Exact quantification of cellular robustness in genome-scale metabolic networks

Matthias P. Gerstl1,2, Steffen Klamt3, Christian Jungreuthmayer1,2 and Jürgen Zanghellini1,2,*

1Austrian Centre of Industrial Biotechnology, Vienna, Austria, 2Department of Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria, and 3Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany

*To whom correspondence should be addressed.

Associate Editor: Igor Jurisica

Received on June 5, 2015; revised on October 6, 2015; accepted on October 28, 2015

Abstract

Motivation: Robustness, the ability of biological networks to uphold their functionality in spite of perturbations, is a key characteristic of all living systems. Although several theoretical approaches have been developed to formalize robustness, it still eludes an exact quantification. Here, we present a rigorous and quantitative approach for the structural robustness of metabolic networks by measuring their ability to tolerate random reaction (or gene) knockouts.

Results: In analogy to reliability theory, based on an explicit consideration of all possible knockout sets, we exactly quantify the probability of failure for a given network function (e.g. growth). This measure can be computed if the network’s minimal cut sets (MSCs) are known. We show that even in genome-scale metabolic networks the probability of (network) failure can be reliably estimated from MSCs with lowest cardinalities. We demonstrate the applicability of our theory by analyzing the structural robustness of multiple Enterobacteriaceae and Blattibacteriaceae and show a dramatically low structural robustness for the latter. We find that structural robustness develops from the ability to proliferate in multiple growth environments consistent with experimentally found knowledge.

Conclusion: The probability of (network) failure provides thus a reliable and easily computable measure of structural robustness and redundancy in (genome-scale) metabolic networks.

Availability and implementation: Source code is available under the GNU General Public License at https://github.com/mpgerstl/networkRobustnessToolbox.

Contact: juergen.zanghellini@boku.ac.at

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

A major characteristic of biological systems is their ‘robustness’, i.e. the ability to perform normally under the presence of perturbations (Kitano, 2002). Despite its fundamental importance, robustness is hard to quantify and a comprehensive, quantitative understanding of robustness has yet to be developed (Kitano, 2007). The problem is augmented by the fact that robustness is an extremely general concept which obstructs efforts for an exact definition (Stelling et al., 2004). Clearly, robustness arises as a consequence of the interactions between the components in a system. It therefore requires a network-based approach for its analysis (Larhlimi et al., 2011).

In the following, we study a key aspect of cellular robustness, namely structural robustness in metabolic networks (Wilhelm et al., 2004), which is a measure of the apparent redundancy in metabolic networks. Redundancy is one reason why specific
elementary flux modes (EFMs) (Schuster distributions in a metabolic network can be achieved in terms of 2014).

We consider a metabolic network made up of published protocols (Thiele and Palsson, 2010; Vlassis et al., 2014).

An unbiased way to characterize all topologically feasible flux distributions in a metabolic network can be achieved in terms of elementary flux modes (EFMs) (Schuster et al., 2000). EFM s are steady-state pathways that use a minimal set of reactions (enzymes) in the appropriate thermodynamic directions. Biologically, they can be interpreted as simple, indivisible functional building blocks, which represent all feasible metabolic capabilities of the organism. Thus, EFM s are likely candidates to be used in a definition of structural robustness. In fact, Wilhelm et al. (2004) were the first to define structural robustness by counting the number of EFM s in different situations.

An alternative approach to EFMs is minimal cut sets (MCSs) (Klamt and Gilles, 2004). These are minimal sets of reaction (enzyme) knockouts that completely disable a particular network functionality. MCSs are intimately connected to EFMs. Any network function can be represented by a set of EFM s. Thus, MCSs can equivalently be defined as the minimal set of reaction knockouts that prevent a set of target EFM s from carrying a steady-state flux. An MCS is therefore a so-called minimal hitting set of structural robustness. In fact, Wilhelm et al. (2004) showed in the Supplementary material that the PoF can be calculated for any number of deletions d by

$$ R(d) = \frac{1}{|E^d|} \sum_{i=1}^{n} \frac{|E^d|}{|E^i|}. $$

We show in the Supplementary material that the network's structural robustness only depends on the cardinality of the network's EFM s and can easily be calculated for any number of deletions d by

$$ R(d) = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{r - |E_i|}{d} \right). $$

Thus, network structural robustness is merely determined by the EFM s' length.

2.2 PoF: F(d)

For a given number of deletions, d, we define the network's PoF, $F(d)$, as the ratio of the number of CSs that disable a particular network objective, obj, to the total number of all possible combinations of exactly d knockouts, i.e.

$$ F_{\text{obj}}(d) := \frac{\#d_{\text{obj}}}{|E^d|}. $$

CSs that disable a particular objective (i.e., the elements of $\#d_{\text{obj}}$) are either MCSs or supersets of MCSs disabling 'obj'. We show in the Supplementary material that the PoF can be calculated by

$$ F_{\text{obj}}(d) = \sum_{\#d \subseteq \{1, \ldots, d\}} (-1)^{|\#d|-1} \left( \frac{r - |M_j|}{d - |M_j|} \right), $$

where $M_j = \bigcup M_i$ and $j$ is a (multi)-index over all power sets of the indices $\{1, \ldots, n\}$ of the set $\#d$ of MCSs blocking 'obj'.

Note that the definition of the PoF requires specification of an objective, that can be the operation of the whole system or of a particular function. A typical cellular function is the production of biomass (BM). Thus, $F_{BM}(d)$ measures the probability that the deletion of d reactions is lethal. In what follows we will drop the index for the objective and implicitly assume that the production of BM is targeted, unless stated otherwise.
2.3 Overall structural robustness \( R \) and overall PoF (OPoF) \( \bar{F} \)

Both \( R(d) \) and \( F(d) \) are defined for a specific number of reaction deletions. Following Behre et al. (2008), we define the overall structural robustness, \( \bar{R} \), if at most (all) \( r \) reactions are knocked out, as the weighted sum of all possible \( R(d) \) (Behre et al., 2008),

\[
\bar{R} := \sum_{d=1}^{r} w_d^R R(d). \tag{5}
\]

Analogous to the structural robustness we define the OPoF, \( \bar{F} \) as

\[
\bar{F} := \sum_{d=1}^{r} w_d^F F(d). \tag{6}
\]

Behre et al. (2008) suggested to use the following weights

\[
w_d^R := \binom{r}{d} p^d (1-p)^{r-d}, \tag{7}
\]

However, their choice is difficult to interpret in terms of a probability distribution. That is why we opted for an appropriately selected probability distribution function for the calculation of \( F \). \( w_d^F \) may be naturally modeled by a binomial distribution,

\[
w_d^F := \binom{r}{d} p^d (1-p)^{r-d}. \tag{8}
\]

which estimates the number of loss-of-function mutations with exactly \( d \) reaction deletions assuming a constant (reaction) mutation rate \( p \). In typical metabolic networks, the number of reactions \( r \) is large and the mutation rate is small. We therefore approximate the binomial distribution with the simpler Poisson distribution,

\[
w_d^F := \lambda^d \exp(-\lambda)/(d!), \tag{9}
\]

with \( \lambda = r p \). \( R(d), \bar{R}, F(d) \) and \( \bar{F} \) all range between 0 and 1. However, while \( \bar{R} = 1 \) indicates a totally robust network, \( \bar{F} = 1 \) denotes the reverse, a completely fragile network.

3 Implementation and computation

The computation of the structural robustness \( R(d) \) according to Equation (2) is straightforward as long as the complete set of EFMs is known. The computation of the PoF is computationally more demanding as the number of summands in the sum of Equation (4) explodes combinatorially. In the following, we will outline countermeasures to estimate the PoF even in GSMMs.

**Algorithm 1. Recursive CS count**

**Require:** \( mcs, deletions d, reactions r \)

1. for \( i = 1 \) to \( \text{LENGTH}(mcs) \) do
2. \( \text{csCount}[i] \leftarrow \text{GET_CS_COUNT}(i, mcs[i], 0, \text{false}) \)
3. end for
4: end function

5: function GET_CS_COUNT(index, cs, csCount, inRecursion)
6: \( \text{foundSubset} \leftarrow \text{false} \)
7: \( \text{combCs} \leftarrow \text{mcs[index]} \) \text{ OR cs}
8: \( \text{combCd} \leftarrow \text{CARDINALITY}(\text{combCs}) \)
9: \( \text{combCsCount} \leftarrow \text{CHOICE}(r - \text{combCd}, d - \text{combCd}) \)
10: for \( i = 1 \) to \( (index - 1) \) do
11: \( \text{testCs} \leftarrow \text{mcs[i]} \) \text{ OR combCs}
12: \( \text{testCd} \leftarrow \text{CARDINALITY}(\text{testCs}) \)
13: if \( \text{testCd} < d \) delete then
14: if \( \text{testCd} > \text{combCd} \) then
15: \( \text{combCsCount} \leftarrow \text{GET_CS_COUNT}(i, \text{combCs}, \text{combCsCount}, \text{true}) \)
16: else
17: \( \text{foundSubset} \leftarrow \text{true} \)
18: exit for loop
19: end if
20: end if
21: end for
22: if not \( \text{foundSubset} \) then
23: if \( \text{inRecursion} \) then
24: \( \text{csCount} \leftarrow \text{csCount} - \text{combCsCount} \)
25: else
26: \( \text{csCount} \leftarrow \text{combCsCount} \)
27: end if
28: end if
29: return \( \text{csCount} \)
30: end function

**Table 1.** Major topological properties of the MSMMs and GSMMs for the investigated *Enterobacteriaceae* and *Blattibacteriaceae*

| Model ID | Organism                  | Medium       | \( M \) | \( r \) | Rank(S) | \( n \) | \( n_{BM} \) |
|----------|---------------------------|--------------|--------|------|---------|-------|------------|
| Ecoli_core + glc (Gerstl et al., 2015) | *E. coli* K-12 MG1655 | Glucose     | 70     | 90   | 65      | 169,916 | 121,753    |
| Ecoli_core + gly (Gerstl et al., 2015) | *E. coli* K-12 MG1655 | Glycerol    | 71     | 91   | 66      | 60,495  | 48,944     |
| Ecoli_core + ac (Gerstl et al., 2015) | *E. coli* K-12 MG1655 | Acetate     | 69     | 88   | 64      | 1299    | 736        |
| iJO1366 + glc (Monk et al., 2013)   | *E. coli* K-12 MG1655 | Glucose     | 1165   | 1726 | 1131    | n/a     | n/a        |
| iJO1366 + mel (Monk et al., 2013)   | *E. coli* K-12 MG1655 | Melibiose   | 1163   | 1718 | 1128    | n/a     | n/a        |
| iECs_1301 + mel (Monk et al., 2013) | *E. coli* O157:H7 Sakai | Melibiose  | 1098   | 1666 | 1057    | n/a     | n/a        |
| i5S_1188 + mel (Monk et al., 2013)  | *S. flexneri* 2a 2457T | Melibiose  | 1,026  | 1,517 | 982     | n/a     | n/a        |
| iCG230 + full (González-Domenech et al., 2012) | *B. subtilis* Pan | Full        | 299    | 342  | 292     | n/a     | n/a        |
| iCG238 + full (González-Domenech et al., 2012) | *B. subtilis* Bge | Full        | 306    | 350  | 299     | n/a     | n/a        |

Models were taken from Gerstl et al. (2015), Monk et al. (2013) and González-Domenech et al. (2012). Aerobic growth for *Enterobacteriaceae* was simulated on minimal medium with glucose or melibiose as sole carbon source and for *Blattibacteriaceae* with full medium (for media compositions see Supplementary Table S1). All reactions that could not carry a steady state flux under any circumstances for the given growth media were removed from the original models. \( m \) and \( r \) refer to the (remaining) number of internal metabolites and reactions, respectively, for fully consistent models (represented by the respective internal stoichiometric matrix, \( S \)). \( n \) and \( n_{BM} \) refer to the total number and the number of BM producing EFMs in the respective models. An EFM analysis was not applicable (n/a) for GSMMs.
Algorithm 1 implements Equation (4) by calculating the total number of CSs for given deletions \( d \) in a recursive way. The number of reactions in the model, as well as all mcs with a cardinality up to the current number of reaction deletions are needed for the calculation. The algorithm calculates the number of all possible CSs for a single MCS by the function choose, which computes the binomial coefficient for \( \binom{r - combCd}{d - combCd} \), where combCd is the number of deletions resulting from the combination of MCs. In order to avoid the consideration of a CS multiple times those CSs that have already been calculated by other MCSs, indicated by their index, are subtracted. This is done by creating an intermediate CS by a logical OR operation and then following the inclusion–exclusion principal (Ryser, 1963) in a recursive procedure.

Algorithm 1 was implemented as multi-threaded C program which is available at https://github.com/mpgerstl/networkRobustnessToolbox.

### 4 Results

In the following analysis, we used \( \lambda = 0.5 \), unless stated otherwise.

#### 4.1 \( \bar{R} \) reliably identifies structural fragility

We used a medium-scale metabolic model (MSMM) that describes the core metabolism of *Escherichia coli* (see Table 1 for an overview of the topological properties of the models used) to study the influence of the carbon source on the structural robustness and the PoF.

![Graph](image)

**Table 2. Structural robustness and PoF in a core metabolic model of *E. coli* growing on minimal media (see Supplementary Table S1) and three different carbon sources (glc, glucose; gly, glycerol; ac, acetate).**

| model ID         | \( R(1) \) | \( R(2) \) | \( \bar{R} \) | \( F(1) \) | \( F(2) \) | \( \bar{F} \) |
|------------------|------------|------------|-------------|------------|------------|-------------|
| EColi_core + glc | 46.32      | 21.54      | 37.98       | 21.11      | 40.45      | 10.33       |
| EColi_core + gly | 46.19      | 21.69      | 37.95       | 31.87      | 54.63      | 14.84       |
| EColi_core + ac  | 50.86      | 26.98      | 42.75       | 42.05      | 67.42      | 19.06       |

We used all EFMs for the calculation of \( \bar{R} \) and all synthetic lethals of groups of up to \( d_0 = 9 \) reactions for the estimation of \( \bar{F} \) with \( \lambda = 0.5 \). All values are listed in %

![Graph](image)

**Fig. 1.** Number of recursions in Algorithm 1 as function of the number of reaction deletions, \( d \), evaluated for the model EColi_core + glc in different computation scenarios. In addition, we plotted the number of MCS (dotted line) as function of their cardinality for the model used (see Table 2). In all cases, a complete EFM and MCS analysis was feasible.

We found that both measures indicated a strong impact of the carbon sources although with opposing trends. The OPoF was lowest for growth on glucose and highest for growth on acetate. We observed that in all three growth environments the PoF \( F(d) \) could be well approximated (coefficient of determination, \( R^2 > 0.999 \)) by \( a[1 - \exp(-\gamma d)] \) with three free parameters \( a, \beta \), and \( \gamma \) (see Supplementary Fig. S1). \( \bar{R} \) however ranked growth on acetate to be most robust and growth on glucose or glycerol similarly robust but far behind growth on acetate. According to Equation (2) this implied that for acetate EFMs were on average shorter then for glucose (see Supplementary Fig. S2). Finally, the simple ratio between the number of BM producing EFMs and the total number of EFMs revealed yet another ranking with the highest ratio for growth on glycerol and the lowest ratio for growth on acetate (see Supplementary Table S2). Comparing the different measures, we found that only the OPoF scored the (lack of) structural robustness in *E. coli* consistently to previous analysis (Klamt and Gilles, 2004; Stelling et al., 2002).

#### 4.2 \( \bar{R} \) and \( \bar{F} \) can be calculated exactly in MSMMs

Judged by the number of EFMs, the three growth models significantly differed in size (see Table 1). Yet all structural robustness-values were calculated within few seconds for any number of reaction deletions. As the number of summands in Equation (4) grew excessively for high number of reaction deletions, \( d \), the calculation of \( F(d) \) was more demanding. We found an exponential increase in the number of recursions for Algorithm 1 (see Fig. 1).

Our optimized implementation scaled consistently better than that and allowed us to evaluate \( F(d) \) in the largest model, EColi_core + glc, up to \( d = 9 \) in \(< 10 \text{ min} \) using 10 threads on an Intel® Core™i7-3930 K with two CPUs à six cores and 3.20 GHz operated with Ubuntu 12.04.

#### 4.3 \( \bar{R} \) and \( \bar{F} \) are determined by the first few deletions

We analyzed the contribution of higher-order deletions to the overall structural robustness and OPoF, respectively. In the best case higher-order terms \( R(d) \) and \( F(d) \) vanish for \( d > d_0 \) causing no error in the sums of Equations (5) and (6). In the worst case higher-order...
Table 3. PoFs in GSMMs of different Enterobacteriaceae growing on minimal media (see Supplementary Table S1) and glucose (glc) or melibiose (mel) or melibiose (mel) or melibiose (mel)

| Model ID          | F(1)  | F(2)  | F(3)  | F(4)  | F | ε(4) |
|-------------------|-------|-------|-------|-------|---|------|
| iJO1366 + glc     | 16.74 | 30.71 | 42.35 | 52.06 | 8.02 | 0.02 |
| iJO1366 + mel     | 17.17 | 31.42 | 43.24 | 53.04 | 8.22 | 0.02 |
| iECs_1301 + mel   | 17.59 | 32.11 | 44.10 | 53.98 | 8.41 | 0.02 |
| iS_1188 + mel     | 19.97 | 35.99 | 48.83 | 59.12 | 9.51 | 0.00 |

The PoF was estimated by considering all synthetic lethals of groups of up to d = 4 reactions and λ = 0.5. All values are listed in %.

For fixed λ, the PoF F depends only on the smallest MCS and consistently converges toward its true value with increasing MCS cardinality. In the E. coli example above (see Table 2), all MCSs up to cardinality 4 were sufficient to estimate F within an error of 0.02 percentage points. However, a similar argument for the overall structural robustness R is not available. The knowledge of the d shortest EFMs does not allow one to accurately estimate R. Supplementary Figure S5 illustrates the error in the overall structural robustness as function of the d shortest EFMs for the E. coli model used above. In that case, >55% of all EFMs were required to get within a 10% error margin.

4.4 For small λ, F is well approximated in GSMMs

According to Equation (11), the PoF could be estimated if all low cardinality MCs were known. Recently, von Kamp and Klamt (2014) showed that it is in fact possible to enumerate the smallest MCs in GSMMs. In the following, we used their approach to calculate low cardinality MCs.

We calculated all lethal MCs up to a cardinality of 4 in GSMMs of three E. coli strains growing on minimal medium and glucose or melibiose as the sole carbon source and estimated the PoF (see Table 3). The models represented E. coli K-12 MG1655 (iJO1366), the pathogenic enterohemorrhagic E. coli O157:H7 Sakai (iECs_1301) and the Shigella flexneri 2a 2457T (iS_1188).

Comparing the three organisms against each other revealed that growth on glucose is more failsafe than growth on the alternative carbon source melibiose. In fact, growth on glucose (and glucose 1-phosphate) was found to be more robust than on any other single, nitrogen-free carbon source (see Supplementary Fig. S4). The two E. coli strains were similarly robust (with a small advantage for E. coli K-12 MG1655), while the S. flexneri was most fragile. Although the difference in the PoF was small, it is indicative as the difference in F for any two models is larger than λ (4). We observed the same trend not only for F but also for F for all tested cardinalities d ≤ 4, as well. The GSMMs were also found to be more robust than the E. coli core model. Again, in all models, F(d) could be nicely approximated by $\lambda[1 - \exp(-d)]$ (see Supplementary Table S4).

Finally, we calculated all synthetic lethal reactions in groups of up to d = 6 reaction deletions and evaluated the PoF in the GSMMs of the endosymbiotic Blattabacterium cuenoti Bge (iCG238 + full) and B. cuenoti Pam (iCG230 + full), see Table 4, for full growth media containing all nutrients that were possibly taken up, see Supplementary Table S1 for detailed media compositions. Both strains were found to be extremely fragile with PoFs > 26% and a maximal inaccuracy of 1 × 10⁻⁴ percentage points. Any three reaction deletions almost certainly killed these two strains (with a small chance of survival of <5%).

Table 4. PoFs in GSMMs of B. cuenoti Bge (iCG238 + full) and B. cuenoti Pam (iCG230 + full) growing on full media (see Supplementary Table S1)

| Model ID          | F(1)  | F(2)  | F(3)  | F(4)  | F(5)  | F(6)  | F | ε(6) |
|-------------------|-------|-------|-------|-------|-------|-------|---|------|
| iCG230 + full     | 64.62 | 87.68 | 95.81 | 98.61 | 99.55 | 99.86 | 27.63 | 10⁻⁴ |
| iCG238 + full     | 62.57 | 86.36 | 95.16 | 98.31 | 99.43 | 99.81 | 26.89 | 10⁻⁴ |

The PoF was estimated by considering all synthetic lethals of groups of up to d = 6 reactions and λ = 0.5. All values are listed in %.

Table 5. GC and RC PoF for various MSMMs and GSMMs

| Model ID          | F(1)  | F(2)  | F | RC |
|-------------------|-------|-------|---|----|
| EColi_core + glc  | 5.00  | 21.11 | 10.01 | 40.45 | 2.50 | 10.31 |
| EColi_core + gly  | 5.71  | 31.87 | 11.46 | 54.63 | 2.86 | 14.84 |
| EColi_core + ac   | 15.00 | 42.05 | 29.10 | 67.42 | 7.38 | 19.06 |
| iJO1366 + glc     | 16.94 | 16.74 | 31.03 | 50.71 | 8.11 | 8.02 |
| iJO1366 + mel     | 17.55 | 17.17 | 32.05 | 31.42 | 8.39 | 8.22 |
| iECs_1301 + mel   | 19.14 | 17.59 | 34.65 | 32.11 | 9.12 | 8.41 |
| iS_1188 + mel     | 21.82 | 19.97 | 38.92 | 35.99 | 10.33 | 9.51 |
| iCG230 + full     | 66.34 | 64.62 | 88.96 | 87.68 | 28.24 | 27.63 |
| iCG238+full       | 62.26 | 62.57 | 86.33 | 86.36 | 26.79 | 26.89 |

For the GSMMs the PoFs were estimated by considering all synthetic lethals of groups of up to d = 4 reactions and λ = 0.5. For these settings the estimation error was <0.02 percentage points. All values are listed in %.

4.5 Gene-centric PoFs showed same behavior as reaction-centric PoFs

For simplicity so far we analyzed structural robustness from a reaction-centric (RC) viewpoint. However, it is possible to change from an RC to a gene-centric (GC) viewpoint while retaining our formalism for the PoF calculation. PoF only requires the calculation of any three MCSs. These can be calculated in an RC model as well as in a GC model provided that a gene–reaction mapping is available as a Boolean function of the genes (as it is typically the case). GC MCSs consider the effect of gene knock-outs on reactions based on the evaluation of the provided Boolean rules. If the complete set of EFMs is known, GC MCSs can be calculated using an established integer programming procedure (Jungeuthmayer and Zanghellini, 2012). In GSMMs, where the complete set of EFMs is unavailable, the dual systems approach (von Kamp and Klamt, 2014) can be
adopted to compute GC MCSs. The common dual approach calculates RC MCSs (ordered by ascending cardinality) by iteratively solving a mixed integer linear program (MILP) (De Figueiredo et al., 2009). In order to obtain a GC MCS instead of an RC MCS the gene–reaction association has to be integrated into the MILP, which can be achieved by means of the so-called indicator constraints. These concepts are provided by modern MILP solvers, which allow for the Boolean coupling of reactions and their associated genes.

We evaluated GC PoFs for models listed in Table 1 and compared them against the corresponding RC PoFs (see Table 5). We found comparable (O)PoF values except for the MSMM of E. coli, where we observed larger differences. Yet the absolute ranking of the organisms with respect to structural robustness remained unchanged no matter if the GC or RC PoF was used.

5 Discussion

Here, we developed a new general measure to quantify the structural robustness (or fragility) of metabolic networks based on the (O)PoF. Both, PoF $F(d)$ as well as OPoF $\mathcal{F}$ are based on MCSs and share analogies with the previously defined structural robustness measures $R(d)$ and $\mathcal{R}$ which rely on EFMs (Behre et al., 2008). In contrast to $\mathcal{R}$, the OPoF remains computationally feasible even in GSMMs for two reasons: (i) the OPoF can be well estimated from the shortest MCSs with up to $d_0$ reaction cuts; and (ii) these $d_0$ shortest MCSs can be calculated with a MILP even in GSMMs (von Kamp and Klamt, 2014). Previously, it was shown that MCSs in the (primal) network are EFMs in a network dual to the original network (Ballerstein et al., 2012). Thus, finding the shortest EFMs in this dual network is equivalent to finding the low cardinality MCSs in the primal GSMM (von Kamp and Klamt, 2014). A high performance computing infrastructure is not required for this task. Thus, the OPoF is in principle an easily computable measure. However, due to the combinatorial explosion of the number of summands in $F(d)$, see Equation (4), the evaluation is practically limited to low cardinalities. This does not cause practical problems for the OPoF $\mathcal{F}$ as the impact of high-cardinality MCSs on $\mathcal{F}$ quickly decreases (see Fig. 2).

The slope of this decrease depends on the selected probability density that models the number of deletions that occur per mutation. Here, we used a Poisson distribution which is typically used for the study of rare events, like mutations (Lee et al., 2011). We were able to show that up to $\lambda \leq 1$ our analysis is computationally feasible on standard computer infrastructure. This is sufficient for the study of naturally occurring mutations. For instance, in E. coli a mutation rate of about $10^{-3}$ per genome and generation was found (Lee et al., 2012), which is far below our limit of $\lambda \leq 1$, as we can assume that not all mutations lead to a loss of function of metabolic enzymes. All our conclusions remained valid (data not shown) if we used other probability densities like a binomial distribution or the one used by Behre et al. (2008), see Equation (7).

Previously, MCSs were already used to compute a so-called fragility coefficient which characterized the sensitivity of a metabolic network to deletions (Klamt, 2006; Klamt and Gilles, 2004). Fragility was defined for each reaction as the inverse of the average cardinality of all MCSs that were supported by that reaction. Upon averaging the reaction-specific fragility over all reactions one obtained a measure for the overall network fragility. In contrast to the here introduced concept of the OPoF, network fragility lacks a rigorous probabilistic definition and more importantly requires the complete set of MCSs for its calculation. Network fragility is therefore not applicable to GSMMs.

A clear difference between the PoF and structural robustness is that the former is function oriented while the latter characterizes the network. $\mathcal{F}$ is calculated with respect to some particular function that fails. Typically, networks perform multiple functions at the same time, which may be differentially robust. Thus, the OPoF is not only a topological property but also a property of the associated function. For instance, it is interesting to ask (and scope of further analysis) whether two essential functions like cellular energy production and BM production are similarly robust. $\mathcal{R}$ on the other hand characterizes the network topology by essentially measuring the proportion of short EFMs in a network. This definition is independent of any functions performed by the network. $\mathcal{R}$ counterintuitively identified growth on acetate to be more robust than growth on glucose which indicates that the length of EFMs may not be an appropriate proxy for structural robustness. Even if we selected only those EFMs that support a particular target function (like BM production), and calculated $\mathcal{R}$ for the reduced set of EFMs, we had misidentified growth on acetate as the most robust growth environment as growth supporting EFMs on acetate were on average ‘shorter’ than on glucose (see Supplementary Fig. S2).

For better understanding we illustrated this effect in a toy network in Supplementary Figure S3. Thus, we propose to use the complement of the OPoF: $\mathcal{R}_{new} = 1 - \mathcal{F}$ as a new measure for structural robustness in metabolic networks.

PoFs can be calculated equally well on the basis of reaction deletions as well as on the basis of gene deletions. Some reactions in an organism are not associated with a gene (e.g. non-catalyzed reactions, reactions with unknown annotations) and can therefore not be knocked out. Other reactions are catalyzed by several different enzymes and are therefore more difficult to knock-out. However, our analysis indicates that the question if an organism or a growth condition is more robust than another can be identically answered in both approaches.

It was criticized that structural robustness (as defined by Behre et al., 2008) lacks the ability to identify critical enzymes, such as knot enzymes (Min et al., 2011). This is true for the OPoF as well. For instance, according to our measure, network A in Figure 3 was correctly identified to be more robust than network B. The cause for the lack of structural robustness is not apparent from the numerical value of the OPoF. Min et al. (2011) proposed to solve this dilemma by considering the number of independent (i.e. non-overlapping) EFMs, which are not available in GSMMs. However, by measuring the specific PoF $f_R(d)$ for each individual reaction (enzyme) $R$ we are at least able to identify critical enzymes. For this purpose, we define the (reaction) specific PoF $f_R(d)$

$$f_R(d) := \frac{|c^d_R|}{|G^d|} \tag{12}$$

to be the ratio between all specific CSs, $c^d_R$, that are supported by reaction $R$ and the number of all possible CSs, $|G^d|$ (and similarly for $\mathcal{R}_R$). By ranking reactions according to their impact on the network we are able to identify critical reactions which may be useful drug targets.

We calculated OPoFs in MSMMs and GSMMs of multiple Enterobacteriaceae growing on minimal media to illustrate our concept. Among these, the ordinary lab strain E. coli K-12 MG1655 growing on glucose was found to be most robust. This was consistently observed in MSMMs and GSMMs in agreement with expectations (Stelling et al., 2002). Moreover, catabolite repression mechanisms support this observation of glucose as a preferential growth source.

Escherichia coli K-12 MG1655 was also found to be more robust than the pathogenic E. coli O157:H7 Sakai, and the least
Here, we analyzed the structural robustness of multiple organisms in several growth environments. Consistent with previous findings our results confirm that structural robustness is correlated with the ability to utilize multiple carbon sources (Papp et al., 2004). They concluded that structural robustness is not a trait selected by evolution but a byproduct of environmental flexibility. However, very recently Yang et al. (2015) showed that structural redundancy also impacts a cell factory’s ability to produce a product of interest in the presence of perturbations. Based on predictions from GSMMs, they argue that pathway diversification leads to robust production capabilities under large perturbations. For instance, we show in Supplementary Table S5 two strain designs with different OPoFs but identical production/growth characteristics. Thus, calculating the OPoF will identify promising designs for cell factories.

Interestingly, we found that the raise in $F(d)$ with increasing $d$ was well approximated by an exponential function in all investigated models and independent of the investigated organism. Whether this is an expression of some general topological properties of metabolic networks or a coincidence will be scope of further work.

### 6 Conclusion

We developed a consistent theory of cellular redundancy based on a rigorous probabilistic definition of failure in metabolic networks. The new measure, called OPoF, allows quantification of the structural robustness of biochemical reaction networks with respect to certain functionalities. OPoF can be reliably estimated—even in GSMMs—from low cardinality MSCs. In contrast to other measures, the OPoF quantifies structural robustness and fragility of cellular metabolism concurrent with current biological paradigms and experimental findings. More specifically we showed that the number of growth-supporting nutrients indirectly correlates with the organisms’ OPoF. As the OPoF is easily computable, we expect that it will be useful in increasing our understanding of key properties of metabolic networks.

### Funding

M.P.G., C.J. and J.Z. acknowledge the support by the Austrian BMWFW, BMVIT, SFG, Standortagentur Tirol, Government of Lower Austria and ZIT through the Austrian FFG-COMET-Funding Program. S.K acknowledges funding from the German Federal Ministry of Education and Research [projects CYANOSYS II (FKZ: 0316183D) and CASCOO (FKZ: 031A180B)] and the Federal State of Saxony-Anhalt (Research Center “Dynamic Systems: Biosystems Engineering”).

**Conflict of Interest:** none declared.

### References

Ballenstein, K. et al. (2012) Minimal cut sets in a metabolic network are elementary modes in a dual network. *Bioinformatics*, 28, 381–387.

Behre, J. et al. (2008) Structural robustness of metabolic networks with respect to multiple knockouts. *J. Theor. Biol.*, 252, 433–441.

Berge, C. (1989) *Hypergraphs, Volume 45: Combinatorics of Finite Sets*, 1st edn. North Holland, Amsterdam.

Bliven, K.A. and Maurelli, A.T. (2012) Antivirulence genes: insights into pathogen evolution through gene loss. *Infect. Immun.*, 80, 4061–4070.

De Figueiredo, L.F. et al. (2009) Computing the shortest elementary flux modes in genome-scale metabolic networks. *Bioinformatics*, 25, 3158–3165.

Gerstl, M.P. et al. (2015) Metabolomics integrated elementary flux mode analysis in large metabolic networks. *Sci. Rep.*, 5, 8930.
