Available online at www.sciencedirect.com
ScienceDirect
BioSystems 88 (2007) 163–172
BioSystems
www.elsevier.com/locate/biosystems

Alternative routes and mutational robustness in complex regulatory networks

Andreas Wagner*, Jeremiah Wright

Department of Biology, MSC03 2020, The University of New Mexico, Albuquerque, NM 87131-0001, United States

Received 22 November 2005; received in revised form 4 May 2006; accepted 12 June 2006

Abstract

Alternative pathways through a gene regulation network connect a regulatory molecule to its (indirect) regulatory target via different intermediate regulators. We here show for two large transcriptional regulation networks, and for 15 different signal transduction networks, that multiple alternative pathways between regulator and target pairs are the rule rather than the exception. We find that in the yeast transcriptional regulation network intermediate regulators that are part of many alternative pathways between a regulator and target pair evolve at faster rates. This variation is not solely explicable by higher expression levels of such regulators, nor is it solely explicable by their variable usage in different physiological or environmental conditions, as indicated by their variable expression. This suggests that such pathways can continue to function despite amino acid changes that may impair one intermediate regulator. Our results underscore the importance of systems biology approaches to understand functional and evolutionary constraints on genes and proteins.

Keywords: Molecular evolution; Systems biology; Gene networks

1. Introduction

Genetic changes in the smallest parts of molecular networks – genes and proteins – can affect the structure of these networks. Conversely, this structure may itself constrain properties of genes and proteins, and the kinds of mutations they can tolerate. To ask how biological networks constrain their parts has become possible with the availability of experimental data on genome-scale metabolic, transcriptional regulation, and protein interaction networks (Forster et al., 2003; Ito et al., 2001; Lee et al., 2002; Uetz et al., 2000; von Mering et al., 2002). If network structure constrains network parts, then an understanding of the evolution of genes and proteins will require an understanding of molecular networks. We here focus on one aspect of network organization, alternative pathways through a genetic network, and how such pathways affect genes in a transcriptional regulation network. The importance of alternative pathways is hinted at by systematic studies on metabolic networks, where alternative pathways of metabolite flow can make a network robust against loss-of-function mutations in enzymes (Edwards and Palsson, 2000; Segre et al., 2002). However, no comparable information exists for any regulatory network.

The evidence for alternative or 'redundant' pathways through regulatory gene networks is mostly anecdotal or circumstantial (Bi et al., 2000; Ho and Satoh, 2003; Kolodner et al., 2002; Lefers et al., 2001; LeRoith, 2000; Morris et al., 1995; Passalaris et al., 1999; Vance and Wilson, 2002; Wang et al., 2002a). This evi-

* Corresponding author. Tel.: +1 505 277 2021; fax: +1 505 277 0304.
E-mail address: wagnera@unm.edu (A. Wagner).

0303-2647/$ – see front matter © 2006 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.biosystems.2006.06.002

dence typically comes from molecular biological studies where different regulatory pathways, sometimes involving overlapping sets of regulators, can influence the same genes or biological processes. For example, the RAD17 gene product of the yeast *Saccharomyces cerevisiae* is involved in detecting DNA damage during different stages of the yeast cell cycle. This protein interacts with different sets of other regulatory proteins, which are part of several redundant regulatory pathways that ensure genome stability (Kolodner et al., 2002). Another such example comes from the degradation of cholesterol in mice. Bile acids, the degradation products of cholesterol, indirectly repress the genes necessary for cholesterol degradation. They do so through several incompletely characterized alternative pathways that involve different transcription factors (Wang et al., 2002a).

Available data on genome-scale biological networks elucidated by currently available functional genomic techniques is by necessity incomplete. It may incorporate only one mode of regulation, such as transcriptional regulation; it may not contain information about all relevant regulatory molecules; and it may contain substantial experimental noise. Even with these caveats, however, such data allows a more systematic exploration of important systems – biological questions than small-scale data on individual pathways. Is the existence of alternative pathways between a regulator and its (indirect) regulatory target the exception or the rule? What are the consequences of such alternative pathways for the function and evolution of intermediate regulators, the regulatory molecules that stand between the regulator and its target? These are some of the questions we will address here.

2. Results and discussion

2.1. Alternative regulatory pathways are not rare

We find that in regulatory networks of even moderate complexity, many pairs of regulatory molecules and their targets are connected by more than one regulatory pathway. Fig. 1a illustrates this notion with data from the transcriptional regulation network of the yeast *S. cerevisiae* (Lee et al., 2002). In this network, a directed edge (link) connects two genes A and B if A encodes a transcriptional regulator that can regulate the expression of B, as indicated by its binding to the regulatory region of B. Regulatory interactions can also be more indirect. That is, a transcriptional regulator can affect the expression of a downstream, regulatory target gene through one or more intermediate regulators. Source and target gene may be connected through one or more alternative pathways involving these intermediate regulators. In the yeast transcriptional regulatory network, the number of such pathways between any origin–target pair varies widely from 1 to over 20 (Fig. 1). There are fewer gene pairs which are connected by many pathways than gene pairs connected by few alternative pathways. The mean number of pathways between any source and target pair is 2.01 (S.D. 2.13). A similar abundance of alternative pathways between genes exists in other regulatory networks, including the transcriptional regulation network of the bacterium *Escherichia coli* and 15 manually curated signal transduction networks (see supplementary methods) with highly diverse functions in processes ranging from cell proliferation and cell death, to the homeostasis of metabolic functions, embryonic development and the vertebrate immune system (Fig. 1b). The mean number of alternative pathways between any two network nodes ranges from 1.12 to 79.6 in these networks. The majority (13/17) of examined networks have a mean number of alternative paths greater than two, and in all of the remaining networks, the coefficient of variation in the number of alternative paths is greater than 1.3 (Fig. 1b).

2.2. Alternative pathways and evolutionary rate

We next examine one candidate possibility for the biological significance of having alternative regulatory pathways between pairs of genes in a cellular network: by providing multiple alternative regulatory routes through a network, such pathways may endow a network with robustness to some mutations (Wagner, 2000). A precedent for this possibility comes from metabolic networks, where some loss-of-function mutations in individual enzymes may have small phenotypic effects because alternative routes around a blocked pathway may exist (Edwards and Palsson, 2000; Papp et al., 2004; Stelling et al., 2002). Similar principles may hold for regulatory networks. Consider a transcriptional regulator whose action on a target gene is mediated by one or more intermediate regulators. If this transcriptional regulator is linked to the target gene through only one regulatory pathway (Fig. 1a, left inset), then mutations that impair the intermediate regulators will abolish proper regulation of the target gene. If this target gene's proper expression is important, then such mutations are likely to reduce fitness. Conversely, if the regulator is linked to the target gene through many alternative pathways, then mutations in intermediate regulators that impair one of these pathways may be less harmful, because proper regulation of the target gene may still be achieved through alternate pathways. The following evidence shows that this link between alternative pathways and mutational robustness indeed exists.

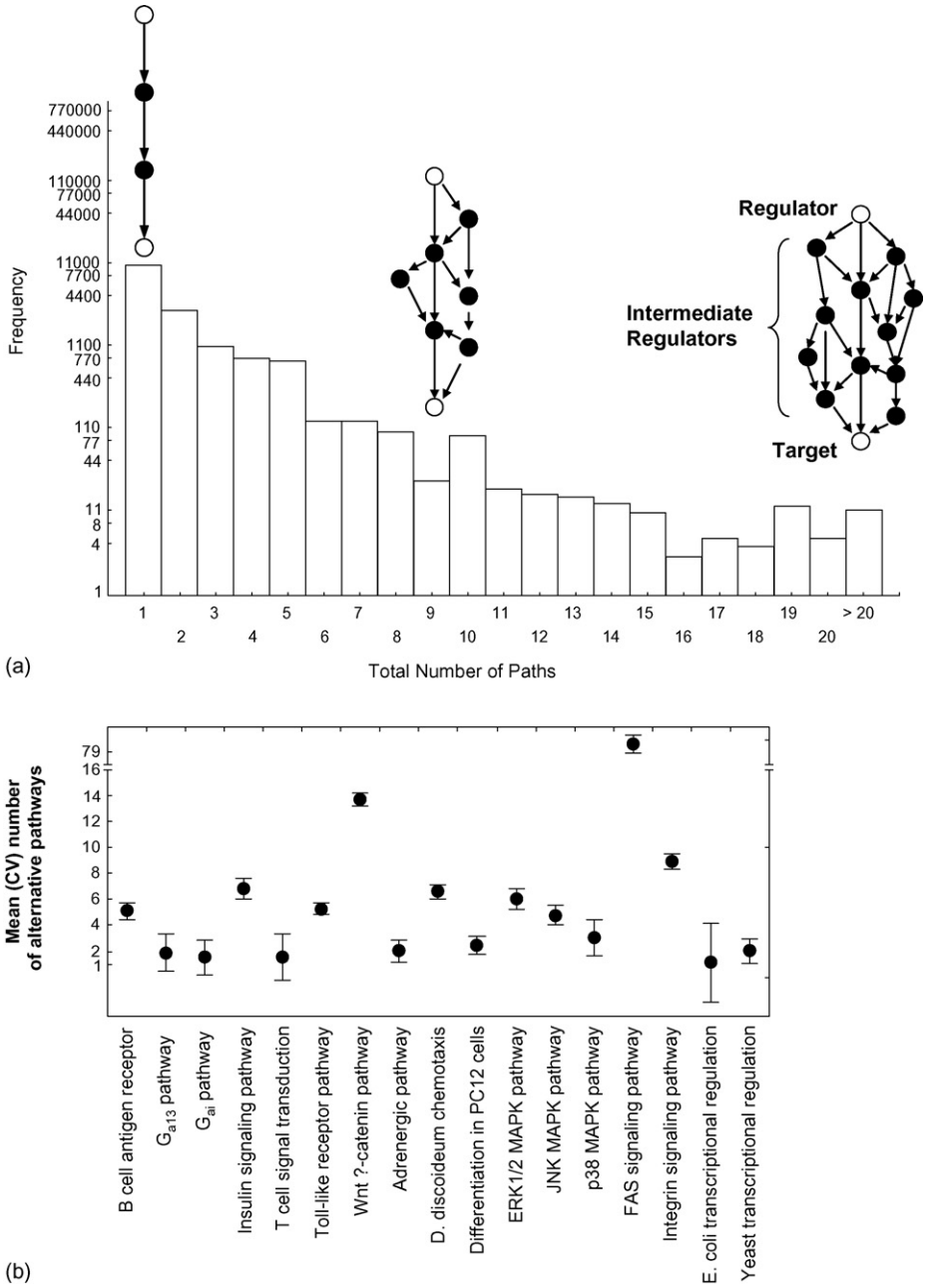

Fig. 1. Many regulator–target pairs in a regulatory network are connected by multiple pathways. (a) Data from the yeast transcriptional regulation network (Lee et al., 2002). The vertical axis shows the number of regulator–target gene pairs connected via the number of alternative pathways indicated on the horizontal axis. The three insets illustrate three hypothetical scenarios of increasing numbers of alternative pathways between a regulator–target gene pair. We refer to the proteins mediating the interaction between a regulator and its target as intermediate regulators. (b) Mean and coefficient of variation of the number of alternative pathways between any two network nodes for two transcriptional regulation networks (Lee et al., 2002; Shen-Orr et al., 2002) and 15 signal transduction networks, as described in greater detail in the supplementary online methods.

The best available indicator of mutational robustness takes into account the mutations that occurred in the evolutionary past of a gene, and that have been tolerated and thus preserved in the evolutionary record. This indicator is the ratio N/S of the fraction N of non-synonymous nucleotide changes, changes that lead to amino acid changes in a gene's product, to the fraction S of silent nucleotide changes, changes that did not lead to amino acid changes. For the vast majority of genes, $N/S < 1$. The smaller this ratio is, the fewer amino acid

changing substitutions a gene can tolerate, and the more evolutionary constrained and less mutationally robust its protein product is (Li, 1997). If mutational robustness is linked to the number of alternate paths between any two-network genes, then intermediate regulators that connect two genes should tolerate more amino acid substitutions, if there are many alternate paths between the two genes. Fig. 2a shows the average evolutionary constraint N/S among all intermediate regulators connecting two gene pairs as a function of the number of paths between the gene pairs. It clearly shows that intermediate regulators evolve faster, and are thus more robust to mutations, if more alternative pathways exist (Kendall's $\tau = 0.58$, $P = 4 \times 10^{-5}$). For this analysis, we used data on the ratio N/S of *S. cerevisiae* genes to their unambiguous orthologs in three other, closely related *Saccharomyces* species (Kellis et al., 2003). Fig. 2b shows a complementary analysis that separates gene pairs according to the shortest path connecting them. It shows the average evolutionary constraint N/S for all intermediate regulators connecting two gene pairs with a given shortest path length, as a function of the total number of paths between the two genes. (The length of a regulatory path between a regulator and target gene pair is one if the regulator regulates the target directly, two if the interaction is indirect and involves one intermediate regulator, three if it involves two intermediate regulators, etc.) Again, we observe a highly significant positive association between evolutionary constraint N/S of intermediate regulators and the number of alternate pathways for the data shown in the figure (Kendall's $\tau = 0.29$, $P = 5 \times 10^{-6}$). This positive association is also statistically significant if shortest distance-categories are considered separately, as long as more than 10 data points are available, that is, for shortest

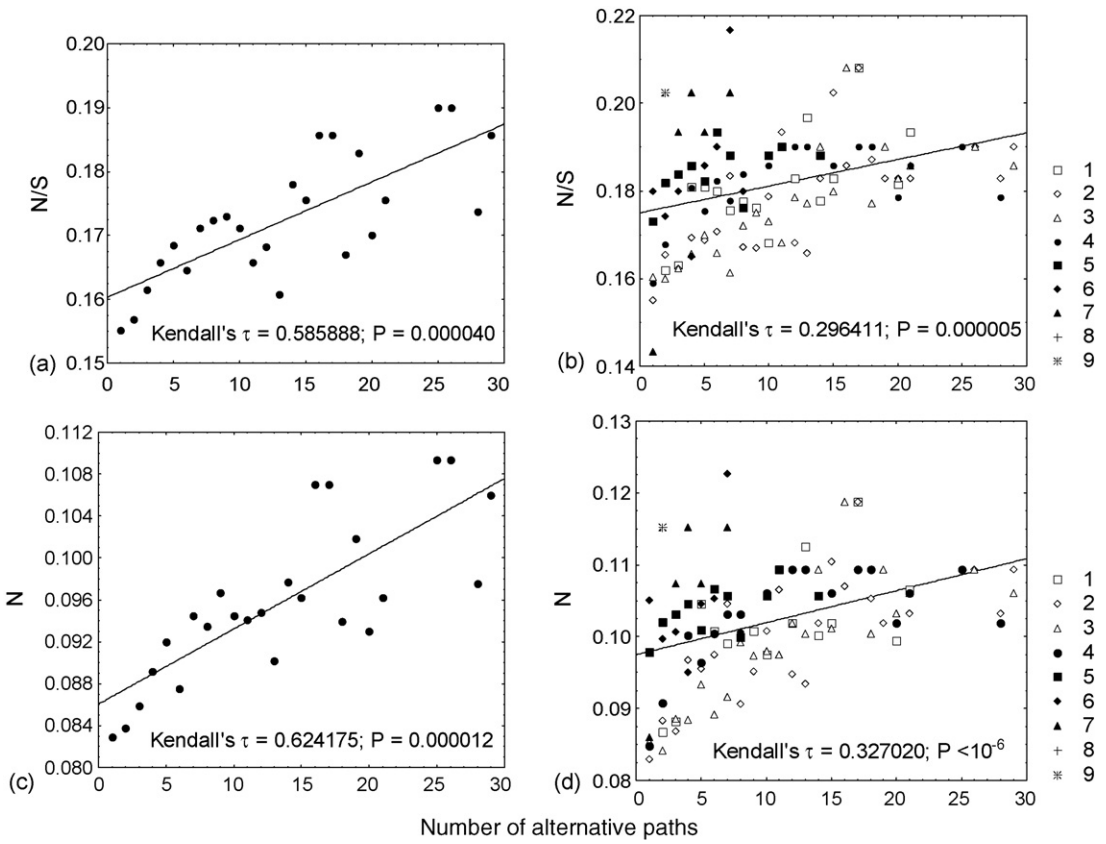

Fig. 2. Alternative pathways and evolutionary rate. The total number of paths between regulator and target gene pairs plotted against (a) the ratio N/S of amino acid to silent divergence as an indicator of mutational robustness, (b) the same ratio N/S, but now regulator–target pairs separated by shortest paths of different lengths are shown separately. That is, the numbered symbols to the right of the panel indicate the shortest path length that separates a regulator–target pair. (c) like (a) but for the number of amino acid substitutions N instead of N/S; (d) like (b) but for the number of amino acid substitutions N instead of N/S. Data on the rates N and S of amino acid and silent divergence represent the average divergence of yeast genes and their unambiguous orthologues in the three closely related *Saccharomyces* species *S. mikatae*, *S. paradoxus*, and *S. bayanus*, as reported in (Kellis et al., 2003).

distances 1 ($\tau = 0.49$, $P = 0.008$), 2 ($\tau = 0.5$, $P = 0.0006$), 3 ($\tau = 0.74$, $P = 2 \times 10^{-6}$), 4 ($\tau = 0.43$, P = 0.008), and 5 ($\tau = 0.49$, $P = 0.04$). If we use non-synonymous (amino acid divergence) N instead of the ratio N/S as an indicator of evolutionary constraint, we obtain the same results (Fig. 2c and d; $\tau = 0.33$ ($P = 0.075$), $\tau = 0.54$ ($P = 0.0002$), $\tau = 0.75$ ($P = 1 \times 10^{-6}$), $\tau = 0.49$ ($P = 0.002$), $\tau = 0.49$ ($P = 0.037$) for shortest distances 1–5, respectively).

Aside from rates of molecular evolution, one can also use the effects of gene deletions (synthetic null mutations) on cell growth as an indicator of mutational robustness. This indicator has multiple disadvantages, because the effect of a gene deletion can usually only be assessed in a small number of environments, and because factors difficult to assess in the laboratory may also influence fitness. We nonetheless carried out a complementary analysis, using the results of a large-scale gene deletion experiment that eliminated more than 4000 yeast genes and tested the growth rate effect of each deletion in five different media (Steinmetz et al., 2002). We asked whether the deletion of intermediate regulators had, on average, a lower maximal effect on cell growth in the five media, if there were many detours around the regulators. If we pool all data on alternative pathways into categories corresponding to the number of alternative pathways between a regulator and target pair, then we find that the deletion of intermediate regulators embedded into 10 or more paths had significantly lower effects on cell growth than the elimination of regulators with fewer than 10 paths ($n_1 = 16$, $n_2 = 9$, Mann–Whitney $U = 140$, $P < 10^{-3}$). However, the result is no longer significant if we analyze each regulator target–pair separately ($n_1 = 248$, $n_2 = 209$, Mann–Whitney $U = 2.7 \times 10^{-4}$, $t = 0.99$, $P > 0.10$). This weak statistical support may not be surprising if one considers the inherent problems of using gene deletion data to measure fitness effects.

We next address three potentially confounding factors in this analysis. In some cellular networks, such as protein interaction and metabolic networks, evidence exists that highly connected proteins evolve at rates different from lowly connected proteins (Dunn and Fraser, 1958; Fraser et al., 2003; Hahn et al., 2004; Jordan et al., 2003). This observation raises the possibility that the evolutionary pattern we observe results from a systematic association between evolutionary rate, regulator connectivity, and alternate pathway number. Specifically, we might see an association between path number and evolutionary constraint, if two genes connected by many alternate pathways preferentially involve intermediate regulators that have many regulatory targets, and if such highly connected regulators show fewer evolutionary constraints (higher N/S) than other regulators. However, highly connected transcriptional regulators – regulating the expression of many target genes – do not evolve more rapidly than less highly connected regulators (Kendall's $\tau = -0.03$, $P = 0.77$; and (Evangelisti and Wagner, 2004)).

A second possible confounding factor is the relation between gene expression and evolutionary rate: genes with a high mRNA expression level evolve more slowly than other genes (Pal et al., 2001). If gene pairs connected by a small number of paths preferentially involve intermediate regulators that are highly expressed, then the association between alternate path number and evolutionary rate might be due to these expression differences. However, we find that the association between the total number of alternate paths between any two nodes and the average mRNA expression level (Wang et al., 2002b) of intermediate regulators in these pathways is not statistically significant (Fig. 3a). In addition, the associ-

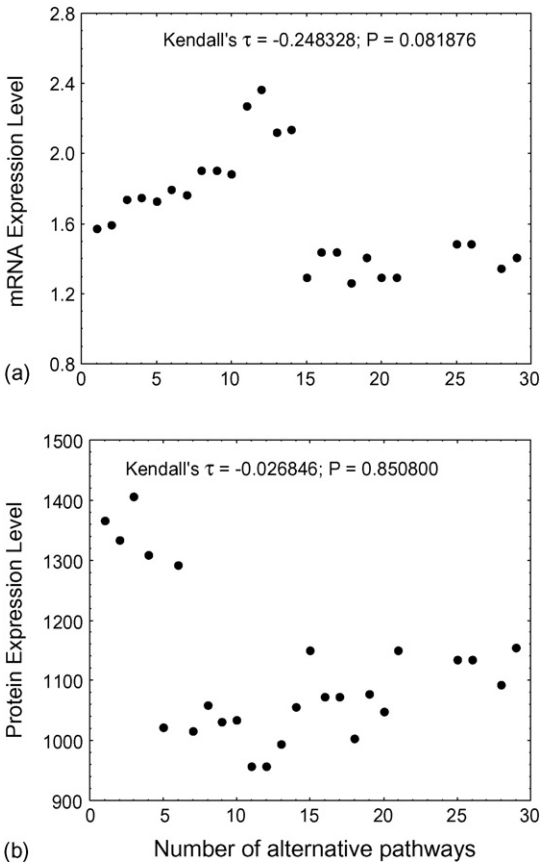

Fig. 3. No significant association between the total number of paths between a regulator and target gene pair and average (a) mRNA (Wang et al., 2002b) or (b) protein (Ghaemmaghami et al., 2003) expression levels of intermediate regulators. Expression levels are expressed in molecules per cell.

ation between protein abundance (Ghaemmaghami et al., 2003) and the number of alternate paths is also not statistically significant (Fig. 3b). Thus, the association between alternate paths and evolutionary rate is not merely the result of differential expression levels of intermediate regulators.

A third possible confounding factor is that alternative pathways may mediate differential regulation of the terminal gene in different environments and physiological conditions. If so, the results of Fig. 2 – a positive association between number of alternative pathways and evolutionary rate – may be merely caused by variation in regulator expression. Specifically, intermediate regulators in pathways with many alternatives may preferentially be expressed or function only under certain environments or physiological conditions. One could view such regulators as more "specialized" than others. The results of Fig. 2 would follow if such specialized regulators can tolerate a higher number of mutations than others. However, it is straightforward to assess whether this is the case: we can ask whether the transcriptional regulators in our analysis do evolve faster if they show higher expression variation, as indicated by their coefficient of variation in expression. This is not the case if N/S is used as an indicator of evolutionary rate, whether we use stress-related variation in expression (Gasch et al., 2000) or normal physiological variation in expression (Chu et al., 1998; Eisen et al., 1998; Spellman et al., 1998) (stress-related: $\tau = -0.11$, $P = 0.56$; normal physiological: $\tau = 0.31$, $P = 0.09$). It also does not hold if amino acid divergence N instead of N/S is used to gauge the rate of evolution (stress-related: $\tau = -0.14$, $P = 0.44$; normal physiological: $\tau = 0.26$, $P = 0.16$). If standard deviation instead of the coefficient of variation is used as a measure of expression variation, then an association is absent if N/S is used as an indicator of evolutionary rate (stress-related: $\tau = 0.31$, $P = 0.09$; physiological: $\tau = 0.33$, $P = 0.07$). An association is present if N instead of N/S is used to gauge the rate of evolution (stress-related: $\tau = 0.48$, $P = 0.007$; physiological: $\tau = 0.46$, $P = 0.01$). However, the use of the standard deviation in this kind of analysis is problematic, because it does not take into account that average expression changes in different environments can vary manifold among different genes.

Taken together, these latter analyses suggest that the variation in evolutionary rates observed in Fig. 2 is not solely due to differential usage of alternative regulators in different environments. However, we note that the apparent mutational robustness suggested by Fig. 2 may be a by-product of other kinds of robustness. For example, genes are subject to substantial expression noise, which can cause large fluctuations in the concentration of a gene product even for moderately to highly expressed genes (Blake et al., 2003; Elowitz et al., 2002; Rao et al., 2002). Robustness to such gene expression noise may entail robustness to mutations that change a gene product's activity. However, currently available data on gene expression noise is too limited to assess whether an association between expression noise of intermediate regulators and the number of paths connecting two genes in a network exist.

2.3. Alternative pathways as an adaptation or a by-product of network organization?

The association between mutational robustness and pathway organization we observe could have two principal ultimate causes. According to the first of them, the increased evolutionary rate of transcriptional regulators reflects an evolutionary adaptation. Specifically, many alternative pathways between a regulator and target gene pair may exist wherever proper regulation of the target gene is especially important to the organism. If mutations or gene expression noise change the activity or concentration of one of the intermediate regulators, then an alternative regulatory pathway could compensate for this change, and still assure proper regulation of the target gene. This hypothesis might explain why not all regulator–target gene pairs have large numbers of alternate pathways between them. If correct, the hypothesis suggests that such target genes should have unique properties that distinguish them from other genes. For example, mutations in such target genes might have more serious consequences for the organism. We do not find any such association: The target genes of a regulator–target gene pair at which many alternative pathways end do not show larger effects of a gene deletion ($\tau = 0.01$, $P = 0.39$), nor do they evolve more slowly than other target genes. ($\tau = -0.02$, $P = 0.26$). There may, however exist other properties of such target genes that have eluded our analysis. An alternative is that in any biological network of minimum size and complexity, some regulator–target gene pairs will be connected by more than one alternative pathway. Wherever a regulator–target gene pair is connected by multiple pathways, the consequence may be mutational robustness of intermediate regulators. In other words, robustness of intermediate regulators to mutations may be a simple consequence of complex network organization.

One assumption in our analysis is that alternative pathways function in similar ways. This is certainly not always true. For example, a pathway responsible

for target repression and one responsible for target induction would both constitute alternatives in a regulatory network. The extent of functional differences between alternative pathways is poorly understood for the genome-scale transcriptional regulation network in *S. cerevisiae*. These differences are dependent on a variety of factors, such as regulator expression levels and environmental conditions. Only detailed experimental work will be able to elucidate the specific functional roles of alternative pathways in a given biochemical regulatory network. We hope that the results presented here will motivate future research in this area.

3. Conclusions

We show that alternative pathways are abundant in more than a dozen biochemical regulatory networks. For the transcriptional regulation network of *S. cerevisiae*, molecular evolution data suggest that such alternative pathways may provide robustness to mutation. Mutational robustness in a network may either be an adaptation in and by itself, or it may have emerged as a by-product of other evolutionary processes. To distinguish between these possibilities remains an important task for future work. Our results demonstrate that systems biology approaches and the analysis of network structure may be essential in understanding the functional and evolutionary constraints that the smallest parts of networks, genes and proteins, are subject to.

4. Methods

4.1. Signal transduction networks

The science signal transduction knowledge environment (http://stke.sciencemag.org/cgi/cm) contains a collection of signal transduction pathways manually assembled by experts on these networks. We analyzed the structure of all 15 signal transduction networks with more than 30 nodes that were available in this repository in May 2004. These networks are the adrenergic pathway (http://www.stke.org/cgi/cm/; CMP_8762), a network that mediates the responses of cells to epinephrine and norepinephrine; the *Dictyostelium discoideum* cAMP chemotaxis network (CMP_7918), which is involved in the aggregation of cells in response to starvation; the differentiation pathway in PC12 cells (CM_8038), a network that mediates the differentiation of a rat adrenal tumor cell line under the influence of nerve growth factor; the extracellular signal regulated kinase 1 and 2 (ERK1/2, or mitogen associated kinase [MAPK] p42 and p44) network (CMP_10705), the c-Jun N-terminal MAPK network (CMP_10827), and the p38 MAPK network (CMP_10958), which are activated by a variety of mitogenic stimuli, differentiation signals, and cellular stresses; the B and T lymphocyte receptor signaling network (CMP_6909 and CMP_7019), which mediate the response of B and T cells to antigens and antigen-presenting cells; the networks that mediate the action of $G_{\alpha 13}$ (CMP_8809) and $G_{\alpha i}$ (CMP_7430), two variants of the α-subunit of heterotrimeric guanine nucleotide binding proteins (G-proteins), which have innumerable functions in cell biological processes; the insulin signaling network (CMP_12069), which modulates the storage and release of energy after nutrient deprivation and nutrient uptake; the mammalian Toll-like receptor networks (CMC_8644), which are involved in the inflammatory response of tissues to microbial infections; the Wnt/β-catenin network (CMP_5533), which influences cell proliferation and other aspects of cell behavior in vertebrates and invertebrates through Wnt proteins, which are secreted glycoproteins; the FAS signal transduction network (CMP_7966), one of whose functions is to trigger apoptosis; and the Integrin signaling network (CMP_6880), which senses the environment in the extracellular matrix and are necessary for cell migration, growth, and survival. Note that, as opposed to transcriptional regulation networks, nodes in all of these networks are heterogeneous: they can represent proteins, small molecules, or ions. A directed edge links node A to node B if A influences the concentration or activity of B.

4.2. Transcriptional regulation networks

For our analysis of the transcriptional regulation network of the yeast *S. cerevisiae*, we used data on likely transcriptional regulatory interactions obtained from a genome-scale chromatin immunoprecipitation analysis (Knop et al., 1999; Lee et al., 2002). In this experiment, 106 epitope-tagged transcriptional regulators were used in three replicate chromatin immunoprecipitation experiments to identify genomic DNA to which these regulators bound (Ren et al., 2000). The immunoprecipitated DNA was hybridized to DNA microarrays containing the regulatory regions upstream of known yeast genes. The fluorescence intensity of a spot (regulatory region) on the array indicates the binding strength of a transcriptional regulator to the regulatory region. This indication of binding is quantitative, but for many analyses, a qualitative (all-none) indication of binding and transcriptional regulation is more useful. The authors thus developed an error model of binding that allowed them to assign a probability or *P*-value of binding for each transcriptional regulator to a gene's regulatory region (Lee et al., 2002). This *P*-value indicates the confidence one has in a factor's binding to a specific DNA region. We here generally follow the authors' suggestion of equating bona fide binding of a transcriptional regulator to a target gene if this *P*-value is smaller than 10^{-3}. This value minimizes the number of false–positive binding interactions, while maximizing the number of true positive regulator–target binding interactions (Lee et al., 2002). Data of this kind is subject to experimental noise and ascertainment bias. However, our results are robust to variation in *P* that corresponds to a 10-fold variation in the number of regulatory interactions (data not shown).

For the transcriptional regulation network of *Escherichia coli*, we used a database of direct transcriptional interactions published by Shen-Orr et al. (2002). This database was compiled from an existing database (RegulonDB) and an extensive literature search; it contains 578 transcriptional regulation interactions among 423 genes or operons, of which 116 encode regulators.

4.3. Gene expression changes in different stressful environmental conditions

To assay the differential expression response of yeast genes to environmental stresses, we used data published by Gasch et al. (2000) for the following conditions: heat shock (25–37 °C, after 30 min), reverse heat shock (37–25 °C, 30 min), H_2O_2 and Menadione exposure, both of which generate reactive oxygen species (60 and 80 min, respectively), dithiothreitol, a reducing agent interfering with protein folding (90 min), diamide, an agent oxidizing sulfhydryl groups (40 min), hyperosmotic shock mediated by 1 M sorbitol (60 min), hypo-osmotic shock mediated by transfer of cells from 1 M sorbitol to medium lacking sorbitol (30 min), amino acid starvation (2 h), nitrogen depletion (1 d), and stationary phase (7 d). Because the expression response to most environmental stresses is transient, we chose a time point (indicated above in parentheses) approximately halfway through the measured response time series as an indicator of change in expression of a gene for any given environmental stressor. We then calculated, for each gene, coefficients of variation (CV, standard deviation/mean expression change × 100) and standard deviations of fold-expression change across the environmental stressors.

4.4. Gene expression changes in different physiological conditions during

To determine the differential expression of yeast genes during different physiological stages, we pooled data that assessed expression changes in yeast genes during three different phases of the life cycle of a yeast cell. These are the diauxic shift – the transition from fermentable to non-fermentable carbon sources (DeRisi et al., 1997) (http://cmgm.stanford.edu/pbrown/explore/array.txt, seven data points) – sporulation (Chu et al., 1998) (http://cmgm.stanford.edu/pbrown/sporulation/additional/spospread.txt, seven data points per gene), and the yeast cell-cycle (Spellman et al., 1998) (http://genome-www.stanford.edu/cellcycle/data/rawdata/combined.txt 72 data points). These data sets consist of relative mRNA expression changes (ratios) as ascertained from cDNA microarray experiments. For the sporulation and diauxic shift data, we normalized expression levels of each gene such that the average logarithmically ($\log_2$) transformed ratio was equal to 0. The cell-cycle data was already in this form. We then pooled the three data sets, and calculated from this data for each gene standard deviations and the coefficients of variation (CV, standard deviation/mean expression change × 100). The CV was calculated by using the absolute value of each data point.

4.5. Growth rates of mutant yeast strains

We utilized results from a genome-scale experiment conducted by Steinmetz and collaborators, which assayed the growth rates of 4706 homozygous diploid yeast deletion strains.

Briefly (Steinmetz et al., 2002), the authors generated a pool containing cells from each deletion strain, and allowed cells in this pool to grow in a variety of media.

These included the rich medium YPD, YPDGE (0.1% glucose, 3% glycerol and 2% ethanol), YPE (2% ethanol), YPG (3% glycerol), and YPL (2% lactate).

The investigators assayed the growth rate of individual strains by hybridizing DNA tags that identified each strain to a suitably designed oligonucleotide microarray. The growth rate thus measured is a growth rate relative to the pool's average growth rate.

We here discuss our analysis of publicly available data from one of two replicate experiments (file 'Regression_Tc1_hom.txt' at http://www-deletion.stanford.edu/YDPM/YDPM_index.html) that reported the growth of homozygous mutant strains grown in the five different media listed above. We here consider two indicators of a gene's deletion effect, the maximum growth rate difference a deletion strain shows between the five media and the pool's average growth rate (Steinmetz et al., 2002), as well as measures of dispersion (standard deviation and coefficient of variation) of the growth rate change in a mutant strain.

4.6. Statistical procedures

We define the set of pathways in a network to be the union of the sets of simple paths and simple cycles. We enumerated all network pathways with a depth-first search algorithm (Mehlhorn and Naher, 1999), and used simple variations of this algorithm to identify intermediate regulators, and to determine the shortest distance between regulator and target pairs.

To determine the statistics of interest here (N/S, etc.), we first grouped regulator–target pairs by the total number of paths connecting them. We then collected the intermediate regulators within each of the groups, and determined the average statistic of interest for these regulators. It should be noted that alternative paths between regulator–target gene pairs are typically not independent of one another. Stated differently, a single regulator can participate as an intermediate in multiple alternative paths. This raises the question: should regulators that occur as intermediates in many alternative paths be weighed differently than regulators that occur as intermediates in few alternative paths? If the number of pathways where a regulator occurs as an intermediate is a biologically important factor, then the answer to this question would be yes. We find however that no significant association exists between degree of pathway involvement as an intermediate and N/S ($\tau = 0.18$, $P = 0.15$). The same is true for maximal deletion effect ($\tau = -0.14$, $P = 0.20$). This evidence warrants our decision to weigh each intermediate regulator equally when calculating averages for alternative path classes.

Acknowledgments

AW would like to acknowledge support through NIH grant GM63882 to the University of New Mexico, the continued support of the Santa Fe Institute, as well as the support of the Institut des Hautes Etudes Scientifique for a sabbatical stay.

References

Bi, E., Chiavetta, J., Chen, H., Chen, G., Chan, C., Pringle, J., 2000. Identification of novel, evolutionarily conserved Cdc42p-interacting proteins and of redundant pathways linking Cdc24p and Cdc42p to actin polarization in yeast. Mol. Biol. Cell 11, 773–793.

Blake, W., Kaern, M., Cantor, C., Collins, J., 2003. Noise in eukaryotic gene expression. Nature 422, 633–637.

Chu, S., Derisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P.O., Herskowitz, I., 1998. The transcriptional program of sporulation in budding yeast. Science 282, 699–705.

DeRisi, J.L., Iyer, V.R., Brown, P.O., 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278, 680–686.

Dunn, R.B., Fraser, A.S., 1958. Selection for an invariant character – 'vibrissa number' – in the house mouse. Nature, 181.

Edwards, J.S., Palsson, B.O., 2000. The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proc. Natl. Acad. Sci. USA 97, 5528–5533.

Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D., 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA 95, 14863–14868.

Elowitz, M., Levine, A., Siggia, E., Swain, P., 2002. Stochastic gene expression in a single cell. Science 297, 1183–1186.

Evangelisti, A., Wagner, A., 2004. Molecular evolution in the transcriptional regulation network of yeast. J. Exp. Zool./Mol. Dev. Evol. 302B, 392–411.

Forster, J., Famili, I., Fu, P., Palsson, B., Nielsen, J., 2003. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. Genome Res. 13, 244–253.

Fraser, H.B., Wall, D.P., Hirsh, A.E., 2003. A simple dependence between protein evolution rate and the number of protein–protein interactions. BMC Evol. Biol. 3, 11.

Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., Brown, P.O., 2000. Genomic expression programs in the response of yeast cells to environmental change. Mol. Biol. Cell 11, 4241–4257.

Ghaemmaghami, S., Huh, W., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., Weissman, J.S., 2003. Global analysis of protein expression in yeast. Nature 425, 737–741.

Hahn, M., Conant, G.C., Wagner, A., 2004. Molecular evolution in large genetic networks: does connectivity equal importance? J. Mol. Evol. 58, 203–211.

Ho, E., Satoh, M., 2003. Repair of single-strand DNA interruptions by redundant pathways and its implication in cellular sensitivity to DNA-damaging agents. Nucleic Acids Res. 31, 7032–7040.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., Sakaki, Y., 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc. Natl. Acad. Sci. USA 98, 4569–4574.

Jordan, I.K., Wolf, Y.I., Koonin, E.V., 2003. Correction: no simple dependence between protein evolution rate and the number of protein-protein interactions: only the most prolific interactors evolve slowly. BMC Evol. Biol. 3, 5.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B., Lander, E., 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423, 241–254.

Knop, M., Siegers, K., Pereira, G., Zachariae, W., Winsor, B., Nasmyth, K., Schiebel, E., 1999. Epitope tagging of yeast genes using a PCR-based strategy: more tags and improved practical routines. Yeast 15, 963–972.

Kolodner, R., Putnam, C., Myung, K., 2002. Maintenance of genome stability in *Saccharomyces cerevisiae*. Science 297, 552–557.

Lee, T., Rinaldi, N., Robert, F., Odom, D., Bar-Joseph, Z., Gerber, G., Hannett, N., Harbison, C., Thompson, C., Simon, I., Zeitlinger, J., Jennings, E., Murray, H., Gordon, D., Ren, B., Wyrick, J., Tagne, J., Volkert, T., Fraenkel, E., Gifford, D., Young, R., 2002. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. Science 298, 799–804.

Lefers, M., Wang, Q., Holmgren, R., 2001. Genetic dissection of the drosophila *Cubitus interruptus* signaling complex. Dev. Biol. 236, 411–420.

LeRoith, D., 2000. Editorial: insulin-like growth factor I receptor signalling—overlapping or redundant pathways? Endocrinology 141, 1287–1288.

Li, W.-H., 1997. Molecular Evolution. Sinauer, Massachusetts.

Mehlhorn, K., Naher, S., 1999. LEDA: A Platform for Combinatorial and Geometric Computing. Cambridge University Press, Cambridge, UK.

Morris, T., Reed, K., Cronan, J., 1995. Lipoic acid metabolism in *Escherichia coli*—the lplA and lipB genes define redundant pathways for ligation of lipoyl groups to apoprotein. J. Bacteriol. 177, 1–10.

Pal, C., Papp, B., Hurst, L.D., 2001. Highly expressed genes in yeast evolve slowly. Genetics 158, 927–931.

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

Passalaris, T., Benanti, J., Gewin, L., Kiyono, T., Galloway, D., 1999. The G(2) checkpoint is maintained by redundant pathways. Mol. Cell. Biol. 19, 5872–5881.

Rao, C., Wolf, D., Arkin, A., 2002. Control, exploitation and tolerance of intracellular noise. Nature 420, 231–237.

Ren, B., Robert, F., Wyrick, J., Aparicio, O., Jennings, E., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T., Wilson, C., Bell, S., Young, R., 2000. Genome-wide location and function of DNA binding proteins. Science, 290.

Segre, D., Vitkup, D., Church, G., 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. USA 99.

Shen-Orr, S., Milo, R., Mangan, S., Alon, U., 2002. Network motifs in the transcriptional regulation network of *Escherichia coli*. Nature Gen. 31, 64–68.

Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D., Futcher, B., 1998. Comprehensive identification of cell-cycle regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. Mol. Biol. Cell 9, 3273–3297.

Steinmetz, L., Scharfe, C., Deutschbauer, A., Mokranjac, D., Herman, Z., Jones, T., Chu, A., Giaever, G., Prokisch, H., Oefner, P., Davis, R., 2002. Systematic screen for human disease genes in yeast. Nature Gen. 31, 400–404.

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

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., QureshiEmili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M.J., Johnston, M., Fields, S., Rothberg, J.M., 2000. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature 403, 623–627.

Vance, J., Wilson, T., 2002. Yeast Tdp1 and Rad1-Rad10 function as redundant pathways for repairing Top1 replicative damage. Proc. Natl. Acad. Sci. USA 99, 13669–13674.

von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S., Bork, P., 2002. Comparative assessment of large-scale data sets of protein–protein interactions. Nature 417, 399–403.

Wagner, A., 2000. Mutational robustness in genetic networks of yeast. Nature Gen. 24, 355–361.

Wang, L., Lee, Y., Bundman, D., Han, Y., Thevananther, S., Kim, C., Chua, S., Wei, P., Heyman, R., Karin, M., Moore, D., 2002a. Redundant pathways for negative feedback regulation of bile acid production. Dev. Cell 2, 721–731.

Wang, Y.L., Liu, C.L., Storey, J.D., Tibshirani, R.J., Herschlag, D., Brown, P.O., 2002b. Precision and functional specificity in mRNA decay. Proc. Natl. Acad. Sci. USA 99, 5860–5865.