irreversible reactions (enzymes) and of internal and external metabolites.

For example, for the reaction scheme shown in Fig. 1 in Schuster *et al*. (2000), which represents part of monosaccharide metabolism, the input file for METATOOL has the following form (Pgi, Ald, etc. are abbreviations of enzymes such as phosphoglucoisomerase

and aldolase; E4P, S7P etc. are abbreviations of metabolites such as erythrose 4-phosphate and sedoheptulose 7-phosphate):

```
-ENZREV
Pgi Ald Tpi Rpi Rpe TktI TktII Tal Gpm Eno Pgl

-ENZIRREV
Pfk Fbp Zwf Gnd Pgk Pyk Gap Prs_DeoB

-METINT
E4P S7P X5P Ru5P GO6P GL6P
DHAP FDP F6P GAP D13PG P3G P2G PEP R5P

-METEXT
CO2 ATP ADP NAD NADH NADP NADPH
G6P PYR R5Pex Pi

-CAT
Pgi : G6P = F6P .
Pfk : F6P + ATP = FDP + ADP .
Fbp : FDP = F6P + Pi .
Ald : FDP = DHAP + GAP .
Tpi : DHAP = GAP .
Gap : GAP + NAD + Pi = D13PG + NADH .
Zwf : G6P + NADP = GO6P + NADPH .
Pgl : GO6P = GL6P .
Gnd : GL6P + NADP = Ru5P + NADPH + CO2 .
Rpi : Ru5P = R5P .
Rpe : Ru5P = X5P .
TktI : X5P + R5P = GAP + S7P .
TktII : E4P + X5P = F6P + GAP .
Tal : S7P + GAP = E4P + F6P .
Pgk : D13PG + ADP = P3G + ATP .
Gpm : P3G = P2G .
Eno : P2G = PEP .
Pyk : PEP + ADP = PYR + ATP .
Prs_DeoB : R5P = R5Pex .
```

As in several other metabolic simulators, this list is automatically translated into a stoichiometry matrix (for an explanation of this and related terms, see Heinrich and Schuster, 1996). Transposing this matrix and augmenting it with the identity matrix gives a matrix called the initial tableau. From this, further tableaux are consecutively computed by pair-wise linear combination of rows so that the columns of the transposed stoichiometry matrix become null vectors successively. This procedure corresponds to ensuring the steady-state condition is fulfilled for each metabolite taken in turn.

The above example includes 15 internal metabolites and 19 reactions. Accordingly, its stoichiometry matrix has dimensions 15 x 19. Before computing the elementary modes, it is convenient (but not necessary) to reduce this matrix by lumping those reactions that necessarily operate together. In the considered system, the sets {Gap, Pgk, Gpm, Eno, Pyk} and {Zwf, Pgl, Gnd} constitute such sequences. An algorithm for detecting such sets of

enzymes in systems of any complexity has been developed and included into the program METATOOL (Pfeiffer *et al*., 1999). Applying this algorithm to the system considered reveals another two sequences: {Fba, TpiA} and {2 Rpe, TktI, Tal, TktII}. “2 Rpe” means that the flux through Rpe is, in any steady state of the network, twice as large as the flux through TktI, Tal and TktII. Lumping the reactions in any one sequence gives the reduced system shown in Fig. 3 in Schuster *et al*. (2000). It encompasses fewer metabolites than the original system because the substances located within a reaction sequence can be omitted. The initial tableau of the reduced system reads

$$\mathbf{T}^{(0)} = \left(\begin{array}{ccccc|cccc} 0 & 0 & 1 & 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 0 & 2 & 0 & 0 & 1 & \cdots & 0 \\ -1 & 0 & 0 & 0 & 1 & 0 & 0 & \cdots & 0 \\ -2 & 0 & 2 & 1 & -1 & 0 & 0 & \cdots & 0 \\ \hline 0 & 0 & 0 & -1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & \cdots & 1 \end{array}\right) \begin{array}{l} \left.\begin{array}{l} \\ \\ \\ \\ \end{array}\right\} \text{reversible reactions} \\ \left.\begin{array}{l} \\ \\ \\ \\ \\ \end{array}\right\} \text{irreversible reactions} \end{array} \qquad (1)$$

The five columns on the left-hand side correspond to the metabolites Ru5P, FP_2, F6P, GAP, and R5P. The nine rows correspond to the reactions or reaction sequences Pgi, {Fba, TpiA}, Rpi, {2Rpe, TktI, Tal, TktII}, {Gap, Pgk, Gpm, Eno, Pyk}, {Zwf, Pgl, Gnd}, Pfk, Fbp, and Prs_DeoB, of which the latter five are irreversible (lower part of the matrix).

The entries with row numbers 1, 2, 5, 7, 8, and 9 in the first column of $\mathbf{T}^{(0)}$ are zeros. Therefore, they need not be combined with other rows. Instead, they can be copied into the "reversible" and "irreversible" parts of the next tableau. In addition, a "reversible" row in the next tableau arises from subtracting the 4^{th} row from twice the 3^{rd} row. Moreover, appropriate linear combinations of the 3^{rd} and 6^{th} rows and of the 4^{th} and 6^{th} rows give "irreversible" rows. In general, linear combinations of two rows belonging to the same type of directionality (reversible or irreversible) go into the part of the respective type in the next tableau, while linear combinations of rows corresponding to different types go into the "irreversible" part because they include at least one irreversible reaction. Note that "irreversible" rows can enter a linear combination only with a positive coefficient in order that all modes use the irreversible reactions in the appropriate direction.

For the system under study, the following tableau $\mathbf{T}^{(1)}$ is obtained:

$$\mathbf{T}^{(1)} = \left(\begin{array}{ccccc|cccccccccc}
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 2 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -2 & -1 & 3 & 0 & 0 & 2 & -1 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 2 & 1 & -1 & 0 & 0 & 0 & 1 & 0 & 2 & 0 & 0 & 0
\end{array}\right) \tag{2}$$

The right-hand side part of the tableau keeps track of the linear combinations performed.

Now rows are combined so as to ensure the second column of the result will be zero. The rows numbered 1, 3, 4, 7, 8, and 9 are copied straight into the next tableau because their respective second elements are zero already. The next tableau reads

$$\mathbf{T}^{(2)} = \left(\begin{array}{ccccc|cccccccccc}
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -2 & -1 & 3 & 0 & 0 & 2 & -1 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 2 & 1 & -1 & 0 & 0 & 0 & 1 & 0 & 2 & 0 & 0 & 0 \\
0 & 0 & -1 & 2 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & -2 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0
\end{array}\right) \tag{3}$$

In the course of the algorithm, calculation of duplicate modes, non-elementary modes, and flux modes violating the sign condition for the irreversible reactions is avoided by checking three conditions. First, a pair of rows is combined only if it fulfills the condition

$$\mathrm{S}(\mathbf{m}_{i\cdot}^{(j)}) \cap \mathrm{S}(\boldsymbol{m}_{k\cdot}^{(j)}) \not\subseteq \mathrm{S}(\mathbf{m}_{l\cdot}^{(j+1)}) \tag{4}$$

for all row indices l belonging to the respective part (reversible or irreversible) of the new tableau as it has been compiled until that stage. $\mathbf{m}_{i\cdot}^{(j)}$ stands for the ith row in the right-hand

side submatrix of tableau $\mathbf{T}^{(j)}$ and $S(\mathbf{m}_{i\cdot}^{(j)})$ is the set of positions of zeroes in this row. This set harbours information about which enzymes are not used in the respective mode. For example, tableau $\mathbf{T}^{(3)}$ includes, in its right-hand side part, the row

$$\boldsymbol{m}_{i\cdot}^{(3)} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \tag{5}$$

as this row is also involved in tableau $\mathbf{T}^{(2)}$ and has a zero in its third position. For this row,

$$S(\boldsymbol{m}_{i\cdot}^{(3)}) = \{1, 2, 4, 5, 7, 8, 9\} \, . \tag{6}$$

When constructing $\mathbf{T}^{(3)}$, a candidate pair for linear combination comprises the 2nd and 6th rows of $\mathbf{T}^{(2)}$. However, as $S(\boldsymbol{m}_{2\cdot}^{(2)}) \cap S(\boldsymbol{m}_{6\cdot}^{(2)}) = \{1, 2, 5, 7, 8, 9\}$, which is a subset of the set given in Eq. (6), these row vectors must not be combined. If we did combine them, we would obtain a row which, after normalization, equals the row given in Eq. (5). Linear combination of the 7th and 8th rows is forbidden for a similar reason. An example where condition (4) prevents non-elementary modes from being computed occurs in the compilation of tableau $\mathbf{T}^{(4)}$. For the rows $\left(\begin{array}{ccccc|ccccccccc} 0 & 0 & 0 & -1 & 3 & 2 & 0 & 2 & -1 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$ and $\left(\begin{array}{ccccc|ccccccccc} 0 & 0 & 0 & 5 & -1 & 0 & 2 & 0 & 1 & 0 & 2 & 2 & 0 & 0 \end{array}\right)$ situated in $\mathbf{T}^{(3)}$, the intersection set $S(\boldsymbol{m}_{i\cdot}^{(3)}) \cap S(\boldsymbol{m}_{k\cdot}^{(3)})$ reads $\{5, 8, 9\}$ and a row transferred earlier into tableau $\mathbf{T}^{(4)}$ reads $\left(\begin{array}{ccccc|ccccccccc} 0 & 0 & 0 & 0 & 6 & 5 & 1 & 4 & -2 & 0 & 0 & 1 & 0 & 0 \end{array}\right)$. If we did combine the two rows, we would obtain $\left(\begin{array}{ccccc|ccccccccc} 0 & 0 & 0 & 0 & 7 & 5 & 1 & 5 & -2 & 0 & 1 & 1 & 0 & 0 \end{array}\right)$, which involves fewer zeros than the row already situated in $\mathbf{T}^{(4)}$.

The second condition says that "irreversible" rows can only be added rather than subtracted (cf. above). For example, rows numbered 6 and 8 in $\mathbf{T}^{(2)}$ must not be combined because they both comprise some entries belonging to irreversible reactions and contain positive entries in the 3rd position. For the system considered, these two conditions are relevant only from tableau $\mathbf{T}^{(3)}$ on. In total, seven linear combinations are allowed, giving rise to a tableau $\mathbf{T}^{(3)}$ comprising 11 rows.

Upon constructing a new tableau, it may occur that some row that has been correctly computed turns out to be non-elementary because some other row, which is calculated later,

comprises more zero positions. This can be avoided by using a third test criterion. If any pair of rows pass condition (4) and are combined and added to the tableau, all the rows $\mathbf{m}_{l\cdot}^{(j+1)}$ previously added to the new tableau are checked to ensure that:

$$\mathrm{S}(\mathbf{m}_{l\cdot}^{(j+1)}) \not\subset \mathrm{S}(\mathbf{m}_{i\cdot}^{(j)}) \cap \mathrm{S}(\boldsymbol{m}_{k\cdot}^{(j)}) \,. \tag{7}$$

According to our experience, this criterion is rarely violated. To save computational time, it is therefore sufficient to apply it only upon computing the final tableau. For the glycolysis/PPP example, criterion (7) is never violated, whereas condition (4) helps us avoid computing 11 and 9 irrelevant rows in $\mathbf{T}^{(4)}$ and $\mathbf{T}^{(5)}$, respectively. The final tableau reads

$$\mathbf{T}^{(5)} = \left(\begin{array}{cccc|cccccccc} 0 & \cdots & \cdots & 0 & 1 & 1 & 0 & 0 & 2 & 0 & 1 & 0 & 0 \\ \vdots & & & \vdots & -2 & 0 & 1 & 1 & 1 & 3 & 0 & 0 & 0 \\ & & & & 0 & 2 & 1 & 1 & 5 & 3 & 2 & 0 & 0 \\ & & & & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ & & & & 5 & 1 & 4 & -2 & 0 & 0 & 1 & 0 & 6 \\ \vdots & & & \vdots & -5 & -1 & 2 & 2 & 0 & 6 & 0 & 1 & 0 \\ 0 & \cdots & \cdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{array}\right) \tag{8}$$

This example shows that the number of rows may increase or decrease in the course of the algorithm. The row vectors in the right-hand side submatrix of the final tableau represent the elementary modes, all of which are irreversible. Note that some of their entries correspond to lumped enzyme sequences. These modes can, however, be easily translated into the modes in terms of the original set of 19 reactions. The relevant part of the output file of METATOOL has the following form. (Note that METATOOL also computes other structural properties of metabolic networks, such as enzyme subsets, conservation relations, and the convex basis, see Pfeiffer *et al*., 1999).

```
Elementary modes (original, Number: 7 ):
 1  1  1  0  0  0  0  0  2  2  0  1  0  0  0  2  2  2  0
-2  0  0  1  2  1  1  1  1  1  3  0  0  3  3  1  1  1  0
 0  2  2  1  2  1  1  1  5  5  3  2  0  3  3  5  5  5  0
 0  0  0  1  0  0  0  0  0  0  1  0  0  1  1  0  0  0  1
 5  1  1  4 -4 -2 -2 -2  0  0  0  1  0  0  0  0  0  0  6
-5 -1 -1  2  4  2  2  2  0  0  6  0  1  6  6  0  0  0  0
 0  0  0  0  0  0  0  0  0  0  0  1  1  0  0  0  0  0  0

1:     Pgi Ald Tpi 2 Gpm 2 Eno Pfk 2 Pgk 2 Pyk 2 Gap irreversible
2:     -2 Pgi Rpi 2 Rpe TktI TktII Tal Gpm Eno 3 Pgl 3 Zwf 3 Gnd Pgk Pyk
Gap irreversible
```

```
3:     2 Ald 2 Tpi Rpi 2 Rpe TktI TktII Tal 5 Pgm 5 Eno 3 Pgl 2 Pfk 3 Zwf
3 Gnd 5 Pgk 5 Pyk 5 Gap irreversible
4:     Rpi Pgl Zwf Gnd Prs_DeoB irreversible
5:     5 Pgi Ald Tpi 4 Rpi -4 Rpe -2 TktI -2 TktII -2 Tal Pfk 6 Prs_DeoB
irreversible
6:     -5 Pgi -1 Ald -1 Tpi 2 Rpi 4 Rpe 2 TktI 2 TktII 2 Tal 6 Pgl Fbp
6 Zwf 6 Gnd irreversible
7:     Pfk Fbp irreversible

Overall reaction of elementary modes:
1:    2 Pi + G6P + 2 NAD + 3 ADP = 2 PYR + 2 NADH + 3 ATP
2:    Pi + G6P + 6 NADP + NAD + 2 ADP = PYR + 6 NADPH + NADH + 2 ATP + 3
CO2
3:    5 Pi + 3 G6P + 6 NADP + 5 NAD + 8 ADP = 5 PYR + 6 NADPH + 5 NADH + 8
ATP + 3 CO2
4:    G6P + 2 NADP = R5Pex + 2 NADPH + CO2
5:    5 G6P + ATP = 6 R5Pex + ADP
6:    G6P + 12 NADP = Pi + 12 NADPH + 6 CO2
7:    ATP = Pi + ADP
```

The overall reactions indicate the overall stoichiometry in terms of the external metabolites. This information is very helpful in determining optimal yields.

References

Clarke, B.L. Complete set of steady states for the general stoichiometric dynamical system. *J. Chem. Phys.* **75** (1981) 4970-4979.

Heinrich, R. and Schuster, S. *The Regulation of Cellular Systems*. Chapman & Hall, New York 1996.

Nožička, F., Guddat, J., Hollatz, H. and Bank, B. *Theorie der linearen parametrischen Optimierung*. Akademie-Verlag, Berlin 1974.

Pfeiffer, T., Sánchez-Valdenebro, I., Nuño, J. C., Montero, F. & Schuster, S. METATOOL: For studying metabolic networks, *Bioinformatics* **15** (1999) 251-257.

Schuster, R. & Schuster, S. Refined algorithm and computer program for calculating all non-negative fluxes admissible in steady states of biochemical reaction systems with or without some flux rates fixed. *Comp. Appl. Biosci.* **9** (1993) 79-85.

Schuster, S. & Hilgetag, C. On elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Syst.* **2** (1994) 165-182.

Schuster, S., Hilgetag, C., Woods, J. H. & Fell, D. A. in *Computation in Cellular and Molecular Biological Systems* (eds. Cuthbertson, R., Holcombe, M. & Paton, R.) pp. 151-165 (World Scientific, Singapore, 1996).

S. Schuster, D. Fell and T. Dandekar: A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks, *Nature Biotechnol.* **18** (2000) 326-332.