Detection of stoichiometric inconsistencies in biomolecular models

Albert Gevorgyan, Mark G. Poolman and David A. Fell*

Cell Systems Modelling Group, School of Life Sciences, Oxford Brookes University, Oxford, OX3 0BP, UK

Received on December 14, 2007; revised on July 11, 2008; accepted on August 8, 2008

ABSTRACT

Motivation: Metabolic modelling provides a mathematically rigorous basis for system-level analysis of biochemical networks. However, the growing sizes of metabolic models can lead to serious problems in their construction and validation. In this work, we describe a relatively poorly investigated type of modelling error, called stoichiometric inconsistencies. These errors are caused by incorrect definitions of reaction stoichiometries and result in conflicts between two fundamental physical constraints to be satisfied by any valid metabolic model: positivity of molecular masses of all metabolites and mass conservation in all interconversions.

Results: We introduce formal definitions of stoichiometric inconsistencies, inconsistent net stoichiometries, elementary leakage modes and other important fundamental properties of incorrectly defined biomolecular networks. Algorithms are described for the verification of stoichiometric consistency of a model, detection of unconserved metabolites and inconsistent minimal net stoichiometries. The usefulness of these algorithms for effective resolving of inconsistencies and for detection of input errors is demonstrated on a published genome-scale metabolic model of Saccharomyces cerevisiae and one of Streptococcus agalactiae constructed using the KEGG database.

Availability: http://mudshark.brookes.ac.uk/index.php/Albert_Gevo
rgyan

Contact: dfell@brookes.ac.uk

1 INTRODUCTION

In recent decades, the growing interest in system-level understanding of biochemical networks resulted in the integration of experimental data from biochemistry and genetics with methods of mathematics and computer science, in the field of metabolic modelling. The first published metabolic models usually covered separate pathways, such as glycolysis in Garfinkel and Hess (1964) and Richter et al. (1975), electron transport in Laisk and Walker (1989) and photosynthesis in Poolman et al. (2001); such models can be effectively built manually, using available literature or experimental data. The sizes of models, however, tended to grow, exceeding several hundreds of reactions in genome-scale models (Forster et al., 2003; Schilling and Palsson, 2000). Despite the ease with which models can be constructed from public databases, serious challenges remain: the quality of databases is insufficient (Poolman et al., 2006) and testing the accuracy of all reaction definitions manually is impractical. The modeller’s task could be simplified by automation of error detection and model validation processes.

Incorrect definition of metabolic systems is likely to give rise to internal inconsistencies in the resulting models, violating physical or biological constraints. Detection of certain types of such inconsistencies is possible by interrogation of the model structure, without testing the correctness of input data; this is an important part of model validation. The present work focuses on conflicts between two constraints to be fulfilled in any metabolic system: positivity of molecular masses of all metabolites and mass conservation in all reactions.

In the network shown below, the reader would probably spot an error:

\[ R_1 : A \leftrightarrow B \]
\[ R_2 : A \leftrightarrow B + C \] (1)

Under the assumption of mass conservation it is impossible to assign any positive molecular mass to C; the only admissible mass is zero. The presence of the error does not depend on the actual chemical species, denoted A, B and C; it only depends on the reaction stoichiometries. Therefore, we call such errors stoichiometric inconsistencies. A metabolic network is stoichiometrically consistent if all metabolites can be assigned some positive molecular masses without violating mass conservation, and stoichiometrically inconsistent otherwise.

Although in this work we consider only structural models (based solely on reaction stoichiometries), the concept of stoichiometric consistency is also applicable to kinetic models of metabolism and, more generally, to any systems of mass-conserving stoichiometric interconversions. We will show that this concept is closely related to two well-known concepts of metabolic modelling: those of elementary flux modes and moiety conservation relationships (Heinrich and Schuster, 1996). In particular, stoichiometric inconsistencies always result in the presence of unconserved metabolites (Nikolaev et al., 2005), defined as those not involved in any non-negative conservation relationships. An algorithm for the detection of unconserved metabolites has been proposed by Nikolaev et al. (2005). It involves a linear program which must be invoked for each metabolite in a model. We later describe an alternative algorithm, based on a mixed-integer linear program (MILP).

In realistic models, where each metabolite is assumed to represent some chemical substance, stoichiometric inconsistencies are always
caused by the presence of atomically unbalanced reactions. Hence, if a model is known to be atomically balanced it must be also stoichiometrically consistent. Unfortunately, checking atomic balances requires the knowledge of chemical formulae, which are often unavailable [e.g. in the KEGG database (Ogata et al., 1999), we found that >50% of metabolite entries do not contain any formula] or variable (in polymers). Further, some metabolite definitions are ambiguous (e.g. 'Primary alcohol') or do not refer to any substances (in conceptual models). In this work, we propose a more reliable and universal solution of the problem.

In the next section, we introduce mathematical definitions of some concepts needed for the understanding of the work. Further, we describe basic algorithms for the solution of the problems stated above. These are followed by the results of the application of our methods to genome-scale metabolic models of Escherichia coli (Reed et al., 2003), Saccharomyces cerevisiae (Forster et al., 2003) and Streptococcus agalactiae (constructed in-house using the KEGG database).

2 PROBLEM FORMULATION

2.1 Basic concepts

We refer to a vector \( \mathbf{d} \) as positive (denoted \( \mathbf{d} > 0 \)) if it contains positive components only, semipositive \( \mathbf{d} \geq 0 \) if it contains at least one positive and no negative components and seminegative \( \mathbf{d} \leq 0 \) if it contains at least one negative and no positive components.

We consider a biochemical system of \( n \) reactions and \( m \) metabolites. The sets of reactions and metabolites, denoted \( \mathcal{R} \) and \( \mathcal{M} \), respectively, comprise a metabolic network \( \mathcal{N} \). All metabolites are assumed to be internal, i.e. produced and consumed only by the reactions of the network. We associate each metabolite with a numerical value specifying its molecular mass; these values comprise a column vector \( \mathbf{m} \) of dimension \( m \). Molecular masses of natural substances are positive:

\[
\mathbf{m} > 0
\]  

(2)

The structure of \( \mathcal{N} \) is described by its \( m \times n \) external stoichiometry matrix \( \mathbf{N} \) (Heinrich and Schuster, 1996). From the mass conservation law, it follows that in any reaction, the sum of molecular masses of metabolites weighted by their stoichiometric coefficients is equal to zero:

\[
\sum_{i=1}^{m} m_i N_{ij} = 0: \quad 1 \leq j \leq n
\]  

(3)

This equation can be written in a simpler manner:

\[
N^T \mathbf{m} = 0
\]  

(4)

We refer to any solution of Equation (4) as a mass conservation vector (Heinrich and Schuster, 1996), which may be considered as a possible distribution of molecular masses which fulfill the mass conservation constraint. A network is called consistent if it has at least one positive mass conservation vector and inconsistent otherwise.

2.2 Inconsistent net stoichiometries, unconservable metabolite sets and unconserved metabolites

As mentioned above, a stoichiometric inconsistency always results in the presence of unconserved metabolites. Let us investigate the relation between these concepts. According to Equation (3), the product of any column of a stoichiometry matrix and an arbitrary conservation vector is equal to zero. We can generalize this observation and define a net stoichiometry as any vector \( \mathbf{y} \) satisfying Equation (5):

\[
y^T \mathbf{m} = 0, \quad N^T \mathbf{m} = 0
\]  

(5)

For each metabolite \( m \in \mathcal{M} \), with the molecular mass \( m_i \), we refer to \( y_i \) as the stoichiometric coefficient of \( m \). The following proposition and corollary state an important property of inconsistent networks:

**Proposition 1** (Dines, 1926). An equation \( y^T \mathbf{m} = 0 \) has no positive solution \( \mathbf{m} \) iff \( \mathbf{y} \) is either semipositive or seminegative.

**Proof.** Necessary condition: Let us assume that \( \mathbf{y} \) is neither semipositive nor seminegative. If \( \mathbf{y} = 0 \) then the equation is solvable with arbitrary \( \mathbf{m} \). Otherwise, let us denote the positive components of \( \mathbf{y} \) by \( y_i, \quad i = i_1, ..., i_p \) and the negative ones by \( y_j, \quad j = j_1, ..., j_N \); \( P + N = m \). Then, the given equation can be written in the form:

\[
\sum_{i=i_1}^{i_p} y_i m_i = - \sum_{j=j_1}^{j_N} y_j m_j
\]

and a positive solution is obtained by taking

\[
m_i = - \sum_{j=j_1}^{j_N} y_j: \quad i = i_1, ..., i_p
\]

\[
m_j = \sum_{i=i_1}^{i_p} y_i: \quad j = j_1, ..., j_N
\]  

(6)

**Sufficient condition:** Let us assume that \( \mathbf{m} > 0 \). If \( \mathbf{y} \geq 0 \) then \( y^T \mathbf{m} > 0 \) and if \( \mathbf{y} \leq 0 \) then \( y^T \mathbf{m} < 0 \). \( \square \)

A net stoichiometry is inconsistent, if it is either semipositive or seminegative.

**Corollary 1.1.** A network is inconsistent, if it has at least one inconsistent net stoichiometry.

A metabolite set is conservative, if its elements can be assigned positive molecular masses simultaneously in some mass conservation vector and unconservable otherwise [e.g. \{A, C\} in Equation (1)]. Since only the terms with zero coefficients can be excluded from an equation \( y^T \mathbf{m} = 0 \) without changing its solutions set, according to Proposition 1, the metabolites with non-zero coefficients in an inconsistent net stoichiometry comprise an unconservable set. The full metabolite set of an inconsistent network is unconserved, but it may contain conservable subsets [e.g. \{A, B\} in Equation (1)].

A minimal net stoichiometry also satisfies the linear undecomposability criterion. An unconservable set is minimal if all of its proper subsets are conservative [e.g. \{C\} in Equation (1) and \{A’, B’, C’\} in Fig. (1)]. Undecomposability criterion ensures that each minimal unconservable set is comprised of the metabolites having non-zero coefficients in exactly one inconsistent minimal net stoichiometry. The following corollary is implied by the criterion of solvability of the equation \( y^T \mathbf{m} = 0 \) with a positive \( \mathbf{y} \) according to Proposition 1.
Detection of stoichiometric inconsistencies

Corollary 1.2. If \( M \) is a minimal unconservable metabolite set, and \(|M| = 1\), then the metabolite has zero mass. If \(|M| > 1\), then all metabolites in the mass conservation vector have zero mass, or if any have a positive mass, then at least one metabolite has a negative mass.

Let us consider a semipositive conservation vector. Corollary 1.2 implies that in this vector, the components corresponding to the elements of minimal unconservable sets can be only equal to zero. Since each of the remaining metabolites can be assigned a positive molecular mass in at least one conservation vector, there must exist one in which all metabolites, not contained in minimal unconservable sets, are assigned positive molecular masses. We refer to such a vector as a maximal conservation vector. The fact that it is non-negative helps to prove the following proposition.

Proposition 2. A metabolite \( \text{met} \in M \) is unconserved if it is an element of at least one minimal unconservable set \( \mathcal{M}' \subseteq M \).

Proof. Necessary condition: If \( \text{met} \) is not an element of any minimal unconservable set, then the corresponding component in the maximal conservation vector of \( \mathcal{N} \) is positive and \( \text{met} \) is not unconserved. Sufficient condition: According to Corollary 1.2, \( \text{met} \in \mathcal{M}' \) cannot obtain a positive value in any non-negative conservation vector, therefore, \( \text{met} \) is unconserved.

Corollary 2.1. A metabolite is unconserved if the mass is not positive in a maximal conservation vector.

### 2.3 Inconsistent reaction sets

Localization of inconsistencies requires the detection of inconsistent subnetworks. In the simplest case, an inconsistent network consists of only one reaction; such a reaction is also called inconsistent. An example is shown below:

\[
A \leftrightarrow A + B
\]

We define a reaction stoichiometry as the vector of its stoichiometric coefficients in a reaction. Let us consider the equation \( y^T \mathbf{m} = 0 \) assuming that \( y \) is the stoichiometry and \( \mathbf{m} \) is the vector of molecular masses of the reactants. Proposition 1 implies that the reaction is inconsistent if it has either an empty set of substrates \( (y \geq 0) \) or an empty set of products \( (y \leq 0) \), but not both \( (y = 0) \). In other words, the stoichiometry of an inconsistent reaction is an inconsistent net stoichiometry as defined in Section 2.2. Metabolites occurring on both sides with equal coefficients are for most purposes represented by zeroes in a stoichiometry matrix, since their net concentration change is zero. Therefore, the reaction shown in Equation (7) is equivalent to \( \emptyset \leftrightarrow B' \) and hence inconsistent.

The more complex network shown below contains no inconsistent reactions. However, by subtracting the stoichiometry of \( R_2 \) from that of \( R_1 \) we obtain the inconsistent net stoichiometry \( \emptyset \leftrightarrow C \). Further, the linear combinations \( (R_1,-R_2)^T \) and \( (R_1,-R_2,-R_3)^T \) result in the net stoichiometries \( \emptyset \leftrightarrow B \) and \( A \leftrightarrow \emptyset \), respectively.

\[
\begin{align*}
R_1 : A & \leftrightarrow A' + B \\
R_2 : B + B' & \leftrightarrow C \\
R_3 : C + 2C' & \leftrightarrow A
\end{align*}
\]

Likewise, in the network shown in Figure 1a, the sum of all three reaction stoichiometries is \( A' + B' + 2C' \leftrightarrow \emptyset \). These observations are explained in Proposition 3.

Proposition 3. A network \( \mathcal{N} \) is inconsistent iff a vector \( \mathbf{v} \) exists, such that:

\[
\mathbf{N} \mathbf{v} = y
\]

where \( y \) is some inconsistent net stoichiometry in \( \mathcal{N} \).

Proof. From Proposition 1 it follows, that \( \mathcal{N} \) is inconsistent iff for each \( \mathbf{m} \), \( N^Ty = 0 \), there exists some semipositive or seminegative \( y \), \( y^T \mathbf{m} = 0 \). The equalities \( N^Ty = 0 \) and \( y^T \mathbf{m} = 0 \) imply that \( y^T \) is a linear combination of rows in \( N^T \). i.e. there exists some \( v \), \( v^TN^T = y^T \). Equivalently, \( \mathbf{N} \mathbf{v} = y \).

Proposition 3 demonstrates that in any inconsistent network, some linear combination of reaction stoichiometries is possible, such that the resulting net stoichiometry describes an inconsistent reaction. The solutions of the system \( \mathbf{N} \mathbf{v} = y \) for an inconsistent \( y \) are called leakage modes and the undecomposable leakage modes are called elementary, e.g. the aforementioned linear combinations in Equation (8) and Figure 1. Each inconsistent minimal net stoichiometry corresponds to an elementary leakage mode.

In contrast to the closely related concept of flux modes, the reaction directions and reversibilities in leakage modes are irrelevant. But similarly to elementary flux modes, elementary leakage modes represent minimal subnetworks able to 'leak' (i.e. to produce in a steady state without consuming anything or vice versa) certain metabolite sets. Therefore, they can be defined as precise locations of stoichiometric inconsistencies and their identification is helpful for the detection of input errors.
3 METHODS

3.1 Verification of stoichiometric consistency

The stoichiometric consistency of a metabolic network can be proved by finding a positive solution of the system $N^Ty = 0$ and disproved by demonstrating that such a solution does not exist. Equations (2) and (4) state a linear programming problem, which can be defined as follows:

Minimize $\sum_{i=1}^{m} m_i$

Subject to $N^Tm = 0$

Where $m_i \geq 1; 1 \leq i \leq m$

The program is solvable if the system is consistent. The third line ensures that the molecular masses are positive (since strict inequalities are not valid in linear programming, the expression $>0$ must be replaced by $\geq \alpha$ where $\alpha$ is an arbitrary positive number).

3.2 Detection of unconserved metabolites

According to Corollary 2.1, conserved and unconserved metabolites can be distinguished in a maximal conservation vector, which can be calculated by means of the following MILP:

Maximize $\sum_{i=1}^{m} k_i$

Subject to $N^Tm = 0$

Where $0 \leq k_i \leq m_i, k_i \in [0, 1]; 1 \leq i \leq m$

The number of positive components of $m$ is maximized. The third line ensures that each component $k_i$ of the vector of integers $k$ can be set to 1 if $m_i \geq 1$ and is 0 otherwise. Since no upper bound is defined for the components of $m$, the non-zero components are increased until all of them become greater than or equal to one. At this point, the number of unit components of $k$ reaches the possible maximum, thus satisfying the termination criterion.

3.3 Detection of inconsistent minimal net stoichiometries

By solving Equation (4) for unknown masses, we can determine if they are consistent. The algorithm involves calculation and analysis of the nullspace matrix $K$ of the system $N^Tm = 0$ (Heinrich and Schuster, 1996), also called the left nullspace matrix of $N$. The left nullspace keeps track of dependent masses. If it is empty, then the system has the trivial solution only. Hence, the only conservative molecular mass of each metabolite is zero and the network is inconsistent. Otherwise, each column in $K$ corresponds to one of the metabolites serving as independent variables for nullspace calculation. Let us denote by $d$ the vector of molecular masses of these metabolites, taken in the same order as in the columns of $K$. Then, each row in $K$ describes the molecular mass of one of the metabolites as a weighted sum of molecular masses of independent metabolites, according to the following equation:

$$m_i = \sum_{j=0}^{d} K_{ij}d_j, 1 \leq i \leq m$$

where $d$ is the dimension of the nullspace (Fig. 1d).

**PROPOSITION 4.** Let $N$ be a network with a left nullspace matrix $K$. Any solution $y$ of the system $y^TK = 0$ is a net stoichiometry of $N$.

**PROOF.** According to Equation (12), each $i$-th element in $m$ can be replaced by the sum of the $i$-th row in $K$. Hence, the equation $y^Tm = 0$ has the same solutions as the system $y^TK = 0$.

Proposition 4 implies that each inconsistent minimal net stoichiometry is an undecomposable semipositive or seminegative solution of the system $y^TK = 0$. As a special case, a singleton minimal unconservable set is always represented by an all-zero row in the left nullspace matrix. Further, an inconsistent minimal net stoichiometry involving a given unconserved metabolite with a non-zero coefficient can be found using the following MILP:

Minimize $\sum_{i=1}^{m} k_i$

Subject to $y^TK = 0, y_j \geq \epsilon$

Where $0 \leq y_i \leq k_i, k_i \in [0, 1]; 1 \leq i \leq m$

The program minimizes the number of positive components in the vector $y$, thus ensuring its undecomposability. The third line ensures that each component $k_i$ of the vector of integers $k$ can be set to 0 if $y_j = 0$ and is equal to 1 otherwise. The second line states that the mass of the metabolite corresponding to the $j$-th row is positive in the solution (since the upper bound of the components of $y$ is set to 1, the lowest possible positive value must be a sufficiently small number, denoted here as $\epsilon$). However, the same metabolite can be an element of more than one minimal unconservable set (given that none of them is singleton). In order to find the next solution, the one already found must be excluded from the further iterations. This can be done by complementing the Equation (13) by an additional constraint called an integer cut (Nikolaev et al., 2005):

$$\sum_{i=1}^{ip} k_i \leq P - 1,$$

where $k_i = i_1, \ldots, i_p$ are the unit components of a previously found solution $k$. The integer cut ensures that these components will not obtain unit values simultaneously in any further solutions, since their sum is forced to be less than their number. The program must be invoked iteratively until no more solutions can be found (see Algorithm 1):

**Algorithm 1** Identify the set $Y$ of inconsistent minimal net stoichiometries in a stoichiometry matrix $N$

$Y := []$

$U := \text{unconserved}_\text{metabolites}(N^T) /* Eq. 11 */$

$K := \text{nullspace}_\text{mtx}(N^T)$

**for all** $m \in U$ **do**

**if** $K_{m} = 0$ **then**

$y := [0, \ldots, 0]$

$y_{\text{met}} := 1 /* all-zero row \rightarrow singleton set*/$

$Y := Y \cup \{y\}$

**else**

$\text{prog} := \text{mixed}_\text{integer}_\text{program}(K)$

$y := \text{minimal}_\text{solution}(\text{prog}, \text{met}) /* Eq. 13 */$

**while** $\text{is feasible}(\text{prog})$ **do**

$Y := Y \cup \{y\}$

$\text{set}_\text{integer}_\text{cut}(\text{prog}, \text{y}) /* Eq. 14 */$

$y := \text{minimal}_\text{solution}(\text{prog}, \text{met})$

**end while**

**end if**

**end for**

3.4 Detection of elementary leakage modes

The detection of inconsistent minimal net stoichiometries enables the identification of the corresponding elementary leakage modes by solving the system of equations $Nv = 0$. A given inconsistent minimal net stoichiometry $y$ can be augmented at the right side of the stoichiometry matrix. This is equivalent to including an additional inconsistent reaction into the network; this reaction may be also considered as a transport reaction in an open metabolic system and will be further referred as the transporter (Fig. 2b). In terms of metabolic modelling, the transporter delivers the molecules which are then lost in the leakage modes, thus satisfying the steady-state condition and enabling the calculation of elementary flux modes in the system $(N|y)v = 0$, considering all reactions as reversible. The elementary flux modes involving the transporter have an inconsistent net stoichiometry.
The above described algorithms were applied to genome-scale models of E.coli, S.cerevisiae and S.agalactiae (built automatically using the annotation and the biochemical data available in the KEGG database). It is worth emphasizing that the quality of the underlying annotations is irrelevant, since stoichiometric inconsistencies are caused by errors on the level of biochemical reaction definitions.

The model of E.coli (931 reactions, 625 metabolites) is stoichiometrically consistent. The model of S.cerevisiae (1172 reactions, 809 metabolites), in contrast, contains 359 unconserved metabolites [very similar results describing these models have been published by Nikolaev et al. (2005)]. Further, in the model of S.cerevisiae, 4030 minimal unconservable sets have been found, out of which 144 are singleton, 2651 comprise two and 1235 comprise three metabolites. Nullspace calculation has been used for the detection of elementary leakage modes. An example is shown in Figure 3(a and c): the same reaction set comprises two elementary leakage modes, which differ in the flux rate coefficients and produce [H₂O₂] and [O₂], respectively. The inconsistency is caused by the incorrect definition of the reaction YGR088W, where two molecules of water are missing on the right-hand side. After correcting this reaction, the subnetwork shown becomes consistent.

The model of S.agalactiae (718 reactions, 904 metabolites) contains 215 unconserved metabolites and 214 minimal unconservable subsets; 212 out of them are singleton and 2 comprise 2 metabolites. For each minimal unconservable set, one elementary leakage mode has been detected by nullspace calculation; the numbers of involved reactions vary between 4 and 96. The analysis of detected modes revealed that most of them involve reactions, in whose definitions the same polymeric molecules appear on both sides and are cancelled out in the
stoichiometry matrix. Two examples are shown below:

R07284: ATP ↔ Pyrophosphate
R07283: CTP ↔ Pyrophosphate

According to these reactions, ATP and CTP are isomers of pyrophosphate and of each other (in KEGG, they are defined as follows: ATP + tRNA ↔ Pyrophosphate + tRNA and ATP + tRNA ↔ Pyrophosphate + tRNA). These reactions are involved in 204 and 158 elementary leakage modes, respectively, out of the detected 214 ones. The whole model contains 18 reactions involving a polymer on both sides; after removing all of them, only two metabolites remain unconserved: water and proton.

An elementary leakage mode producing a proton is shown in Figure 3(b and d); this mode has been obtained using mixed-integer programming. The inconsistency is caused by the incorrect definition of the reaction R04326, where one hydrogen atom is missing on the right-hand side. The model contains 71 reactions with incorrect atomic balances; however, even after removing all of them, proton still remains unconserved. This is apparently due to the further 170 reactions, whose atomic balance cannot be tested because of missing or ambiguous empirical formulae. Since the concentrations of water and protons are likely to be irrelevant for biochemical analysis (water is present in a large excess and protons are buffered), they can be safely declared external, thus resolving the inconsistencies in the internal stoichiometry matrix.

5 DISCUSSION AND CONCLUSION

In the present work, we introduced the concept of stoichiometric inconsistencies—a common structural error in metabolic models. We described fundamental properties of incorrectly defined metabolic networks, such as inconsistent net stoichiometries and elementary leakage modes and proposed algorithms identifying these properties. In this section, we will make some remarks about the practical applicability and the general implications of the concepts and methods introduced.

Practical applications: We proposed an algorithm detecting the complete set of unconserved metabolites in a network using a MILP. Although the complexity of the latter is much higher than the complexity of the linear program used in the algorithm proposed by Nikolaev et al. (2005), the MILP needs to be invoked only once for the whole model. Thus the full computation takes a significantly shorter time. The detection of unconserved metabolites enables a simple method for resolving all inconsistencies in a network, namely by externalizing these metabolites, i.e. by removing the corresponding rows from the internal stoichiometry matrix, which is used for the analysis of reaction fluxes. This solution may be acceptable if the concentrations and conversions of unconserved metabolites are not relevant for the investigated problem, such as the case with water and proton discussed in Section 4. However, in most cases, more sophisticated solutions are needed, for which the identification of unconserved metabolites is a necessary step.

A more precise characterization of inconsistent networks is possible by detection of inconsistent minimal net stoichiometries. This is enabled by the analysis of the left nullspace $\mathbf{K}$ in Algorithm 1. Although this algorithm solves the non-polynomial problem of finding a complete set of undecomposable solutions in the system $\mathbf{y}^{T}\mathbf{K} = \mathbf{0}$, it appears to be feasible even in large genome-scale models due to: (a) the relatively small size of the left nullspace matrix, compared to a whole network and (b) usually small numbers of inconsistent minimal net stoichiometries involving a given metabolite.

In contrast, the calculation of a complete set of elementary leakage modes is subject to combinatorial explosion and is currently not feasible in genome-scale models. On the other hand, the total number of elementary leakage modes can be very large. So, in a model of a potato tuber (constructed manually in our group) containing 30 reactions and 31 metabolites, 7 out of them unconserved, we found 123 elementary leakage modes, all caused by a single input error. Such amounts of data are hardly analysable and apparently superfluous. Therefore, we propose the calculation and analysis of a limited population of elementary leakage modes using Gaussian elimination. As demonstrated in Section 4, this result can be used in two ways: inspection of individual modes (Fig. 3) and of those reactions which are involved in large percentages of the calculated modes.

The smallest possible leakage mode with a given net stoichiometry can be identified using a modified version of the mixed-integer program shown in Equation (13). The advantage of this method is the maximally precise localisation of input errors (Fig. 3(a and c)), the disadvantage is the high computational cost of using integer constraints and especially integer cuts in large models. Hence, the next mode should be calculated after the error in an already found one is corrected.

General remarks: stoichiometric inconsistencies may be considered in the context of graph theory. The fact, that elementary leakage modes can carry (invalid) steady-state fluxes in a closed network implies, that the reactions comprising them are organized in cycles. These cycles have either of the properties described below:

(a) Some metabolites enter the cycle, but nothing leaves it, or vice versa [see Equations (1) and (8) and Figs 1 and 3d]. So, after each turnover, the total mass of the system increases or decreases, assuming that the entering or leaving metabolites have positive molecular masses. Therefore, these metabolites comprise a minimal unconserving set. Elementary leakage modes of this type can be detected using elementary modes analysis, given that the entering or leaving metabolites are externalized. In contrast, the algorithms proposed in the present work do not depend on the selection of external metabolites.

(b) Inconsistently defined stoichiometric coefficients. So, considering these coefficients as the edge weights in a graph, the product of all edge weights in a cycle is greater or less than one, thus increasing or decreasing the total mass of the system. A simple example is shown below:

$$R_1: A \leftrightarrow B$$
$$R_2: A \leftrightarrow 2B$$

Interestingly, the reactions involved in such a cycle may comprise more than one elementary leakage mode with different net stoichiometries, depending on the coefficients. So, in Equation (15), the leakage modes $(R_2, -R_1)$ and $(R_2, -2R_1)$ have the net stoichiometries $\emptyset \leftrightarrow B$ and $\emptyset \leftrightarrow A$. Another example is shown in Figure 3(a and c).

Conservation relationships (Heinrich and Schuster, 1996) are linear combinations of metabolite concentrations which are invariant over time; they can be formally represented as conservation vectors. The positive components of these vectors correspond to
metabolites containing some common conserved molecular moiety and represent the numbers of the moiety units in the corresponding molecules. In this work, we suggested an alternative interpretation of conservation vectors as distributions of molecular masses admissible under the assumption of mass conservation (hence we called them mass conservation vectors). This interpretation can be justified as follows: the largest non-negative conservation relationship in any connected closed system represents the total mass conservation; if the model is valid, this relationship must involve all metabolites. The common molecular moiety in this case is the largest discrete unit of matter with the molecular mass of 1 (physically these units represent protons and neutrons). Hence, the coefficients can be interpreted as the conservable ratios of the contained mass units, i.e. of the molecular masses.

Although the emphasis in this article has been placed on the detection of errors, the theoretical foundations laid in it may give rise to other interesting methods of analysis of living systems. In contrast to the traditional interpretation of reactions as continuous molecular interconversions, we consider them as transformations of metabolite sets, whose possibilities are defined by the stoichiometries (hence, directions and reversibilities are irrelevant) and subject to the mass conservation constraint. This constraint serves as the basis for Equation (4), where reactions are regarded as linear equations and metabolites as variables. The connection between left and right nullspace analyses, i.e. between the possible distributions of molecular masses and reaction flux rates [Equations (5) and (9)] deserves a further investigation. Leaving these perspectives open for the future developments, the present work provides effective and useful solutions to a range of model validation problems.

ACKNOWLEDGEMENTS

The authors thank Prof. Michael Pidcock, Oxford Brookes University, for the valuable discussions and kind help in the mathematical verification and presentation of the work; Dr Mark Anthony, Birmingham Women’s Health Care Trust, for the help in the construction of S. agalactiae model and the anonymous reviewers of the former version of the manuscript for the useful remarks and suggestions.

Funding: Oxford Brookes University (studentship to A.G.) Biotechnology and Biosciences Research Council (BB/E00203X/1).

Conflict of Interest: none declared.

REFERENCES

Richter, O. et al. (1975) The response of oscillating glycolysis to perturbations in the NAD/NADH system: a comparison between experiments and a computer model. Biosystems, 7, 137–146.