Molecular BioSystems

Cite this: DOI: 10.1039/c1mb05168g

www.rsc.org/molecularbiosystems

PAPER

Patterns of indirect protein interactions suggest a spatial organization to metabolism

Åsa Pérez-Bercoff,[a] Aoife McLysaght[a] and Gavin C. Conant*[bc]

Received 5th May 2011, Accepted 8th August 2011
DOI: 10.1039/c1mb05168g

It has long been believed that cells organize their cytoplasm so as to efficiently channel metabolites between sequential enzymes. This metabolic channeling has the potential to yield higher metabolic fluxes as well as better regulatory control over metabolism. One mechanism for achieving such channeling is to ensure that sequential enzymes in a pathway are physically close to each other in the cell. We present evidence that indirect protein interactions between related enzymes represent a global mechanism for achieving metabolic channeling; the intuition being that protein interactions between enzymes and non-enzymatic mediator proteins are a powerful means of physically associating enzymes in a modular fashion. By analyzing the metabolic and protein–protein interactions networks of *Escherichia coli*, yeast and humans, we are able to show that all three species have many more indirect protein interactions linking enzymes that share metabolites than would be expected by chance. Moreover, these interactions are distributed non-randomly in the metabolic network. Our analyses in yeast and *E. coli* show that reactions possessing such interactions also show higher flux than do those lacking them. On the basis of these observations, we suggest that an important role of protein interactions with mediator proteins is to contribute to the spatial organization of the cell. This hypothesis is supported by the fact that these mediator proteins are also enriched with annotations related to signal transduction, a system where scaffolding proteins are known to limit cross-talk by controlling spatial localization.

Introduction

Although the analogy between enzymes and machines is a common one, it is rather more rare to note that an obvious extension, that of pathways and assembly lines, is also appropriate. Nonetheless, the idea that enzymes physically associate into complexes that reflect pathway structures is an old one.[1–4] One of the primary advantages of these associations is likely to be in facilitating metabolic channeling between sequential enzymes. Metabolic channeling is the collective name for a group of mechanisms that allow metabolites to move between enzymes without being released into the bulk solvent,[5] reducing the degree to which pathway flux is diffusion-limited.[6] One of the earliest examples of channeling was found in the tryptophan biosynthesis pathway, when Yanofsky and Rachmeler were unable to detect one of the pathway's necessary intermediates in cell extracts.[1] It was only after the two catalytic domains in question were crystallized that an enclosed tunnel between the two active sites was identified, accounting for the intermediate's apparent absence.[7]

More recently, it has become clear that macromolecular complexes able to facilitate channeling are common and can associate with specific cellular locations, such as the plasma membrane or mitochondrion.[8–11] Channeling is also widely phylogenetically distributed, with known examples from yeasts,[9,12] plants,[11,13,14] mammals[3,8] and bacteria.[2,7] One of the important roles that these associations play is to partially isolate particular metabolites, such as ATP, so that functional units, such as myofibrils and ion pumps, respond to the local and not the global concentration of that metabolite.[15]

Metabolic channeling is at least in part an emergent property of two other cellular features: cells have very high protein concentrations (macromolecular crowding)[16] and are highly spatially organized.[17,18] Indeed, an elegant recent experiment has shown that even when the plasma membrane has been disrupted much of the cell's protein machinery maintains its organization and function rather than simply diffusing away.[19] Cells use several mechanisms to spatially organize metabolism, including protein localization by signaling peptides[20] and associations between enzymes and membranes.[21] Another potential mechanism for achieving subcellular organization is protein–protein interactions, or PPIs.[22] Here, we explore the

[a] *Smurfit Institute of Genetics, University of Dublin, Trinity College, Dublin 2, Ireland*
[b] *Division of Animal Sciences, University of Missouri, Columbia MO, USA. E-mail: conantg@missouri.edu*
[c] *Informatics Institute, University of Missouri, Columbia MO, USA*

role of PPIs in metabolism, dividing the interactions involved into two types: *direct* and *indirect*.[14] For each interaction type we ask whether it is more common than expected among related reactions in the metabolic network. Our work follows that of Huthmacher and colleagues, who have argued for a global excess of direct protein interactions (dPPIs) between pairs of enzymes with shared metabolites,[23] as well as a correlation in enzyme-enzyme distance in the metabolic and protein interaction networks.[24,25] Analyzing data from *Escherichia coli*, bakers' yeast and humans, we revisit the role of dPPIs in structuring metabolism. We then introduce a role for *indirect* protein interactions (iPPIs) that link pairs of enzymes through nonenzymatic intermediate proteins.

Results

To study metabolic channeling in these three organisms, we employed published metabolic and protein interaction networks (Table 1),[26–28] including two different human metabolic networks, those of Ma *et al.*, (Hs_M networks)[29] and Duarte *et al.*, (Hs_D networks).[30] We studied reaction-centered metabolic networks, consisting of nodes that are metabolic reactions.[31] Two reactions are connected by an edge if they share a metabolite (Fig. 1). Because there are a handful of ubiquitous metabolites such as hydrogen ions and water, we first counted the number of reactions each metabolite participates in (n). We then excluded the most highly connected metabolites from the network at five stringencies: $n \geq 10$, 15, 25, 50, and 100. We have named the networks based on this exclusion stringency. Thus the Hs_M_25 network is derived from the human network of Ma *et al.*, with metabolites participating in 25 or more reactions excluded. Table 2 gives selected network statistics for the twenty networks examined. Except where noted, all of our conclusions hold across all five exclusion stringencies and all three taxa.

Any two proteins that catalyze distinct reactions but nonetheless share both a metabolite and a protein–protein interaction are defined to possess a direct protein–protein-interaction (dPPI). Huthmacher *et al.*,[23] detected an excess of dPPIs in yeast and *E. coli* compared to the number expected when randomizing either the metabolic network or both the metabolic and protein interaction networks. The authors randomized the metabolic network by exchanging reaction identities across the network and the protein interaction network by creating new network edges under an assumed probability distribution. Whether an excess of dPPIs was observed depended on the randomization approach used.[23]

While network randomization appears to be the most appropriate way to assess whether there are more dPPIs than would be expected by chance, the exact method of constructing the required random networks involves controlling for a number of effects. We describe the issues associated with previous analyses and the precise approach used here in Appendix 1. Briefly, we kept the metabolic network unaltered and compared it to 1000 randomized protein-interaction networks in which the number of edges for each protein was kept constant but the identity of those edges was randomized. We then calculated the number of dPPIs found when comparing the original metabolic network to each randomized PPI network (Appendix 1). One important refinement that we introduce to the randomization is the exclusion of *intra*-reaction protein interactions.[28,32] For example, in yeast, the enzymes Bat1 and Bat2 are encoded by genes produced by the yeast genome duplication.[33] These two proteins both catalyze the same two distinct reactions, namely the final steps of isoleucine and valine production.[34] They also interact with each other according to the Database of Interacting Proteins (DIP).[35] A naïve approach to identifying dPPIs would include this interaction, an inclusion that we would dispute, since the more parsimonious explanation is that the two proteins have descended from a single self-interacting protein through the genome duplication. To avoid these types of false signals, we therefore did not allow PPIs joining proteins from the same reaction to form part of a dPPI.

In *E. coli*, when we excluded metabolites involved in at least 10 or at least 15 reactions, we find an excess of dPPIs (Ec_10, Ec_15; $P \leq 0.003$; Table 2), but this enrichment is not observed for the three larger cutoffs (Ec_25, Ec_50, Ec_100; $P > 0.05$). In yeast, there are not significantly more dPPIs than would be expected in any of the five networks shown in Table 2 ($P > 0.05$). Finally, we observe a significant overabundance of dPPIs for all ten human networks in Table 2 ($P < 0.001$). Our conclusion of no dPPI enrichment in yeast may differ from that of Huthmacher *et al.*, because these authors do not appear to have excluded intra-enzyme interactions shared across multiple reactions (*i.e.*, as in the Bat1/Bat2 example cited). We cannot currently assess why these three organisms show differing evidence of dPPIs.

Table 1 Statistics regarding the networks employed

Species	Metabolic network reference	# Rxns[a]	# Genes[b]	Avg. # genes per rxn[c]	Avg. # rxns per gene[d]	Proteins[e]	Non-self interactions[f]	Proteins in met. network[g]	Interactions in met. network[h]
E. coli	27	873	904	1.7	1.6	1862	7798	294	401
S. cerevisiae	28	810	743	1.7	1.8	4893	17 169	269	329
H. sapiens	29	887	2269	3.3	1.3	8876	32 916	827	1735
H. sapiens	30	2306	1475	1.9	2.9	8876	32 916	232	187

[a] Number of reactions in the metabolic network with annotated genes. [b] Number of genes annotated in the metabolic network. [c] The average number of genes annotated per reaction. [d] The average number of distinct reactions a gene participates in. [e] Number of proteins with annotated protein–protein interactions. [f] Total number of protein–protein interactions in the network involving distinct proteins (*i.e.*, non-self interactions). [g] Number of proteins from the protein interaction network also present in the corresponding metabolic network. Used in only the network randomizations when analyzing dPPIs. [h] The number of distinct protein–protein interactions that involve proteins that are both present in the metabolic network. Used in only the network randomizations when analyzing dPPIs.

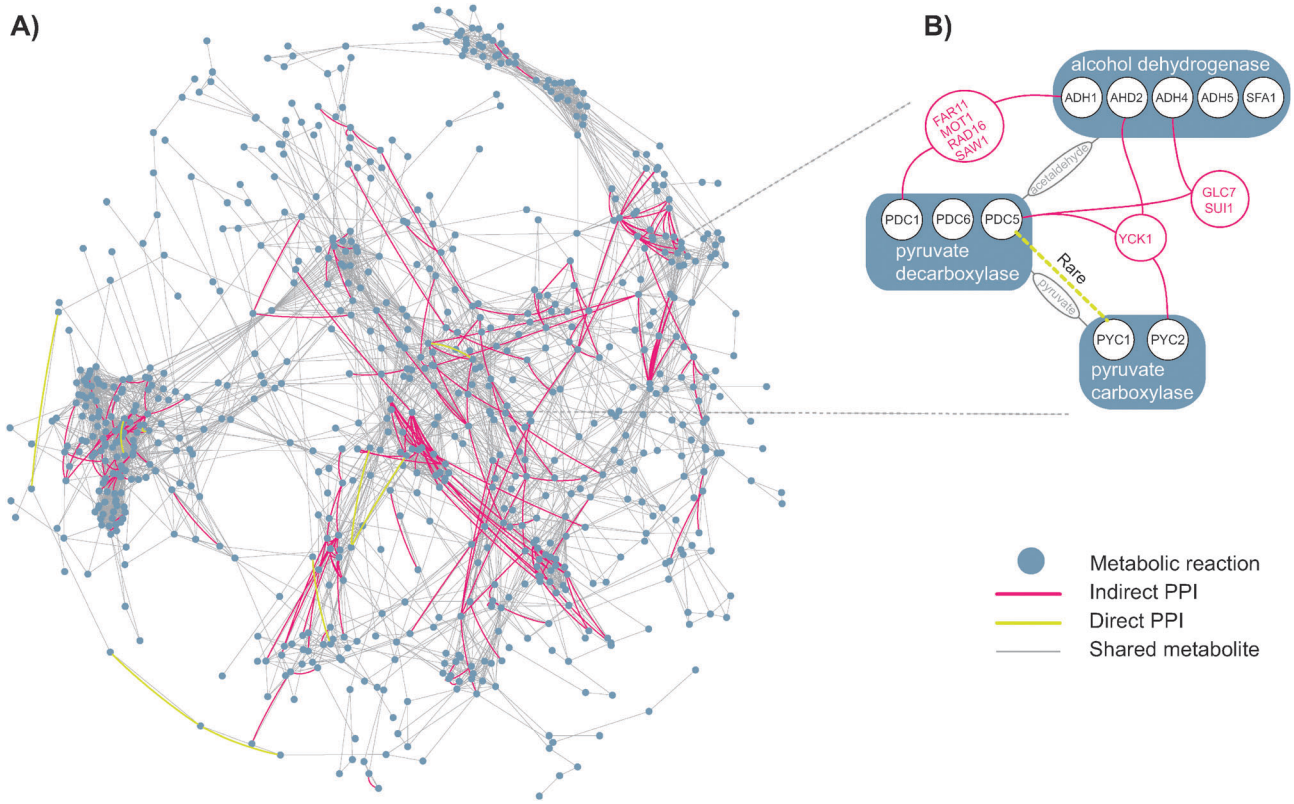

Fig. 1 Indirect protein–protein interactions provide structure to the yeast metabolic network. (A) The central component of the yeast metabolic network described by Duarte and colleagues is shown.[28] Metabolites involved in more than 25 reactions are omitted (Sc_25). Nodes (blue) are reactions, joined by three types of edges. Thin grey lines indicate two reactions that share a common metabolite or metabolites. Light green lines indicate a direct PPI (dPPI) between at least two of the constituent enzyme proteins of a pair of reactions. Pink lines represent an indirect PPI (iPPI; meaning that there is a third, nonmetabolic, protein that both enzymes interact with). **(B)** An enlarged view of a small part of the network, consisting of three enzymes involved in central carbon metabolism. Pyruvate decarboxylase and alcohol dehydrogenase are sequential steps in ethanol fermentation, and isozymes for these two reactions are joined by a total of seven indirect PPIs. Pyruvate decarboxylase and pyruvate carboxylase represent branch points in pyruvate metabolism. Note that the Yck1 protein actually joins all three reactions, although the link joining pyruvate carboxylase and alcohol dehydrogenase is not shown in **A** since these two reactions do not share a common metabolite.

However, we do note that, of the three species, the set of protein interactions inferred for yeast may be the least biased by investigator choices, since many of these interactions were identified with high-throughput approaches. Since protein interactions from *E. coli* and humans are taken from literature studies, a tendency among researchers to look for protein interactions among members of the same pathways could potentially give rise to a spurious indication of a general abundance of dPPIs. Given the ambiguous evidence for dPPIs and the relatively small number of them seen (Table 2), we have also explored the potential for more complex protein interactions that may facilitate channeling.

Indirect interactions between neighboring enzymes

Proteins not themselves part of the metabolic network might also act as mediators between enzymes. Durek and Walther [25] have shown that non-metabolic proteins contribute to bringing reactions with shared metabolites into closer proximity in the protein interaction network. We extend this result by searching for an overabundance of indirect PPIs (iPPIs), where a pair of PPIs to a third non-metabolic protein join two enzymes sharing a metabolite (Fig. 1).

For each metabolic network, we determined m: the number of unique interactions between two enzymes and their mediator protein (the number of unique iPPIs). We then created 1000 randomized protein-interaction networks as described above and calculated m_s: the number of unique iPPIs in each randomized network. We then compared the distribution of m_s to the value of m. For example, in the Sc_50 network, $m = 504$, while the largest value of m_s observed was 253. Indeed, for none of the fifteen eukaryotic networks did we find any values of m_s as large as the respective values of m ($P < 0.001$, Table 3). Moreover, because the yeast protein interaction data derive from different types of high-throughput experiments, we can show that our results in yeast are robust to the method of protein interaction detection (*Methods*). In *E. coli* the situation is slightly more complex, as significant iPPI enrichment was observed only for the two most stringent thresholds for currency metabolite exclusion ($P \leq 0.001$; Ec_10 and Ec_15, Table 3). Given that an iPPI implies physical proximity between its three members, we conclude that iPPI enrichment

Table 2 Metabolic network structure and dPPI prevalence. Selected statistics regarding the twenty metabolic networks used are given. dPPIs are defined as shared protein interactions between two enzymes that also share a metabolite

Network	*Edges*[a]	*Path length*[b]	*Clust. Coeff.*[c]	l^d	Average l_s^e	Maximum l_s^f	P^g
Ec_10	2225	7.7	0.65	13	4.4	16	0.003
Ec_15	3168	5.3	0.63	15	6.0	16	0.002
Ec_25	5006	4.3	0.64	20	13.5	27	**0.052**
Ec_50	8355	3.6	0.66	25	21.4	36	**0.25**
Ec_100	15 807	2.9	0.69	51	41.7	61	**0.08**
Sc_10	2004	7.1	0.66	6	4.2	11	**0.23**
Sc_15	2959	5.5	0.66	7	6.6	16	**0.49**
Sc_25	4375	5.0	0.67	7	7.6	18	**>0.5**
Sc_50	8749	3.9	0.70	12	14.1	30	**>0.5**
Sc_100	19 196	2.8	0.73	33	29.3	47	**0.27**
Hs_D_10	4292	12.9	0.71	22	2.9	9	<0.001
Hs_D_15	6555	9.3	0.69	27	3.2	10	<0.001
Hs_D_25	11 018	8.0	0.68	30	4.7	14	<0.001
Hs_D_50	22 844	5.4	0.74	34	5.9	20	<0.001
Hs_D_100	44 642	4.2	0.76	38	8.6	20	<0.001
Hs_M_10	2726	4.5	0.71	174	83.0	104	<0.001
Hs_M_15	3797	3.9	0.70	431	238.3	273	<0.001
Hs_M_25	5139	3.5	0.69	435	241.5	292	<0.001
Hs_M_50	7962	3.2	0.70	448	244.4	281	<0.001
Hs_M_100	16 213	2.7	0.73	459	250.8	286	<0.001

[a] Number of edges in the undirected metabolic network considering only reactions with annotated genes. [b] Average minimum path length: the average over all nodes of the average for each node of the minimum number of edges needing to be traversed to reach any other node, calculated with Dijkstra's algorithm.[60] [c] Clustering coefficient of each network.[61] [d] Number of unique dPPIs, *i.e.*, unique interactions between two enzymes, in the actual metabolic network. [e] Average number of unique dPPIs seen in 1000 randomizations of the protein interaction network. [f] Maximal number of unique dPPIs observed in any of the 1000 randomizations. [g] *P*-value for the test of the hypothesis that there are no more dPPIs than would be expected given the respective structures of the two networks. *Non-significant* tests are shown in bold.

Table 3 Real metabolic networks have many more iPPIs than do randomized networks. iPPIs are defined as a shared protein interaction between two enzymes that share a metabolite and a third mediator protein. The number of such interactions in real metabolic networks is much larger than can be explained by chance ($m > m_s$)

Network	m^a	Average m_s^b	Maximum m_s^c	P^d
Ec_10	157	109.4	156	<0.001
Ec_15	208	156.1	216	0.001
Ec_25	394	383.3	476	**0.35**
Ec_50	486	591.7	708	**>0.5**
Ec_100	870	1236.6	1464	**>0.5**
Sc_10	147	57.5	88	<0.001
Sc_15	206	84.0	119	<0.001
Sc_25	259	99.8	141	<0.001
Sc_50	504	187.5	253	<0.001
Sc_100	2192	466.1	605	<0.001
Hs_D_10	352	62.9	96	<0.001
Hs_D_15	369	69.5	113	<0.001
Hs_D_25	386	73.2	109	<0.001
Hs_D_50	399	78.6	120	<0.001
Hs_D_100	444	90.5	126	<0.001
Hs_M_10	2776	879.5	1050	<0.001
Hs_M_15	6381	3277.8	3787	<0.001
Hs_M_25	6411	3315.2	3706	<0.001
Hs_M_50	6464	3356.4	3906	<0.001
Hs_M_100	6564	3464.8	3980	<0.001

[a] Number of unique iPPIs, *i.e.*, unique interactions between two enzymes and a mediator protein, in the actual metabolic network. [b] Average number of unique iPPIs seen in 1000 randomizations of the protein interaction network. [c] Maximal number of unique iPPIs observed in any of the 1000 randomizations. [d] *P*-value for the test of the hypothesis that there are no more iPPIs than would be expected given the respective structures of the two networks. *Non-significant* tests are shown in bold.

will drive physical associations between related enzymes in a nonrandom fashion.

If the functional role of iPPIs is to physically associate metabolic enzymes, we would expect that their distribution in that network would be non-random. A consequence of our definition of iPPIs is that they represent a subset of the edges in their respective metabolic networks. We thus asked if the metabolic edges possessing iPPIs were unusual by examining the number of network components those iPPI-possessing edges define. We compared this number to the number of components observed when the same number of metabolic edges were selected at random. (Note that, as with our previous analysis, this randomization was performed only on the set of metabolic nodes with annotated genes, *Methods*). In all 17 real networks for which we identified iPPI enrichment, the number of components seen was greater than that seen for any randomly selected sets of edges ($P < 0.001$, Fig. 2). Further analyses suggested that the primary driver of this phenomenon was that a limited number of reactions possess iPPIs, meaning that the iPPI-induced metabolic networks had many more isolated nodes than did random ones ($P < 0.001$).

Analysis of yeast pathways for dPPIs and iPPIs

To explore whether the patterns we observed also existed at the level of metabolic pathways, we obtained a set of 183 metabolic pathways from the *Saccharomyces* Genome Database.[36] Using a similar network approach, we asked whether there was an excess of dPPIs and iPPIs between members of the same metabolic pathway (*Methods*). For both dPPIs and iPPIs,

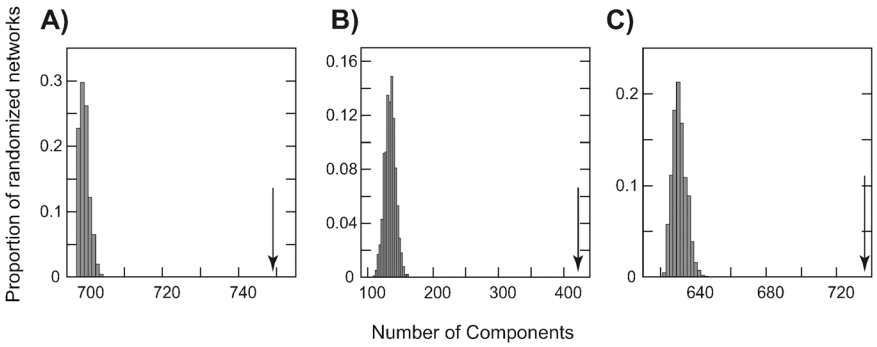

Fig. 2 Indirect protein–protein interactions (iPPIs) produce more distinct metabolic subnetworks than would be expected by chance. On the *x*-axis is the number of components observed when a number of metabolic edges equal to the number of iPPIs is sampled at random from the metabolic network. The *y*-axis gives the proportion of simulations having that number of components (1000 simulations). The arrows in each panel give the number of network components produced in each metabolic network when only the metabolic edges corresponding to iPPIs are retained (749, 422 and 736, respectively). For each species, we selected the network with the most liberal threshold that still showed an excess of iPPIs. **(A)** The *E. coli* Ec_15 network. **(B)** The yeast Sc_100 network. **(C)** The human Hs_M_100 network.

we observe such an excess, with the number of dPPIs or iPPIs in the actual network being at least 2.5 times greater than the largest number seen in the randomized networks ($P < 0.001$). Although the annotation status of these pathways introduces some error into this analysis (*Methods*), it is encouraging that the general trends are similar to those seen when analyzing metabolic networks.

Functional annotation of mediator proteins

To understand the biological role of the mediator proteins (pink gene names in Fig. 1B), we searched for Gene Ontology (GO) terms enriched among the mediator proteins.[37] To do so, we used the largest eukaryotic networks because the other networks contain subsets of the mediator genes in these three. In all three cases (Sc_100, Hs_D_100 and Hs_M_100), the most significantly enriched molecular function term was "protein binding" (Bonferroni-corrected $P \leq 0.009$ across three hypothesis tests). In yeast, the most enriched biological process term was "localization" (Bonferroni-corrected $P < 10^{-11}$), while in the Hs_D_100 and Hs_M_100 networks, the most enriched terms were "signaling" and "regulation of cellular process" respectively (Bonferroni-corrected $P < 10^{-34}$ and $P < 10^{-101}$). However, "localization" was also significantly enriched in these datasets (Bonferroni-corrected $P < 10^{-17}$).

Next we asked whether the yeast mediator proteins were more likely to be essential than the average gene. To do so, we asked what proportion of the mediator protein genes from the largest yeast network, Sc_100, were essential according to the Munich Information Center for Protein Sequences (MIPS).[38] Of these genes, 26% were deemed essential, as compared to 18% of the genome at large and of genes with at least one protein interaction, both significant differences (χ^2 tests, $P < 0.01$).

Association of iPPI presence and high metabolic flux

We hypothesized that the reactions possessing iPPIs might be biased toward those carrying high flux, accounting for their non-random distribution in the network. Durek and Walther have shown that enzymes with high protein interaction degree show high flux:[25] we asked if this result could be partly attributed to higher flux among enzymes with iPPIs. We thus computed the maximal flux through each *E. coli* and yeast reaction across several growth conditions using flux-balance analysis (FBA; *Methods*).[39] For both *E. coli* networks with iPPI enrichment (Ec_10 and Ec_15), the mean flux through reactions with an iPPI was at least 1.5-fold greater than that through reactions without one (Wilcoxon rank sum tests, $P < 0.04$). In yeast, no network showed a significant difference between reactions with and without iPPIs. However, many of the reactions in these networks show no flux in our FBA analyses, either because the set of environment conditions where that reaction is used was not tested or because the biomass reaction used did not include that reaction's contribution (*e.g.*, for micronutrients).[40] The lack of these reactions reduces the sample size and hence power of our flux analysis. To partly compensate for this effect, we introduced a second analysis where we asked whether the proportion of reactions having both an iPPI and high flux was greater than would be expected. We defined "high flux" according to a sliding threshold t, measured relative to the flux through the biomass reactions (*Methods*). Thus, when $t = 1$, a high flux reaction is a reaction with at least the flux of the biomass reaction. For a wide range of values of t ($0 \leq t \leq 128$), all five yeast networks show a greater proportion of reactions having both high flux and an iPPI than would be expected ($P < 0.05$; Fig. 3A). As Fig. 3A indicates, even for values of t where no significant excess of high-flux reactions with iPPIs is observed (due to the small number of total reactions with flux $> t$), the proportion of reactions having iPPIs that are high flux is always greater than that of reactions without iPPIs. Because of the clear difference in flux between reactions with and without iPPIs in *E. coli*, no equivalent test was carried out in that species.

We also computationally limited the flux through each reaction in the two metabolic networks and calculated the resulting change in biomass flux. We defined the fitness effect of this knockdown as the ratio of the knockdown to original biomass flux (*Methods*). For Sc_15, Sc_25 and Sc_50, average knockdown fitness is *higher* for those reactions associated with an iPPI ($P < 0.05$; Fig. 3B). No significant difference was observed in Sc_10, Sc_100, Ec_10 or Ec_15.

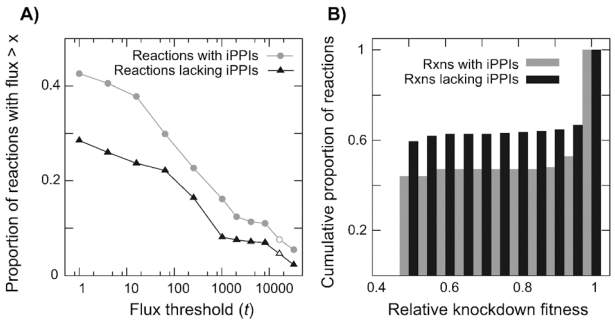

Fig. 3 Reactions with iPPIs have higher flux and weaker knockdown effects. (A) Reactions with iPPIs represent a greater proportion of high flux reactions than expected. On the *x*-axis is the threshold used to define high flux (t). On the *y*-axis is the proportion of reactions either having (grey) or lacking (black) iPPIs that have flux of at least t. Open symbols indicate comparisons where the difference in proportion between the reactions with and without iPPIs is not statistically significant ($P > 0.05$). **(B)** Reactions with iPPIs have smaller knockdown effects than other reactions. On the *x*-axis is our measure of relative knockdown fitness: the ratio of biomass flux when a reaction is limited to one half its optimal flux to "wild-type" flux. On the *y*-axis is the cumulative proportion of reactions with relative knockdown fitness $\leq x$ for reactions that have (grey) or lack (black) iPPIs. Both panels show results from Sc_50, but other yeast networks are qualitatively similar.

While it would be interesting to examine the metabolic roles of the iPPIs in humans, the goal of maximal biomass production used in yeast and *E. coli* is not an evolutionarily meaningful objective function in humans, since rapid cell division does not generally confer a fitness benefit in multicellular species. Shlomi and colleagues[41,42] have instead provided some more computationally challenging methods for analyzing the human metabolic network which may provide interesting insights into this problem in the future.

Discussion

We offer several new observations supporting the contention that protein interactions play a global role in spatially structuring metabolism.[23–25] First, were iPPIs simply a result of noise in the protein interaction data, it would be difficult to understand our observations that iPPIs involve a distinct subset of the reactions in the metabolic network and involve mediators that are more likely to be essential. Likewise, the association between high flux reactions and the presence of iPPIs suggests that the physical association of sequential enzymes may be reserved for reactions that require elevated flux. It is certainly the case that a number of existing examples of channeling involve such reactions.[2,3,8,9]

More generally, the advantages of spatially structuring cellular functions extend beyond simple increases in metabolic efficiency. For instance, micro-compartmentalization may help to sequester metabolic intermediates with undesirable side effects.[9,43] Likewise, in signal transduction, spatial co-localization increases efficiency and minimizes cross-talk.[44,45] In fact, there are scaffolding proteins responsible for maintaining this co-localization,[45] and it is thus intriguing that a number of the mediator proteins here also have signaling annotations. We speculate that similar scaffolding structures may be at work in metabolism. Given these Gene Ontology results, however, it is not yet clear how functionally specialized a given mediator protein will be.

One obvious question is why we occasionally obtain differing results depending on the stringency of currency metabolite removal. Thus, only Ec_10 and Ec_15 show significant iPPI enrichment in *E. coli*. We note that excluding more potential currency metabolites should *reduce* the number of spurious associations between reactions because only reactions with unique shared metabolites will be connected. Networks with more stringent currency metabolite removal (Ec_10, Sc_10 *etc.*) should therefore give better sensitivity in detecting dPPIs and iPPIs (at a cost of potentially excluding real iPPIs and a concomitant loss of statistical power). Importantly, unlike yeast and humans, *E. coli* lacks mitochondria, nuclei and other membrane-bound compartments, facts reflected in the respective metabolic networks. We hypothesize that currency metabolite removal is therefore more important for this organism, since spurious associations between reactions that would be removed by compartmentalization in eukaryotes (where, for instance, ATP in the cytoplasm is distinguished from that in the mitochondria) are not eliminated in the prokaryotic network. A similar tradeoff between statistical power and sensitivity probably explains why significant associations between iPPIs and knockdown effects are seen only for Sc_15, Sc_25, and Sc_50.

Our observations may also provide insights on other biological phenomena. For instance, the relative scaling between metabolic rate and body size is surprisingly similar across groups of related organisms and organ systems; a result that appears to be due to the fractal branching patterns of the transport networks involved.[46] It has recently been suggested that this scaling might also extend to the subcellular level.[47] If so, similar hierarchical structures within the cell, such as those implied above,[5,17,43] might be one source of that scaling. Likewise, the associations of iPPIs with high flux reactions also reaffirms the modular nature of the metabolic network,[48] with groups of enzymes working in functional isolation from each other. It is of course terribly tempting to employ the network structures of Fig. 1 to infer this set of metabolic assembly lines. However, such an effort is likely premature, given the incomplete nature of the protein interaction data.[49,50] Nonetheless, there is certainly more information embedded in the protein interaction and metabolic networks than has been analyzed here. We suggest that the addition of protein abundance data might help infer the stoichiometry of the metabolic clusters relative to other parts of the cellular substructure, and we speculate that this organization might show fractal patterns.[51]

Methods

Protein interaction data

Statistics on the protein interaction networks used and their overlap with the metabolic networks are given in Table 1. Data on human protein–protein interactions (PPIs) was obtained from the Human Protein Reference Database (HPRD)

release 7[52,53] and matched against Ensembl release 50,[54] as previously described.[55] Yeast and *E. coli* protein interaction data were obtained from the Database of Interacting Proteins (DIP).[35]

Gene and protein identifiers used

Uniprot identifiers taken from the DIP database[35] were mapped to gene identifiers from the *E. coli* genome and metabolic network[27,56] using custom Perl scripts that queried the EBI uniprot database (http://www.ebi.ac.uk/Tools/dbfetch/). Yeast gene and protein identifiers were presented in standard format both by DIP[35] and the metabolic network,[28] such that no identifier mapping was required. Human protein identifiers from both the protein interaction[52,53] and the metabolic networks[29,30] were mapped to Ensembl identifiers as previously described.[55,57]

Mapping of metabolic and protein interaction networks

All three metabolic networks used are provided with gene identifiers for reactions with known enzymes. For the purposes of this study, any reaction without an annotated gene was omitted from all analyses except the flux balance analysis. For the remaining nodes, the proteins corresponding to said genes were mapped to the proteins identified in the protein interaction network using Perl. Table 1 gives the resulting statistics.

Robustness of iPPI enrichment to method of PPI detection

We repeated our analysis of the yeast metabolic network excluding all interactions determined by mass spectrometry and including only interactions obtained from yeast two-hybrid experiments. In neither case did we observe a value of m_s in the randomized networks as large as m in any of the five yeast networks ($P < 0.001$).

Pathway analysis

From the *Saccharomyces* Genome Database (SGD),[36] we downloaded pathway annotations for 531 yeast metabolic genes (a total of 181 pathways). We defined a network such that pairs of genes in the same pathway were joined by an edge. We then counted all instances where two genes in the same pathway shared a protein–protein interaction (*i.e.*, a dPPI) or shared an interaction to a nonmetabolic protein (*i.e.*, an iPPI). Cases where two genes were annotated as being involved in the same reaction by SGD were not counted toward the total of dPPIs and iPPIs. In other words, *intra*-reaction/*intra*-enzyme protein interactions were excluded. Using the network randomization approach of Appendix 1, we asked whether the same number of dPPIs and iPPIs could be expected by chance. In neither case did the randomized networks show as many dPPIs or iPPIs as observed in the real network ($P < 0.001$). However, we note that, especially for the dPPIs, the pathways used are somewhat imperfect, since a gene may be annotated into a pathway without a corresponding reaction annotation (meaning that we could again mistake an intra-enzyme protein interaction for a dPPI).

GO analyses

Gene ontology (GO) analyses for eukaryotic iPPIs were conducted as described by Boyle *et al.*[37] using the website http://go.princeton.edu/cgi-bin/GOTermFinder.

Flux balance analysis

Flux balance analysis (FBA) is a common technique for bounding the space of potential metabolic fluxes using reaction stoichiometry and input metabolic constraints.[39] Briefly, the approach uses linear programming to bound the solution space of a system of homogenous linear equations such that a defined biomass objective function is maximized. Our custom software uses the GNU Linear programming toolkit (http: //www.gnu.org/software/glpk/) to perform this computation: we note that our results are numerically identical to those produced by publically available FBA packages (*e.g.*, the Systems Biology Research Tool[58]). We thus inferred the combination of reaction fluxes that yielded maximal biomass production for the yeast and *E. coli* metabolic networks used here.[27,28] In yeast, we performed our analysis under six sets of input nutrient conditions, including aerobic growth with glucose, fructose, glycerol, ethanol and glutamine as the primary carbon source and anaerobic growth with glucose as the primary carbon source. In *E. coli* we used growth on glucose under anaerobic and aerobic conditions. (Software and data files available upon request.) For each condition we normalized all fluxes to the overall biomass flux under that condition. Finally, for each set of inputs, we also computed the change in biomass flux that resulted from individually constraining the flux through each reaction to half of its maximal value (a computational knockdown). We estimated knockdown fitness f as the ratio of the constrained biomass flux to the unconstrained flux (Fig. 3).

Appendix 1 Comparison of metabolic and protein interaction networks by randomization

In order to properly assess whether there is an excess of dPPIs or iPPIs, there are several potential confounding factors that any randomization approach needs to avoid:

1. Different proteins can have very different numbers of protein interactions, and a randomization scheme that fails to preserve this skew in the number of interaction partners might have unpredictable effects on the assessment of dPPI and iPPI excess.
2. One metabolic reaction might involve enzymes coded for by multiple different genes, either as isoenzymes or multi-enzyme complexes.
3. One enzyme may catalyze more than one reaction.
4. Points 2 and 3 may be confounded, such that one multi-enzyme or isoenzyme complex can catalyze more than one reaction (*cf.*, the Bat1 and Bat2 enzymes in *S. cerevisiae*).[34]
5. It is desirable to preserve the interaction degree of the enzymes of the metabolic network, something that one obvious randomization scheme, that of randomizing the identities of the proteins in the protein interaction network relative to the metabolic network, fails to do.

We suggest that the randomization approaches used by Huthmacher *et al.*,[23] do not preserve interaction degree in the protein interaction network, nor do they ensure that the isoenzyme/enzyme complex structure in the metabolic network is retained. We instead sought a randomization approach that avoids all five of the issues listed in order to provide a clear assessment of dPPI and iPPI enrichment.

To do so, it is helpful first to more precisely define the question being asked. First, take the existing metabolic network M and an ensemble of protein interaction networks P_r where, across the entire ensemble of P_rs, each node has the same degree as it does the real protein interaction network P. We define V^M as the edge set of M and V^P as the edge set of P such that x is the number of shared nodes between M and P. The expected number of dPPIs between M and a random network P_r is given by:

$$\sum_{V^M} P(V_{i,j}^{P_r} \in V^{P_r} | V_{i,j}^{M} \in V^{M}) \qquad (1)$$

where $i < x$ and $j < x$. In other words, given each edge i, j in the metabolic network ($V_{i,j}^{M}$), we ask what the probability is that there is an equivalent edge in the ensemble of random protein interaction networks ($V_{i,j}^{P_r}$). When we take the sum over the entire metabolic network of these probabilities, we get the expected number of dPPIs. Obviously, we can calculate the equivalent number for the real network P by replacing the probabilities in (1) with an indicator variable that takes on value 1 if a given edge exists and 0 otherwise.

However, for the random networks it is clear that:

$$P(V_{i,j}^{P_r} \in V^{P_r} | V_{i,j}^{M} \in V^{M}) = P(V_{i,j}^{P_r} \in V^{P_r}) \qquad (2)$$

since the P_r networks do not depend on M. Thus, when we compute (1) for all P_r and compare that distribution to the real number of dPPIs, if we find that the real number of dPPIs falls within range seen in the P_rs, we cannot reject the conclusion that (2) is true of the real network as well. In other words, in that case our null hypothesis, that there are no more shared edges (dPPIs) than would be expected given the structures of M and P, cannot be rejected.

We have previously described a network rewiring procedure for creating the P_rs that *exactly* maintains the interaction degrees of all nodes: only the interaction identities are changed.[59] This algorithm starts with an existing protein interaction network with all self-interactions removed. Each edge in the network is then broken, leaving a "stub" for each of the two nodes it connects (thus there are $2V^P$ such stubs, 2 per edge in the original network). To create a randomized network, we use the following algorithm:

1. Select at random two stubs, A and B, to join.
2. If stub A is from the same protein as is stub B, goto 5
3. If we have already created an interaction A/B in a previous step, goto 5
4. Goto 8
5. If A and B are the first stubs chosen, goto 1
6. Else, break a random existing interaction C/D created at a previous step 8
7. For the pair A/C and for the pair B/D, goto 2
8. Create interaction A/B
9. Goto 1

The result of this algorithm is a randomized network where each node has exactly the same number of edges as it did in the original network: steps 2 and 3 of our procedure prevent the randomization procedure from becoming trapped with a partial network where all remaining potential interactions are forbidden. In the worst case, the randomization procedure will return the original network as the only possible network that allows the observed set of node degrees (*e.g.*, in the case of a fully connected network). Thus, our approach is conservative in the sense that a failure to randomize the PPI network would result in our failure to reject the null hypothesis of no more dPPIs or iPPIs than expected by chance.

Using this protocol to analyze the ten yeast and *E. coli* networks shown in Table 2 leads us to infer an excess of dPPIs ($P < 0.05$), similar to the results of Huthmacher and colleagues.[23,24] However, this approach has a flaw. The protein interaction network contains many non-metabolic proteins. By randomizing the entire network, we implicitly assume that the random networks will have just as many protein interactions between pairs of metabolic proteins as did the original network. If however the fact that two proteins are both enzymes increases the probability that they will share a protein interaction (violating eqn (2) above), the randomized networks will have fewer interactions between pairs of metabolic proteins than did the original network, as is in fact observed for all three taxa ($P < 0.01$). Since dPPIs can only exist for proteins that are both members of the metabolic network, any tendency of the randomized networks to have fewer such interactions will be mistaken for evidence of more dPPIs in the real network. This problem is easily overcome by limiting the randomizations to interactions where both proteins are members of the metabolic network. Importantly, if there were no bias in interactions for metabolic proteins, this subset-based approach would be equivalent to randomizing the entire network.

Conclusions

Our analyses suggest that one role for the cell's protein interactions is to define the appropriate spatial configuration of metabolism. That this process involves indirect interactions suggests that such structures might have a modular nature (*i.e.*, one localization protein can co-localize a number of enzymes with similar roles).

Abbreviations used

PPI	protein–protein interaction
dPPI	direct protein–protein interaction
iPPI	indirect protein–protein interaction
FBA	flux-balance analysis
GO	Gene Ontology

Acknowledgements

We would like to thank Karsten Hokamp for computational support and Michaël Bekaert, Patrick Edger, Corey Hudson and Chris Pires for helpful discussions. ÅPB is supported by Gålöstiftelsen Stipendium för högre utlandsstudier. AMcL is supported by Science Foundation Ireland. GCC is supported by the Reproductive Biology Group of the Food for the 21st Century program at the University of Missouri.

References

1 C. Yanofsky and M. Rachmeler, *Biochim. Biophys. Acta*, 1958, **28**, 640–641.
2 J. Mowbray and V. Moses, *Eur. J. Biochem.*, 1976, **66**, 25–36.

3 B. Räis, F. Ortega, J. Puigjaner, B. Comin, F. Orosz, J. Ovadi and M. Cascante, *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.*, 2000, **1479**, 303–314.
4 F. H. Gaertner, *Trends Biochem. Sci.*, 1978, **3**, 63–65.
5 A. S. Verkman, *Trends Biochem. Sci.*, 2002, **27**, 27–33.
6 P. A. Srere, *Trends Biochem. Sci.*, 2000, **25**, 150–153.
7 C. C. Hyde, S. A. Ahmed, E. A. Padlan, E. W. Miles and D. R. Davies, *J. Biol. Chem.*, 1988, **263**, 17857–17871.
8 E. K. Seppet, T. Kaambre, P. Sikk, T. Tiivel, H. Vija, M. Tonkonogi, K. Sahlin, L. Kay, F. Appaix, U. Braun, M. Eimre and V. A. Saks, *Biochim. Biophys. Acta, Bioenerg.*, 2001, **1504**, 379–395.
9 I. Brandina, J. Graham, C. Lemaitre-Guillier, N. Entelis, I. Krasheninnikov, L. Sweetlove, I. Tarassov and R. P. Martin, *Biochim. Biophys. Acta, Bioenerg.*, 2006, **1757**, 1217–1228.
10 M. E. Campanella, H. Chu and P. S. Low, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 2402–2407.
11 J. W. Graham, T. C. Williams, M. Morgan, A. R. Fernie, R. G. Ratcliffe and L. J. Sweetlove, *Plant Cell*, 2007, **19**, 3723–3738.
12 D. Araiza-Olivera, J. G. Sampedro, A. Mujica, A. Pena and S. Uribe-Carvajal, *FEMS Yeast Res.*, 2010, **10**, 282–289.
13 B. S. Winkel, *Annu. Rev. Plant Biol.*, 2004, **55**, 85–107.
14 S. D. Chuong, A. G. Good, G. J. Taylor, M. C. Freeman, G. B. Moorhead and D. G. Muench, *Mol. Cell. Proteomics*, 2004, **3**, 970–983.
15 V. Saks, N. Beraud and T. Wallimann, *Int. J. Mol. Sci.*, 2008, **9**, 751–767.
16 R. J. Ellis, *Trends Biochem. Sci.*, 2001, **26**, 597–604.
17 G. J. Pielak, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 5901–5902.
18 J. Ovádi and P. A. Srere, *Cell Biochem. Funct.*, 1996, **14**, 249–258.
19 A. Hudder, L. Nathanson and M. P. Deutscher, *Mol. Cell. Biol.*, 2003, **23**, 9318–9326.
20 P. Dolezal, V. Likic, J. Tachezy and T. Lithgow, *Science*, 2006, **313**, 314–318.
21 R. A. Stuart, *J. Bioenerg. Biomembr.*, 2008, **40**, 411–417.
22 J. Ovádi and V. Saks, *Mol. Cell. Biochem.*, 2004, **256–257**, 5–12.
23 C. Huthmacher, C. Gille and H. G. Holzhutter, *Genome Inf. Ser.*, 2007, **18**, 162–172.
24 C. Huthmacher, C. Gille and H. G. Holzhutter, *J. Theor. Biol.*, 2008, **252**, 456–464.
25 P. Durek and D. Walther, *BMC Syst. Biol.*, 2008, **2**, 100.
26 X. Zhu, M. Gerstein and M. Snyder, *Genes Dev.*, 2007, **21**, 1010–1024.
27 J. L. Reed, T. D. Vo, C. H. Schilling and B. O. Palsson, *GenomeBiology*, 2003, **4**, R54.
28 N. C. Duarte, M. J. Herrgård and B. Ø. Palsson, *Genome Res.*, 2004, **14**, 1298–1309.
29 H. Ma, A. Sorokin, A. Mazein, A. Selkov, E. Selkov, O. Demin and I. Goryanin, *Mol. Syst. Biol.*, 2007, **3**, 135.
30 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.
31 A. Wagner and D. A. Fell, *Proc. R. Soc. London, Ser. B*, 2001, **268**, 1803–1810.
32 C. Gancedo and C. L. Flores, *Microbiol. Mol. Biol. Rev.*, 2008, **72**, 197–210.
33 K. P. Byrne and K. H. Wolfe, *Genome Res.*, 2005, **15**, 1456–1461.
34 D. Voet and J. G. Voet, *Biochemistry*, John Wiley & Sons, Inc., Hoboken, NJ, 2004.
35 L. Salwinski, C. S. Miller, A. J. Smith, F. K. Pettit, J. U. Bowie and D. Eisenberg, *Nucleic Acids Res.*, 2004, **32**, D449–451.
36 J. M. Cherry, C. Adler, C. Ball, S. A. Chervitz, S. S. Dwight, E. T. Hester, Y. Jia, G. Juvik, T. Roe, M. Schroeder, S. Weng and D. Botstein, *Nucleic Acids Res.*, 1998, **26**, 73–80.
37 E. I. Boyle, S. Weng, J. Gollub, H. Jin, D. Botstein, J. M. Cherry and G. Sherlock, *Bioinformatics*, 2004, **20**, 3710–3715.
38 H. W. Mewes, K. Heumann, A. Kaps, K. Mayer, F. Pfeiffer, S. Stocker and D. Frishman, *Nucleic Acids Res.*, 1999, **27**, 44–48.
39 J. D. Orth, I. Thiele and B. Ø. Palsson, *Nat. Biotechnol.*, 2010, **28**, 245–248.
40 J. Förster, I. Famili, P. Fu, B. Ø. Palsson and J. Nielsen, *Genome Res.*, 2003, **13**, 244–253.
41 T. Shlomi, M. N. Cabili, M. J. Herrgard, B. O. Palsson and E. Ruppin, *Nat. Biotechnol.*, 2008, **26**, 1003–1010.
42 L. Jerby, T. Shlomi and E. Ruppin, *Mol. Syst. Biol.*, 2010, **6**, 401.
43 J. Ovádi, F. Orosz and S. Hollán, *Mol. Cell. Biochem.*, 2004, **256–257**, 83–93.
44 D. Mochly-Rosen, *Science*, 1995, **268**, 247–251.
45 T. Pawson and J. D. Scott, *Science*, 1997, **278**, 2075–2080.
46 G. B. West, J. H. Brown and B. J. Enquist, *Science*, 1999, **284**, 1677–1679.
47 G. B. West, W. H. Woodruff and J. H. Brown, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**(Suppl 1), 2473–2478.
48 E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai and A. L. Barabasi, *Science*, 2002, **297**, 1551–1555.
49 A. C. Gavin, P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L. J. Jensen, S. Bastuck, B. Dumpelfeld, A. Edelmann, M. A. Heurtier, V. Hoffman, C. Hoefert, K. Klein, M. Hudak, A. M. Michon, M. Schelder, M. Schirle, M. Remor, T. Rudi, S. Hooper, A. Bauer, T. Bouwmeester, G. Casari, G. Drewes, G. Neubauer, J. M. Rick, B. Kuster, P. Bork, R. B. Russell and G. Superti-Furga, *Nature*, 2006, **440**, 631–636.
50 T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori and Y. Sakaki, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 4569–4574.
51 M. A. Aon, B. O'Rourke and S. Cortassa, *Mol. Cell. Biochem.*, 2004, **256–257**, 169–184.
52 T. S. Keshava Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, L. Balakrishnan, A. Marimuthu, S. Banerjee, D. S. Somanathan, A. Sebastian, S. Rani, S. Ray, C. J. Harrys Kishore, S. Kanth, M. Ahmed, M. K. Kashyap, R. Mohmood, Y. L. Ramachandra, V. Krishna, B. A. Rahiman, S. Mohan, P. Ranganathan, S. Ramabadran, R. Chaerkady and A. Pandey, *Nucleic Acids Res.*, 2009, **37**, D767–772.
53 S. Peri, J. D. Navarro, R. Amanchy, T. Z. Kristiansen, C. K. Jonnalagadda, V. Surendranath, V. Niranjan, B. Muthusamy, T. K. Gandhi, M. Gronborg, N. Ibarrola, N. Deshpande, K. Shanker, H. N. Shivashankar, B. P. Rashmi, M. A. Ramya, Z. Zhao, K. N. Chandrika, N. Padma, H. C. Harsha, A. J. Yatish, M. P. Kavitha, M. Menezes, D. R. Choudhury, S. Suresh, N. Ghosh, R. Saravana, S. Chandran, S. Krishna, M. Joy, S. K. Anand, V. Madavan, A. Joseph, G. W. Wong, W. P. Schiemann, S. N. Constantinescu, L. Huang, R. Khosravi-Far, H. Steen, M. Tewari, S. Ghaffari, G. C. Blobe, C. V. Dang, J. G. Garcia, J. Pevsner, O. N. Jensen, P. Roepstorff, K. S. Deshpande, A. M. Chinnaiyan, A. Hamosh, A. Chakravarti and A. Pandey, *Genome Res.*, 2003, **13**, 2363–2371.
54 P. Flicek, B. L. Aken, K. Beal, B. Ballester, M. Caccamo, Y. Chen, L. Clarke, G. Coates, F. Cunningham, T. Cutts, T. Down, S. C. Dyer, T. Eyre, S. Fitzgerald, J. Fernandez-Banet, S. Graf, S. Haider, M. Hammond, R. Holland, K. L. Howe, K. Howe, N. Johnson, A. Jenkinson, A. Kahari, D. Keefe, F. Kokocinski, E. Kulesha, D. Lawson, I. Longden, K. Megy, P. Meidl, B. Overduin, A. Parker, B. Pritchard, A. Prlic, S. Rice, D. Rios, M. Schuster, I. Sealy, G. Slater, D. Smedley, G. Spudich, S. Trevanion, A. J. Vilella, J. Vogel, S. White, M. Wood, E. Birney, T. Cox, V. Curwen, R. Durbin, X. M. Fernandez-Suarez, J. Herrero, T. J. Hubbard, A. Kasprzyk, G. Proctor, J. Smith, A. Ureta-Vidal and S. Searle, *Nucleic Acids Res.*, 2008, **36**, D707–714.
55 Å. Pérez-Bercoff, T. Makino and A. McLysaght, *BMC Evol. Biol.*, 2010, **10**, 160.
56 F. R. Blattner, G. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Colladovides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau and Y. Shao, *Science*, 1997, **277**, 1453–1462.
57 M. Bekaert and G. C. Conant, *Mol. Biol. Evol.*, 2011, **28**, 1111–1121.
58 J. Wright and A. Wagner, *BMC Syst. Biol.*, 2008, **2**, 55.
59 G. C. Conant and K. H. Wolfe, *PLoS Biol.*, 2006, **4**, e109.
60 J. Yoon, A. Blumer and K. Lee, *Bioinformatics*, 2006, **22**, 3106–3108.
61 D. J. Watts and S. H. Strogatz, *Nature*, 1998, **393**, 440–442.