BIOINFORMATICS

Vol. 20 no. 2 2004, pages 226–234
DOI: 10.1093/bioinformatics/btg395

Minimal cut sets in biochemical reaction networks

Steffen Klamt and Ernst Dieter Gilles*

Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr.1, D-39106 Magdeburg, Germany

Received on April 28, 2003; revised on July 8, 2003; accepted on July 18, 2003

ABSTRACT

Motivation: Structural studies of metabolic networks yield deeper insight into topology, functionality and capabilities of the metabolisms of different organisms. Here, we address the analysis of potential failure modes in metabolic networks whose occurrence will render the network structurally incapable of performing certain functions. Such studies will help to identify crucial parts in the network structure and to find suitable targets for repressing undesired metabolic functions.

Results: We introduce the concept of minimal cut sets for biochemical networks. A minimal cut set (MCS) is a minimal (irreducible) set of reactions in the network whose inactivation will definitely lead to a failure in certain network functions. We present an algorithm which enables the computation of the MCSs in a given network related to user-defined objective reactions. This algorithm operates on elementary modes. A number of potential applications are outlined, including network verifications, phenotype predictions, assessing structural robustness and fragility, metabolic flux analysis and target identification in drug discovery. Applications are illustrated by the MCSs in the central metabolism of *Escherichia coli* for growth on different substrates.

Availability: Computation and analysis of MCSs is an additional feature of the FluxAnalyzer (freely available for academic users upon request, special contracts for industrial companies; see web page below).

Contact: klamt@mpi-magdeburg.mpg.de

Supplementary information: http://www.mpi-magdeburg.mpg.de/projects/fluxanalyzer

INTRODUCTION

Cellular life is characterized by a vast number of direct or indirect interactions of cellular components. Thus, a holistic understanding of cellular processes means understanding the structure, dynamic behavior and regulation of cellular networks. Depending on how the network nodes interact, different types of cellular networks—and abstractions thereof—can be identified, such as metabolic networks (transport, conversion, synthesis and degradation of material), signal transduction networks (signal flow and integration) or gene and regulatory networks (activation and inhibition of genes or gene derivatives). Interfaces between these networks provide exchange of material and information.

Systems biology (Kitano, 2002) aims at investigating cellular networks by combining experiments and mathematical modeling and computer simulations. Knowledge of the network elements and their interactions is a fundamental prerequisite for 'forward' modeling. So far, of all network types, metabolic networks are structurally best-characterized as they can now be reconstructed for many organisms up to genome-scale (Ouzonis and Karp, 2000; Ma and Zeng, 2003a; Förster *et al*., 2003). Studies relying on the stoichiometry of (metabolic) reaction systems have demonstrated that the underlying network structure limits the possible overall behavior. Thus, a number of physiologically important results can be derived solely from the well-known structure without knowledge of kinetic mechanisms and parameters. For example, mass-action theory (Clarke, 1988; Feinberg, 1995; Bailey, 2001) allows for predictions of the existence, multiplicity and stability of steady states in reaction networks. Studies on the large-scale topology have yielded deeper insights into the global organization of metabolic networks (e.g. Fell and Wagner, 2000; Ma and Zeng, 2003a,b). Flux Balance Analysis has been used to study and predict optimal flux patterns (Ibarra *et al*., 2002; Price *et al*., 2003) and mutant phenotypes (Edwards and Palsson, 2000). Pathway analysis relying on elementary modes or the very closely related extreme pathways (Schuster *et al*., 2000; Schilling *et al*., 2000; Klamt and Stelling, 2003a) enables the assessment of structural robustness and redundancy (e.g. Papin *et al*., 2002; Stelling *et al*., 2002). In (Stelling *et al*., 2002), pathway analysis has been used to predict mutant phenotypes and to give rough estimates of gene expression ratios taking into account network flexibility and efficiency.

In this work we introduce the concept of minimal cut sets (MCSs). They can be considered as the smallest 'failure modes' in the network that render the correct functioning of a cellular reaction impossible. Analyzing MCSs is somewhat opposite to the approaches mentioned above, which determine rather the capabilities of a network. However, we will illustrate

*To whom correspondence should be addressed.

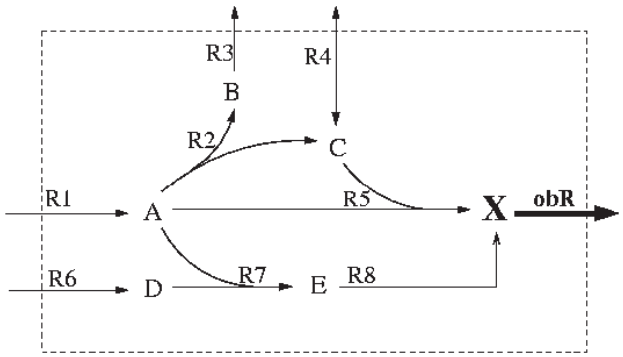

Fig. 1. Network example (NetEx).

a number of potential applications, including target identification in metabolic engineering and drug discovery, network verifications, structural fragility analysis and observability in metabolic flux analyses. An algorithm for computing MCSs in arbitrary biochemical reaction networks will be described and the applicability of MCSs will be illustrated for the central carbon metabolism in *Escherichia coli*.

Definition of MCSs

For illustration, we first consider a simple fictitious reaction network called NetEx (Fig. 1). Reactions crossing the system boundaries may be thought of as coming from/leading to buffered metabolites. The only reversible reaction is R4. In NetEx we are particularly interested in the flux through reaction obR (exporting synthesized metabolite X), which we therefore call the objective reaction. It is easy to see that there are different possibilities for synthesizing X. One approach for finding qualitatively distinct pathways is to calculate the elementary (flux) modes (EMs) of the network (Schuster *et al*., 2000). An EM describes a feasible and balanced (steady-state) flux distribution through the network, which is minimal with respect to utilized reactions (enzymes). The condition of being 'minimal' (elementary, non-decomposable, irreducible) is fundamental and means that removing any reaction in an EM impedes the realization of a non-zero steady-state flux distribution (and therefore of a functionality) by means of the remaining reactions of the EM. It is important to see that the condition of being 'minimal' is related to the exclusion of involved reactions and not to their total number. We will see that a similar property is also characteristic for MCSs. Computation of EMs in a given network only requires the stoichiometric matrix and the (practically relevant) reversibilities of the reactions (Schuster *et al*., 2000, 2002). In NetEx, one finds four EMs (Table 1). Three of them (EM2–EM4, shaded region in Table 1) allow the production of metabolite X, i.e. they involve the objective reaction exporting the produced X.

Now assume that we want to prevent the production of metabolite X. Using a structural approach we demand that

Table 1. Elementary modes (EMx) and minimal cut sets (MCSx) in the network example (NetEx, Fig. 1)

	R1	R2	R3	R4	R5	R6	R7	R8	obR
Elementary modes									
EM1	1	1	1	−1	0	0	0	0	0
EM2	1	0	0	0	0	1	1	1	1
EM3	2	1	1	0	1	0	0	0	1
EM4	1	0	0	1	1	0	0	0	1
Minimal cut sets (objective reaction: obR)									
MCS0									×
MCS1	×								
MCS2					×	×			
MCS3					×		×		
MCS4					×			×	
MCS5		×		×		×			
MCS6			×	×		×			
MCS7		×		×			×		
MCS8			×	×			×		
MCS9		×		×				×	
MCS10			×	×				×	
F_i	1	1/3	1/3	1/3	1/2	3/8	3/8	3/8	1

The three EMs involving the objective reaction (obR) are highlighted. F_i: fragility coefficient of reaction i.

there is no balanced flux distribution possible, which involves the objective reaction obR. One strategy would be to inactivate (cut) one or several reactions in the network, e.g. by deleting the genes of certain enzymes or by other manipulations resulting in an inhibition of enzyme activity. This leads us to the definition of a cut set:

DEFINITION. *We call a set of reactions a cut set* (*with respect to a defined objective reaction*) *if after the removal of these reactions from the network no feasible balanced flux distribution involves the objective reaction.*

Obviously, a trivial cut set is the objective reaction itself: C0 = {obR}. However, there are different reasons why we might be interested in other potential cut sets. For example, from an engineering point of view, it might be desirable to cut reactions at the beginning of a pathway (whereas obR, e.g. is at the end). Another example is biomass synthesis, which in structural network analyses often is considered as one lumped 'reaction'. This reaction is not related to a single gene or enzyme and can therefore not be directly inactivated. Furthermore, from a physiological point of view, it might be interesting to see under which distinct conditions the synthesis of metabolite X is definitely impossible. Finally, one can also define a set of several objective reactions (see below), whose simultaneous failure might be achieved more efficiently by cutting away other reactions.

An extreme choice is the removal of all reactions except obR. However, this would not be an efficient and intelligent cut set. For example, C1 = {R5, R8} is a cut set already sufficient

for preventing the production of X. Moreover, removing only R5 or only R8 cannot exclude the possibility that the objective reaction works in a balanced manner. Thus, we have found a minimal cut set because no subset of C1 would be a cut set anymore.

DEFINITION. *A cut set C (related to a defined objective reaction) is a minimal cut set (MCS) if no proper subset of C is a cut set.*

Another cut set not that simple is C2 = {R2, R4, R6}, which is also minimal. A third cut set is C3 = {R2, R5, R7}. However, this cut set is not minimal, because {R5, R7} is already a cut set. Thus, C3 contains a redundant deletion which makes this cut set sub-optimal (non-minimal) for manipulating the system. Finally, C4 = {R1} is the only MCS with only one reaction (except the trivial C0). This would be an especially suitable candidate if we want to prevent production of X. We would say that R1 is essential for synthesizing X.

An algorithm for computing MCSs is given below. All MCSs in NetEx related to obR are given in Table 1.

REMARK 1. *The removal of all reactions contained in an MCS ensures a dysfunction of the objective reaction from a perspective of the network structure. Thus, as the structural integrity of the network is the fundamental prerequisite for its function, an MCS always guarantees dysfunction as long as the assumed network structure is correct. However, it might be possible that even a proper subset of an MCS is a cut set as the often unknown regulatory circuits or capacity restrictions may impose further constraints with respect to network function. For example, if removal of one reaction of an MCS leads, by means of cellular control, to a repression of another reaction from this MCS, then this MCS would not be minimal anymore. Therefore, when in the following we speak about MCSs and network dysfunction it should always be seen from a purely structural point of view. Extending the concept by considering also regulatory rules is possible, as will be indicated below. Finally, tolerating the inactivation of reactions not forming a cut set can, nevertheless, be accompanied by an altered performance of the objective reaction (e.g. lower rate or yield). This is also not further considered here.*

REMARK 2. *Removing a complete MCS from the network represses the functioning of the objective reaction by definition. However, other pathways might still be active: in NetEx, the metabolites B and C can still be produced from A after removal of MCS2 = {R5, R6}. In some applications, it is just the goal to find those MCSs that repress the objective reaction but spare certain other pathways.*

REMARK 3. *When looking at MCS4 = {R5, R8} it is clear that metabolite X cannot be produced in any case as all pathways to it are interrupted. However, for the MCS6 = {R3, R4, R6} it seems that metabolite X might be still produced via R1, R2, R5. However, an MCS always ensures that this would only be possible in an unbalanced (and therefore physiologically problematic) fashion. For MCS6, metabolite B would accumulate and a continuous operation of the objective reaction is therefore physiologically impossible.*

REMARK 4. *In our example we consider only one objective reaction. However, one can easily generalize the above definition of (minimal) cut sets by using a set of several objective reactions. It is then demanded that after the removal of the (minimal) cut set each feasible balanced flux distribution does not involve any of the objective reactions. For example, all reactions in the network are objective reactions, if any balanced flux in the network is to be prevented.*

Minimal cut sets in risk assessment and graph theory An MCS can be considered as a minimal set of events (loss of reactions) which—if these events occur together—leads to system failure, i.e. that the objective reaction cannot operate in a balanced fashion anymore. A very similar definition of MCSs exists for fault trees studied in reliability and risk assessment of industrial systems (Fard, 1997; Sinnamon and Andrews, 1997). A fault tree is a non-recursive boolean network and consists of a number of events, which are combined by logic gates leading to other events. Basic events are the 'entries' at the lowest level (leafs of the tree) and intermediate events are those obtained by binary operations (e.g. AND, OR, XOR) of other events. At the top of the fault tree is the top event which represents a usually undesired system failure. A cut set in a fault tree is a set of basic events, whose occurrence will cause the top event, and an MCS possesses no proper subset which can cause the top event. The approach pursued here is completely analogous: our top event is that the objective reaction cannot operate correctly. A basic event is the removal/inhibition of one reaction. An intermediate event is a set of several inactivated reactions. The difference between fault trees and our approach is that we cannot directly construct a fault tree, as we—at least for metabolic networks—do not know which combinations of removed reactions cause our top event. Therefore, for calculating MCSs in metabolic networks, we cannot directly apply algorithms used for fault trees.

A similar definition of (minimal) cut sets does also exist in graph theory (Bollabas, 1998) where they ensure a disconnection of a given graph. However, as far as we can see, these graph-theoretical concepts (even if applied in bipartite graphs) do not fit into the definition of MCSs as defined here and would, in general, lead to other results. The reason is that we need an explicit consideration of the hypergraphical nature of metabolic networks. Hypergraphs are generalized graphs, where an edge (reaction) can link k nodes (reactants) with l nodes (products), whereas in graphs only 1 : 1-relations are allowed. For illustration, consider the simple network shown in Figure 2 (first row). Metabolite A is an available substrate and we are interested in inhibiting the production of E. Thus, R4 is our objective reaction. It is easy to see, that E can no

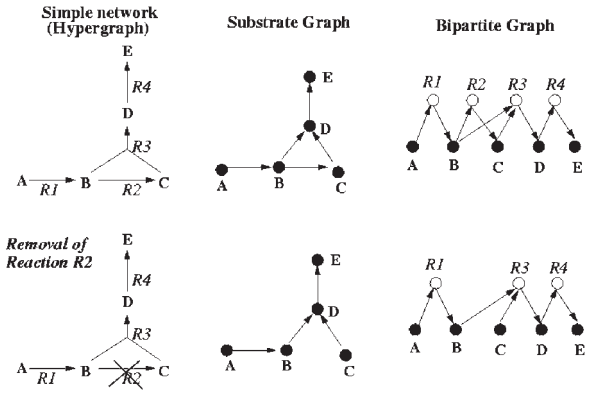

Fig. 2. Hypergraph, substrate and bipartite graph representations of a simple reaction network (first row). After removing reaction R2 (second row) product E can no longer be produced from substrate A in the hypergraph, although there is still a path from A to E in both graph representations.

longer be produced if reaction R2 is removed from the network (because C cannot be provided for driving reaction R3). Thus, {R2} is an MCS. However, R2 would not be an MCS in terms of graph theory, neither in the substrate nor in the bipartite graph representation of this network. All metabolites are still connected when R2 is removed and it seems, that E can still be produced from A via {R1, R3, R4} (Fig. 2, second row). The hypergraphical feature of reaction R3—it needs B and C simultaneously—is not taken into account. One obtains the same result when assuming all reactions to be reversible (leading to undirected graphs).

Algorithm for computing MCSs

The MCSs for a given network and objective reaction are members of the power set of the set of reaction indices and are uniquely determined. In NetEx, it is relatively easy to determine all MCSs. Obviously, for larger networks we need a systematic computation scheme. This scheme must ensure that the calculated MCSs are

(1) cut sets ('destroying' all possible balanced flux distributions involving the objective reaction) and
(2) that the MCSs are really minimal and
(3) that all MCSs are found.

Concerning condition (1) our developed algorithm exploits the fact that any feasible steady-state flux distribution in a given network—expressed by a vector **r** of the q net reaction rates—can be represented by a non-negative linear combination of elementary modes:

$$\mathbf{r} = \sum_{i=1}^{N} \gamma_i \, \mathrm{EM}_i \quad \text{with } \gamma_i \geq 0,$$

N = number of elementary modes

Thus, if we want to ensure that the rate r_k of the objective reaction is zero in all feasible and balanced flux distributions **r**, then each EM must contain a zero at the k-th place. Therefore, if C is a proper cut set the following condition must hold:

Cut set condition (CSC): For each EM involving the objective reaction (with a non-zero value), there is at least one reaction in C also involved in this EM.

The condition guarantees that all EMs, in which the objective reaction participates, will vanish when the reactions in the cut set are removed from the network. We say that a cut set 'covers' all EMs where the objective reaction is involved. The CSC explicitly exploits the important conservation property of EMs: if a reaction is removed from a network, all EMs not involving this reaction build up the complete set of EMs in the new (smaller) network without need of recalculation (Schuster *et al*., 2002; Klamt and Stelling, 2003a). Utilizing the CSC, computing a cut set can be done by successively building up a set of reactions (preliminary cut set) until all EMs involving the objective reaction are covered by this reaction set. Keeping also in mind conditions (2) and (3), the pseudo-code of our iterative algorithm is as follows:

ALGORITHM:

(1) Calculate the EMs in the given network
(2) Define the objective reaction obR
(3) Choose all EMs where reaction obR is non-zero and store it in the binary array *em_obR* (*em_obR*[*i*][*j*]==1 means that reaction *j* is involved in EM *i*)
(4) Initialize arrays *mcs* and *precutsets* as follows (each array contains sets of reaction indices): append {*j*} to *mcs* if reaction *j* is essential (*em_obR*[*i*][*j*]=1 for each EM *i*), otherwise to *precutsets*
(5) FOR i=2 TO MAX_CUTSETSIZE
  (5.1) *new_precutsets*=[ ];
  (5.2) FOR $j = 1$ TO q (q: number of reactions)
    (5.2.1) Remove all sets from *precutsets* where reaction *j* participates
    (5.2.2) Find all sets of reactions in *precutsets* that do not cover at least one EM in *em_obR* where reaction *j* participates; combine each of these sets with reaction *j* and store the new preliminary cut sets in *temp_precutsets*
    (5.2.3) Drop all *temp_precutsets* which are a superset of any of the already determined minimal cut sets stored in *mcs*
    (5.2.4) Find all retained *temp_precutsets* which do now cover all EMs and append them to *mcs*; append all others to *new_precutsets*
  ENDFOR

(5.3) If isempty(*new_precutsets*)
(5.3.1) Break
ELSE
(5.3.2) *precutsets=new_precutsets*
ENDIF
ENDFOR

(6) result: *mcs* contains the MCSs

Additional remarks:

- All used arrays contain sets of the reaction indices $(1 \cdots q)$ that can efficiently be stored bit-wise and allow for quick bit-operations.
- An important property of MCSs—which can easily be seen—is that at most one reaction per enzyme subset is contained in an MCS. Enzyme subsets are sets of reactions that must always operate together in steady state, i.e. either all or none of the reactions of an enzyme subset participate in an EM (Pfeiffer *et al*., 1999). In NetEx two enzyme subsets occur: ES1 = {R2, R3} and ES2 = {R6, R7, R8}. For example, if one reaction of ES2 appears in an EM, then also the other two do. In contrast, no pair of this triple occurs together in any MCS. On the other hand, for each MCS in which R6 occurs, there are equivalent MCSs in which R6 is exchanged by R7 or R8. This property can be exploited in a preprocessing step (before the main loop): all reactions (columns in *em_obR*) of an enzyme subset are removed except for one representative. After computation, the equivalent sets can easily be built up for each enzyme subset and be appended to *mcs*. Thus, if an enzyme subset comprises e reactions, and if the representative is contained in z MCSs after computation, then $z \cdot (e-1)$ equivalent sets are appended to *mcs*. This can drastically reduce memory demand and computation time.
- Reactions catalyzed by a common multifunctional enzyme (like transketolase) must be considered properly because a removal of one enzyme removes all these reactions simultaneously. This can easily be achieved by considering only one 'reaction' (column) in *em_obR* containing a '1' for each EM, if any of the reactions of that multifunctional enzyme is involved in the respective EM.
- Knowledge about regulatory rules can be incorporated by eliminating those EMs (before step 3) which can, by regulatory actions, never occur (Covert and Palsson, 2003). For example, uptake of glucose and lactate does not occur simultaneously in *Escherichia coli* due to catabolite repression and, hence, all EMs involving both uptake reactions might be deleted. This reduces the set of EMs each MCS has to cover.
- For considering a set of several objective reactions (see Remark 4) one selects all EMs, where any of the objective reactions participates (step 3).

Table 2. Overview on computed MCSs in the central metabolism of *E.coli* for growth on four different substrates

	Acetate	Succinate	Glycerol	Glucose
No. of EMs with growth	363	3421	9479	21 592
No. of MCSs (objective reaction: growth)	245	1255	2970	4225
Maximal number of preliminary MCSs (during computation)	3563	69 628	344 196	902 769
Computation time (Intel Pentium, 1 MHZ; 4 GB RAM)	7 s	20 min	5.42 h	29.67 h
F_i values (in parentheses: size of the smallest MCS in which the reaction occurs)				
F16P-bisphosphatase	1 (1)	1 (1)	1 (1)	0.102 (6)
ATP-synthase	1 (1)	0.325 (3)	0.141 (3)	0.149 (3)
SuccCoA-synthetase	0.207 (2)	0.145 (2)	0.125 (2)	0.131 (2)
PEP-carboxylase	0.128 (2)	0.117 (2)	0.120 (2)	0.143 (2)
Malic enzyme	0.5 (2)	0.5 (2)	0.114 (2)	0.123 (2)
Rl5P-X5P (epimerase)	0.198 (2)	0.135 (2)	0.128 (2)	0.148 (2)
F	0.783	0.718	0.699	0.643

The computation time does not involve the time needed for computing the elementary modes. F_i: fragility coefficient of reaction i; **F**: network (overall) fragility coefficient. See text for further explanations.

- Step 5.2.1 avoids redundant combinations of reactions for finding the MCSs.
- The current coverage (of EMs) of each preliminary cut set may be saved accelerating step 5.2.2. However, this can drastically increase the memory demand.
- Step 5.2.3 ensures that the computed MCSs are really *minimal*. This check is analogous to the one used for calculating elementary modes (cf. Schuster *et al*., 2000). This emphasizes the conceptual similarities of EMs and MCSs with respect to the 'minimal' (non-redundant, elementary, non-decomposable) property.
- Calculating the elementary modes is known to be a combinatorial complex problem, as the number of EMs grows rapidly with network size (Klamt and Stelling, 2002b). Computing the MCSs is similarly complex due to a very high number of possible permutations. Although in a larger example (see below and Table 2), the final number of MCSs is lower than the number of EMs, the number of preliminary MCSs occurring during the calculation is up to 100 times higher than the number of EMs. On the other hand—in contrast to EMs—computing MCSs only involves bit arrays and fast bit operations. In the large example, the required memory space was more limiting than the computation time.

  Therefore, the opportunity to limit the maximal size of the MCSs to be calculated is a very useful feature of our algorithm (MAX_CUTSETSIZE, step 5). The algorithm ensures that after finishing the i-th outer loop, all existing

MCSs with equal to or less than i elements have been found and stored in *mcs*.

- Computation, display and subsequent analysis of MCSs [subject to arbitrary objective reaction(s)] are incorporated as additional features in the FluxAnalyzer—a MATLAB package for studying metabolic networks on interactive flux maps (Klamt *et al*., 2003b).

Minimal cut sets in the central metabolism of *E.coli*

We calculated the MCS in the central metabolic network of *E.coli* with objective reaction 'biomass synthesis'. The network comprises 110 reactions and 89 metabolites and has already been investigated by elementary-mode analysis (Stelling *et al*., 2002). The catabolic part was modeled in detail, whereas for the anabolic part predominately lumped reactions were considered. The set of catabolic reactions includes pseudo-reactions enabling excretion of five metabolites and the uptake of four substrates (glucose, acetate, glycerol and succinate). As in the prior work, growth is considered on each of the four substrates individually. Thus, actually four networks are investigated differing only in the respective substrate uptake/transport reactions. Using FluxAnalyzer we determined the EMs and then the MCSs for each case (Table 2). Note that the number of EMs is smaller than in the original paper as we consider here only the growth-related modes according to our objective reaction. The computed MCSs are discussed below in connection with potential applications for MCSs.

Applications of MCSs

The conceptual properties of MCSs offer a number of potential applications both for obtaining a deeper understanding of structural fragility of cellular (sub)networks as well as for finding targets that efficiently repress cellular functions.

Target identification and repressing cellular functions The concept of MCSs is an excellent theoretical tool for target identification in drug discovery as well as for metabolic engineering towards rational strain design (Stephanopoulos *et al*., 1998; Wiechert, 2002). The MCSs define all efficient (irreducible) sets of interventions that will lead to an intended dysfunction in the manipulated network. Here, it becomes clear why we have chosen the term 'objective reaction' instead of 'target reaction': each MCS represents a set of targets to prevent the functioning of the objective reaction (the MCS containing the objective reaction itself is the trivial target set). A screening of all MCSs allows for the identification of the best suitable manipulation from a structural point of view, which, for practical reasons, will probably have to fulfill certain conditions:

- Usually, a small number of interventions is desirable (i.e. small size of the MCS).
- Other pathways in the network should only be weakly affected, i.e. one searches for an MCS where these routes are still—at least structurally—functioning (see Remark 2). Whether an MCS fulfills this condition can easily be checked: the set of remaining EMs [not covered (destroyed) by the MCS] must contain at least one EM involving the desired pathway. For example, in NetEx, the MCSs 0, 2, 3, 4 do not contain any reaction involved in EM1. Hence, when removing one of these MCSs, EM1 would still be intact and metabolites B and C could be still produced.
- Some of the cellular functions might be difficult to turn off genetically or by inhibition, e.g. if many isozymes for a reaction exist. The set of MCSs contains all alternatives not involving these network edges.

Network verification and mutant phenotype predictions Cutting away an MCS from the network can be predicted to be definitely intolerable with respect to certain cellular reactions/processes. These predictions, derived purely from network structure, might be suitable for verification of hypothetical or reconstructed networks.

Applying this procedure to metabolic networks with 'growth' as the objective reaction is straightforward. If a set of gene deletions (single, double, triple, ... mutants) completely contains an MCS then (case 1) it should lead to a non-viable phenotype, otherwise (case 2) growth is structurally possible. A wrong prediction for case (1) would be a false negative prediction and is then a proof for an incorrect or incomplete network structure. A wrong prediction for case (2) would be a false positive prediction and is—due to the explanation given in Remark 1—a clue but not necessarily a proof of a false assumption in the network structure. Especially, 'marginally' structurally tolerable deletions will probably lead to false positive predictions where all reactions except one of a large MCS are inactivated. For instance, several MCSs with 12 elements occur for *E.coli* (substrate: glucose) and it seems questionable—but would be interesting to check—whether 11 deletions in such a set would be tolerated. Note that each of the 12 deletions in these MCSs contribute to the dysfunction and are therefore non-redundant. In contrast, in a non-essential linear pathway with 11 reactions deleting all 11 reactions could also be tolerated but only the deletion of a single reaction is a non-redundant deletion set.

Using the non-redundant MCSs for such experiments is efficient and important. For illustration, assume that in NetEx we have not yet recognized that there is another reaction Rx: $B + C + D = X$. Deleting all reactions (except obR and the unknown Rx) and looking whether obR is functioning would not be an intelligent choice. Besides the high experimental effort, we would also indirectly destroy the pathway from Rx to obR and could then not falsify our network structure. In contrast, applying MCS4 could reveal that there is another reaction in the network not recognized so far.

Phenotype predictions for single mutants of *E.coli* have already been conducted by means of Flux Balance Analysis (FBA; Edwards and Palsson, 2000) and elementary-mode analysis (Stelling *et al*., 2002). In these works, the viability of mutants could be predicted with high agreement with real mutants. The prediction of whether a single mutation is lethal or not is equivalent to whether the respective reaction is essential (MCS with only one element) and depends on the chosen substrate. Owing to mathematical equivalence, the predictions of FBA, elementary-mode analysis and MCSs are the same. However, the MCSs enable a systematic search for reasonable single, double, triple, . . . mutants. For example, a triple mutant where the subset of two of the three knocked out genes are already structurally intolerable has no corresponding MCS. Hence, the MCSs calculated for *E.coli* could be a basis for systematically falsifying the catabolic part of the network by mutants growing on different substrates. For instance, {Rl5P-epimerase, transhydrogenase, Succ-CoA-Synthase} is an MCS for all four substrates. Thus, if only a proper subset of this MCS would be removed, then *E.coli* would structurally be able to grow. {PEP carboxylase, isocitrate lyase} is an MCS for growth on glucose and glycerol, it is a cut set but not a minimal one for acetate (already a mutant missing the isocitrate lyase cannot grow on acetate) and it is not a cut set for succinate. Hence, these two simultaneous deletions could only be tolerated by *E.coli* for growth on succinate.

Structural fragility and robustness As described above, one important application of MCSs in risk assessment is to evaluate the reliability of a system and to find those combinations of events that have the highest probability to cause a system failure. Similarly, we can use the MCS for assessing *structural* fragility and robustness in metabolic networks, which are here inversely related (cf. Csete and Doyle, 2002). For simplicity, we assume that each reaction in a metabolic network has the same probability to fail. This directly implies that small MCSs are most probable to be responsible for a failing objective reaction. This is in line with the suggestion that the number of elementary modes occurring in a network is a measure of the (structural) robustness (Papin *et al*., 2002; Stelling *et al*., 2002). The more EMs—and thus available pathways—exist in which the objective reaction participates, the more network elements have to fail for a guaranteed dysfunction of the objective reaction. Thus, the larger the number of EMs, the larger will be the size of the resulting MCSs. This is confirmed by the MCSs in *E.coli*: the highest number of EMs and MCSs occurs for growth on glucose (Table 2) and one also finds the largest MCSs for this substrate (Fig. 3). The ranking of the other substrates with respect to number and size distribution of the MCSs corresponds to the ranking of the number of EMs. Accordingly, for growth on acetate one finds predominately small MCSs (with the highest number of essential reactions: 12 more than for glucose) and only few MCSs with more than five elements. These results clearly support what

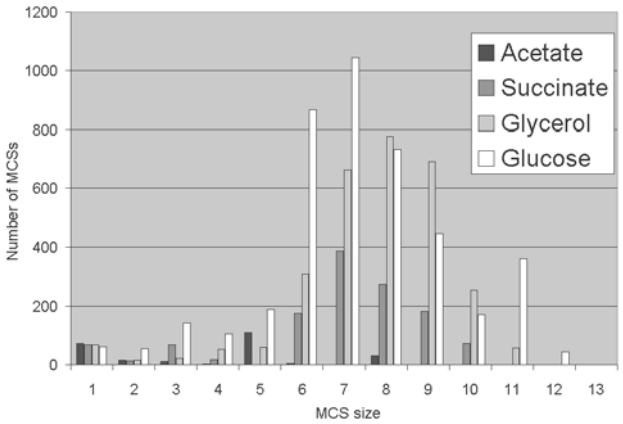

Fig. 3. Size distribution of the MCSs in the central metabolism of *E.coli* for four different substrates.

has been claimed by Stelling *et al*. (2002): growth on glucose is structurally less fragile (more robust) than growth on acetate. However, the size distribution of the MCSs gives a somewhat more subtle view on network fragility and allows a direct identification of the small and therefore most 'dangerous' MCSs with respect to the objective reaction.

Besides, the set of MCSs allows not only for assessment of the overall fragility of the network structure but also to investigate the importance of each single reaction in the network. If a reaction is predominately part of larger MCSs, then a malfunction of this reaction will be less crucial for the operation of the objective reaction. As a quantitative measure we define for each reaction a fragility coefficient F_i as the reciprocal of the average size of all MCSs in which reaction i participates. The minimal value for F_i is zero (defined for the case where reaction i is not member in any of the MCSs) and reaches the highest value of one for the most crucial edges in the network, namely for essential reactions. The F_i for all nine reactions in NetEx are given in Table 1 indicating that—beside the essential reaction R1—reaction R5 is most crucial for the objective process. The loss of its function makes reactions R2, R3 and R4 automatically meaningless for obR. As expected, an inverse correlation between F_i and N_i (the number of modes in which the objective reaction AND the reaction i participate) is apparent: the higher N_i the smaller F_i. However, the ranking is not always the same (e.g. $N_{R2} = N_{R6} = 1$ but $F_{R2} = 1/3 < F_{R6} = 3/8$).

The fragility coefficients for some selected reactions in the *E.coli* network (Table 2) strongly depends on the respective substrate. For example, as is intuitively clear, the fructose-bisphosphatase is not crucial for growth on glucose (F_i = 0.102) but becomes essential for the other three substrates. The relatively high F_i of malic enzyme for growth on acetate and succinate can be explained by the necessary carbon efflux from the TCA cycle fed by these substrates into the cycle.

Alternative fragility coefficients may also be defined. For example, the smallest MCS in which a certain reaction occurs might also be important (Table 2). In *E.coli*, the F_i for the malic enzyme under growth on glucose is low. However, there is one crucial combination (MCS with only two elements): if the malic enzyme AND the malate dehydrogenase fail together, then this will result in a non-viable phenotype.

One might also think about a rigorous probabilistic definition of the fragility coefficient. We would then need for each reaction a probability value for its failure. In case reaction i is deleted in the network, one might calculate the probability of a failure in the objective reaction. However, this would require a precise analysis of the underlying complex combinatorics.

Finally, for an overall quantification of the structural fragility, we propose a network fragility coefficient **F**, which is the average fragility coefficient over all q reactions: $\mathbf{F} = \sum_i^q F_i/q$. For NetEx we have $\mathbf{F} = 0.514$, the values for *E.coli* for the different substrates are given in Table 2.

Observability of reaction rates in metabolic flux analyses
In (Klamt *et al*., 2002a) it has been demonstrated how elementary-mode analysis supports the identification of those sets **M** of known (measured) reaction rates, which enable the calculation of a certain unknown reaction rate r_u in a steady-state flux distribution in a given network. The procedure is as follows: (1) calculate the elementary modes in the network and thereby consider all reactions as reversible. (2) Select for the next step (3) only those modes where r_u is unequal to zero. (3) Construct the set **M** of rates to be measured in such a way, that of all the reactions participating in an EM at least one is a member of **M**. Comparing this rule with our computation scheme shows that a suitable set **M** is a cut set with respect to r_u. Thus, if we determine the MCSs for the (completely reversible) network then we get all possible sets of measurements being minimal and non-redundant. Screening these MCSs enables one to find the best-suited sets of measurements and is therefore also useful for preparing metabolic flux analyses.

The reader might be a bit confused why the reaction reversibilities must not be taken into account. For illustration, assume the rate of obR in NetEx at steady state is to be calculated. We know that {R1} is an essential reaction in NetEx and, hence, an MCS. However, if we measured only the rate of R1 we would not be able to calculate the rate of obR uniquely—except in the very special case where the reaction rate of R1 is zero. In this case, we can directly conclude that the flux through obR must be zero—this is simply the condition to be a cut set. However, in the general case with a non-zero rate for R1, we need more information for determining the rate of obR. Indeed, first calculating the (9) EMs for the network with all reactions considered as reversible and then the (16) MCSs with respect to obR, one finds out that R1 alone is not an MCS anymore (results not shown). The smallest MCS—and therefore the minimal number of necessary measurements—is two. For instance, {R5, R8} would remain valid. Usually, only the rates crossing the boundaries can be measured (here R1, R3, R4, R6). A screening of the MCSs reveals that the best set of measurements would then be {R1, R3}. Even in this simple example it is not easy to see that measuring these rates is sufficient for computing the stationary rate of obR, which is the difference of R1 and R3. All other MCSs comprising only the measurable rates have a size of 3 (note: there are three degrees of freedom in the system).

CONCLUSION

The concept of MCSs is promising for studying and predicting the non-decomposable failure modes in biochemical reaction networks. An MCS is a irreducible combination of network elements whose simultaneous inactivation leads to a guaranteed dysfunction of certain cellular reactions or processes. MCSs are inherent and uniquely determined structural features of metabolic networks similar to EMs. Analyzing the EMs proved to be a useful tool for assessing a variety of structural and functional network properties. They are here also needed for computing the MCSs. Both MCSs as well as EMs possess the property of being minimal or irreducible. Whereas an EM is a minimal set of reactions which can perform a function, an MCS is a minimal set of reactions whose removal impedes a certain function. The computation of both features becomes challenging in large networks.

We illustrated a number of far-reaching applications. Analyzing the MCSs yields deeper insights in the structural fragility of a given metabolic network. MCSs are useful for identifying target sets for an intended repression of network functions. Such target sets are also relevant for a verification of a given network structure by systematic falsification. This is important especially for network topologies that are supposed to be incomplete. MCSs are also a suitable concept to find the necessary information one needs to make unknown stationary network fluxes observable.

Applying the concept of MCSs also to gene and signal transduction networks is very appealing. However, stoichiometric matrices do not seem to be an adequate approach for describing the flow and processing of information in these networks. Therefore, we need appropriate representations for such network topologies that allow for searching for functional or failure modes similar to EMs and MCSs, respectively. Boolean networks is one suitable approach for simulating genetic control circuits (Thomas, 1973) and could be also one for describing signal transductions (cf. Genoud and Metraux, 1999). In non-recursive boolean networks, MCSs can be identified as in the above mentioned fault trees emerging in risk assessment.

ACKNOWLEDGEMENT

We thank Jörg Stelling for critical reading of the manuscript and helpful discussions.

REFERENCES

Bailey,J.E. (2001) Complex biology with no parameters. *Nat. Biotechnol.*, **19**, 503–504.

Bollabas,B. (1998) *Modern Graph Theory*. Springer-Verlag, New York.

Clarke,B.L. (1988) Stoichiometric network analysis. *Cell Biophys.*, **12**, 237–253.

Covert,M. and Palsson,B.O. (2003) Constraints-based models: regulation of gene expression reduces the steady-state solution space. *J. Theor. Biol.*, **221**, 309–325.

Csete,M.E. and Doyle,J.C. (2002) Reverse engineering of biological complexity. *Science*, **295**, 1664–1669.

Edwards,J.S. and Palsson,B.O. (2000) The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl Acad. Sci. USA*, **97**, 5528–5533.

Fard,N.S. (1997) Determination of minimal cut sets of a complex fault tree. *Comput. Ind. Eng.*, **33**, 59–62.

Feinberg,M. (1995) The existence and uniqueness of steady states for a class of chemical reaction networks. *Arch. Rat. Mech. Anal.*, **132**, 311–370.

Fell,D.A. and Wagner,A. (2000) The small world of metabolism. *Nat. Biotechnol.*, **18**, 1121–1122.

Förster,J., Famili,I., Fu,P., Palsson,B.O. and Nielsen,J. (2003) Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.*, **13**, 244–253.

Genoud,T. and Metraux,J.-P. (1999) Crosstalk in plant cell signaling: structure and function of the genetic network. *Trends Plant Sci.*, **4**, 503–507.

Ibarra,R.U., Edwards,J.S. and Palsson,B.O. (2002) *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature*, **420**, 186–189.

Kitano,H. (2002) Systems biology: a brief overview. *Science*, **295**, 1662–1664.

Klamt,S., Schuster,S. and Gilles,E.D. (2002a) Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria. *Biotechnol. Bioeng.*, **77**, 734–751.

Klamt,S. and Stelling,J. (2002b) Combinatorial complexity of pathway analysis in metabolic networks. *Mol. Biol. Rep.*, **29**, 233–236.

Klamt,S. and Stelling,J. (2003a) Two approaches for metabolic pathway analysis? *Trends Biotechnol.*, **21**, 64–69.

Klamt,S., Stelling,J., Ginkel,M. and Gilles,E.D. (2003b) FluxAnalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps. *Bioinformatics*, **19**, 261–269.

Ma,H. and Zeng,A. (2003a) Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics*, **19**, 270–277.

Ma,H. and Zeng,A. (2003b) The connectivity structure, giant strong component and centrality in metabolic networks. *Bioinformatics*, **19**, 1423–1430.

Ouzonis,A.C. and Karp,P.D. (2000) Global properties of the metabolic map of *Escherichia coli*. *Genome Res.*, **10**, 568–576.

Papin,J.A., Price,N.D., Edwards,J.E. and Palsson,B.O. (2002) The genome scale metabolic extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy. *J. Theor. Biol.*, **215**, 67–82.

Price,N.D., Papin,J.A., Schilling,C.H. and Palsson,B.O. (2003) Genome-scale microbial *in silico* models: the constraints-based approach. *Trends Biotechnol.*, **21**, 162–169.

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

Schilling,C.H., Letscher,D. and Palsson,B.O. (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.*, **203**, 229–248.

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

Schuster,S., Hilgetag,J.C., Woods,J.H. and Fell,D.A. (2002) Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism. *J. Math. Biol.*, **45**, 153–181

Sinnamon,R.M. and Andrews,J.D. (1997) New approaches to evaluating fault trees. *Reliab. Engng. System Safety*, **58**, 89–96.

Stelling,J., Klamt,S., Bettenbrock,K., Schuster,S. and Gilles,E.D. (2002) Metabolic network structure determines key aspects of functionality and regulation. *Nature*, **420**, 190–193.

Stephanopoulos,G.N., Aristidou,A.A. and Nielsen,J. (1998) *Metabolic Engineering*. Academic Press, San Diego.

Thomas,R. (1973) Boolean formalization of genetic control circuits. *J. Theor. Biol.*, **42**, 563–585.

Wiechert,W. (2002) Modeling and simulation: tools for metabolic engineering. *J. Biotechnol.*, **94**, 37–64.