Functional alignment of metabolic networks

Arnon Mazza[1], Allon Wagner[1,2], Eytan Ruppin[1,3,4], and Roded Sharan[1]

[1] Blavatnik School of Computer Science,
Tel Aviv University, Tel Aviv 69978, Israel.
Email: `roded@post.tau.ac.il`.
[2] Department of Electrical Engineering and Computer Science,
University of California, Berkeley, Berkeley, CA 94720-1770, USA.
[3] The Sackler School of Medicine,
Tel Aviv University, Tel Aviv 69978, Israel.
[4] Department of Computer Science, Institute of Advanced Computer Sciences (UMIACS) & the Center for
Bioinformatics and Computational Biology,
University of Maryland, College Park, MD 20742, USA.

Abstract. Network alignment has become a standard tool in comparative biology, allowing the inference of protein function, interaction and orthology. However, current alignment techniques are based on topological properties of networks and do not take into account their functional implications. Here we propose, for the first time, an algorithm to align two metabolic networks by taking advantage of their coupled metabolic models. These models allow us to assess the functional implications of genes or reactions, captured by the metabolic fluxes that are altered following their deletion from the network. Such implications may spread far beyond the region of the network where the gene or reaction lies. We apply our algorithm to align metabolic networks from various organisms, ranging from bacteria to humans, showing that our alignment can reveal functional orthology relations that are missed by conventional topological alignments.

1 Introduction

With the ever growing high throughput measurements of biological entities and relations, there is considerable interest in methods that go beyond sequence analysis to compare and contrast different species, conditions or time points. Network alignment methods present a promising alternative as they are able to capture topological similarities or differences that cannot be gleaned from sequence alone, improving the prediction of protein function, interaction and evolution [1].

Network alignment was originally applied to metabolic pathways [2] and soon thereafter to protein-protein interaction networks [3]. Over the last decade a plethora of methods were developed for the comparison of networks, including local alignment efforts [1,4] and global alignment methodologies [5,6]. All these methods are based on comparing the topology of the networks in question.

A metabolic network can be modeled by a hypergraph, whose nodes represent metabolites and hyperedges represent metabolic reactions. Many alignment algorithms transform this representation into a directed graph, where nodes represent reactions, or enzymes, and an edge is directed from

reaction A to B if A produces a substrate of B. This transformation allows the application of generic network alignment methods to metabolic networks, aiming to maximize a combined similarity measure that is based on enzyme homology and topology [7–10].

Another common form of modeling metabolism is via a constraint-based model, which allows expressing the space of fluxes of metabolic reactions under steady state assumptions (see a detailed definition in the next section). Such models can add functional information on the networks being compared, which could be exploited for alignment computation. The approach of [11] aligns two metabolic networks by comparing their elementary flux modes (EFMs), defined as minimal sets of reactions that can operate at steady state [12]. The concept of EFMs is also used by [13], where the similarity between two reaction sets is measured according to the impact incurred by their inhibition on the flux cone.

In this paper we propose a novel algorithm to align metabolic networks by taking advantage of their constraint-based models. These models allow us to assess the functional implications of genes or reactions, captured by the metabolic fluxes that are altered following their deletion from the network. Such implications may spread far beyond the region of the network where the gene or reaction lies, enabling the discovery of functional orthology relations that cannot be gleaned from topology alone. In the context of a human network, finding these nonobvious relations may reveal novel proteins in some model species that are inferred to be functionally similar to a disease causing protein and, hence, may allow new models for the disease in question.

The paper is organized as follows: In Section 2 we describe constraint-based modeling and present the metabolic network alignment problem. In Section 3 we present our alignment algorithm. In Section 4 we apply our algorithm to align metabolic networks from various organisms, ranging from bacteria to humans, and demonstrate its utility over topology-based approaches.

2 Preliminaries

2.1 Metabolic modeling

A genome-scale metabolic model (GSMM) describes a metabolic network by terms of metabolites (nodes) and biochemical reactions (hyperedges) on these metabolites. A GSMM can be represented by a stoichiometric matrix S, whose rows correspond to metabolites and columns to reactions.

It is accepted to assume that in a living cell, the concentrations of all metabolites are kept constant over time, signifying that a metabolite's production rate equals its consumption rate; this is known as the steady-state assumption. Some of the metabolites are continuously being taken up from the environment (a.k.a. growth medium), while others are being secreted to it; a special type of reactions, called *exchange reactions*, take care of these types of transport.

Every reaction is assigned with a flux, which measures the flow rate of compounds through the reaction. Flux capacities are naturally limited by availability of nutrients and capabilities of enzymatic activity. These presumptions are combined by applying a constraint-based modeling

approach (CBM), representing mass balance and flux directionality and capacity constraints on the space of possible fluxes in a metabolic network through a set of linear equations:

$$Sv = 0 \qquad (1)$$

$$v_{min} \leq v \leq v_{max} \qquad (2)$$

where v is the flux vector (a.k.a. flux distribution) of the reactions in the model. Flux balance analysis (FBA) is then applied to study various properties of the model [14]. FBA methods typically assume that the metabolic model attempts to optimize some objective function and use linear programming to compute the optimal solution of that function. A common biological optimization goal is the maximization of the amount of biomass that can be produced. The biomass function is an organism-dependent combination of metabolites which reflects its growth rate [15]. Often, the biomass optimal value may be achieved through many possible flux distributions, studied using a flux variability analysis (FVA) approach [16].

Another part of a GSMM is the gene-protein-reaction (GPR) associations, describing which genes and proteins catalyze which reactions, as well as the logical rules between the proteins required for catalyzation. This information allows simulating gene knockouts by inferring the affected reactions from the GPR and constraining the flux through them to 0. This approach was successfully used in numerous studies, for example in distinguishing viable from lethal yeast gene deletion strains by testing whether the optimal biomass production rate was severely damaged under a given knockout [17].

2.2 Problem definition

We focus here on the global alignment of two metabolic networks, representing the metabolic reactions (nodes) in two species. In this problem, one roughly seeks a one-to-one correspondence (or alignment) between the nodes of the two networks so that the similarity between matched nodes is maximized.

Different variants of the network alignment problem can be derived depending on the definition of node similarity and on whether one-to-many or many-to-many relations are allowed. Extant approaches to tackle this problem integrate sequence-based similarity data (on the genes catalyzing the reactions) as well as topology-based comparisons to construct a weighted bipartite graph, in which a plausible matching is sought using various types of algorithms.

In this paper we address a problem that is similar in flavour, which we solve using a maximum matching approach. Specifically, we assume that the input consists of two GSMMs that are to be compared, representing two species. Our goal remains to align them so as to maximize node similarity. The crucial difference from standard alignment is that the similarity measure that we use relies on the input metabolic models. Briefly, we represent every metabolic reaction by the functional impacts that its deletion induces on the model's metabolites. The derived similarity

measure is, hence, functional in nature. In our alignment we allow many-to-many relations. A detailed description of the algorithm appears next.

3   The alignment algorithm

We devised a metabolic network alignment algorithm that takes as input two metabolic models (GSMMs) and outputs a many-to-many alignment of their reactions/genes. The algorithm proceeds in three main phases: (i) computing the effect of knocking out each of the model's reactions/genes across randomized media; (ii) deducing the pairwise reaction/gene similarities in each of the media; and (iii) computing maximum matchings over the similarity graphs to obtain the alignment. We describe these steps in detail below. It is worth noting that while in this work we focused on evaluating alignments over reaction knockouts, all steps of the algorithm are applicable to genes as well with minor adjustments.

3.1   Similarity computation

We represent the functional (model-based) properties of each reaction in a given GSMM by a vector denoting the effect of its deletion on the species' ability to produce each of the metabolites in the model. To "delete" a reaction we constrain its flux to 0. To test the ability of the resulting model to produce a certain metabolite we apply linear programming to maximize the flux through a fictive reaction that secretes only that metabolite. We then record, per metabolite, the ratio between its maximal production rate under the knockout and the corresponding maximal rate in the wild-type (no knockout). We consider a metabolite to be affected by the deletion of a reaction if the obtained ratio is smaller than 99%, denoting some minimal effect that is not due to a numerical error. We call the set of metabolites that are affected by the deletion of a reaction, its *functional profile*.

We exclude from the analysis reactions that are considered to be *deadends*. These are defined as reactions that are unable to carry nonzero flux even in the richest growth medium (with all exchange reactions open) and, thus, do not affect the computational model [18].

To create a common ground to compare the deletion effects in the two species, only metabolites that are common to the two input models are examined (see the Conclusions section for a discussion on how to remove this limitation). The similarity between two reactions, one per species, is then defined as the Jaccard index of their functional profiles, i.e., the number of metabolites that are jointly affected by the reactions over the total number of metabolites affected by them.

3.2   Alignment computation

Given all pairwise similarities between the reactions of two models, we represent them using a weighted bipartite graph. In this graph, each side represents the reactions of a different species, and edges connect similar reactions, weighted by the corresponding similarity values (defined above). A

maximum matching algorithm is then applied to find an alignment between the reactions. Precisely, we transform the similarities into distances (with the transformation $d = 1 - s$) and apply the Hungarian method [19], yielding a collection of reaction pairs with total minimum distance, or maximum similarity.

To account for different possible matches that are equally likely, we add small random noise (a Gaussian function with parameters $\mu = 0, \sigma = 0.02$) to the computed distances and recalculate the matching. We repeat this procedure four times, and keep only the stable matches, that is, the reaction pairs that are returned in all four repetitions.

3.3 Media selection and the final alignment

A growth medium in a GSMM is characterized by the set of exchange reactions that are allowed to carry incoming (negative) flux. The alignment thus far represented the reaction similarities computed under two fixed media (one per model). Depending on the application, it is often desired to compare two metabolic models under a variety of media, exploring the metabolic spaces spanned by the different uptake constraints [20, 21]. We restrict our computations to biologically relevant media, i.e., media under which the species can plausibly grow. We define a medium to be *viable* if the biomass flux under this medium is at least 10% of the flux under the richest conditions.

To extend our comparison to different media, we repeat the similarity computation and alignment derivation in 100 random viable growth media. Each medium is randomly generated by allowing only a small fraction (25%) of the exchange reactions to carry inbound flux, in addition to enabling uptake reactions which are essential to survival, i.e., reactions whose deletion reduces the biomass flux to less than 10% of the maximal one (for all the species we tested, the same essential reactions were found for all thresholds in the range 10-50%). The benefit of working with such poor media is that when only a small part of the network is activated, the deletion of a reaction has potentially more impact due to shortage in backups. In order to activate similar regions of the two metabolic networks in each medium, we limit the pool of exchange reactions that could be enabled to reactions that exist in both species (i.e., reactions that transport the same metabolite).

Applying the basic alignment algorithm in all media, we achieve 100 different sets of reaction matches. We gather all reaction pairs from all the matchings and score each pair by the percentage of matchings it appears in. The result is a collection of aligned reaction pairs, each with its associated *confidence score*. This weighted collection comprises the final alignment which the algorithm outputs.

4 Results

We implemented the alignment algorithm in Matlab. Linear optimizations were performed using the Tomlab Cplex optimization tool. The knockout simulations were executed using grid computing over five Intel Xeon X5650 servers with 24 cores each. A complete alignment between two species over

5

100 media was generated in approximately 20 hours. We applied the algorithm to align the networks of several species pairs with varying evolutionary distances. We describe below our performance evaluation measures, the algorithms we compared to and the alignment instances we processed.

4.1 Performance evaluation

To evaluate our method and compare it to extant ones, we estimate the accuracy of the predicted reaction pairs with respect to a ground-truth set. The latter includes reaction pairs whose input and output metabolites are identical (determined based on name similarity). The evaluation is summarized in a precision-recall (PR) curve. For a given confidence threshold, *precision* is the percent of aligned pairs that are part of the ground-truth set, while *recall* is the percent of ground-truth pairs that occur in the alignment.

We compared our algorithm to two state of the art network alignment methods. The first, IsoRankN [22], is a leading approach for aligning protein-protein and other molecular interaction networks. This algorithm performs topological-based alignment and is able to exploit prior information on node similarity (e.g., sequence based). To create the input for IsoRankN, we constructed a reaction graph per model, in which a node represents a reaction and an undirected (unweighted) edge connects two reactions if one of them produces a substrate of the other (ignoring very abundant metabolites that "contribute" to that graph more than 150 edges). The prior node similarity scores were computed based on the EC number categorization. Specifically, for two reactions, one from each network, the similarity score was set as the Jaccard index between the sets of EC numbers associated with the genes that catalyze the reactions. Finally, we experimented with two values for the parameter α that balances between the prior information and network-derived match scores – $\alpha = 0.5$, which gives equal weight to both, and $\alpha = 0.99$ that emphasizes the topology-based score.

We also aimed to compare our algorithm to a recent metabolic network alignment method, CAMPways [10], whose code was readily available. However, when applying this method to any of the alignment instances described below, it did not finish processing any of them (nor produced a progress log) within a 96-hour time frame. Hence, we could not report its results.

4.2 Aligning similar models

As a basic validation of our approach we applied it to align the yeast metabolic network of [17] with itself (randomly permuting the reactions). The model contains 1577 reactions over 1228 metabolites. Out of 1024 non-deadend reactions, 690 reactions were correctly aligned (to themselves, 67% recall), with 85% precision. The PR curve is displayed in Figure 2a. Notably, the maximal recall further increases to 95% when considering in the ground-truth only the 722 reactions whose knockout affected the production of at least one metabolite in the model. This gap between the number of non-deadend reactions and the number of reactions having nonempty functional profiles can be explained by the existence of alternative pathways in the metabolic network and may indicate

that higher-order deletions may improve the recall (these numbers are summarized in Figure 1; see discussion in the Conclusions Section).

The predicted matches also contain 126 non-identical reaction pairs over 129 distinct reactions. While these pairs represent different reactions, they admit perfect similarity in some media and, thus, are indistinguishable from their real matches in those media (60% are neighbouring reactions, $p = e^{-153}$ by a hypergeometric test). For example, the sn Glycero 3 phosphocholine exchange reaction was matched to itself in roughly half of the media, affecting tens of metabolites. But in the other half of the media, it affected only two metabolites (2-Hydroxy-hexadecanal and Hexadecanal) – exactly like the phosphoethanolamine cytidyltransferase reaction – raising this couple as an alternative prediction in some of those media. Indeed, the majority (118) of the above 129 reactions also appear in the predicted matching with their real match.

We compared the performance of our approach to the IsoRankN algorithm, executed over the induced reaction graph (4610 edges). Given two reactions, their prior similarity score was computed as the overlap percentage between the sets of EC numbers corresponding to their associated genes in the model (see precise description in Section 4.1). The results for the two parameter choices (0.5 and 0.99) are shown in Figure 2b. To allow a fair comparison, these plots relate to the complete set of reactions, including also deadend ones. Evidently, our algorithm is much more precise than both competitors, producing higher-quality alignments.

As a second validation test, we aligned successive versions of the *E. coli* metabolic model that were published by the same lab, iAF1260 [23] (2382 reactions, 2159 of them non-deadend; 1668 metabolites) and iJO1366 [24] (2583 reactions, 2351 non-deadend; 1805 metabolites). These models have 2114 (non-deadend) reaction pairs in common, identified by comparing their internal codes. Notably, with a precision of 90% (corresponding to a matching score threshold of 4%), half of the ground-truth pairs are discovered by our algorithm (Figure 2a). The maximal recall achieved was nearly 60%, corresponding to 1235 true positive pairs. Of the remaining 879 reaction pairs that were not discovered, 407 do not have an observable effect in the model when knocked out. Figure 2c demonstrates a clear advantage of our algorithm over the topology-based IsoRankN. This result also shows that our approach is robust to the inherent incompleteness of metabolic networks, as the aligned models have a large core of common reactions but also considerable differences (Figure 1).

4.3   Aligning the yeast and human models

To test our algorithm on distant species, we applied it to align GSMMs of yeast and human. We used the yeast iMM904 model [17] (1577 reactions, 1024 non-deadend; 1228 metabolites) and the human Recon1 model [25] (3788 reactions, 2515 non-deadend; 2766 metabolites). Our alignment algorithm was limited to 657 yeast reactions and 766 human reactions whose knockout affected the production of at least one of the 663 metabolites that are common to the two models (Figure 1; common metabolites were identified based on name similarity). Figure 2a displays the PR curve of the predictions with respect to a ground-truth set of 421 non-deadend reaction pairs. The maximal

measure/model	yeast	human	ecoli iAF1260	ecoli iJO1366
#reactions	1577	3788	2382	2583
#non-deadend reactions	1024	2515	2159	2351
#reactions with nonempty functional profile	722	1072	1860	2337
#reactions with nonempty functional profile over the shared metabolites	657	766	1707	2323

Fig. 1. Reaction and knockout statistics. This table displays reaction statistics for each tested model, including the number of reactions (all or non-deadend only) and the number of reactions having some observable perturbation effect, (i) when considering the full functional profile of a reaction, and (ii) when considering only the functional profile's metabolites that are shared between the aligned species.

recall obtained by our algorithm is 24%, corresponding to a low precision of 11%. For a precision of 60%, the recall mildly decreases to 15%. The relatively inferior performance compared to the alignments computed in the previous section can be related to the small number of human reactions that have nonempty functional profiles and to the small number of shared metabolites (Figure 1). Indeed, when considering only ground-truth reaction pairs having nonempty functional profiles, the maximal recall increases to 36%; further, at 60% precision the recall increases to 22%. In comparison to IsoRankN (Figure 2d), the PR plot of our algorithm dominates that of the topology-based variant (α=0.99). The second variant, which balances sequence and topology information ($alpha$=0.5) has higher recall for low precision values (up to 45%), but its recall drops to almost zero for higher precision values (above 55%), while our algorithm maintains relatively stable recall values even in this range (15% recall at 60% precision).

To systematically evaluate the quality of the predicted matches that are not in the ground-truth alignment, we calculated the functional similarity between the gene sets catalyzing the reactions in each of these pairs. We defined the functional similarity between two genes as the maximum semantic similarity [26] between their annotated gene ontology (GO) terms. We extended this definition to gene sets by defining their similarity as the maximum value obtained over any two members of these sets. Using these definitions, we computed the average functional similarity between 136 predicted reaction pairs with a confidence score of at least 2% (and at least one associated gene in both models). The average score was 4.45, corresponding to an empirical p-value of 0.005 (permutation test).

To demonstrate the utility of our approach in identifying nonobvious orthology relations, we first computed how many yeast or human reactions have long-range impacts. To this end, we counted for each perturbed reaction the ratio between the number of affected metabolites that are
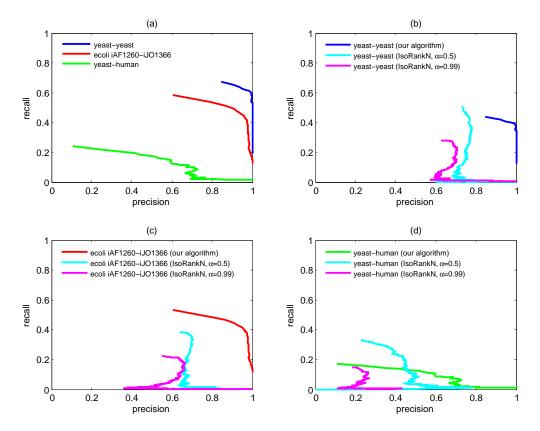
Fig. 2. Performance evaluation and comparison to IsoRankN. (a) Precision-recall curves for our functional alignment algorithm over the three case studies: yeast vs. yeast, E. coli iAF1260 vs. E. coli iJO1366, and yeast vs. human. The recall is measured with respect to the number of non-deadend ground-truth reaction pairs between the metabolic networks. (b)-(d) These subplots compare the performance of our algorithm and the IsoRankN algorithm over the three case studies. Here the complete ground-truth set is considered, including matches of deadend reactions.

not part of the reaction and the total number of metabolites affected by that reaction. Figure 3 shows that the majority of the reactions that have nonempty functional profiles also have long-range effects in the network. To exploit this property for identifying nonobvious disease models, we looked for a human reaction which (i) is associated with a disease from the Online Mendelian Inheritance in Man database (OMIM, [27]), and (ii) has a ground-truth yeast match that is distant in the reaction graph from the suggested match in our alignment. As an example, our method functionally aligned the human reaction catalyzed by the enzyme cystathionine-β-synthase (CBS) with the exchange reaction that imports sulfate into the yeast cell. CBS deficiency in humans leads to a severe disease due to disruption of sulfur metabolism, homocystinuria, in which the inability of CBS to convert homocysteine leads to its excessive accumulation in the blood and urine. CBS is part of the transsulfuration pathway, which FBA analysis suggests is used for homocysteine degradation to ultimately (and indirectly) increase the availability of sulfur to the cell. It is thus plausible that some of the phenotypes of CBS deficiency can be modeled by blocking sulfate uptake.

Our algorithm detected this functional alignment even though yeast has a close sequence ortholog to the human CBS enzyme and the reaction catalyzed by that enzyme is topologically distant from sulfate exchange.

Moreover, this coupling correctly reflects pathologies associated with CBS deficiency in humans: the alignment is due to the similar effects incurred by the deletion of the yeast/human reaction over a set of eight sulfur-containing metabolites, all of them derived from cysteine or glutathione; the latter is an antioxidant with key cellular functions, which is thought to be produced in considerable quantities by the transsulfuration pathway through the intermediate cysteine [28]. As this process depends on CBS, glutathione deficiency may partly account for homocystinuria's symptoms (cf. [29]).
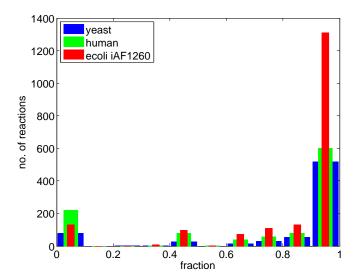

Fig. 3. Long-range knockout effects. For different fraction values t (in bins of size 0.1), shown is the number of reactions for which a fraction of t of the affected metabolites following a knockout of the reaction are not part of it.

5   Conclusions

We presented a model-based alignment strategy to align metabolic networks. Our strategy employs GSMMs to compute the global functional implications of metabolic reactions, thereby aligning them. We applied our strategy to align different metabolic models, demonstrating its utility over topological approaches. Importantly, our method is applicable to current large scale metabolic models.

One limitation of our alignment apporach is the need to restrict attention to metabolites that are shared between the two models being compared. Such metabolites serve as fixed "anchors" according to which the functional profiles (deletion effects) of different reactions are compared.

To circumvent this limitation, one can define the similarity between reactions based on anchors that are not known in advance. The basic idea is to compare the deletion effects by the set of matched reaction pairs. Initially, the reactions are randomly paired. This pairing allows comparing two reactions, one per species, by their functional profiles over the paired reactions. The pairing can be then scored based on the computed similarities. In subsequent iterations, the pairing is greedily changed to optimize its score, until a local maximum is attained.

Another limitation of our approach is its applicability to reactions whose deletion has some observable effect. The majority of the reactions do not exhibit an effect when deleted in isolation, suggesting that better results can be obtained if extending the functional profiles to knockouts (deletions) of higher order.

References

1. Sharan, R., Suthram, S., Kelley, R.M., Kuhn, T., McCuine, S., Uetz, P., Sittler, T., Karp, R.M., Ideker, T.: Conserved patterns of protein interaction in multiple species. Proceedings of the National Academy of Sciences of the United States of America **102**(6) (2005) 1974–9
2. Ogata, H., Fujibuchi, W., Goto, S., Kanehisa, M.: A heuristic graph comparison algorithm and its application to detect functionally related enzyme clusters. Nucleic acids research **28**(20) (2000) 4021–4028
3. Kelley, B.P., Sharan, R., Karp, R.M., Sittler, T., Root, D.E., Stockwell, B.R., Ideker, T.: Conserved pathways within bacteria and yeast as revealed by global protein network alignment. Proceedings of the National Academy of Sciences of the United States of America **100**(20) (2003) 11394–9
4. Flannick, J., Novak, A., Srinivasan, B.S., McAdams, H.H., Batzoglou, S.: Graemlin: general and robust alignment of multiple large interaction networks. Genome research **16**(9) (2006) 1169–81
5. Zhenping, L., Zhang, S., Wang, Y., Zhang, X.S., Chen, L.: Alignment of molecular networks by integer quadratic programming. Bioinformatics (Oxford, England) **23**(13) (2007) 1631–9
6. Singh, R., Xu, J., Berger, B.: Global alignment of multiple protein interaction networks with application to functional orthology detection. Proceedings of the National Academy of Sciences of the United States of America **105**(35) (2008) 12763–8
7. Pinter, R.Y., Rokhlenko, O., Yeger-lotem, E., Ziv-ukelson, M.: Alignment of metabolic pathways. Bioinformatics **21**(16) (2005) 3401–3408
8. Li, Y., de Ridder, D., de Groot, M.J., Reinders, M.J.: Metabolic pathway alignment (M-Pal) reveals diversity and alternatives in conserved networks. In: Advances in Bioinformatics and Computational Biology. Volume 6., Imperial College Press (2008) 273–285
9. Ay, F., Kellis, M., Kahveci, T.: SubMAP: aligning metabolic pathways with subnetwork mappings. Journal of computational biology **18**(3) (2011) 219–35
10. Abaka, G., Biyikoglu, T., Erten, C.: CAMPways: constrained alignment framework for the comparative analysis of a pair of metabolic pathways. Bioinformatics **29**(13) (2013) i145–53
11. Baldan, P., Cocco, N., Simeoni, M.: Comparison of metabolic pathways by considering potential fluxes. In: BioPPN2012 - 3rd International Workshop on Biological Processes and Petri Nets. (2012) 2–17

12. Schuster, S., Hilgetag, C.: On elementary flux modes in biochemical reaction systems at steady state. Journal of Biological Systems **2**(2) (1994) 165–182

13. Ay, F., Kahveci, T.: Functional similarities of reaction sets in metabolic pathways. Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology - BCB '10 (2010) 102–111

14. Orth, J.D., Thiele, I., Palsson, B.O.: What is flux balance analysis? Nature biotechnology **28**(3) (2010) 245–8

15. Feist, A.M., Palsson, B.O.: The biomass objective function. Current opinion in microbiology **13**(3) (2010) 344–9

16. Mahadevan, R., Schilling, C.: The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab. Eng. **5**(4) (2003) 264–276

17. Mo, M.L., Palsson, B.O., Herrgård, M.J.: Connecting extracellular metabolomic measurements to intracellular flux states in yeast. BMC systems biology **3** (2009) 37

18. Burgard, A.P., Nikolaev, E.V., Schilling, C.H., Maranas, C.D.: Flux coupling analysis of genome-scale metabolic network reconstructions. Genome Res. **14**(2) (2004) 301–312

19. Munkres, J.: Algorithms for the assignment and transportation problems. Journal of the Society for Industrial and Applied Mathematics **5**(1) (1957) 32–38

20. Bilu, Y., Shlomi, T., Barkai, N., Ruppin, E.: Conservation of expression and sequence of metabolic genes is reflected by activity across metabolic states. PLoS Comput Biol **2**(8) (2006) e106

21. Guimerà, R., Sales-Pardo, M., Amaral, L.A.N.: A network-based method for target selection in metabolic networks. Bioinformatics **23**(13) (2007) 1616–22

22. Liao, C.S., Lu, K., Baym, M., Singh, R., Berger, B.: IsoRankN: spectral methods for global alignment of multiple protein networks. Bioinformatics **25**(12) (2009) i253–8

23. Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., Palsson, B.O.: A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Molecular systems biology **3**(121) (2007) 121

24. Orth, J.D., Conrad, T.M., Na, J., Lerman, J.a., Nam, H., Feist, A.M., Palsson, B.O.: A comprehensive genome-scale reconstruction of Escherichia coli metabolism–2011. Molecular systems biology **7**(535) (2011) 535

25. Duarte, N.C., Becker, S.a., Jamshidi, N., Thiele, I., Mo, M.L., Vo, T.D., Srivas, R., Palsson, B.O.: Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proceedings of the National Academy of Sciences of the United States of America **104**(6) (2007) 1777–82

26. Resnik, P.: Semantic similarity in a taxonomy: an information-based measure and its application to problems of ambiguity in natural language. Journal of Artificial Intelligence Research **11** (1999) 95–130

27. Amberger, J., Bocchini, C.A., Scott, A.F., Hamosh, A.: McKusick's Online Mendelian Inheritance in Man (OMIM). Nucleic acids research **37** (2009) D793–6

28. Mosharov, E., Cranford, M.R., Banerjee, R.: The quantitatively important relationship between homocysteine metabolism and glutathione synthesis by the transsulfuration pathway and its regulation by redox changes. Biochemistry **39**(42) (2000) 13005–13011

29. Robert, K., Nehmé, J., Bourdon, E., Pivert, G., Friguet, B., Delcayre, C., Delabar, J., Janel, N.: Cystathionine β synthase deficiency promotes oxidative stress, fibrosis, and steatosis in mice liver. Gastroenterology **128**(5) (2005) 1405–1415