

Network motifs provide signatures that characterize metabolism†

Cite this: *Mol. Biosyst.*, 2013, **9**, 352

Erin R. Shellman,^{ab} Charles F. Burant^a and Santiago Schnell^{*cb}

Motifs are repeating patterns that determine the local properties of networks. In this work, we characterized all 3-node motifs using enzyme commission numbers of the International Union of Biochemistry and Molecular Biology to show that motif abundance is related to biochemical function. Further, we present a comparative analysis of motif distributions in the metabolic networks of 21 species across six kingdoms of life. We found the distribution of motif abundances to be similar between species, but unique across cellular organelles. Finally, we show that motifs are able to capture inter-species differences in metabolic networks and that molecular differences between some biological species are reflected by the distribution of motif abundances in metabolic networks.

Received 27th August 2012,
Accepted 18th December 2012

DOI: 10.1039/c2mb25346a

www.rsc.org/molecularbiosystems

Introduction

Life can be studied at many different strata ranging from the molecular level to the ecosystem. Regardless of the stratum, a fundamental characteristic of life is a high degree of order which is divided into hierarchical levels of organization and function.¹ Because metabolism is a fundamental process shared among all organisms, it influences multiple strata of biological function. At the molecular level, proteins and other macromolecules are activated by metabolites. Likewise, changes in the metabolic function at the organism level can provide adaptation to novel environments.² Understanding the emergent, organizational properties of metabolism is one way to unravel molecular evolution,³ and is thus a crucial goal in the field of systems biology.⁴

As a consequence of advances in the field of molecular biology, particularly in sequencing technology, it is now possible to assemble genome-level metabolic networks by integrating known biochemical pathways with genomic annotation.⁵ These large biochemical networks are often referred to as metabolic network reconstructions.⁶ Many characteristics of biological networks have been described as a result of the availability of network reconstructions, and commonalities among transcriptional, signalling and metabolic networks have emerged.^{1,7–9} For example,

biological networks share patterns of enrichment (network motifs) that are unlike those of engineered networks.¹⁰ A network motif (or just “motif”) is a repeated pattern or subgraph that is over-represented in a network compared to its expected abundance in a collection of random graphs.^{11,12} Motifs are of interest in systems biology because their patterns of enrichment may determine the dynamical properties of whole networks.¹³ In addition, motifs provide a reduced, simplified framework in which to describe the organization of large networks without losing resolution.^{14–17} Vázquez *et al.*¹⁸ demonstrated that motif abundances could be used to predict organizational properties in the transcription and metabolic networks of *E. coli* and *S. cerevisiae*, and likewise that the global organizational properties could predict motif abundances.

Since organizational properties of networks can be described using motif distributions (the collection of relative abundances of all motifs), we applied the methods of motif mining and analysis to characterize the metabolic networks of 21 species. Most biological reactions require enzymes for catalysis and as a result, the collection of enzymes active within a cellular compartment partially characterize the range of biochemical function within the compartment.

In this work we characterized all 3-node motifs using Enzyme Commission (EC) numbers from the International Union of Biochemistry and Molecular Biology to show that, in metabolism, motif abundance is associated with a specific collection of biochemical reactions. Further, we present a comparative analysis of the distributions of 3-node motifs in the metabolic networks of 21 organisms^{19–39} by compartmentalizing the metabolism in the cellular organelles: cytosol, endoplasmic reticulum (ER), Golgi, mitochondrion, nucleus

^a Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, USA

^b Department of Computational Medicine & Bioinformatics, University of Michigan Medical School, Ann Arbor, USA

^c Department of Molecular & Integrative Physiology, University of Michigan Medical School, Ann Arbor, USA. E-mail: schnells@umich.edu

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c2mb25346a

and peroxisome. Using compartmentalized metabolic networks enabled us to test whether the motif distribution is unique for each structure in the hierarchical organization of the cell. We found that each organelle has a unique metabolic signature. Finally, we show that motifs are able to capture known metabolic differences between some biological species.

Results

Prior to analysis we mined for motifs in the metabolic networks of 21 species (see Methods). Motifs are numbered as previously presented in the literature¹⁰ and are roughly in order of increasing edge density. For example, motif 1 (concurrent reactions) has only two, non-reversible edges, while motif 13 (reversible cycle) has six edges (or three fully reversible edges). Motif names briefly describe the biochemical relationship between the three nodes, and motif names and numbers will be used interchangeably.

Motifs can be characterized by their enzyme associations

The first two digits of EC numbers in the networks yielded a total of 47 enzyme classes. For each enzyme class, we calculated the proportion of reactions associated with each motif and found that each of the 13 motifs was associated with a distinct catalog of enzymes (Fig. 1). Note that the motifs described in this work are derived from substrate graphs; they describe reactive associations rather than mechanisms of enzyme action. Therefore, double edge motifs do not necessarily represent reversible enzyme catalyzed reactions.

The number and type of enzymes associated with the concurrent, trapping and consecutive reactions and feed-forward reaction motifs (motifs 1–3 and 7) was wide-reaching. In motifs one through three, 43 of the 47 total enzyme classes had non-zero proportions. This result implies that these motifs represent a breadth of enzymatic functions, perhaps serving as intermediates between other motifs. The feed-forward reaction motif (motif 7) had EC proportions similar to those of motifs 1–3, but was distinguished by increased proportions of EC 2.4 (Glycosyltransferases) and EC 6.3 (forming carbon–nitrogen bonds).

Consecutive reactions with reversible step (motif 4), trapping reactions with reversible step (motif 5) and reversible consecutive reactions (motif 6) had EC distributions similar to one another. Enzymes that were key in characterizing these motifs include 1.1 (acting on CH–OH group of donors), 1.6 (acting on NADH or NADPH), 1.8 (acting on a sulfur group of donors) and 5.4 (intramolecular transferases). Although many of these enzymes are modestly represented, their presence remains a relevant characteristic. For example, EC 1.8 constitutes 1% of all enzymes associated with motifs 4–6, but was rarely found in all other motifs.

The EC distributions of concurrent reaction with exchange (motif 9) and trapping reaction with exchange (motif 10) vary in key enzymes with respect to the other 11 motifs, but are similar to one another. A distinguishing characteristic of concurrent reaction with exchange is the high level of EC 2.7 (transferring phosphorus-containing groups) which comprises 20% of its total collection of enzymes, and nearly double that of motif 10.

This is also true for ECs 1.17 (acting on CH or CH₂ group) and 3.6 (acting on acid anhydrides) which are doubled in the concurrent reaction with exchange motif *versus* the trapping reaction with exchange motif. The enzyme that distinguishes motif 10 from motif 9 and all others is EC 6.1 (forming carbon–oxygen bonds) which comprises 9% of the enzymes associated with motif 10 and is twice to ten times the amount seen in all other motifs.

One way cycle with one reversible step (motif 11) and one way cycle with two reversible steps (motif 12) were similar in their enzyme proportions. Enzymes that distinguish these two from all other motifs are high proportions of glycotransferases (EC 2.4) and enzymes catalyzing carbon–oxygen bonds (EC 6.2). They differ primarily in the amounts of EC 2.3 (acyltransferase) which is four times greater in motif 12 compared to motif 11.

Motifs closed cycle (motif 8) and reversible cycle (motif 13) are particularly interesting because their EC distributions are sparser than the other motifs, suggesting a narrower range of enzymatic function. The closed cycle motif had non-zero proportions in just 19 of the 47 enzyme classes and reversible cycle had non-zero proportions in only 17 of the 47. These two motifs also displayed enzymes distributions that were unlike all other motifs. The closed cycle (motif 8) shows proportions of ECs 3.5 (acting on carbon–nitrogen bonds, other than peptide bonds) and 1.7 (acting on other nitrogenous compounds as donors) that are at least twice the amount in all other motifs. Motif 13 is lacking any glycosyltransferases (EC 2.4) which are ubiquitous in every other motif.

It is worth mentioning some motifs are rarer than others in metabolic networks. For example, motif 13 is much less abundant than motif 1. However, this does not necessarily imply that motif 13 has narrower range of enzymatic function than motif 1. It is not possible to draw a direct relationship between the proportions of motifs shown in Fig. 1 and the range of enzymatic function of a motif.

Enzyme association mirrors motif structure

Using the proportions of enzyme classes associated with each motif (Fig. 1), we calculated a pairwise distance metric to quantify the level of similarity of enzyme spectrum between motifs. In agreement with the previous section, we found that motifs with similar structural features have similar proportions of enzyme classes in metabolic networks (Fig. 2). The feed-forward structures (motifs 1–3 and 7) fall within their own cluster with motifs 1 and 2 showing more similarity with each other and less similarity with motifs 3 and 7. Motifs 4–6 cluster together, but the motifs that share the structural property of one reversible edge and one non-reversible edge (4 and 5) cluster more closely to each other than to motif 6. This finding shows that the addition of one edge to a motif will create a motif comprised of a different spectrum of enzymes (this is seen also in the clustering of motifs 3 and 7).

The findings depicted in Fig. 2 suggest that motifs have chemical function associated with the spectrum of enzymes of reactions in which they participate. The spectrum of enzyme is quantified with the EC numbers. Further, the similarity of the EC distributions is related to the structural features of the motifs such

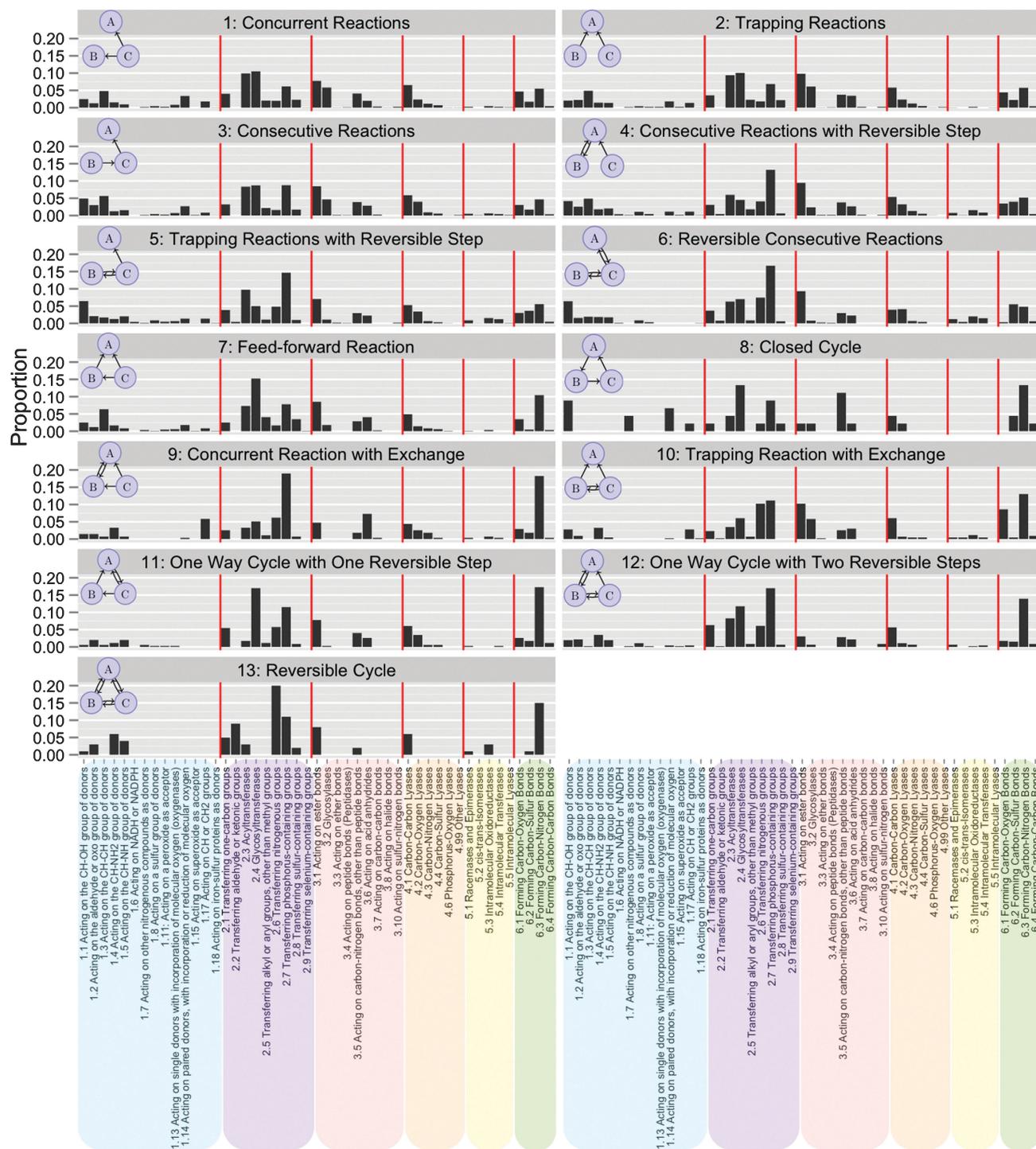


Fig. 1 Proportions of two-digit Enzyme Commission (EC) numbers of the International Union of Biochemistry and Molecular Biology associated with each motif. EC data were obtained from the metabolic network reconstructions of *E. coli*, *H. sapiens*, *M. barkeri* and *S. cerevisiae*. Red vertical lines separate the six classes of EC number: oxidoreductases, transferases, hydrolases, lyases, isomerases and ligases.

that motifs which are structurally similar also share similarities in the composition of enzymes associated to those motifs.

Each cellular organelle displays a unique motif distribution

We used fully compartmentalized metabolic networks from 21 organisms to describe the average motif distribution for

each organelle. Each of the six organelles displayed a unique pattern of motif enrichment and suppression, determined using normalized z -scores calculated by comparing each metabolic network with a collection of random graphs (Fig. 3). Motifs with normalized z -scores greater than zero were considered enriched and motifs with normalized z -scores less than zero

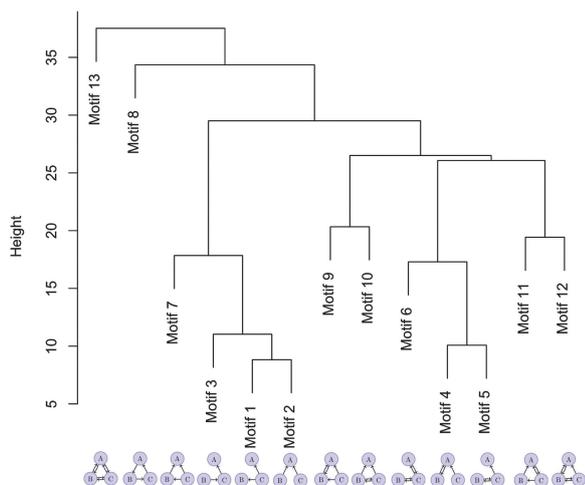


Fig. 2 Dendrogram showing the distance between the proportions of EC numbers each of the 13 motifs.

were considered suppressed. Motifs with z -scores equal to zero have abundances equal to what could be expected at random (null hypothesis). Confidence intervals not containing the null, $z = 0$, are statistically significantly enriched or suppressed at 95% confidence (see Methods).

In the cytosolic compartment, 11 of the 13 motifs achieved statistical significance with the exception of trapping reactions with reversible step (motif 5) and reversible consecutive reactions (motif 6). The tightness of the confidence intervals indicates relatively small variance between organisms and

suggests that the local structure in the cytosol is well conserved across all kingdoms of life in our sample.

The ER had only two motifs that reached statistical significance, concurrent reaction with exchange (motif 9) and trapping reaction with exchange (motif 10), both of which were suppressed. This is due primarily to inter-species variation in motif enrichment as seen from the points in Fig. 3.

The Golgi showed enrichment in only one motif, the feed-forward reaction (motif 7), and suppression or absence in all others. This profile suggests that, unlike the cytosol, the Golgi performs a narrow set of enzymatic functions, for example protein glycosylation, and therefore one type of motif is sufficient.

The nuclear motif distributions also displayed significant enrichment of the feed-forward reaction motif (motif 7) and had high levels of inter-species variation (as seen from the points in Fig. 3).

An intriguing finding is the similarity of the cytosol, mitochondrion and peroxisome motif distributions. The profiles are remarkably similar with motifs 1 to 7 displaying the same pattern of enrichment and suppression (though not the same pattern of statistical significance) among all three organelles.

It is notable that the trapping reaction (motif 2) and consecutive reaction (motif 3) motifs are enriched in cytosol, mitochondrion and peroxisome but suppressed or non-significant in the ER, Golgi and nucleus. Recall, that these motifs were associated with a wide range of enzyme classes and had non-zero proportions for nearly all 47 enzyme classes. Because the cytosol, mitochondria and peroxisomes contain a more varied and complex set of metabolic reactions and roles, it is reasonable that we see this pattern of enrichment.

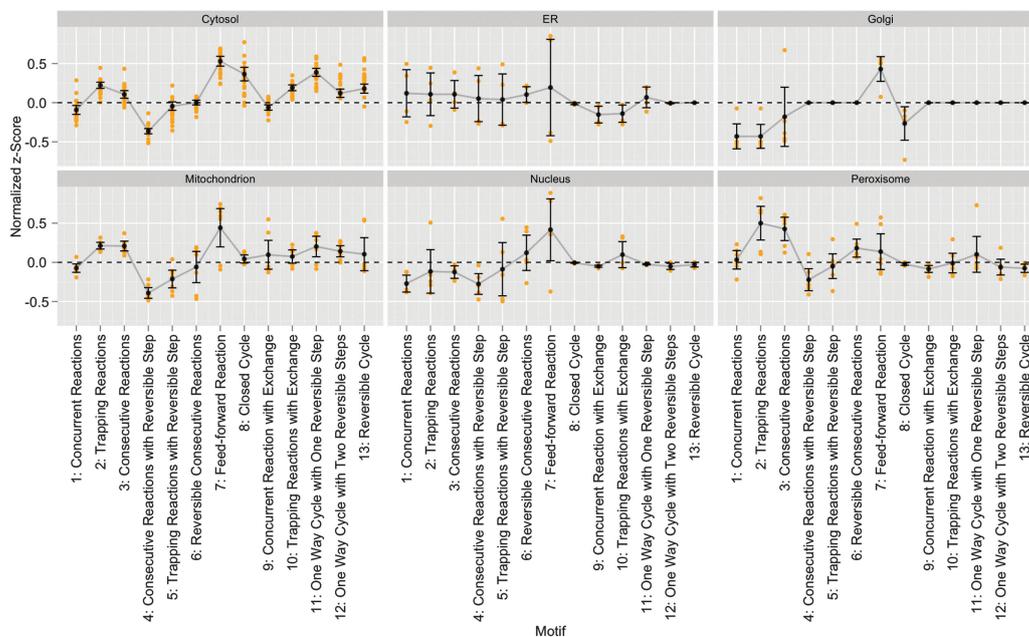


Fig. 3 Average significance profiles of 3-node motifs by organelle. Black points (connected for visualization of relative abundances) indicate the average normalized z -score for each motif. Orange dots show the scatter of the normalized z -scores corresponding to each of the 21 organisms. Standard errors of the means were estimated from 1000 bootstrap samples of the normalized z -score and used to calculate 95% confidence intervals. Confidence intervals not containing the null, $z = 0$, are statistically significantly enriched or suppressed compared to the expected abundance in 1000 random graphs.

Differences in evolutionary origin and metabolic functionality are reflected in motif enrichment

As we have seen in the previous sections, metabolic network motifs are associated with a wide array of enzymatic functions. Although it is difficult to fully predict an organelle metabolism from the network motifs, the mitochondrial motif distribution provides an interesting example to evaluate variation in motif enrichment between organisms. The mitochondrial data included only seven species because the prokaryotes in our sample do not contain mitochondria. There is very little variation in mitochondrial motif distributions between the two species in Animalia, *H. sapiens* and *M. musculus* (Fig. 4). Likewise, the two Fungi, *S. cerevisiae* and *P. pastoris*, show identical distributions to one another, and to those of Animalia. The two plants, *A. thaliana* and *Z. mays*, have motif distributions unlike those of any of the other kingdoms, showing enrichment of both concurrent reaction with exchange (motif 9) and reversible cycle (motif 13) even while those motifs are primarily suppressed in the other kingdoms. Similarly, reversible consecutive reactions (motif 6), feed-forward reaction (motif 7), and one way cycle with one reversible step (motif 11) are suppressed in plants but primarily enriched in other kingdoms. The motif distribution of the protist *C. reinhardtii* is somewhat of a hybrid of the plant and the animal distributions. These results show that gradual divergence in metabolism between organisms is reflected in the discrepancies in the motifs distributions.

We should expect some variation in motif distributions between plants and other organisms because the evolutionary origin of plant mitochondria differs markedly from that of bacteria, fungi and animalia.⁴⁰

The reversible cycle (motif 13) is enriched in plant mitochondria, but suppressed in animals, fungi and protista. In the previous section, the reversible cycle motif was found to be characterized by the transferral of aldehyde or ketonic groups (EC 2.2) and intramolecular oxidoreductases (EC 5.3). Oxidoreductases are a class of enzymes that catalyze the transfer of electrons from one molecule to another, and they

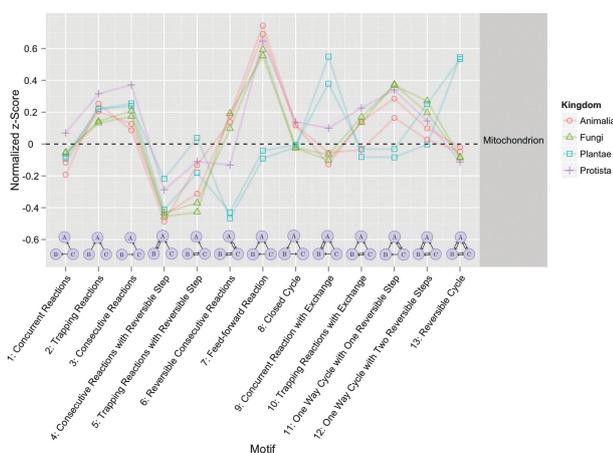


Fig. 4 Mitochondrial significance profile. Each line corresponds to an organism, and the line color denotes the kingdom of life to which the organism belongs.

are common in the pathways of glycolysis and gluconeogenesis. In most organisms, the pathways of glycolysis and gluconeogenesis occur in the cytoplasm, however in plants these pathways are contained within mitochondria.⁴¹ Approximately 10% of the reactions of the reversible cycle motif are considered part of glycolysis/gluconeogenesis, and this 10% constitutes the largest proportion for that motif (Fig. 5). The remaining pathways of the reversible cycle motif take place primarily outside of the mitochondria, which could explain the suppression of reversible cycle in all other kingdoms. Again we can see how the differences in metabolic function between organisms are reflected in the motif distribution.

The reversible consecutive reactions (motif 6) is suppressed in plants while enriched in animals and fungi mitochondrion. Similarly, the concurrent reaction with exchange (motif 9) is enriched in plants while suppressed in animals and fungi. Interestingly *C. reinhardtii*, a photosynthetic alga, follows the same pattern of suppression and enrichment as plants. Once more this illustrates that evolutionary divergences in photosynthetic organisms are reflected in these two motifs. Both motifs are associated with biochemical reactions in alternate carbon metabolism pathways (Fig. 5), which vary between plants and animals due to the presence of chloroplasts in photosynthetic organisms. Chloroplasts are a crucial component of oxidation–reduction reactions, and provide the cell with the ability to synthesize oxygen and carbohydrates from energy derived from sunlight.⁴²

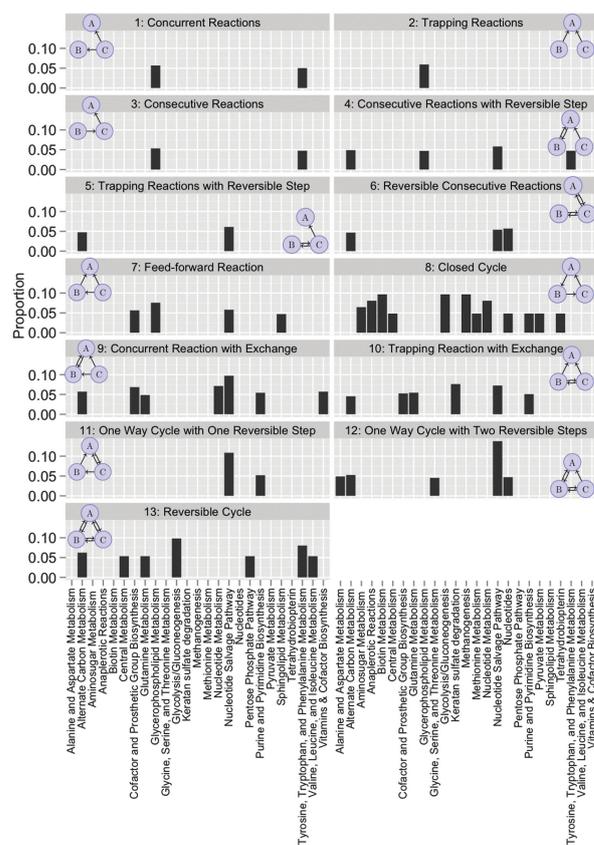


Fig. 5 Barplots showing proportions of pathway participation for each motif. A minimum threshold of 5% was used for clarity. See Fig. S2 (E51) for all pathways.

One of the many functions of mitochondria is fatty acid oxidation, which occurs less in plants than in other organisms.⁴³ The feed-forward reaction (motif 7) is associated with fatty acid oxidation through the EC numbers 1.3, which refers to various types of oxidases and hydrogenases used in the beginning steps of fatty acid oxidation. Also, EC 1.1.1.35 and 1.1.1.211 which are dehydrogenases and 2.3.1.16 an acyltransferase involved in the conversion of coenzymes to acetyl-CoA. Because fatty acid oxidation is relatively rare in plants, we expect less enrichment of the feed-forward reaction motif in plants, which was the case here.

Discussion

We characterized network motifs in terms of their enzyme associativities, and estimated motif abundances in the metabolic networks of 21 organisms and 6 organelles. Evaluating the properties of metabolic networks is only as useful as the reconstructions are valid. Many reconstructions are built using previous versions as starting points and thus perpetuate errors and biases that may have been present in previous incarnations of the networks. However, we expect that those biases are consistent across most reconstructions because of their high degree of relatedness.

There is also a disconnect between the ever-growing number of fully sequenced genomes and the number of validated, usable network reconstructions to accompany these genomes. Currently, network reconstruction is massively time-consuming and largely done *via* manual curation. As a result of the time-intensive process of creating metabolic network reconstructions, our sample contained relatively few eukaryotes. Despite this, we expect that while the ensemble of enzymes and pathways associated with each motif will likely change as more reconstructions become available, motifs will still be enzyme-specific.

Notwithstanding the previously mentioned limitations, the findings presented here improve on previous work^{12,44} on metabolic network motifs in two key ways. First, our analyses were restricted to include only manually-curated metabolic network reconstructions. We conducted a small analysis comparing the motif distributions of *in silico* versus manually generated reconstructions and found that *in silico* reconstructions systematically underestimate the number of reversible reactions in metabolic networks (unpublished data). Underestimation of reversibility results in underestimation of motifs with reversible edges (motifs 9–13) and overestimation of simpler motifs (motifs 1–3). Second, as a consequence of analyzing high-quality reconstructions that were fully compartmentalized, we were able to present motif distributions for six distinct Eukaryotic organelles, which to our knowledge, is a novel contribution.

In this work we also show that the feed-forward structural motifs (motifs 1–3 and 7) displayed wide-ranging enzymatic associations (Fig. 1). We proposed that these motifs might be intermediaries connecting motifs of greater complexity (in terms of edge connectivity) into modules. In networks, modules are semi-autonomous units that can function primarily independently. It has been demonstrated in previous work^{18,45,46} that motifs

aggregate into functional modules in metabolic networks. Kashtan *et al.*⁸ showed that network modularity and motif aggregation evolve spontaneously in *in silico* networks exposed to external perturbations, and that the consecutive reaction (motif 3) and feed-forward reaction (motif 7) in particular aggregate in modules. We found that in metabolic networks, consecutive reaction and feed-forward reaction motifs have a breadth of enzyme associativity, perhaps because they aggregate within many metabolic modules. This is also supported by the motif enrichment levels observed in Fig. 3. Besides the ER and peroxisome, all organelles showed enrichment of the feed-forward reaction motif (motif 7) and the cytosol, mitochondrion and peroxisome showed enrichment of the consecutive reaction motif (motif 3), suggesting high abundances of these motif structures which may contribute to the network modularity and perhaps the benefits conferred by that feature such as stability and robustness.⁴⁷

In contrast with motifs 1–3 and 7, motifs 8 and 13 displayed the narrowest range of enzymatic associativity with non-zero proportions in only 36–40% of all enzyme classes (Fig. 1). Interestingly these cyclic motifs were only significantly enriched in the cytosol and no other organelle. Cyclic motifs like 8 and 13 have been shown to have dynamically unstable properties in biological networks (transcription, signal transduction and neuronal signaling)⁴⁸ and to be unreliable in the context of information processing.⁴⁹ This could explain the lack of enrichment of these motifs in networks where metabolites are used as signaling molecules.

Conclusions

In this work we have shown that, in metabolic networks, motifs can be characterized by their enzymatic associations. Further we found that similarities in enzyme class proportions were explained by similarity in the structural features of the motifs. We also showed that cellular organelles displayed motif distributions that are distinct from one another and likely reflect differences in metabolic function in the cell.

Enzyme Commission of the International Union of Biochemistry and Molecular Biology numbers allowed us to uncover motif specificity at the enzymatic level, and pathway data allowed us to supplement the chemical information within a functional context. In this way we have been able to make inferences about higher-level biochemical function based on motifs distribution of metabolic networks. Our analysis suggests that metabolic differences between organisms are reflected network motifs.

Methods

Metabolic networks

Metabolic networks were built from reconstructions of overall reactions between metabolites. Criteria for inclusion in this study was (1) that the reconstruction was curated in the Systems Biology Mark-up Language (SBML) and (2) readable by the COBRA toolbox in Matlab.⁵⁰ Neither COBRA nor Matlab were used for analysis, but this criterion insured that the reconstructions

were curated using similar protocols, adequately formatted, and vetted for typographical errors. Once each reconstruction was read into Matlab, we exported relevant data as plain text files to use for motif mining. Specifically, we extracted the stoichiometric matrix, the reaction and metabolite names, a dummy variable indicating the reversibility of each reaction and the subsystem to which the reaction belonged (e.g. “Folate Biosynthesis,” “TCA Cycle,” “Salvage Pathway of ATP”).

Cofactors such as water (H₂O), protons (H⁺), ammonia (NH₃), CO₂, SO₄, thioredoxin (oxidized and reduced form), organic phosphate (Pi) and pyrophosphate (PPi), the metabolites acetyl coenzyme A (acetyl-CoA), adenosine triphosphate (ATP), adenosine diphosphate (ADP), nicotinamide adenine dinucleotide (NAD), as well as its phosphorylated and reduced forms, were omitted in our network analysis. Transport reactions across membranes were also removed, because they are not of interest in investigating the function of metabolic pathways.

In this work we analyzed 21 metabolic network reconstructions from 21 species and 6 kingdoms of life. Metabolic networks were divided according to 6 distinct physiological organelles. See Table S1 in the ESI† for a complete list of the organisms used in our analysis, including the sources for each network reconstruction. The stoichiometric matrices and metabolic networks used in our study are available in the upon request, and all are freely downloadable.

Graph construction

We generated a list of overall reactions from the stoichiometric matrices of our metabolic network reconstructions. We used substrate graphs to represent the metabolic networks. Substrate graphs represent associativity of nodes, rather than mechanistic relationships between the nodes. In our case, they represent the interactions between chemical species, rather than elementary steps describing the reaction mechanism. For example, the Concurrent Reaction motif () represents two possible patterns of interactions: (C → A and C → B) or C → A + B. In the graphical representation we included reversible reactions when they were present in the stoichiometric matrix. We generated network graph reconstruction as FANMOD input files, following the FANMOD specifications.⁵¹ FANMOD is a software tool for fast network motif detection in graph representations of networks.

Network motif analysis

To estimate the abundance of motifs in our metabolic networks, we employed FANMOD,⁵² which identifies motifs present in a network by enumerating every possible motif combination. FANMOD determines enrichment by generating a collection of random networks of equal node and edge size for comparison. If the motif appears more often in the metabolic network than in the random networks, it is considered enriched. From the random graphs it is possible to estimate how many occurrences of a particular motif could be expected purely by chance. In this study, all comparisons were made

against 1000 random graphs generated by FANMOD, which minimized the variation of results (results not shown). Upon completion of the motif mining process, normalized z-scores were calculated in order to make valid comparisons across organisms as well as organelles.¹⁰ The z-score is calculated in the standard way:

$$z_i = N_{\text{met}_i} - \text{mean}(N_{\text{random}_i}) / \text{std}(N_{\text{random}_i}) \quad (1)$$

where N_{met_i} is the number of occurrences of motif i in the metabolic network and N_{random_i} is the number of occurrences of motif i in a random network. To make comparisons between networks of differing sizes, the resultant z-scores were normalized to generate the motif distribution (SP):

$$\text{SP}_i = z_i / \sqrt{\sum z_i^2} \quad (2)$$

Normalized z-scores range from -1 to 1 and any motif with a z-score greater than 0 is considered over-represented (or enriched) compared to what could be expected by chance. Likewise, any motif with a z-score less than 0 is considered under-represented (or suppressed). Motifs with z-scores equal to 0 appear in the network as often as could be expected by chance. To assess whether motifs were statistically significantly over- or under-represented we calculated the standard error of the normalized z-score for each motif with 1000 bootstrap samples and constructed 95% confidence intervals. Confidence intervals not containing the null, $z = 0$, are statistically significantly enriched or suppressed at p -value ≤ 0.05 .

Note that the results of any motif mining procedure can be sensitive to the choice of the random background used to generate the random graphs. In our case, the random graphs were generated using the method of edge switching along with the “global constant” randomization model.⁵¹ Global constant randomization holds the total number of bidirectional edges constant, but any particular node may gain or lose a bidirectional edge. A small comparison of the three randomization models (‘local,’ ‘global’ and ‘no regard’) was done, and the results did not change appreciably.

Enzyme commission and subsystem data

To test whether motifs are uniquely characterized by their biochemical functionality, we identified the collection of reactions that, when taken together, form each motif. From this list of reactions, we retrieved the corresponding Enzyme Commission (EC) number and pathway (or “subsystem”) annotation from the reconstructions. We restricted our analysis of the EC numbers to the metabolic networks of *E. coli*, *H. sapiens*, *M. barkeri* and *S. cerevisiae*, because they provide representation of a diverse set of organisms and are derived from high-quality reconstructions. To reduce complexity we evaluated only the first two digits of the EC numbers. Truncated EC numbers retain chemical meaning, but represent more general categories of enzymes (e.g. EC 2.1.1.1: nicotinamide *N*-methyltransferase versus EC 2.1: transferring one-carbon groups).

Enzyme Commission and subsystem participation data was retrieved from the annotation contained in the metabolic network reconstructions of *E. coli*, *H. sapiens*, *M. barkeri*, and *S. cerevisiae*. Once every motif was identified by FANMOD, we extracted the biochemical reaction that form each motif. From each set of reaction we retrieve the corresponding pathway and enzyme associations in the reconstructions.

Enzyme commission and motif structure

Similarities in enzyme associations were measured using the Canberra distance metric, a measure similar to Euclidean or Manhattan distances but better suited for data containing zeros, as was the case for the proportions of enzyme associations.

Computer and software resources

Motif mining was conducted on the high performance computing cluster Flux (Intel Nehalem/I7 64-bit cluster with over 4016 Nehalem cores with an average of 4GB RAM/core) at the University of Michigan Center for Advanced Computing. All data were analyzed in *R* (v 2.15.0) and visualized with the package “ggplot2.” Motif visualizations were created in LaTeX with the package “TikZ.”

Acknowledgements

We thank Marcio Mourao, Conner Sandefur and Bryan Mayer for editing and organizational advice. This work was partially supported by the National Institute of Diabetes and Digestive and Kidney Diseases (R01 DK079084), the Michigan Nutrition Obesity Research Center (P30 DK089503), the Robert C. and Veronica Atkins Foundation, the A. Alfred Taubman Research Institute, the James S. McDonnell Foundation under the 21st Century Science Initiative, Studying Complex Systems Program. In addition, ERS acknowledges support from the Systems and Integrative Biology Training Grant (T32 GM008322).

References

- 1 E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai and A. L. Barabási, *Science*, 2002, **297**, 1551–1555.
- 2 M. R. Berenbaum, C. Favret and M. A. Schuler, *Am. Nat.*, 1996, **148**, S139–S155.
- 3 F. Li, T. Long, Y. Lu, Q. Ouyang and C. Tang, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 4781–4786.
- 4 A.-L. Barabási and Z. N. Oltvai, *Nat. Rev. Genet.*, 2004, **5**, 101–113.
- 5 C. Francke, R. J. Siezen and B. Teusink, *Trends Microbiol.*, 2005, **13**, 550–558.
- 6 M. A. Oberhardt, B. Ø. Palsson and J. A. Papin, *Mol. Syst. Biol.*, 2009, **5**, 320.
- 7 G. Basler, S. Grimbs, O. Ebenhöf, J. Selbig and Z. Nikoloski, *J. R. Soc. Interface*, 2012, **71**, 1168–1176.
- 8 N. Kashtan and U. Alon, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 13773–13778.
- 9 P. François and V. Hakim, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 580–585.
- 10 R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer and U. Alon, *Science*, 2004, **303**, 1538–1542.
- 11 R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii and U. Alon, *Science*, 2002, **298**, 824–827.
- 12 Y.-H. Eom, S. Lee and H. Jeong, *J. Theor. Biol.*, 2006, **241**, 823–829.
- 13 B. Mirzasoleiman and M. Jalili, *PLoS One*, 2011, **6**, e20512.
- 14 N. A. Shah and C. A. Sarkar, *PLoS Comput. Biol.*, 2011, **7**, e1002085.
- 15 U. Alon, *Nat. Rev. Genet.*, 2007, **8**, 450–461.
- 16 S. S. Shen-Orr, R. Milo, S. Mangan and U. Alon, *Nat. Genet.*, 2002, **31**, 64–68.
- 17 D. M. Wolf and A. P. Arkin, *Curr. Opin. Microbiol.*, 2003, **6**, 125–134.
- 18 A. Vázquez, R. Dobrin, D. Sergi, J.-P. Eckmann, Z. N. Oltvai and A.-L. Barabási, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 17940–17945.
- 19 C. G. de Oliveira Dal’Molin, L.-E. Quek, R. W. Palfreyman, S. M. Brumbley and L. K. Nielsen, *Plant Physiol.*, 2010, **152**, 579–89.
- 20 R. L. Chang, L. Ghamsari, A. Manichaikul, E. F. Y. Hom, S. Balaji, W. Fu, Y. Shen, T. Hao, B. Ø. Palsson, K. Salehi-Ashtiani and J. A. Papin, *Mol. Syst. Biol.*, 2011, **7**, 518.
- 21 S. B. Roberts, C. M. Gowen, J. P. Brooks and S. S. Fong, *BMC Syst. Biol.*, 2010, **4**, 31.
- 22 M. Ahsanul Islam, E. A. Edwards and R. Mahadevan, *PLoS Comput. Biol.*, 2010, **6**, e1000887.
- 23 A. M. Feist, C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis and B. Ø. Palsson, *Mol. Syst. Biol.*, 2007, **3**, 121.
- 24 I. Thiele, T. D. Vo, N. D. Price and B. Ø. Palsson, *J. Bacteriol.*, 2005, **187**, 5818–5830.
- 25 O. Gonzalez, S. Gronau, M. Falb, F. Pfeiffer, E. Mendoza, R. Zimmer and D. Oesterhelt, *Mol. Biosyst.*, 2008, **4**, 148–159.
- 26 N. C. Duarte, S. A. Becker, N. Jamshidi, I. Thiele, M. L. Mo, T. D. Vo, R. Srivas and B. Ø. Palsson, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 1777–1782.
- 27 V. S. Kumar, J. G. Ferry and C. D. Maranas, *BMC Syst. Biol.*, 2011, **5**, 28.
- 28 A. M. Feist, J. C. M. Scholten, B. Ø. Palsson, F. J. Brockman and T. Ideker, *Mol. Syst. Biol.*, 2006, **2**, 2006.0004.
- 29 M. I. Sigurdsson, N. Jamshidi, E. Steingrímsson, I. Thiele and B. Ø. Palsson, *BMC Syst. Biol.*, 2010, **4**, 140.
- 30 X. Fang, A. Wallqvist and J. Reifman, *BMC Syst. Biol.*, 2010, **4**, 160.
- 31 B. K. S. Chung, S. Selvarasu, A. Camattari, J. Ryu, H. Lee, J. Ahn, H. Lee and D.-Y. Lee, *Microb. Cell Fact.*, 2010, **9**, 50.
- 32 S. A. Becker and B. Ø. Palsson, *BMC Microbiol.*, 2005, **5**, 8.
- 33 N. C. Duarte, M. J. Herrgård and B. Ø. Palsson, *Genome Res.*, 2004, **14**, 1298–1309.
- 34 I. Thiele, D. R. Hyduke, B. Steeb, G. Fankam, D. K. Allen, S. Bazzani, P. Charusanti, F.-C. Chen, R. M. Fleming,

- C. A. Hsiung, S. C. De Keersmaecker, Y.-C. Liao, K. Marchal, M. L. Mo, E. Özdemir, A. Raghunathan, J. L. Reed, S.-I. Shin, S. Sigurbjörnsdóttir, J. Steinmann, S. Sudarsan, N. Swainston, I. M. Thijs, K. Zengler, B. O. Palsson, J. N. Adkins and D. Bumann, *BMC Syst. Biol.*, 2011, **5**, 8.
- 35 Y. Zhang, I. Thiele, D. Weekes, Z. Li, L. Jaroszewski, K. Ginalski, A. M. Deacon, J. Wooley, S. A. Lesley, I. A. Wilson, B. Palsson, A. Osterman and A. Godzik, *Science*, 2009, **325**, 1544–1549.
- 36 H. U. Kim, S. Y. Kim, H. Jeong, T. Y. Kim, J. J. Kim, H. E. Choy, K. Y. Yi, J. H. Rhee and S. Y. Lee, *Mol. Syst. Biol.*, 2011, **7**, 460.
- 37 R. Saha, P. F. Suthers and C. D. Maranas, *PLoS One*, 2011, **6**, e21784.
- 38 R. Mahadevan, D. R. Bond, J. E. Butler, A. Esteve-Nuñez, M. V. Coppi, B. O. Palsson, C. H. Schilling and D. R. Lovley, *Appl. Environ. Microbiol.*, 2006, **72**, 1558–1568.
- 39 O. Resendis-Antonio, J. L. Reed, S. Encarnación, J. Collado-Vides and B. Ø. Palsson, *PLoS Comput. Biol.*, 2007, **3**, e192.
- 40 P. Kenrick and P. R. Crane, *Nature*, 1997, **389**, 33–39.
- 41 P. Giegé, J. L. Heazlewood, U. Roessner-Tunali, A. H. Millar, A. R. Fernie, C. J. Leaver and L. J. Sweetlove, *Plant Cell*, 2003, **15**, 2140–2151.
- 42 G. Noctor, R. De Paepe and C. H. Foyer, *Trends Plant Sci.*, 2007, **12**, 125–134.
- 43 R. R. Sederoff, *Am. Nat.*, 1987, **130**, S30–S45.
- 44 P. van Nes, D. Bellomo, M. Reinders and D. de Ridder, *EURASIP J. Bioinf. Syst. Biol.*, 2009, **2009**, 630695.
- 45 R. Dobrin, Q. K. Beg, A.-L. Barabasi and Z. N. Oltvai, *BMC Bioinf.*, 2004, **5**, 10.
- 46 T. Michoel, A. Joshi, B. Nachtergaele and Y. Van de Peer, *Mol. BioSyst.*, 2011, **7**, 2769–2778.
- 47 I. A. Kovács, R. Palotai, M. S. Szalay and P. Csermely, *PLoS One*, 2010, **5**, e12528.
- 48 R. J. Prill, P. A. Iglesias and A. Levchenko, *PLoS Biol.*, 2005, **3**, e343.
- 49 K. Klemm and S. Bornholdt, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 18414–18419.
- 50 S. A. Becker, A. M. Feist, M. L. Mo, G. Hannum, B. Ø. Palsson and M. J. Herrgard, *Nat. Protoc.*, 2007, **2**, 727–738.
- 51 F. Rasche and S. Wernicke, *Fast Network Motif Detection: Manual*, Institut für Informatik, Friedrich-Schiller-Universität Jena, Ernst-Abbe-Platz 2, D-07743 Jena, Germany, 2006.
- 52 S. Wernicke and F. Rasche, *Bioinformatics*, 2006, **22**, 1152–1153.