Available online at www.sciencedirect.com
ScienceDirect
Journal of Theoretical Biology 252 (2008) 433–441
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Structural robustness of metabolic networks with respect to multiple knockouts

Jörn Behre[a,*], Thomas Wilhelm[b,1], Axel von Kamp[a], Eytan Ruppin[c,d], Stefan Schuster[a]

[a]*Faculty of Biology and Pharmaceutics, Section of Bioinformatics, Friedrich Schiller University Jena, Ernst-Abbe-Platz 2, D-07743 Jena, Germany*
[b]*Leibniz Institute for Age Research—Fritz Lipmann Institute, Theoretical Systems Biology Group, Beutenbergstrasse 11, D-07745 Jena, Germany*
[c]*School of Computer Sciences, Tel Aviv University, Tel Aviv 69978, Israel*
[d]*School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel*

Received 31 May 2007; received in revised form 21 September 2007; accepted 26 September 2007
Available online 9 October 2007

Abstract

We present a generalised framework for analysing structural robustness of metabolic networks, based on the concept of elementary flux modes (EFMs). Extending our earlier study on single knockouts [Wilhelm, T., Behre, J., Schuster, S., 2004. Analysis of structural robustness of metabolic networks. IEE Proc. Syst. Biol. 1(1), 114–120], we are now considering the general case of double and multiple knockouts. The robustness measures are based on the ratio of the number of remaining EFMs after knockout vs. the number of EFMs in the unperturbed situation, averaged over all combinations of knockouts. With the help of simple examples we demonstrate that consideration of multiple knockouts yields additional information going beyond single-knockout results. It is proven that the robustness score decreases as the knockout depth increases.

We apply our extended framework to metabolic networks representing amino acid anabolism in *Escherichia coli* and human hepatocytes, and the central metabolism in human erythrocytes. Moreover, in the *E. coli* model the two subnetworks synthesising amino acids that are essential and those that are non-essential for humans are studied separately. The results are discussed from an evolutionary viewpoint. We find that *E. coli* has the most robust metabolism of all the cell types studied here. Considering only the subnetwork of the synthesis of non-essential amino acids, *E. coli* and the human hepatocyte show about the same robustness.

Keywords: *Escherichia coli* metabolism; Elementary flux modes; Erythrocyte metabolism; Hepatocyte metabolism; Robustness measure

1. Introduction

A general feature of living cells is their robustness to varying environmental conditions. Moreover, internal perturbations (e.g. knockout mutations or enzyme deficiencies) can be tolerated to a certain extent. Experimental and theoretical analyses of robustness have attracted increasing interest in recent years. For example, the virtual independence of biological oscillations on temperature (Ruoff et al., 2003) has been analysed. Besides such cases of dynamic robustness (Schuster and Holzhütter, 1995; Stelling et al., 2004; Wolf et al., 2005), structural robustness also has been intensively studied (Stelling et al., 2002; Çakır et al., 2004; Wilhelm et al., 2004; Lemke et al., 2004; Fong and Palsson, 2004; Klamt and Gilles, 2004; Papp et al., 2004; Blank et al., 2005; Ghim et al., 2005; Shlomi et al., 2005; Kaufman et al., 2005; Klamt, 2006; Deutscher et al., 2006). Structural robustness refers to the tolerance against changes in the structure of cellular networks. Knockout mutations and (complete) enzyme deficiencies obviously affect network structure. The analysis of structural robustness is part of the general trend of network-based approaches in which kinetic parameters are not included, motivated by the fact that kinetic parameters are often not perfectly known.

*Corresponding author. Tel.: +49 3641 949584.
E-mail addresses: jbehre@minet.uni-jena.de (J. Behre), thomas.wilhelm@bbsrc.ac.uk (T. Wilhelm), kamp@minet.uni-jena.de (A. von Kamp), ruppin@post.tau.ac.il (E. Ruppin), schuster@minet.uni-jena.de (S. Schuster).
[1]Present address. Institute of Food Research, Norwich Research Park, Colney, Norwich NR47UA, UK.

0022-5193/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2007.09.043

In a previous study we analysed the structural robustness of metabolic systems with respect to single knockouts (Wilhelm et al., 2004). We introduced various robustness measures, all of them averaging the effect of single knockouts over all enzymes in the system or all enzymes leading to a specific product. As a proof of concept, we applied these measures to metabolic networks of human erythrocytes and *Escherichia coli* central metabolism. In agreement with biochemical experience, we obtained lower robustness values for the erythrocyte network.

However, often organisms are affected by double and multiple knockouts or enzyme deficiencies. Multiple knockouts are of importance in biotechnology and medicine, for example, to suppress pathogenic bacteria. For instance, for a pair of synthetic lethal genes the knockout of two genes is fatal for the organism while a single knockout of either gene is not (Lemke et al., 2004; Schuldiner et al., 2005; Harrison et al., 2007). The most simple example is a pair of isoenzymes (Pál et al., 2005; Kuepfer et al., 2005), catalyzing the same essential reaction. A systematic experimental screening of double knockout mutants is complicated due to the large number of combinations. The most advanced screening of double mutants has been done on *Saccharomyces cerevisiae*: Tong et al. (2004) have analysed 4700 viable gene yeast deletion mutants, and Wong et al. (2004) have presented a comprehensive method to predict synthetic lethal gene pairs. First attempts have also been made for cultured human cells (Simons et al., 2001).

In the present paper, we generalise our theoretical network robustness studies by taking into account double and multiple knockouts and propose appropriate generalised robustness measures. As in our previous paper (Wilhelm et al., 2004), the analysis is based on the concept of elementary flux modes (EFMs). These are minimal sets of enzymes that can operate at steady state with all irreversible reactions used in the appropriate orientation (Schuster et al., 2000). In recent years, various biochemical systems (Stelling et al., 2002; Çakır et al., 2004; Schwender et al., 2004; Schuster and Kenanov, 2005; Krömer et al., 2006) have been studied using the concept of EFMs.

We apply the generalised robustness measures to reaction schemes of the central metabolism in human erythrocytes and amino acid synthesis in *E. coli* and human hepatocytes. The latter two systems are known to be quite redundant. Thus, it is interesting to compare their robustness. Moreover, in the *E. coli* model the two subnetworks synthesising (i) amino acids that are essential for humans and (ii) amino acids that are non-essential for humans are studied individually. The results will be discussed from an evolutionary viewpoint.

This paper is dedicated to the memory of Reinhart Heinrich, who was the highly respected academic teacher of two of the authors (S.S. and T.W.). He taught us the theoretical apparatus needed to better understand the behaviour of intracellular networks. Reinhart became famous (together with others) for establishing Metabolic Control Analysis. In that theoretical framework, less input information is necessary than in dynamic simulation. (Note that the elasticity coefficients harbour less information than the full set of kinetic parameters, cf. Heinrich and Schuster, 1996.) Later, he and his coworkers, as well as other people in the field, became interested in structural approaches, in which only a minimum input information is used, motivated by the unfortunately imperfect knowledge of kinetic parameters. Reinhart was very open to elementary-modes analysis, as witnessed by the monograph Heinrich and Schuster (1996). Besides his interest in dynamic robustness (Ruoff et al., 2003; Wolf et al., 2005), he also worked on the structural robustness of metabolism. In Ebenhöh and Heinrich (2003), a metabolic network is defined to be strongly robust against a knockout, exchange or even addition of an enzyme if it can still produce the same products. It is weakly robust if it can still produce at least one product. In the beginning of this millenium, he established the "scope" approach to elucidating the evolution of metabolic networks (Handorf et al., 2005; Ebenhöh et al., 2006). In that approach, information about the network structure and chemical formulas of substances as taken from online databases is used. Robustness issues play a role in that approach as well. Handorf et al. (2005) showed that the outcome of network expansion is in general rather robust against elimination of single or few reactions. There exist, however, crucial reactions the elimination of which leads to a dramatic reduction in the size of the network reachable in evolution.

2. Generalised measures of network robustness

Consider a metabolic network made up of a set of r enzymes $M = \{E_1, E_2, \ldots, E_r\}$. To quantify the structural robustness to the knockout (deficiency) of a subset of M, $K_i = \{E_{i,1}, E_{i,2}, \ldots\}$, the ratio, $z^{(i)}/z$, between the number of EFMs remaining after knockout, $z^{(i)}$, and the number in the unperturbed network, z, is used. The global robustness of the entire network is described by the arithmetic mean of all these numbers for all subsets K_i with the same cardinality, d:

$$R_1(d) = \sum_{i=1}^{c(d)} z^{(i)} \Big/ c(d) z \tag{1}$$

where

$$c(d) = \binom{r}{d}, \tag{2}$$

denotes the total number of these subsets (i.e., the total number of combinations of d knockouts from r reactions). $R_1(d)$ is the robustness with respect to knockout of exactly d enzymes. It is the average number of remaining EFMs divided by the total number of EFMs in the original system. Thus, Eq. (1) can also be written as $R_1(d) = \langle z^{(i)}(d) \rangle / z$. Since each term $z^{(i)}$ in the numerator

of Eq. (1) is as most as large as z, it follows that the quantity $R_1(d)$ is between 0 and 1.

In the special case $d = 1$, the measure $R_1(1)$ coincides with the robustness measure for single knockouts defined earlier (Wilhelm et al., 2004). Moreover, $R_1(r) = 0$ because knocking out all enzymes in a system (of enzyme-catalysed reactions) obviously deletes all EFMs. Generally, the following inequality holds

$$1 = R_1(0) \geqslant R_1(1) \geqslant R_1(2) \geqslant R_1(3) \geqslant \ldots R_1(r-1) \geqslant R_1(r) = 0. \tag{3}$$

This can be rationalised as follows. As mentioned above, the robustness measure is the average number of EFMs remaining after knockout of d enzymes, divided by the total number of EFMs in the original system. For $d = k+1$, we have

$$R_1(k+1) = \frac{\langle z^{(j)}(k+1) \rangle}{z}. \tag{4}$$

Since the number of EFMs cannot increase as the number of out-knocked enzymes increases, we have

$$z^{(i)}(k) \geqslant z^{(j)}(k+1) \text{ for all } i,j \text{ with } K_i(k) \subset K_j(k+1). \tag{5}$$

This implies that also the average number cannot increase, which leads to relation (3). A more detailed proof is given in Appendix A.1.

The measures $R_1(d)$ are defined separately for different knockout depths. It is tempting to define an overall measure by combining them in an appropriate way. We define this in the following way:

$$R_1(\leqslant D) = \sum_{d=1}^{D} R_1(d) p_d \text{ with } D \leqslant r, \tag{6}$$

where the p_d are weighting factors for a knockout of d enzymes together. We focus on the situation where at least one and at most D enzymes have been knocked out. In this case p_d is related to the conditional probability that the knockout which occurred is of depth d, and hence has to fulfill the normalisation condition

$$\sum_{d=1}^{D} p_d = 1. \tag{7}$$

Except for the more thorough analysis in Fig. 1, in this paper we choose $D = 3$ or 5, for two reasons: first, it is computationally too demanding to consider all knockout-combinations for a medium number of deleted enzymes ($d \sim r/2$, cf. Eq. (2)), and secondly, the higher robustness measures $R_1(d > 5)$ usually become very small and therefore do not make an important contribution to the overall measure (see Fig. 1).

The choice of appropriate weighting factors p_d is motivated by two facts: First, multiple knockouts or enzyme deficiencies occur less and less frequently as the knockout depth increases. Second, the number of possible knockout combinations changes according to $\binom{r}{d}$. Taken together this gives the weighting factors

$$p_d = p^d \binom{r}{d}. \tag{8}$$

That means, the factor p should then be calculated by solving the equation:

$$\sum_{d=1}^{D} p^d \binom{r}{d} = 1. \tag{9}$$

To get an idea of the values of the weighting factors we present, in Table 1, some exemplifying p_d values. Interestingly, the p_d converge quickly to finite values in the limit of very large reaction networks.

Moreover, we now generalise our previously defined robustness measures for specific products (Wilhelm et al., 2004) to the case of multiple knockouts. To this end, we consider the subnetwork consisting of all elementary modes producing a certain essential product P_k and apply definition (1), giving $R_1^{(k)}$. If all products of the network are essential, that is, if the mutant is not viable as soon as one product cannot be produced anymore, we define

$$R_2(d) = \min\left\{ R_1^{(1)}(d), R_1^{(2)}(d), \ldots, R_1^{(n)}(d) \right\}, \tag{10}$$

where the superscript refers to the index of the essential product. More generally, no single product might be

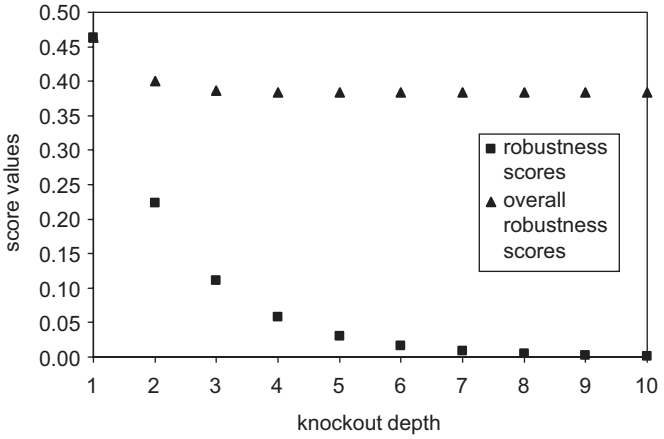


Fig. 1. Plot of the robustness score $R_1(d)$ (squares) and overall robustness score $R_1(\leqslant D)$ (triangles) vs. knockout depth (d resp. D) for the erythrocyte model. The plotted values are listed in the Supplementary Material.

Table 1
Weighting factors for $D = 3$ and the r values used in Table 2 and for $r \to \infty$

r	p_1	p_2	p_3
4	0.757	0.215	0.027
5	0.745	0.222	0.033
8	0.727	0.231	0.042
$r \to \infty$	0.699	0.244	0.057

essential, while a combination of different products is. In such a case each superscript indicates a unique smallest set of essential products.

In contrast, if no particular product (or set of products) is absolutely essential, the average of the individual product robustnesses is another appropriate robustness measure:

$$R_3(d) = \frac{\sum_{i=1}^{n} R_1^{(i)}(d)}{n}. \tag{11}$$

We have here defined the measures $R_2(d)$ and $R_3(d)$ for the sake of completeness. However, we will not further use them because we concentrate on a general comparison of networks without focussing on single products.

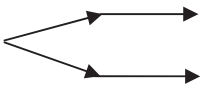
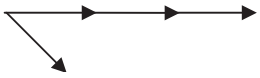
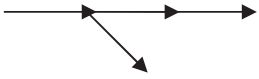
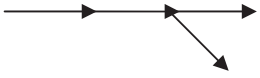
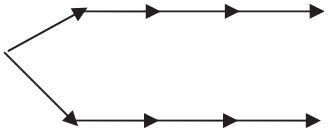
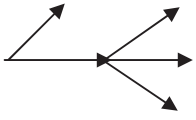
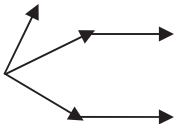
3. Simple examples

To point out differences in robustness to single and multiple knockouts, we consider the simple examples given in Table 2. A comparison of systems 1 and 2 shows that the measures corresponding to multiple knockouts carry additional information: although both systems have the same number of substrates, products, internal metabolites, reactions, EFMs, and the same $R_1(1)$, system 1 is less robust with respect to double and triple knockouts

$$\text{Network 1}: \quad R_1(2) = \frac{1+0+0+0+0+1}{6\times 2} = \frac{1}{6},$$
$$\text{Network 2}: \quad R_1(2) = \frac{1+1+0+1+0+0}{6\times 2} = \frac{1}{4}.$$

In both cases, the numbers in the numerators correspond to the following double knockouts: {1,2}, {1,3}, {1,4}, {2,3}, {2,4}, {3,4}. This means that after knockout of two enzymes, on average, one-sixth and one-fourth of the pathways in systems 1 and 2 are still present. The higher robustness of network 2 against double knockouts can intuitively be understood: the probability that double knockouts affect the same branch and, hence, leave the other branch unperturbed, is higher in network 2 than in network 1. System 2 is more robust than system 1 also with

Table 2
Simple examples of reaction schemes demonstrating different features of the robustness measures[a]

Example system	Number of reactions	Number of elementary modes	$R_1(1)$	$R_1(2)$	$R_1(3)$	$R_1(\leqslant 3)$
1 	4	2	$1/2 = 0.5$	$1/6 \approx 0.167$	0	0.414
2 	4	2	$1/2 = 0.5$	$1/4 = 0.25$	$1/8 = 0.125$	0.436
3 	4	2	$3/8 = 0.375$	$1/12 \approx 0.083$	0	0.302
4 	4	2	$1/4 = 0.25$	0	0	0.189
5 	8	2	$7/16 \approx 0.438$	$3/8 = 0.375$	$5/16 \approx 0.313$	0.418
6 	8	2	$1/2 = 0.5$	$3/14 \approx 0.214$	$1/14 \approx 0.071$	0.416
7 	5	4	$13/20 = 0.65$	$3/8 = 0.375$	$7/40 = 0.175$	0.573
8 	5	3	$2/3 \approx 0.667$	$2/5 = 0.4$	$1/5 = 0.2$	0.592

[a]For the definition of robustness measures, see text. The shown reaction networks involve monomolecular reactions only, and the metabolites at the upper and lower ends of reaction chains are defined external.

respect to triple knockouts: $R_1(3)$ of system 1 is zero, because no EFM remains, regardless of the chosen three out-knocked enzymes, whereas system 2 still has a triple-knockout robustness of $R_1(3) = 1/8$. For comparison, the networks 3 and 4 are even less robust than network 1. It can be seen that for branched systems with parallel pathways the location of the branching point is important. In system 2, which diverges at the upper end, the robustness is positive even for triple knockouts, while a branching point near the lower end can lead to a zero robustness for double and triple knockouts (system 4).

Interestingly, system 6 has a higher robustness with respect to single knockouts, but a lower double- and triple-knockout robustness than system 5. As can be seen by comparing examples 3 and 5 in Table 2, a lumping of enzymes does affect the robustness values. A lumping of all consecutive enzymes in the upper branch (which is an enzyme subset in the terminology of Pfeiffer et al., 1999) leads to the same system in both cases, notably a system consisting of three (super)enzymes and one internal metabolite acting as a branch point. Thus, the robustness values are equalised for the two systems, whereas the values for the original systems are different.

Systems 7 and 8 (both with five reactions) demonstrate that the pure number of EFMs is not an appropriate robustness measure, because system 7 has more EFMs (four), but has nevertheless a lower robustness than system 8 (three EFMs).

4. Robustness of central metabolisms in *E. coli* and human

First we analyse the central metabolism of human erythrocytes, using the network model of Schuster et al. (1998) comprising $n = 36$ internal metabolites, $r = 41$ reactions, and giving rise to 21 EFMs. The calculation procedure is explained in Appendix A.2. The EFMs are computed by the program METATOOL 5.0 (von Kamp and Schuster, 2006). The calculated robustness measures are given in Table 3. As expected, relation (3) is fulfilled for this system. Note that $R_1(\leqslant 3)$ is slightly higher than $R_1(\leqslant 5)$. This is understandable because the system is less robust when the knockout depth is larger. Mathematically, it is due to a change in the weighting factor p when D changes in the normalisation condition (9).

In Fig. 1 the robustness scores are plotted up to $d = 10$. It can be seen that knockout depths $d \geqslant 5$ are practically negligible. The overall robustness hardly changes even for $D \geqslant 3$, due to the monotonically decreasing weighting factors.

Second, we have compiled the amino acid synthesis network of *E. coli*. The reaction equations and the information about their reversibilities were taken from the databases KEGG (http://www.genome.jp/kegg/) and EcoCyc (http://www.ecocyc.org/). In total, the network contains 164 reactions (involving one spontaneous reaction) and 119 internal metabolites. The list of reaction equations including the names of external metabolites is given in the Supplementary Material. We then subdivided that network into two subnetworks: the system comprising the synthesis of amino acids which are essential for humans ($n = 89$, $r = 111$) and the system producing non-essential amino acids ($n = 53$, $r = 81$ comprising one spontaneous reaction). Note that there is a considerable overlap between the two subsystems, notably glycolysis, TCA cycle, etc. Although tyrosine is indicated as non-essential in most textbooks because it can be synthesised from phenylalanine in one step, we have here classified it as essential (in agreement with Voet and Voet (2004)) because phenylalanine is essential. The set of EFMs obtained involves several cycles of transaminase reactions. We have eliminated these because they are thermodynamically infeasible. The whole system gives rise to 65,836 EFMs (excluding such cycles), while for the subnetworks corresponding to essential and non-essential amino acids, 6874 and 11,435 EFMs, respectively, are calculated. Interestingly, the number is higher for the non-essential amino acids although the reaction number is lower (81 versus 111 for the essential amino acids). This is due to the higher degree of interconnectivity.

Table 3
Robustness measures for single and multiple knockouts up to $d = 5$ for the erythrocyte model of Schuster et al. (1998)

d (number of out-knocked enzymes)	$R_1(d)$
1	0.463
2	0.223
3	0.112
4	0.058
5	0.031
$\leqslant 3$	0.386
$\leqslant 5$	0.383

In the knockout studies, we took care that in the case of multifunctional enzymes, all reactions catalysed by a given enzyme are knocked out simultaneously. Spontaneous reactions cannot be knocked out because they are not catalysed by enzymes which could be inhibited. Therefore we skipped them during the calculations of the robustness measures. Table 4 shows the robustness values for the *E. coli* networks. For the subsystems we calculated the robustnesses up to $d = 5$. However, for the entire system, we limited the calculations to $d = 3$ due to memory restrictions. Since biologically, multiple knockouts with $d > 3$ are very rare, we do not lose important information in this way (see also Fig. 1). It can also be seen in Tables 2–5 that the overall robustness measures for the cases $d \leqslant 3$ and $d \leqslant 5$ are similar. As expected, the robustness of the *E. coli* metabolism is higher than that of erythrocyte metabolism (Tables 3 and 4). Moreover, the subnetwork corresponding to non-essential amino acids has a higher robustness at each knockout level than the network producing the essential amino acids. Furthermore, the entire network has robustness values that are above the maximum values

Table 4
Robustness measures for single and multiple knockouts up to $d = 5$ for the amino acid synthesis network of *E. coli*

d (number of out-knocked enzymes)	$R_1(d)$
Entire system	
1	0.776
2	0.602
3	0.468
$\leqslant 3$	0.716
Non-essential amino acids	
1	0.654
2	0.430
3	0.285
4	0.191
5	0.129
$\leqslant 3$	0.579
$\leqslant 5$	0.576
Essential amino acids	
1	0.623
2	0.397
3	0.258
4	0.172
5	0.117
$\leqslant 3$	0.548
$\leqslant 5$	0.544

Table 5
Robustness measures for single and multiple knockouts up to $d = 5$ for the amino acid synthesis network of human hepatocyte

Number, d, of out-knocked enzymes	$R_1(d)$
Non-essential amino acids	
1	0.659
2	0.443
3	0.305
4	0.214
5	0.154
$\leqslant 3$	0.587
$\leqslant 5$	0.584

of the two subnetworks. This is because additional routes starting and ending in one subnetwork and passing the other subnetwork drop out upon decomposition of the network. If tyrosine is taken as non-essential, similar robustness values are obtained (e.g. 0.701 and 0.616 with $d = 1$ for non-essential and essential amino acids, respectively).

Third we analysed the amino acid synthesis network of human hepatocytes (producing only non-essential amino acids, of course). The reaction equations were taken from KEGG (http://www.genome.jp/kegg/) and (mainly for the information about their reversibilities) from HumanCyc (http://www.humancyc.org/). In total, the network contains 82 reactions (involving six spontaneous reactions) and 59 internal metabolites. We considered compartmentation by distinguishing between reactions proceeding in the cytosol and those in mitochondria. We included the exchange reactions between these two compartments. The list of reaction equations including the names of external metabolites is given in the Supplementary Material. Spontaneous (non-enzymatic) reactions, multifunctional enzymes and futile cycles (again cycles of transaminase reactions) are handled as in the *E. coli* networks. For this system we calculated 712 EFMs.

Table 5 shows the robustness values for the hepatocyte network. Like the *E. coli* subnetworks, this system is, on average, still working even after the knockout of five enzymes. As expected, the hepatocyte metabolism is more robust than the erythrocyte metabolism (Tables 3 and 5), because of the minimalist and almost non-redundant metabolism of erythrocytes. Interestingly, the subnetwork corresponding to non-essential amino acids in *E. coli* and the amino acid anabolism in human hepatocytes have nearly the same robustness.

5. Discussion

Here, we have presented generalised robustness measures for metabolic networks, taking into account single, double and multiple knockouts. These measures are based on the ratio of the number of EFMs in the unperturbed situation vs. the number of remaining EFMs after knockout of one enzyme (cf. Wilhelm et al., 2004) or several enzymes, averaged over all combinations of knockouts. Since this is a normalised quantity, it does not depend on the size of the network. Moreover it is only based on network topology, so that the robustness values are the same for different metabolic systems having the same topology. With the help of simple examples (Table 2), we have demonstrated that consideration of double and triple knockouts yields additional information beyond the single-knockout studies.

We have proven that the robustness decreases if the cardinality of knockouts, d, increases. The overall robustness is, for all systems considered, between the robustness against single knockouts and that against double knockouts (Tables 3–5). It would be interesting to prove this in a general way. This observation is likely to be related with the facts that both the robustness measures (cf. Eq. (3)) and the weighting factors in the overall robustness represent monotonically decreasing series.

The examples show that for given numbers of enzymes and EFMs the robustness against double or multiple knockouts is higher if two branches in the network have different lengths (see systems 1 and 2 in Table 2). Interestingly, this is often the case in metabolism. For example, the different amino acids are synthesised on pathways of very different lengths. Although this probably has mainly chemical reasons, it might be that robustness issues also played a role in metabolic network evolution. We were not able to devise two example systems for which $R_1(3)$ is different, although $R_1(1)$ and $R_1(2)$ are equal for

the two systems. It is an interesting question whether or not two such reaction schemes can be found.

To illustrate the applicability of the new concepts, we have analysed networks of the central metabolisms in *E. coli* and in human erythrocytes and hepatocytes. Among these, the erythrocyte model shows the lowest robustness. In contrast to our previous work (Wilhelm et al., 2004) where enzymes were lumped to some extent, here we have considered each enzyme separately. As can be seen by comparing examples 3 and 5 in Table 2, a lumping of enzymes does affect the robustness values. This holds true even if the combined enzymes operate in fixed flux proportions due to structural constraints (enzyme subsets according to the definition in Pfeiffer et al., 1999).

As expected, hepatocyte metabolism is more robust than erythrocyte metabolism. This is because erythrocytes must be as small as possible in order to pass thin capillaries and are densely packed with haemoglobin for oxygen transport. Therefore only the most necessary enzymes have been retained in evolution. Both erythrocytes and hepatocytes are living under relatively homoeostatic conditions with hepatocytes having much more metabolic capabilities. In contrast, *E. coli* must adapt to widely varying situations. Thus it needs to be even more robust than hepatocytes. As a consequence *E. coli* synthesises all amino acids while hepatocytes (as all human cells) can save the metabolic effort for producing those amino acids being essential for human. A comparison of the hepatocyte network with the corresponding subnetwork of *E. coli* (just the non-essential amino acids) shows, interestingly, slightly higher robustnesses for the hepatocyte. However, the entire amino acid network of *E. coli* is significantly more robust. One reason is that the compartmentation in hepatocytes implies transporters forming bottlenecks in the system. Since not every metabolite can cross intracellular membranes, it can be hypothesised that compartmentation reduces structural robustness in many cell types.

Our analysis of the *E. coli* networks shows that amino acids essential for humans are less robustly produced than the non-essential amino acids. It is tempting to speculate that this might be the reason why their synthesis pathways got lost in the evolution towards higher organisms such as humans. The structural background for this difference is that the synthesis pathways of essential amino acids (such as tryptophan or isoleucine) are relatively "straight", that is, they do not involve many branch points. The enzyme genes corresponding to some of these pathways are gathered in operons (e.g. Trp operon), so that a mutational loss of the whole pathway occurs easily. In contrast, the synthesis of the non-essential amino acids runs on pathways with a higher degree of ramification and is embedded in the entangled synthesis network of other compounds, giving rise to much more redundancy.

Our approach is important for future applications in pharmacology and biotechnology. Combination drugs of two or more enzyme inhibitors have recently attracted increasing interest, while the progress in detecting drugs acting on single proteins has slowed down (Huang, 2001; Frantz, 2005). Combination therapies are of interest in treating bacterial infections (Barchiesi et al., 2004), AIDS (Taburet et al., 2004) and others. Similarly, in biotechnology, when inefficient pathways are to be suppressed, often undesired side reactions need to be deleted as well, so that multiple knockouts are necessary.

In metabolic modelling, isoenzymes are often lumped into combined reactions. While this is appropriate for many applications, it is not when robustness to enzyme deletions is studied because particular enzymes rather than particular reactions are knocked out. Thus, for example, succinate dehydrogenase and fumarate reductase in *E. coli* need to be distinguished. Analogously, enzymes with broad substrate specificity require special attention. The knockout of an enzyme catalyzing several reactions implies the deletion of all these reactions (unless they are catalysed by other enzymes simultaneously).

Some of the previous studies on robustness have tackled the question what percentage of enzymes is essential. Estimates range from about 20% (cf. Papp et al., 2004) to 30% (cf. Blank et al., 2005). The question is difficult to answer, though, because variations in external conditions are hard to take into account. Anyway, the percentage of essential enzymes is not the only relevant robustness measure. The knockout of a non-essential enzyme can have widely different effects, depending on which other enzymes are knocked out simultaneously. We have here made an attempt to quantify these effects by the average number of the remaining EFMs. A more detailed approach would consider not only the average but the diversity of effects: Is the knockout of some pair of enzymes lethal to the system and the knockout of another pair completely irrelevant, or is the effect always moderate?

The relationship of our approach to the concept of minimal cut sets (Klamt and Gilles, 2004; Klamt, 2006) is worth discussing. Minimal cut sets are minimal sets of enzymes whose suppression prevents a target reaction under study from operating. When such a set includes one reaction only, this reaction is obviously essential. When the smallest minimal cut set involves, for example, two enzymes, the target reaction can still proceed in any single-knockout mutant while it cannot in the double knockout case corresponding to that cut set. Klamt and Gilles (2004) introduced a fragility coefficient for each enzyme as the reciprocal of the mean size of all minimal cut sets in which this enzyme is included. A network fragility coefficient, F, was defined by averaging over all enzymes, taking into account both single and multiple knockouts. The fragility coefficient is based on whether or not a desired substance can still be produced, while our measures take into account the number of feasible synthesis routes. The mathematical relationship between the two concepts is not straightforward and an interesting subject of future studies. Another interesting challenge is to compare the robustness of the subnetworks of amino acids recruited early in evolution and of amino acids accrued

later (Jordan et al., 2005). It can be presumed that early adopted amino acids are more cross-linked in the network and thus lead to higher robustness.

Moreover it should be possible to extend our concept of structural robustness to gene regulatory and signal transduction networks. In that context it would be interesting to calculate the robustness of the network motifs studied by Alon (2007). Another interesting question is the extensibility of metabolic networks by additional reactions ("knock-ins"). However for this it would be necessary to define a set of plausible additional reactions.

Acknowledgements

Financial support by the BMBF (German Ministry for Education and Research) to J. Behre (HepatoSys Program) and T. Wilhelm (Jena Centre for Bioinformatics) and from the German-Israeli Foundation to A. von Kamp and E. Ruppin is gratefully acknowledged. We are grateful to Sabrina Hoffmann (Charité, Berlin) for providing information about the hepatocyte reaction network.

Appendix A.1. Proof of relation (3)

Let us compare the situations where $d=k$ with that where $d=k+1$. The sum in Eq. (1) may involve a different number of terms in these two situations. To relate these sums, we write each term of the sum in Eq. (1) for the case $d=k$ as a sum of $r-k$ identical terms divided by $r-k$:

$$\sum_{i=1}^{c(k)} z^{(i)} = \frac{\sum_{i=1}^{c(k)}\sum_{j=1}^{r-k} z^{(i)}}{r-k}. \tag{A.1}$$

The decomposition of each term of the sum in Eq. (1) corresponds to the transition from a subset K_i for $d=k$ to the $r-k$ situations of $k+1$ knockouts in which k knockouts are the same as in K_i. For example, when $d=1$, the term describing the knockout of enzyme 1 corresponds to the terms describing the knockout of enzymes {1,2}, {1,3}, etc. in the case $d=2$. In the set of knockout combinations thus generated, each K_i (for $d=k+1$) occurs $k+1$ times because it can come from $k+1$ original situations. For example, the term describing the knockout of enzymes {1,2} corresponds to the terms describing the knockout of enzyme 1 and to the knockout of enzyme 2 in the case $d=1$. Therefore, we can combine the terms of the sum in Eq. (A.1) such that

$$R_1(k) \geqslant \frac{(k+1)\sum_{j=1}^{c(k+1)} \tilde{z}^{(j)}(k)}{(r-k)c(k)z} \quad \text{with} \quad K_i(k) \subset K_j(k+1) \tag{A.2}$$

$\tilde{z}^{(j)}(k)$ denotes the minimum of the remaining EFMs for each of the corresponding knockout combinations of k enzymes. For instance, $\tilde{z}^{(1,2,3)}(2) = \min\{z^{(1,2)}(2), z^{(1,3)}(2), z^{(2,3)}(2)\}$. Making use of the relation

$$\binom{r}{k+1} = \binom{r}{k}\frac{r-k}{k+1}, \tag{A.3}$$

we obtain

$$R_1(k) \geqslant \frac{\sum_{j=1}^{c(k+1)} \tilde{z}^{(j)}(k)}{c(k+1)z}. \tag{A.4}$$

This sum has the same number of terms as the sum describing $R_1(k+1)$. Therefore, it can be compared, term by term, with that sum. Since the number of EFMs cannot increase as the number of out-knocked enzymes increases, we have

$$\tilde{z}^{(i)}(k) \geqslant z^{(j)}(k+1) \text{ for all } i,j \text{ with } K_i(k) \subset K_j(k+1). \tag{A.5}$$

for each term. This leads to relation (3).

Appendix A.2. Calculation of reaction robustnesses

In principle the calculation of the robustness measures is straightforward. However, for d-knockouts in a network with r reactions, $\binom{r}{d}$ combinations (cf. Eq. (2)) have to be calculated. The number of combinations therefore increases drastically with the knockout depth d. In order to keep the calculation practical, the knockout combinations are not calculated on the level of reactions but on the level of enzyme subsets (Pfeiffer et al., 1999) and from these results the robustness is determined. The advantage is that the number of enzyme subsets is often significantly smaller than the number of reactions in the network. In order to calculate $R_1(d)$ the first step is to determine the number of remaining modes for all subset-knockout combinations of all depths from 1 to d. The next step consists of summing the number of remaining modes for depths $k=1,\ldots,d$ whereby each subset-knockout combination at depth k is weighted with the number of different ways it can be knocked out by knocking out exactly d reactions from that subset combination (e.g. a subset combination with two times two reactions can be knocked out in four ways when $d=2$). Note that these weights depend on d. This means that if the subset-knockout combinations of all depths from 1 to d have been calculated, the robustnesses $R_1(1),\ldots,R_1(d)$ can be computed from them by summing over them with different weighting factors. Finally, the sums calculated in this way have to be properly normalised in order to obtain the robustness values.

Appendix B. Supplementary materials

The online version of this article contains additional supplementary data. Please visit doi:10.1016/j.jtbi.2007.09.043.

References

Alon, U., 2007. An Introduction to Systems Biology: Design Principles of Biological Circuits. Chapman & Hall/CRC, Boca Raton, FL.

Barchiesi, F., Spreghini, E., Schimizzi, A.M., Maracci, M., Giannini, D., Carle, F., Scalise, G., 2004. Posaconazole and amphotericin B combination therapy against *Cryptococcus neoformans* infection. Antimicrob. Agents Chemother. 48 (9), 3312–3316.

Blank, L.M., Kuepfer, L., Sauer, U., 2005. Large-scale ^{13}C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. Genome Biol. 6 (R49), 1–16.

Çakır, T., Tacer, C.S., Ülgen, K.Ö., 2004. Metabolic pathway analysis of enzyme-deficient human red blood cells. Biosystems 78 (1–3), 49–67.

Deutscher, D., Meilijson, I., Kupiec, M., Ruppin, E., 2006. Multiple knockout analysis of genetic robustness in the yeast metabolic network. Nat. Genet. 38 (9), 993–998.

Ebenhöh, O., Heinrich, R., 2003. Stoichiometric design of metabolic networks: multifunctionality, clusters, optimization, weak and strong robustness. Bull. Math. Biol. 65 (2), 323–357.

Ebenhöh, O., Handorf, T., Kahn, D., 2006. Evolutionary changes of metabolic networks and their biosynthetic capacities. IEE Proc. Syst. Biol. 153 (5), 354–358.

Fong, S.S., Palsson, B.Ø., 2004. Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. Nat. Genet. 36 (10), 1056–1058.

Frantz, S., 2005. Drug discovery: playing dirty. Nature 437 (7061), 942–944.

Ghim, C.-M., Goh, K.-I., Kahng, B., 2005. Lethality and synthetic lethality in the genome-wide metabolic network of *Escherichia coli*. J. Theor. Biol. 237 (4), 401–411.

Handorf, T., Ebenhöh, O., Heinrich, R., 2005. Expanding metabolic networks: scopes of compounds, robustness, and evolution. J. Mol. Evol. 61 (4), 498–512.

Harrison, R., Papp, B., Pál, C., Oliver, S.G., Delneri, D., 2007. Plasticity of genetic interactions in metabolic networks of yeast. Proc. Natl Acad. Sci. 104 (7), 2307–2312.

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

Huang, S., 2001. Genomics, complexity and drug discovery: insights from Boolean network models of cellular regulation. Pharmacogenomics 2 (3), 203–221.

Jordan, I.K., Kondrashov, F.A., Adzhubei, I.A., Wolf, Y.I., Koonin, E.V., Kondrashov, A.S., Sunyaev, S., 2005. A universal trend of amino acid gain and loss in protein evolution. Nature 433 (7026), 633–638.

Kaufman, A., Keinan, A., Meilijson, I., Kupiec, M., Ruppin, E., 2005. Quantitative analysis of genetic and neuronal multi-perturbation experiments. PLoS Comput. Biol. 1 (6), 500–506.

Klamt, S., 2006. Generalized concept of minimal cut sets in biochemical networks. Biosystems 83 (2–3), 233–247.

Klamt, S., Gilles, E.D., 2004. Minimal cut sets in biochemical reaction networks. Bioinformatics 20 (2), 226–234.

Krömer, J.O., Wittmann, C., Schröder, H., Heinzle, E., 2006. Metabolic pathway analysis for rational design of L-methionine production by *Escherichia coli* and *Corynebacterium glutamicum*. Metab. Eng. 8, 353–369.

Kuepfer, L., Sauer, U., Blank, L.M., 2005. Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. Genome Res. 15 (10), 1421–1430.

Lemke, N., Herédia, F., Barcellos, C.K., dos Reis, A.N., Mombach, J.C.M., 2004. Essentiality and damage in metabolic networks. Bioinformatics 20 (1), 115–119.

Pál, C., Papp, B., Lercher, M.J., 2005. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. Nat. Genet. 37 (12), 1372–1375.

Papp, B., Pál, C., Hurst, L.D., 2004. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. Nature 429 (6992), 661–664.

Pfeiffer, T., Sanchez-Valdenebro, I., Nuño, J.C., Montero, F., Schuster, S., 1999. METATOOL: for studying metabolic networks. Bioinformatics 15 (3), 251–257.

Ruoff, P., Christensen, M.K., Wolf, J., Heinrich, R., 2003. Temperature dependency and temperature compensation in a model of yeast glycolytic oscillations. Biophys. Chem. 106 (2), 179–192.

Schuldiner, M., Collins, S.R., Thompson, N.J., Denic, V., Bhamidipati, A., Punna, T., Ihmels, J., Andrews, B., Boone, C., Greenblatt, J.F., Weissman, J.S., Krogan, N.J., 2005. Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. Cell 123 (3), 507–519.

Schuster, R., Holzhütter, H.G., 1995. Use of mathematical models for predicting the metabolic effect of large-scale enzyme activity alterations. Application to enzyme deficiencies of red blood cells. Eur. J. Biochem. 229 (2), 403–418.

Schuster, S., Kenanov, D., 2005. Adenine and adenosine salvage pathways in erythrocytes and the role of *S*-adenosylhomocysteine hydrolase—a theoretical study using elementary flux modes. FEBS J. 272 (20), 5278–5290.

Schuster, S., Fell, D.A., Pfeiffer, T., Dandekar, T., Bork, P., 1998. Elementary modes analysis illustrated with human red cell metabolism. In: Larsson, C., Påhlman, I.-L., Gustafsson, L. (Eds.), BioThermoKinetics in the Post Genomic Era. Chalmers, Göteborg, pp. 332–339.

Schuster, S., Fell, D.A., Dandekar, T., 2000. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat. Biotechnol. 18 (3), 326–332.

Schwender, J., Goffman, F., Ohlrogge, J.B., Shachar-Hill, Y., 2004. Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. Nature 432 (7018), 779–782.

Shlomi, T., Berkman, O., Ruppin, E., 2005. Regulatory on/off minimization of metabolic flux changes following genetic perturbations. Proc. Natl Acad. Sci. 102 (21), 7695–7700.

Simons, A., Dafni, N., Dotan, I., Oron, Y., Canaani, D., 2001. Establishment of a chemical synthetic lethality screen in cultured human cells. Genome Res. 11 (2), 266–273.

Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., Gilles, E.D., 2002. Metabolic network structure determines key aspects of functionality and regulation. Nature 420 (6912), 190–193.

Stelling, J., Sauer, U., Szallasi, Z., Doyle 3rd, F.J., Doyle, J., 2004. Robustness of cellular functions. Cell 118 (6), 675–685.

Taburet, A.M., Raguin, G., Le Tiec, C., Droz, C., Barrail, A., Vincent, I., Morand-Joubert, L., Chêne, G., Clavel, F., Girard, P.-M., 2004. Interactions between amprenavir and the lopinavir–ritonavir combination in heavily pretreated patients infected with human immunodeficiency virus. Clin. Pharmacol. Ther. 75 (4), 310–323.

Tong, A.H.Y., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G.F., Brost, R.L., Chang, M., Chen, Y., Cheng, X., Chua, G., Friesen, H., Goldberg, D.S., Haynes, J., Humphries, C., He, G., Hussein, S., Ke, L., Krogan, N., Li, Z., Levinson, J.N., Lu, H., Ménard, P., Munyana, C., Parsons, A.B., Ryan, O., Tonikian, R., Roberts, T., Sdicu, A.-M., Shapiro, J., Sheikh, B., Suter, B., Wong, S.L., Zhang, L.V., Zhu, H., Burd, C.G., Munro, S., Sander, C., Rine, J., Greenblatt, J., Peter, M., Bretscher, A., Bell, G., Roth, F.P., Brown, G.W., Andrews, B., Bussey, H., Boone, C., 2004. Global mapping of the yeast genetic interaction network. Science 303 (5659), 808–813.

Voet, D., Voet, J.G., 2004. Biochemistry. Wiley, Hoboken, New Jersey.

von Kamp, A., Schuster, S., 2006. Metatool 5.0: fast and flexible elementary modes analysis. Bioinformatics 22 (15), 1930–1931.

Wilhelm, T., Behre, J., Schuster, S., 2004. Analysis of structural robustness of metabolic networks. IEE Proc. Syst. Biol. 1 (1), 114–120.

Wolf, J., Becker-Weimann, S., Heinrich, R., 2005. Analysing the robustness of cellular rhythms. IEE Proc. Syst. Biol. 2 (1), 35–41.

Wong, S.L., Zhang, L.V., Tong, A.H.Y., Li, Z., Goldberg, D.S., King, O.D., Lesage, G., Vidal, M., Andrews, B., Bussey, H., Boone, C., Roth, F.P., 2004. Combining biological networks to predict genetic interactions. Proc. Natl Acad. Sci. 101 (44), 15,682–15,687.