

Department of Molecular Genetics  
Weizmann Institute of Science, Rehovot, Israel

Thesis submitted for the degree of Doctor of Philosophy

**Subject:**

**The evolution and utilization of genetic redundancy**

**By: Ran Kafri**

**Advisor: Dr. Yitzhak Pilpel,**

**Advisor: Prof. Doron Lancet,**

Submitted to the scientific council of the Weizmann Institute of Science

May 2006

<b>Introduction:</b> .....	4
Redundancy and evolutionary change .....	4
Dispensability of duplicates – a fingerprint of functional redundancy? .....	6
Dissecting duplicate dispensability .....	7
The preservation and utilization of functional redundancy .....	8
Transcriptional reprogramming in genetic backup circuits .....	11
Genetic backup of hubs in the protein interaction network: evidence for evolutionary selection of redundancy .....	13
The regulatory utilization of redundancy .....	14
Statistics of biological and non-biological molecular associations. ....	16
Conclusions .....	18

**Abstract:** Genetic networks demonstrate a remarkable capacity to carry-out precise regulatory programs in the face of habitat variations, stochasticity of the internal cellular environment and genetic perturbations. In parallel, genomes contain tremendous amounts of redundancies, typically associated with duplicated genes (paralogs). It is thus an intriguing possibility that such functional overlaps have specifically evolved for the role of downplaying the effects of mutations and stochasticity and contributing to phenotypic robustness. Following this notion, I applied genome wide analyses in yeast gene duplicates and demonstrated that redundant pairs are often transcriptionally responsive to each others' intactness. Furthermore, I provide evidence suggesting that these '*responsive backup circuits*' are preferentially associated with protein network hubs and master regulators. To interpret these results I formally describe the kinetics of these genetic circuitries and solve both for their dynamics and their steady states. Relying on these I show how '*responsive backup circuits*' may function to minimize the consequence of protein dosage fluctuations that arise from low molecule number and stochasticity of gene expression. Our conclusions, thus, challenge the view that redundancies are simply leftovers of ancient duplications and suggest them as an additional component of the sophisticated machinery of cellular regulation.

## **Introduction:**

The association between gene duplicates and functional redundancy is long considered common knowledge and is continuously encountered in the literature (Nowak, Boerlijst et al. 1997; Krakauer and Nowak 1999; Kafri, Levy et al. 2006). The recurring observation is that knockouts of genes that have close sequence homologs often result with significantly less severe phenotypes than expected given their sometimes crucial functions (Kirschner and Gerhart 1998; Wagner 2000; Hartman, Garvik et al. 2001; Gu, Steinmetz et al. 2003). In fact, because of this, elucidation of gene functions has often required the systematic disruption of whole families of redundant partners (Braun and Arnold 1995; Schwarz, Alvarez-Bolado et al. 1997; Zhang, Tessaro et al. 2003; Pearce, Senis et al. 2004; Enns, Kanaoka et al. 2005). Examples typifying this phenomenon include the gene pair *GSL1* and *GSL5* having an “essential yet redundant role for plant and pollen development” (Enns, Kanaoka et al. 2005) or *Vav1* and *Vav3* reported to have “critical but redundant roles in mediating platelet activation” (Pearce, Senis et al. 2004) (for other examples see REF (Kirschner and Gerhart 1998)). An immediate consequence of redundancy, emphasized by these reports, is the decoupling of the essentiality of *gene functions*, which may be “critical” or “essential”, from the essentiality of the genes themselves (Zhang, Tessaro et al. 2003). In other words, biochemical functions that are vitally essential may be redundantly performed by several genes, each of which appears separately dispensable.

## **Redundancy and evolutionary change**

Currently, there have only been a limited number of published works specifically addressing the roles and implications that may be associated with redundancy in the context of the biological system. In fact, from an evolutionary perspective, the possibility that certain biochemical functions may have evolved to specifically exploit redundancies seems apparently unlikely (Taylor and Raes 2004). The reasons for this derive from the notion that, ideally, functionally backed-up genes should not be associated with any mutation phenotype and must therefore accumulate mutations at random (Lynch and Conery 2000). Based on this, redundancy was suggested as the fuel allowing evolutionary change and consequently “burning out” in the process (Ohno 1970; Taylor and Raes 2004).

Ever since it was initially suggested, the notion associating redundancy with evolutionary novelty has been demonstrated repeatedly (Gerhart and Kirschner 1997; Taylor and Raes 2004). Particularly important is a seminal work by Lynch and Conery (Lynch and Conery 2000) where, by associating the synonymous substitution rate with evolutionary time, an accelerated evolution was demonstrated to transiently occur immediately after the duplication event. Following this observation, the authors then continue and define the three possible fates of gene duplicates as: (i), loss of function of one or both duplicate partners (nonfunctionalization); (ii), divergence leading to the partition of the functions of the ancestral, preduplicated, gene between the newly formed duplicates (subfunctionalization); and (iii), the evolution of new functions (neofunctionalization). These three alternative evolutionary trajectories are more easily conceptualized in cases where discrete gene functions can be enumerated. For example, the partial overlap of duplicates in their sets of cis-regulatory motifs was interpreted to be the result of incomplete subfunctionalization (Papp, Pal et al. 2003). Similar interpretations were also made for downstream targets of transcription factors (Foster, Kauffman et al. 2006), RNA expression profiles of duplicates (Ferris and Whitt 1979; Gu, Nicolae et al. 2002; Makova and Li 2003), and physical interaction partners of proteins (Wagner 2001). In all cases the functional overlap of duplicates was shown to generally decrease with time (Ferris and Whitt 1979; Wagner 2001; Gu, Nicolae et al. 2002; Brookfield 2003; Conant and Wagner 2003; Makova and Li 2003; Taylor and Raes 2004). This is also observed by correlating the dispensability of duplicates with their synonymous substitution rate, a measure routinely used to approximate divergence time (Gu 2003; Gu, Steinmetz et al. 2003). Following this latter approach it was demonstrated that the proportion of dispensable paralogs decreases with time ( $K_s$ ) (Kafri, Bar-Even et al. 2005).

An extreme example of nonfunctionalization describes the complete deletion of one gene copy resulting with a return to a singleton state. The existence of this phenomenon was verified in a seminal study inferring a whole genome duplication followed by massive gene loss to have occurred *S. Cerevisiae* during the last 100 MYR (Wolfe and Shields 1997; Kellis, Birren et al. 2004).

### **Dispensability of duplicates – a fingerprint of functional redundancy?**

While the association of redundancy with evolutionary change is well established, it is still unclear whether divergence is the single and inevitable fate of all redundant pairs that escaped nonfunctionalization. An alternative may be that the functional overlap of some particular gene pairs may be associated with physiological functionalities that provide its evolutionary preservation. A tempting suggestion is that redundancies confer the organism with sustainability to mutations and gene loss. This idea, while serving a working hypothesis of geneticists for many years, has only recently been systematically demonstrated (Brookfield 2003; Gu, Steinmetz et al. 2003; Conant and Wagner 2004). What these authors showed was that knockouts of genes with paralogous partners are less likely deleterious than knockouts of singletons. In other words, sequence similarity of duplicates is associated with apparent gene dispensability. Additionally important, this analysis allowed, for the first time, to calculate a quantitative assessment of the proportion of redundant gene duplicates, resulting with a lower bound estimate of 25% (Gu, Steinmetz et al. 2003). The rationale was that the difference between the dispensability of duplicates to that of singletons is the outcome of redundancies and is quantitatively associated with its prevalence.

These estimates, while may partially explain the robustness of phenotypes to mutations, seem to challenge our fundamental understandings of the clockwork of molecular evolution. Specifically, if redundancy is inevitably instable, how can such a large proportion of redundant pairs be accounted for? One possibility maintains that the source of error lies with the interpretation associating the dispensability of duplicates with overlapping functions. It may be, for example, that duplicates are dispensable simply because dispensable genes are more likely to retain duplicated partners (He and Zhang 2006). In fact, this notion is inspired by examples of genes (mainly protein complex members) (Veitia 2002; Papp, Pal et al. 2003; Veitia 2004) whose protein dosage fluctuations are associated with deleterious phenotypes, which in turn suggests similar results for their duplications. Yet, a drawback of this hypothesis is that it requires the unsupported assumption that essential genes are significantly more intolerant to dosage fluctuations than dispensable ones. An alternative explanation holds that duplicate genes are retained not by the virtue of selection but rather through a mode of copy correction by non-homologous recombination (Pyne, Skiena et al. 2005).

### **Dissecting duplicate dispensability.**

Despite the higher proportions of dispensable genes among the duplicates, most genes (e.g. 80% in *S. Cerevisiae*) appear dispensable regardless of whether they do or do not have a duplicate partner. In fact, it is likely that the majority of duplicates and singletons simply seem expendable because their functions were not required in particular experiments (Gu, Steinmetz et al. 2003; Papp, Pal et al. 2004). Stemming from this is the challenge to distinguish between duplicates whose dispensability is *intrinsically* associated with their duplicated state to those whose dispensability is “*coincidental*”. Progress along these lines has recently been established by work demonstrating that the phenotypes of duplicated genes are correlated with factors related to their wild-type regulation and interactions (Kafri, Bar-Even et al. 2005; Kafri and Pilpel 2006). Specifically, I showed that the dispensability of duplicates is exclusively associated with pairs that are on average regulated differently (Kafri, Bar-Even et al. 2005; Kafri, Levy et al. 2006). In fact, by partitioning the duplicates according to their average co-expression, Kafri et al. have demonstrated that while dissimilar expression corresponded to an enrichment of over 90% non-essential genes, deletions of similarly expressed paralogs almost always results with lethality (Kafri, Bar-Even et al. 2005).

To elaborate on this finding, I separately examined the degree of duplicate co-expression in individual experimental conditions (Kafri, Bar-Even et al. 2005). This more refined examination revealed that co-expression is often a condition dependent phenomenon. More surprisingly, duplicate pairs that display a strong co-expression in some conditions and yet are regulated differently in most others, were those most likely to be dispensable. This result was also supported by *cis*-regulatory motif analysis of the duplicate pair members (Kafri, Bar-Even et al. 2005). Thus, the conclusion was that dispensable duplicates were those that exhibit a condition-dependent rather than a constitutive co-expression, and on average, are differently expressed in wild-type (Kafri, Bar-Even et al. 2005). In contrast to that, deletion phenotypes were found to be condition independent in the sense that similar phenotypes were revealed irrespective of experimental conditions. These observations were interpreted by hypothesizing a transcriptional reprogramming mechanism that allows genes to alter their expression profiles in response to the deletion of their duplicate partners. More intriguingly, these results were exclusively associated with

duplicates corresponding to non-recent duplications, with recent duplicates being almost exclusively dispensable (Kafri, Bar-Even et al. 2005).

The variability of the phenotypes associated with different genes, irrespective of whether they are singletons or duplicates, is largely attributed to a variability in the extent to which their functions are required for cell viability (Papp, Pal et al. 2004). This variability, in turn, is partially reflected by the connectivities of gene products in the protein interaction network (Jeong, Mason et al. 2001). In fact, the number of protein-protein interactions in which a given gene product is involved is highly predictive of its dispensability (Jeong, Mason et al. 2001). This, in turn, in the absence of genetic backup, is correlated with essentiality of functions. In line with this, it was recently demonstrated that the “*intrinsic duplicate dispensability*” is preferentially associated with protein network hubs (Kafri and Pilpel 2006). In fact, the difference between the proportion of dispensable duplicates to that of singletons was shown to steadily increase with increasing connectivity in the protein interaction network (Kafri and Pilpel 2006). For example, duplicates with a degree higher than 10 contain 5 times higher proportions of dispensable genes than do singletons, while those with a low degree are largely as dispensable as singletons. From the perspective of the inferred essentiality of gene functions, this result suggests that among duplicates, dispensability is associated with greater probability with the essential functions.

### **The preservation and utilization of functional redundancy**

To summarize, advances extending from computational dissections of genomic, transcriptional and phenotypic data have recently made it possible to isolate the specific genes giving rise to the increased dispensability of duplicates (Gu, Steinmetz et al. 2003; Kafri, Bar-Even et al. 2005; Kafri, Levy et al. 2006; Kafri and Pilpel 2006). For example, while the dispensability of a given duplicate may be coincidental and unrelated to its duplicated state, this possibility becomes unlikely when specifically considering duplicates constituting protein network hubs (Jeong, Mason et al. 2001; Kafri and Pilpel 2006). Based on this approach it was demonstrated that the majority of genes intrinsically associated with duplicate dispensability are implicated by previous reports to have functionally redundant homologs (Kafri, Bar-Even et al. 2005; Kafri, Levy et al. 2006; Kafri and Pilpel 2006). Furthermore, as anticipated from computational data analysis (Gasch, Spellman et al. 2000; Kafri, Bar-Even et al. 2005), most redundant partners are not co-regulated.



Most interestingly, the association of functional redundancy with protein network hubs (Kafri and Pilpel 2006) almost imperatively suggests its evolutionary conservation. In line with this, it has also been observed that the majority of redundant duplicates are the consequence of most ancient duplications (Kafri, Bar-Even et al. 2005). Furthermore, duplicates have been shown by three independent studies to be associated with slowly evolving genes (Kondrashov, Rogozin et al. 2002; Davis and Petrov 2004; Jordan, Wolf et al. 2004). Together, this evidence suggests that redundancy may have functional implications and serve advantageous functionalities allowing their selection. In other words, in contrast to contemporary thought, it seems that genetic networks have evolved to exploit benefits associated with redundancies. On the other hand, if redundancy is simply more of the same, what beneficial role may it serve and how is it exploited?

The possibility that redundancy has been maintained for “backing-up” against mutations seems unlikely, mainly because the mutation frequency, in most genes, is lower than organism life time (Nowak, Boerlijst et al. 1997; Krakauer and Nowak 1999). In contrast, non-genetic perturbations causing protein dosage fluctuations are highly frequent, reflecting the stochasticity of the intracellular environment (Elowitz, Levine et al. 2002; Swain, Elowitz et al. 2002). While for some biochemical subsystems such fluctuations may have no important implications, for others they may be intolerable (Vilar, Kueh et al. 2002; Fraser, Hirsh et al. 2004; Raser and O'Shea 2004). An appealing possibility is that genetic networks employ redundant pairs to gain control over undesired fluctuations. Enquiring about this possibility, I exhaustively searched the literature for pairs of duplicates with redundancy that may be associated with exploited functionalities (Kafri, Levy et al. 2006). I relied on two clues to indicate a function's direct benefit from existing redundancy. First is the evolutionary conservation of the functional overlap and second is a non-trivial regulatory design that utilizes it (Kafri, Levy et al. 2006). By collecting pairs answering both these criteria I demonstrated that redundant genes are often regulated by a specialized regulatory circuitry which is responsive to their intactness. These genetic circuits, termed *responsive backup circuits*, encode the up-regulatory response of one duplicate, in case its partner was mutationally inactivated. Our conclusion was that RBCs confer regulatory precision by balancing the stochasticity of one gene with a reciprocally regulated redundant partner, thus maintaining constancy of function. Relying on modeling of particular example systems I demonstrated the plausibility of

this possibility by showing that RBC can maintain constancy. I thus suggest that at in some of the cases damping the effect of non-genetic noise may have served as the evolutionary deriving force behind the preservation of redundant duplicates, and backup against mutations in such cases may be obtained as an evolutionary by product in the extreme case where a gene is not just fluctuating, but rather mutated and even deleted altogether.

## **Transcriptional reprogramming in genetic backup circuits<sup>1</sup>.**

Gene duplicates and paralogous gene families have been proposed to acquire organisms with robustness against mutations (Gu, Steinmetz et al. 2003). The assumption is that a functional overlap of these genes enables one paralog to substitute for the function of its partner in case the later was mutated. An essential prerequisite of this model is that the functionally homologous genes respond to the same regulatory inputs, otherwise the reaction they carry out may not be activated in conditions where it is required. In contrast to this, our analyses (Kafri, Bar-Even et al. 2005) and others (Gasch, Spellman et al. 2000) have suggested that isozymes and other functionally redundant genes are not correlated in their expression profiles and are regulated differently.

To resolve this inconsistency and better characterize functional compensatory relationships of redundant genes I comparatively examined in this paper the regulatory associations within the *S. cerevisiae* transcriptional regulatory network together with transcriptional data and knockout phenotypes of yeast genes (Kafri, Bar-Even et al. 2005). Our results indicated that these functionally redundant genes are often cross-regulated in a circuitry that enables one gene to transcriptionally respond to the deletion of the other. Thus the paradigm that emerges is that genes that are functionally redundant are often not independently controlled but rather they are regulated by a system that both monitors and responds to their intactness. There are three general regulatory designs that may provide such a response; the feedforward regulation, feedback regulation and a direct regulatory interaction between the redundant gene products. Motivated by the understanding that such 'Responsive Backup Circuits' could not have evolved for the sake of buffering gene-specific mutations, I reviewed many such circuits in search of common denominators that may explain their functional advantage in wild-type.

I find that responsive backup circuits fall into two main categories which are metabolism, exemplified by the Acs1 and Acs2 isozymes in yeast (van den Berg, de Jong-Gubbels et al. 1996), and developmental pathways, exemplified by the MyoD/Myf-5 (Rudnicki, Braun et al. 1992) regulators of skeletal muscle development.

---

<sup>1</sup> Kafri, R., A. Bar-Even, et al. (2005). "Transcription control reprogramming in genetic backup circuits." *Nat Genet* 37(3): 295-9.

I further suggest that this responsive circuitry regulating genomic redundancies functions to dampen gene dosage fluctuations.

# Transcription control reprogramming in genetic backup circuits

Ran Kafri, Arren Bar-Even & Yitzhak Pilpel

A key question in molecular genetics is why severe mutations often do not result in a detectably abnormal phenotype. This robustness was partially ascribed to redundant paralogs<sup>1,2</sup> that may provide backup for one another in case of mutation. Mining mutant viability and mRNA expression data in *Saccharomyces cerevisiae*, we found that backup was provided predominantly by paralogs that are expressed dissimilarly in most growth conditions. We considered that this apparent inconsistency might be resolved by a transcriptional reprogramming mechanism that allows the intact paralog to rescue the organism upon mutation of its counterpart. We found that in wild-type cells, partial coregulation across growth conditions predicted the ability of paralogs to alter their transcription patterns and to provide backup for one another. Notably, the sets of regulatory motifs that controlled the paralogs with the most efficient backup activity deliberately overlapped only partially; paralogs with highly similar or dissimilar sets of motifs had suboptimal backup activity. Such an arrangement of partially shared regulatory motifs reconciles the differential expression of paralogs with their ability to back each other up.

Functionally redundant gene duplicates are inherently evolutionarily unstable; consequently, in many duplications, one of the duplicates is silenced<sup>3,4</sup>. Retention of duplicates over long evolutionary time scales was therefore suggested to require either degenerative subfunctionalizing mutations or introduction of new functions<sup>3,5-8</sup>. We aim here to understand both the relevance of transcription regulation to duplicate retention in evolution and its role in controlling expression of genes that provide backup in case of mutation.

Mining single gene-knockout phenotype data and annotations of molecular functions of all yeast genes, we found high correlation between the essentiality of genes and the similarity of molecular function between themselves and their paralogs (Supplementary Figs. 1 and 2 online). We also found that only 4% of the dispensable paralogs did not colocalize<sup>9</sup> in the same organelles (Supplementary Fig. 3 online). These observations corroborate the notion that 'dispensability' may be explained by backup between paralogs.

*A priori*, it might seem that backup requires the paralogs' mRNAs to be coregulated. To examine this possibility, we calculated, for each pair

of paralogs, 40 correlation coefficients of mRNA expression corresponding to 40 different experiments. We define the means and standard deviations of such correlations, for each pair, as their mean expression similarity and partial coregulation (PCoR) values, respectively. We refer to the standard deviation of the correlations as PCoR because its value is high for pairs that have interchangeably high and low correlations across different conditions. Figure 1 shows the proportion of dispensable genes in sets of gene pairs versus their mean expression similarity and PCoR. We inspected close and remote paralogous pairs separately and found markedly different trends. Among remote paralogs, we found that the essentiality of coexpressed pairs was very high, implying that there is little backup activity among them. In remote pairs, backup was most efficient among transcriptionally noncorrelated pairs, as their essentiality was substantially lower than that of single genes. Supplementary Figures 4 and 5 online show the increase in protein-protein interaction among paralogs and the decrease in similarity of Gene Ontology-annotated molecular function between them, respectively, as a function of coexpression. These results provide a potential explanation for the observed decrease in backup capacity with increased coexpression. In contrast to remote pairs, close pairs showed an almost opposite, more intuitive trend, in which dispensability increased somewhat with expression similarity (in agreement with refs. 1,10).

Backup among naturally dissimilarly expressed genes A and B may suggest that, upon mutation in gene A, expression of gene B is reprogrammed to acquire a profile that is similar to the wild-type expression profile of gene A. Such reprogramming has been experimentally verified for the Acs1 and Acs2 isoenzymes. Wild-type Acs1 is subject to glucose repression<sup>11</sup> (Fig. 1), but upon deletion of Acs2, the repression of Acs1 is relieved, and Acs1 acquires an Acs2-like responsiveness to glucose<sup>12</sup>. Despite dissimilar expression, the two genes share a promoter motif (CSRE) and also have unique motifs<sup>12</sup>. As befits a genuine backup circuit, Acs1 and Acs2 are synthetically lethal<sup>11</sup>. Additional examples of reprogramming in response to mutations in prokaryotes, yeast and mammals are given in Supplementary Note online.

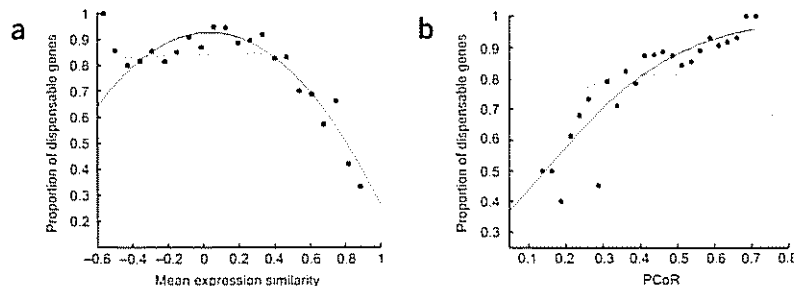
In search for a mechanism that may regulate switching between dissimilar and similar expression in response to mutation, we examined the dependence of gene essentiality on PCoR. We asked whether backup occurs among paralogs that show high PCoR in

Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel. Correspondence should be addressed to Y.P. (pilpel@weizmann.ac.il).

Published online 20 February 2005; doi:10.1038/ng1523



**Figure 1** Dependence of backup on expression similarity between paralogs. Proportion of dispensable genes as a function of the mean expression similarity (a) and PCoR (b) in sets of paralogs. Results are shown only for genes having remote paralogous partners ( $K_s > 1$ ). Results for close pairs ( $K_s < 1$ ) showed a marginally significant opposite trend (Supplementary Figs. 11 and 12 online). The red line represents a quadratic regression fit (a), scoring an adjusted  $R^2$  value of 0.83, and a logistic regression fit (b) with a  $P$  value of  $10^{-25}$ . To exclude the possibility that the trend in a reflects a tendency for genes that belong to major expression clusters to be essential, we repeated the analysis using random pairing of genes and observed a nonsignificant trend (blue). Examination of remnants of whole-genome duplication<sup>26</sup> showed similar trends to that observed with all remote pairs, but with marginal significance.



the wild type. We reasoned that because PCoR represents the ability to switch between similar and dissimilar expression profiles in a condition-dependent manner, it may be predictive of switching between similar and dissimilar expression in response to mutation. We found that PCoR was a very strong predictor of backup (Fig. 1b).

We next investigated the promoter architecture of backup-providing paralogs. The possibility that the partial overlap in the sets of regulatory motifs controlling *Acs1* and *Acs2* accounts for their wild-type differential expression and for reprogramming upon mutation prompted us to inspect the similarity of motif content of all paralogs. To quantify the extent to which promoters of paralogs are arranged to obtain partial coexpression, we defined  $O$ , a normalized measure of the overlap between the sets of promoter motifs that regulate two genes:

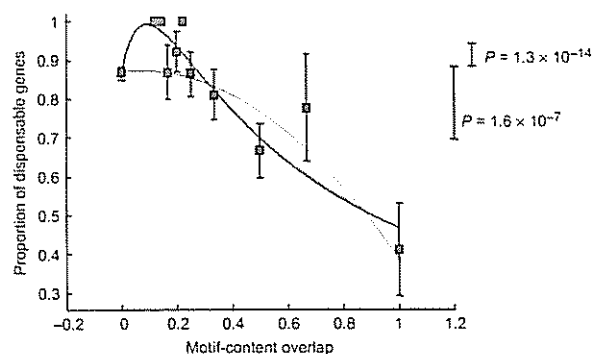
$$O = \frac{|m_1 \cap m_2|}{\max(|m_1|, |m_2|)},$$

where  $m_1$  and  $m_2$  are the sets of motifs that regulate genes 1 and 2, respectively, and  $|x|$  is the size of a set  $x$ . By plotting gene dispensability versus motif-content overlap  $O$ , we found that maximal backup coincided with intermediate levels of motif sharing (Fig. 2). Pairs with high or low promoter similarity had suboptimal backup activity. These observations confirm that optimal backup is obtained when two paralogs share some, but not all, motifs. We propose that the unique motifs of each paralog provide differential expression in the wild type and that the shared motifs allow paralogs to respond to the same conditions. This situation allows for reprogramming in response to mutations. We plotted the number of shared transcription factor binding sites against the rate of substitutions per synonymous position,  $K_s$  (a rough duplication-age surrogate), and found nearly identical average numbers of shared motifs across the entire range of  $K_s$  values ( $R = 0.025$ ,  $P > 0.29$ ; Supplementary Figs. 6 and 7 online). This indicates that sharing of transcription factor binding motifs results either from restricted divergence or from converging and is not an evolutionary artifact that is likely to dissipate on an evolutionary time scale.

To corroborate the hypothesis that PCoR underlies reprogramming and, ultimately, backup, we examined three predictions. First, one member of a pair with high PCoR should be upregulated transcriptionally in response to the deletion of its paralog. To investigate this prediction, we used the Rosetta Compendium<sup>13</sup> containing genome-wide expression response to single-gene deletions. Of the 259 knock-outs in the Compendium, 76 have paralogs in our data set. Of these, 18 share high similarity in molecular function, and another 5 are synthetically lethal. We reasoned that if such potential backup-providing pairs undergo reprogramming then the transcriptional level

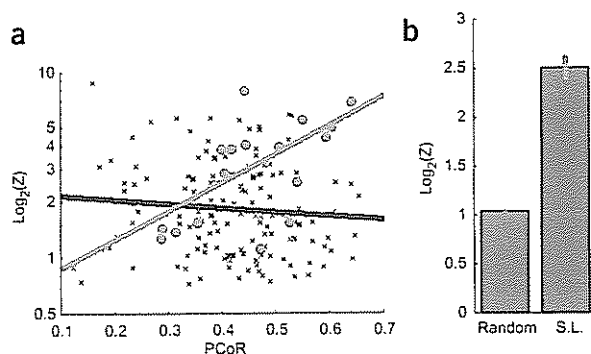
of the intact paralog should increase as a function of the pair's PCoR. In fact, we found a significant correlation between PCoR and the logarithm of transcriptional response to deletion among these backup-providing candidates ( $R = 0.67$ ,  $P = 0.002$ ; Fig. 3). As a negative control, functionally similar nonparalogs and random pairs showed no correlation between PCoR and transcriptional response to deletion ( $R = -0.02$  and  $R = 0.01$ , respectively). Therefore, we conclude that PCoR measured across wild-type conditions predicts backup capacity or the ability of a gene to respond, by upregulation, to deletion of its counterpart.

Our second prediction addresses 478 paralogs in which only one of the two genes is essential. We tested our ability to predict which of the two genes in such asymmetric pairs is essential by inspecting their regulatory motifs. Our reprogramming scenario predicts that the more motifs control a gene, the better its reprogramming and backup-providing capacity will be. Therefore, for paralogous pairs, we expect a negative correlation between number of motifs controlling a gene and its dispensability. As expected, the more essential of the two genes tended to have more motifs (Fig. 4). As a negative control, we repeated the analysis with random pairing of the paralogs to determine whether this observation merely reflected the potential bias that essential genes are regulated by a larger number of motifs. This analysis with random pairing resulted in no signal (Fig. 4).



**Figure 2** Gene dispensability as a function of the regulatory motif-content overlap  $O$  between genes and their closest paralogs. By fitting these data to a linear function (not shown), a quadratic function (red) and the rational function (purple;  $y = (ax + b)/(x^2 + cx + d)$ ), we obtained adjusted  $r^2$  values of 0.56, 0.72 and 0.82, respectively. A binomial test showed that the proportion of dispensable genes with  $O$  values between 0 and 0.25 (blue bars) was significantly higher than that of genes with  $O$  values of either 0 ( $P = 1.6 \times 10^{-7}$ ) or  $>0.25$  ( $P < 1.3 \times 10^{-14}$ ).

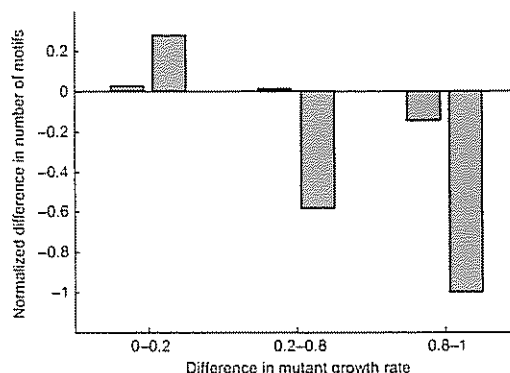




**Figure 3** Transcriptional response of backup-providing genes to the deletion of the counterparts. (a) Transcriptional responses of genes to the deletion of their functionally similar paralogs (red,  $R = 0.67$ ,  $P = 0.002$ ) and functionally similar nonparalogs (black,  $R = -0.02$ ) as a function of their PCoR (data obtained from the Rosetta Compendium<sup>13</sup>). Response is depicted as average  $\log_2$  relative change of the expression level of a gene in the mutant strain lacking its paralog divided by the expression level of the gene in the wild type. Decreased reads in response to deletion may result from artifacts owing to potential cross-hybridization in the wild type. This effect was excluded by analyzing only genes that are upregulated after the deletion. Only functionally similar paralogs were analyzed, defined either as genes encoding enzymes with the same EC classification or, for nonenzymes, as genes with high Gene Ontology-based semantic similarity<sup>27</sup> (where high similarity indicates similarity exceeding that observed at the 90<sup>th</sup> percentile of similarities of all gene pairs in the genome). (b) Average upregulation of functionally similar pairs that are also synthetically lethal (from the BIND database) compared with randomly selected gene pairs.

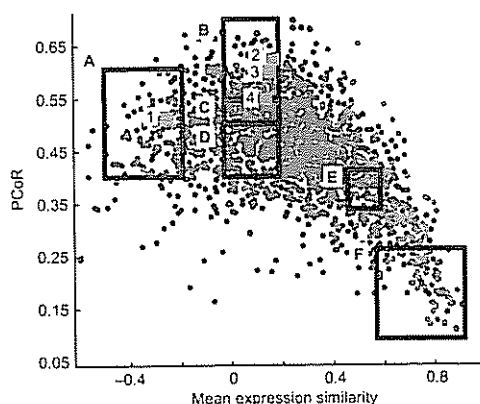
Third, our proposed model predicts synthetic lethal interactions. We embedded the paralogous pairs in a plane spanned by their mean expression similarity and PCoR (Fig. 5). We gathered evidence for synthetic lethality for certain pairs of paralogs and observed that the prevalence of backup depended on both mean expression similarity and PCoR score. Backup was maximal among pairs with high PCoR and low coexpression. Physical interactions between paralogs showed an opposite trend (Supplementary Fig. 4 online), in agreement with previous observations<sup>14,15</sup>. The model also includes verified cases of reprogramming.

A crucial question is what controls reprogramming of a gene upon mutation of its paralog. We propose a kinetic model, or reprogramming switch, consisting of two genes, G1 and G2, that encode enzymes

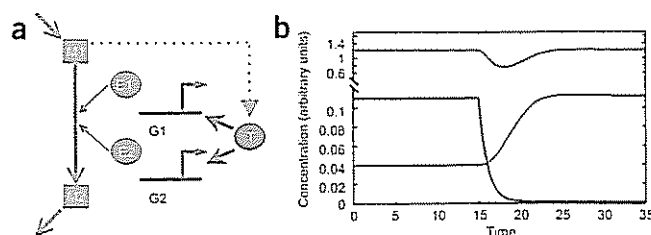


**Figure 4** Difference in the number of motifs regulating paralogous pair members as a function of the difference in the growth rates of mutants lacking them. For each pair of paralogs, the number of motifs contained by the gene with the higher growth rate was subtracted from the number of motifs in the promoter of the gene with the lower growth rate. This difference was normalized to the size of the larger of the two motif sets. All paralogs were then grouped into three categories on the basis of the absolute value of the difference between their growth rates, and for each category, the mean normalized difference in the number of motifs was calculated. The analysis was done separately for all paralogs (blue) and for paralogs with similar molecular functions (red; defined as in the legend to Fig. 3).

E1 and E2, which interconvert metabolite M1 into metabolite M2. In the wild type, only E1 is active. Assuming that the two genes contain binding sites for a shared transcription factor T that is induced by M1, T reprograms (*i.e.*, activates) G2 and hence maintains the level of M2 upon knockout of G1 (Fig. 6). Upon silencing of G1, M1 accumulates and the concentration of T increases, resulting in more efficient activation of G2 (Fig. 6). Consequently, the level of E2 increases and the level of M2 returns to its original value after a transient decrease (Fig. 6). This model provides appropriate control of backup as it couples response of G2 to an environmental condition (*i.e.*, the accumulation of M1) with response to an internal perturbation (*i.e.*, silencing of G1). The model describes enzymatic reactions; enzymes are over-represented in our data set (34%; Supplementary Fig. 8 online). Backup among paralogous transcription factors may use alternative architectures (Supplementary Note online).



**Figure 5** Confirmation and characterization of genetic backup circuits. Paralogous gene pairs are plotted as a function of their mean expression similarity and PCoR. Pairs are colored red if both members are essential or blue if both are dispensable. Black rectangles (A–F) enclose sets of genes whose functional redundancy was confirmed or disputed using the Proteome and BIND databases. In this analysis, pairs were considered to back each other up only if they have similar molecular activities and are synthetically lethal. The number of such backup-providing pairs was divided by the total number of functionally characterized pairs in each of the marked rectangles individually (A:  $10/29 = 0.34$ ; B:  $14/30 = 0.47$ ; C:  $12/35 = 0.34$ ; D:  $10/45 = 0.22$ ; E:  $8/30 = 0.27$ ; F:  $1/42 = 0.02$ ). The highest probability for verified backup coincides with cases where mean expression similarity is  $\sim 0$  and PCoR  $> 0.4$ . The placement of the rectangles reflects our desire to examine how incidence of backup depends on the  $x, y$  coordinates of the pairs. Four examples of paralogs that show transcriptional reprogramming in response to gene deletion are also shown (green squares): 1: Acs1-Acs2 (ref. 11); 2: Hxt2-Hxt10 (ref. 28); 3: Idp1-Idp2 (ref. 29); 4: Fks1-Gsc2 (ref. 30).



**Figure 6** Schematic and dynamics of the reprogramming switch. (a) The reprogramming switch. (b) Simulated dynamics of the switch before and after knockout, at time point 15. The blue and green curves represent the concentrations of E1 and E2, respectively; the red curve represents the concentration of M2. The dynamics were calculated from the differential equations describing the system using the ode23 solver of Matlab's simulink.

The different behavior of close and remote paralogs probably stems from the profoundly different evolutionary regimens acting on them<sup>3</sup>. Focusing first on remote pairs, we propose that preservation of high coexpression in a subset of these pairs was predominantly due to evolutionary pressures that are inconsistent with, and compromise, backup. One such effect is evolving protein-protein interactions between paralogs, which requires coexpression but precludes backup (Supplementary Fig. 4 online). Second, subfunctionalization of proteins may alleviate the pressure to diverge in expression<sup>3</sup>, but that, too, precludes backup between coexpressed pairs (Supplementary Fig. 5 online). Third, quantitative subfunctionalization<sup>16</sup> that may result in regulatory motif degeneration<sup>10,17</sup> accounts for both coexpression and lack of backup (e.g., due to low dosage of each of the coacting paralogs<sup>18,19</sup>).

Why do remote pairs back each other up? Although it is difficult to imagine that backup by duplicates is evolutionarily selectable<sup>3</sup>, we propose that backup-providing duplicates may be retained during evolution if their retention is coupled to other selectable traits, such as acquisition of new regulatory capabilities<sup>10</sup>. Such novelties do not preclude backup, provided that shared functionalities are preserved. Our finding that backup is optimal among pairs that maintain high partial coregulation provides considerable support to this notion. Notwithstanding this, however, backup has a profound impact on an organism's robustness, whether selected for its own sake or not. But apparent dispensability may be partially due to limited coverage of growth conditions tested in the laboratory, and a recent computational study<sup>19</sup> estimated that this factor accounts for 37–68% of dispensable genes, compared with the 15–28% that are estimated to be compensated by a duplicate. Gu *et al.* estimated a similar lower bound of 25% (ref. 2).

Many of the close paralogs that represent recent duplications<sup>3</sup> are assumed to be under free selection, meaning that they have not yet undergone either sub- or neofunctionalization; hence, they are redundantly similarly expressed. This probably explains their somewhat more intuitive behavior (backup increases with coexpression), which does not depend on evolving reprogramming.

## METHODS

**Set of analyzed genes and definition of paralogs.** To ensure that we analyzed genuine genes, we discarded from the list of *S. cerevisiae* open reading frames (ORFs) all entries corresponding to spurious ORFs<sup>20</sup> and all transposon-derived genes as annotated by Saccharomyces Genome Database. This resulted in a list of 5,862 ORFs. We defined paralogs as pairs of ORFs that, by BLASTP with standard parameters, had E valued  $< 10^{-30}$ , provided that the ratio of the length of the long protein to the length of the short protein was not larger than 1.33. For each pair of paralogs, we calculated the number of synonymous and

nonsynonymous substitutions ( $K_s$  and  $K_a$ , respectively)<sup>21</sup>. We defined remote paralogs as pairs with  $K_s > 1$  and close paralogs as pairs with  $K_s \leq 1$ . To avoid potential misclassification of borderline cases, we also adopted an alternative definition in which we regarded remote pairs as those with  $K_s > 1.2$  and close pairs as those with  $K_s < 0.8$  and found that the same trends characterized the two sets (Supplementary Fig. 9 online). Supplementary Figure 9 contains additional cutoff justifications including a systematic assessment of the robustness of the results to changes of threshold value and to use of alternative measures of sequence similarity (e.g.,  $K_a$ ). To remove from the set of close pairs ( $K_s < 1$ ) any paralogs that represent old duplications, we removed close paralogs in which at least one of the genes had a low ( $< 32$ ) effective number of codons<sup>22</sup>.

**Gene essentiality data.** We defined dispensable genes as all genes with a viable gene-deletion phenotype that were not included in the lists of spurious ORFs or transposon-derived ORFs. Additionally, we obtained data on growth rates of mutants lacking each of the ORFs in the genome in five different growth media<sup>23</sup>.

**mRNA expression data.** We obtained whole-genome mRNA expression data of 40 natural and perturbed time series and the Rosetta Compendium data, which measures genome-wide transcription response to gene deletions<sup>13</sup>, from ExpressDB. We normalized all expression profiles of genes in each time series with respect to mean and variance. Detailed descriptions of all analyzed conditions is presented on our project website (see URL below).

We obtained expression data from either Affymetrix chips (seven experiments) or PCR product-based microarrays (33 experiments). Because the latter technology is more prone to cross-hybridization errors, we used only data derived from Affymetrix chips when analyzing close paralogs.

**A nonredundant set of promoter regulatory motifs in *S. cerevisiae*.** We compiled a nonredundant set of 112 yeast regulatory motifs, along with their gene assignments, from three different sources: ChIP-chip (originally augmented with phylogenetic conservation of motifs across multiple yeast species)<sup>24</sup>, expression data<sup>25</sup> and phylogenetic conservation<sup>20</sup>. We included motifs derived from the last two computational methods only if they corresponded to experimentally known motifs and had a significance score higher than the 90<sup>th</sup> percentile in their respective methods.

**Kinetic analysis of the reprogramming switch.** We modeled the concentration of induced transcription factor ( $T^*$ ) and the fractions of time in which genes G1 and G2 were transcribed, denoted as  $G1^*$  and  $G2^*$ , respectively, with three saturation equations:

$$T^* = T^{\text{Tot}} \frac{(M1/K_M)^{n_M}}{(M1/K_M)^{n_M} + 1};$$

$$G1^* = \frac{(T^*/K_1)^n}{(T^*/K_1)^n + 1};$$

$$G2^* = \frac{(T^*/K_2)^n}{(T^*/K_2)^n + 1};$$

and

where  $T^{\text{Tot}}$  is the concentration of total transcription factor;  $K_M$  is the affinity between the transcription factor  $T$  and the inducing metabolite  $M1$ ;  $K_1$  and  $K_2$  are the transcription factor's affinities to  $G1$  and  $G2$ , respectively; and the powers  $n_M$  and  $n$  represent binding cooperativity Hill coefficients of  $M1$  to  $T$  and of  $T^*$  to the two promoters, respectively.

The concentration of the enzymes and the metabolites are described with time-dependent differential equations:

$$\frac{dE1}{dt} = \beta \cdot G1^* - \alpha E1;$$

$$\frac{dE2}{dt} = \beta \cdot G2^* - \alpha E2;$$

$$\frac{dM1}{dt} = \beta_M - \phi(E1 + E2)M1;$$





and

$$\frac{dM_2}{dt} = \phi(E_1 + E_2)M_1 - \alpha_M M_2,$$

where  $\beta$  and  $\beta_M$  are the maximal production rate of  $E_1$  and  $E_2$  and of  $M_1$ , respectively;  $\alpha$  and  $\alpha_M$  represent the degradation and dilution of  $E_1$  and  $E_2$  and of  $M_2$ , respectively; and  $\phi$  represents the conversion rate of  $M_1$  to  $M_2$ . Values of the coefficients, the present simulation, are  $T_{\text{Tot}} = 1$ ,  $K_M = 20$ ,  $K_1 = 0.1$ ,  $K_2 = 0.3$ ,  $n_M = 4$ ,  $n = 1$  (i.e., no cooperativity is assumed) and  $\alpha = \beta = \alpha_M = \beta_M = \phi = 1$ . There are three reasonable assumptions in the model. First, we assume that the binding of  $M_1$  to  $T$  and of  $T^*$  to  $G_1$  and  $G_2$  occurs on a short time scale compared with the other reactions; therefore, these reactions are in a quasi-steady state. Second, we assume that  $M_1 > T$  (the total number of free  $M_1$  molecules is roughly the same as the total number of  $M_1$  molecules). Finally, we assume that  $E_1$  and  $E_2$  work linearly with respect to  $M_1$  as a substrate.

**Statistical analyses.** We computed proportions of dispensable genes as a function of mean expression similarity, PCoR and motif-content overlap  $O$  (Figs. 1 and 2) by binning genes into groups according to each of these three variables and then calculating the frequency of viable mutants in each bin. We counted each gene in each bin only once to avoid repetitions caused by one gene having multiple paralogs. To establish that our results are independent of the particular choice of binning strategy, we verified that the observed trends were valid under any relevant bin-size choice (Supplementary Fig. 10 online).

We tested the significance of trends observed for the proportions of dispensable genes against mean expression similarity (Fig. 1a), PCoR (Fig. 1b) and motif-content overlap (Fig. 2) using logistic regressions analyses (in Fig. 1a, only the declining portion of the curve, with positive expression similarity values, was used). Further statistical analyses of the results are shown in Supplementary Figures 9 and 10 online.

**URLs.** We downloaded gene sequences from the Saccharomyces Genome Database (<http://www.yeastgenome.org/>) and retrieved gene knockout phenotype data from [http://sequence-www.stanford.edu/group/yeast\\_deletion\\_project/Essential\\_ORFs.txt](http://sequence-www.stanford.edu/group/yeast_deletion_project/Essential_ORFs.txt). Functional annotations for all genes came from the Gene Ontology annotation scheme at <http://www.yeastgenome.org/>. We downloaded synthetic lethal interactions and physical interactions between proteins from the BIND database (<http://bind.ca/>). We collected gene expression data from ExpressDB (<http://arep.med.harvard.edu/ExpressDB/>) and used the Proteome database (<http://proteome.incyte.com/>) to collect synthetic lethal pairs manually. Our project website is <http://longitude.weizmann.ac.il/BackUpCircuits/>. We obtained EC classifications from <http://mips.gsf.de/genre/proj/yeast/>.

*Note: Supplementary information is available on the Nature Genetics website.*

#### ACKNOWLEDGMENTS

We thank all members of the laboratory of Y.P. for discussions; I. Pechersky for computational assistance; and Y. Garten, N. Barkai, J. Berman, B. Shilo, A.M. Dudley, I. Yanai, O. Man, S. Shen-Orr, D. Graur, D. Lancet, M. Levy and D. Artzi for critical review of the manuscript. Y.P. is an incumbent of the Aser Rothstein Career Development Chair in Genetic Diseases and is a Fellow of the Hurwitz Foundation for Complexity Sciences. We thank the Leo and Julia Forchheimer Center for Molecular Genetics and the Ben May Foundation for grant support. This paper is dedicated to the memory of I. Kafri.

#### COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 16 November 2004; accepted 25 January 2005  
Published online at <http://www.nature.com/naturegenetics/>

- Conant, G.C. & Wagner, A. Duplicate genes and robustness to transient gene knock-downs in *Caenorhabditis elegans*. *Proc. R. Soc. Lond. B Biol. Sci.* **271**, 89–96 (2004).
- Gu, Z. *et al.* Role of duplicate genes in genetic robustness against null mutations. *Nature* **421**, 63–66 (2003).
- Nowak, M.A., Boerlijst, M.C., Cooke, J. & Smith, J.M. Evolution of genetic redundancy. *Nature* **388**, 167–171 (1997).
- Lynch, M., O'Hely, M., Walsh, B. & Force, A. The probability of preservation of a newly arisen gene duplicate. *Genetics* **159**, 1789–1804 (2001).
- Lynch, M. & Conery, J.S. The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155 (2000).
- Force, A. *et al.* Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531–1545 (1999).
- Wagner, A. The role of population size, pleiotropy and fitness effects of mutations in the evolution of overlapping gene functions. *Genetics* **154**, 1389–1401 (2000).
- Gu, Z., Nicolae, D., Lu, H.H. & Li, W.H. Rapid divergence in expression between duplicate genes inferred from microarray data. *Trends Genet.* **18**, 609–613 (2002).
- Huh, W.K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
- Papp, B., Pal, C. & Hurst, L.D. Evolution of cis-regulatory elements in duplicated genes of yeast. *Trends Genet.* **19**, 417–422 (2003).
- van den Berg, M.A. *et al.* The two acetyl-coenzyme A synthetases of *Saccharomyces cerevisiae* differ with respect to kinetic properties and transcriptional regulation. *J. Biol. Chem.* **271**, 28953–28959 (1996).
- Kratzer, S. & Schuller, H.J. Transcriptional control of the yeast acetyl-CoA synthetase gene, ACS1, by the positive regulators CAT8 and ADR1 and the pleiotropic repressor UME6. *Mol. Microbiol.* **26**, 631–641 (1997).
- Hughes, T.R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).
- Jansen, R., Greenbaum, D. & Gerstein, M. Relating whole-genome expression data with protein-protein interactions. *Genome Res.* **12**, 37–46 (2002).
- Ge, H., Liu, Z., Church, G.M. & Vidal, M. Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. *Nat. Genet.* **29**, 482–486 (2001).
- Lynch, M. & Katju, V. The altered evolutionary trajectories of gene duplicates. *Trends Genet.* **20**, 544–549 (2004).
- Teichmann, S.A. & Babu, M.M. Gene regulatory network growth by duplication. *Nat. Genet.* **36**, 492–496 (2004).
- Kondrashov, F.A., Rogozin, I.B., Wolf, Y.I. & Koonin, E.V. Selection in the evolution of gene duplications. *Genome Biol.* **3**, RESEARCH0008 (2002).
- Papp, B., Pal, C. & Hurst, L.D. Metabolic network analysis of the causes and evolution of enzyme 'dispensability' in yeast. *Nature* **429**, 661–664 (2004).
- Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E.S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**, 241–254 (2003).
- Goldman, N. & Yang, Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**, 725–736 (1994).
- Cavalcanti, A.R., Ferreira, R., Gu, Z. & Li, W.H. Patterns of gene duplication in *Saccharomyces cerevisiae* and *Caenorhabditis elegans*. *J. Mol. Evol.* **56**, 28–37 (2003).
- Steinmetz, L.M. *et al.* Systematic screen for human disease genes in yeast. *Nat. Genet.* **31**, 400–404 (2002).
- Harbison, C.T. *et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99–104 (2004).
- Pilpel, Y., Sudarsanam, P. & Church, G.M. Identifying regulatory networks by combinatorial analysis of promoter elements. *Nat. Genet.* **29**, 153–159 (2001).
- Kellis, M., Birren, B.W. & Lander, E.S. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**, 617–624 (2004).
- Lord, P.W., Stevens, R.D., Brass, A. & Goble, C.A. Investigating semantic similarity measures across the Gene Ontology: the relationship between sequence and annotation. *Bioinformatics* **19**, 1275–1283 (2003).
- Ozcan, S. Two different signals regulate repression and induction of gene expression by glucose. *J. Biol. Chem.* **277**, 46993–46997 (2002).
- McCammon, M.T. & McAlister-Henn, L. Multiple cellular consequences of isocitrate dehydrogenase isozyme dysfunction. *Arch. Biochem. Biophys.* **419**, 222–233 (2003).
- Garcia-Rodriguez, L.J. *et al.* Characterization of the chitin biosynthesis process as a compensatory mechanism in the *fks1* mutant of *Saccharomyces cerevisiae*. *FEBS Lett.* **478**, 84–88 (2000).



## **Genetic backup of hubs in the protein interaction network: evidence for evolutionary selection of redundancy<sup>2</sup>**

Redundancies are, and have been, considered the main source of the remarkable robustness and constancy of phenotypes. Yet, together with that, redundancy is rarely regarded as an evolved functional component of the genetic system. In fact, redundancy is typically regarded as the fuel for evolutionary change and thus consequently, pre-destined for divergence.

In this submitted manuscript we provide, for the first time, evidence suggesting that redundancy itself may be selected for. This, in turn, implies an advantageous utilization of genetic redundancies. Specifically, we address redundancies associated with gene duplicates, yet unlike previous works, we concentrate on the conservation of redundancy rather than the conservation of sequence similarity. Our question stems from the fact that while some duplicate pairs have retained their functional overlap, most have not. To identify functional overlaps of duplicates we rely on a well established correlation, identified by Barabasi et al., stating that genes with more partners in the protein interaction network are significantly more likely to be essential for cell viability. Relying on this correlation among singleton genes, we calculate what would have been the expected proportion of dispensable paralogs in the absence of redundancy. By comparing these to the observed frequencies of dispensable paralogs we estimate the proportion of paralogs that are functionally backed-up by redundant partners. Our results indicate that the frequency of backed-up duplicates, i.e. duplicate pairs that have preserved their functional overlap, increases with degree in the protein interaction network. More significantly, by normalizing to our calculated expected proportions, we show that the probability of a gene to maintain a redundant partner is significantly dependent on its degree.

We conclude by noting that while our results do indicate that redundancy was selected for, it does not necessitate that these were selected for the sake of buffering mutations. In fact, as an alternative, we suggest that redundancies may be utilized to regulate and dampen non-genetic noise in regulatory pathways. This conclusion may also be supported by a large number of high level regulators observed in our dataset of backed-up paralogs.

---

<sup>2</sup> Manuscript in preparation to be sent to *Science*.

# **Genetic backup of hubs in the protein interaction network: evidence for evolutionary selection of redundancy**

Ran Kafri and Yitzhak Pilpel\*

Department of Molecular Genetics

Weizmann Institute of Science

Rehovot, 76100 Israel

\*Email: [Pilpel@weizmann.ac.il](mailto:Pilpel@weizmann.ac.il)

**Genetic redundancy is likely to be a primary source of robustness of organisms' phenotypes. Nevertheless, it is commonly regarded evolutionarily instable, at least on the basis of its role as genetic backup. In contrast, we report here strong evidence for an evolutionary selection and preservation of genetic redundancies in yeast. Specifically, we show a preferential retention of redundant duplicate partners of genes that constitute "hubs" in the protein interaction network. Our results demonstrate that, in contrast to singletons, deletion phenotypes of duplicates are independent of their degree of connectivity within the network. Thus we demonstrate that redundancy itself is selected for and is thus likely utilized by the cellular system, though possibly for other reasons in addition to backing-up against mutations.**

From studies in several model organisms (1-3) it is apparent that the majority of genes are dispensable in the sense that their knockouts have a surprisingly small effect on phenotype. In principle, this dispensability of genes may be either attributed to a dispensability of their functions, at least under tested laboratory conditions (4), or to compensations, e.g. by partially redundant duplicates (5). These two alternatives manifest themselves in two approaches that were shown useful for predicting the phenotypic consequence of mutations. According to the first, known as "centrality and lethality" (6), knockouts of "hubs", i.e. highly connected (or "central") nodes in the protein interaction networks has a higher chance to result with lethal phenotypes, while sparsely connected proteins are more often dispensable. This result was found to be independent of the of the recently suggested calcification of hubs into "date" or "party" types (7). According to the second approach, genes that have duplicates

(paralogs) in the genome are significantly more likely to be dispensable (1), partially due to compensations or “backup”. In these latter cases, the dispensability of genes is decoupled from the dispensability of their *functions*. In other words, a genes’ essential function may be backed-up by a redundant duplicate, rendering the gene itself non-essential. An independent supporting evidence for the essentiality of the biological *functions* carried out by genes that have duplicates is the recent report that showed preferential duplication of conserved genes (8, 9).

We set here to investigate whether or not redundancies, defined by their ability to “backup” against mutations, are distributed randomly among the genetic network. Specifically, we ask whether the probability of a gene to be backed-up is affected by its position in the protein network or the importance of its’ biological function (Fig 1A). To test this we collected the set of *S. cerevisiae* gene duplicates as described in Kafri et al. (10). From these we removed all duplicate pairs that are co-expressed, as it was previously shown (10) that these are highly essential, i.e. non-redundant. Also, since the redundancy of among young duplicates can not be regarded as the consequence of selection we removed these pairs from our analyses and examined only remote pairs ( $K_s > 1$ ) as defined elsewhere (10). We found that in this set of genes, unlike the rest of the genome, “centrality and lethality” does not hold, namely genes with low and high “degree connectivity” (i.e. low and high number of physically interacting protein partners (11)) are equally non-essential (Fig. 1). We reasoned that the disappearance of the “centrality and lethality” trend reflects the consequence of redundancies rather than a dispensability of the *functions* carried out by duplicates with high values of degree connectivity. We, thus, relied on the set of singleton genes, which approximates a backup-free regime, as a background control in which the essentiality of genes is coupled to the essentiality of their *functions*. As expected the singletons strongly manifest “centrality and lethality” as they show a highly significant ( $p\text{-value} = 4 \times 10^{-53}$  see legend for details) decline in dispensability as a function of degree connectivity. This comparison allows us to ask what would have been the expected dependency between dispensability and degree connectivity in the absence of redundancies among the duplicates. A comparison between the trends observed in the two gene sets suggested that the increased essentiality of the biological *functions* of hubs is accompanied, in the set of duplicates, with an increased likelihood for backup by redundant partners. Based on the duplicates and singleton trends we calculated the frequency of duplicates that are backed-up at any degree

connectivity (see Fig. 1 and legend for details). From this we found that the probability of a gene to be backed-up by a redundant duplicate increases with degree in the protein interaction network, culminating in preferential protection of the “hubs” (for example consider the probability for backup for genes with degree connectivity above or below five, a value that was recently suggested to define the “hubs” (7)).

We have also reasoned that if duplications are preferentially retained in evolution when providing protection to the hubs, then the duplication age (estimated by the rate of synonymous substitutions in the coding region, see SI) of duplicated hubs should be higher compared to the age of duplications of lowly connected proteins. Indeed we found exactly that trend (Fig. 1C), consistent with a scenario in which retention of duplicates of hubs over long evolutionary time spans is more likely compared to retention of duplicates of lowly-connected proteins.

In summary, our results demonstrate selection for redundancy on long evolutionary time scales. An open question is why some of hubs have retained a redundant duplicate while many others have not. We propose that the answer involves two separate criteria pertaining to different evolutionary time scales (Fig. 1E). In the short time scale, the immediate consequence of redundancy is a doubling of the dosage of the duplicated protein. This may be intolerable by some dosage-sensitive genes and hence be selected against. We are then left to further consider only those genes for which dosage multiplication was not deleterious. Among these, for redundancies to be retained for long evolutionary time scales, mere lack of deleterious effects is not sufficient. These redundancies may be retained if they provide advantageous and selectable functionalities. Nevertheless, as it is less likely that redundancies have been preserved for the sake of backing-up against mutations, the precise nature of the advantageous utilization of redundancy remains an open question. One intriguing possibility is that functional overlaps contribute to the regulation or dampening of non-genetic noise and random fluctuations, yet other possibilities should also be considered.

#### Figure Legend

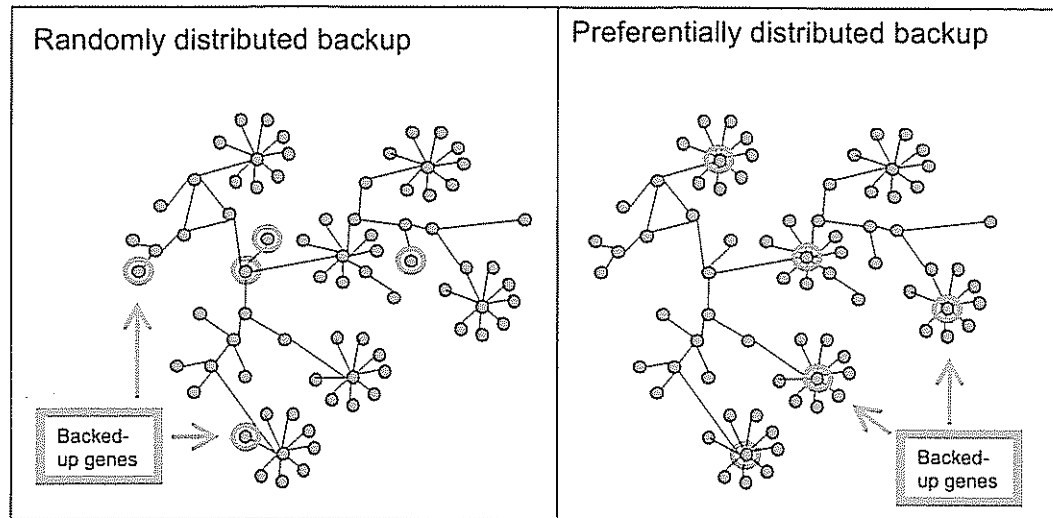
A. A schematic diagram showing two hypothetical scenarios for the allocation of backup in the protein interaction network. In this network nodes are proteins and edges capture their physical interactions. On the left, backup is randomly allocated to genes irrespective of their patterns of interactions with other genes, while on the right

backup is preferentially allocated to highly connected genes **B**. Gene dispensability vs. degree connectivity for singleton genes (black) and for dissimilarly expressed remote paralogs (blue). The left vertical axis represents the fraction of dispensable genes with a degree equal too or greater than that specified by the x-axis. The red line corresponds to the right vertical axis and depicts the proportion of genes that are deduced to be backed-up by a partner duplicate. These proportions were computed similarly to (1), yet for each degree connectivity value separately as:  $P_{DD} - P_{DS}$  (where  $P_{DD}$  – proportion of dispensable paralogs,  $P_{DS}$  – proportion of dispensable singletons, see SI for further details). Statistical significance of the dispensability trends of the singletons and duplicates were calculated by performing logistic regression analyses on the non-cumulative data obtaining a p value of  $4 \times 10^{-53}$  (with a slope on the logistic-transformed curve,  $\beta = -0.26 \pm 0.02$ ) for the set of singletons. On the other hand, in the duplicates the trend is very weak, and in fact in the opposite direction from “centrality and lethality” ( $\beta = 0.12 \pm 0.07$ , p value = 0.067). **C**. The conditional probability for retention of redundancy given the predicted essentiality of gene functions. Conditional probabilities were estimated through  $\frac{P_{DD} - P_{DS}}{1 - P_{DS}}$ . In other words, we normalized the frequency of backed-up genes to the frequency of genes that were predicted, based on singletons, to have essential functions. **D**. Proportion of genes having duplicate partners with a  $K_s > 1$  as a function of their degree connectivity. Statistical significance for the increasing slope was established by performing a logistic regression ( $\beta = 0.158 \pm 0.02$ , p value =  $4 \times 10^{-22}$ ). In order to verify that the trend does not merely result from duplicates with degree connectivity = 0 we repeated the analysis for genes with degree > 0 and obtained significant increasing slope ( $\beta = 0.085 \pm 0.02$ ; p value =  $4 \times 10^{-5}$ ). Number of genes in the categories is given below the figure. **E**. Schematic depiction of the possible fates of redundancy following gene duplication events. The scheme distinguishes between the immediate selective pressures against dosage multiplication following the duplication events and the long time scale consequence of drift in diverging functional overlaps. The conclusion is that redundancy may be preserved if it conforms to these two separate criteria, namely, dosage insensitivity and an option for an advantageous utilization of the redundancy, consequently preventing random drift.

## References

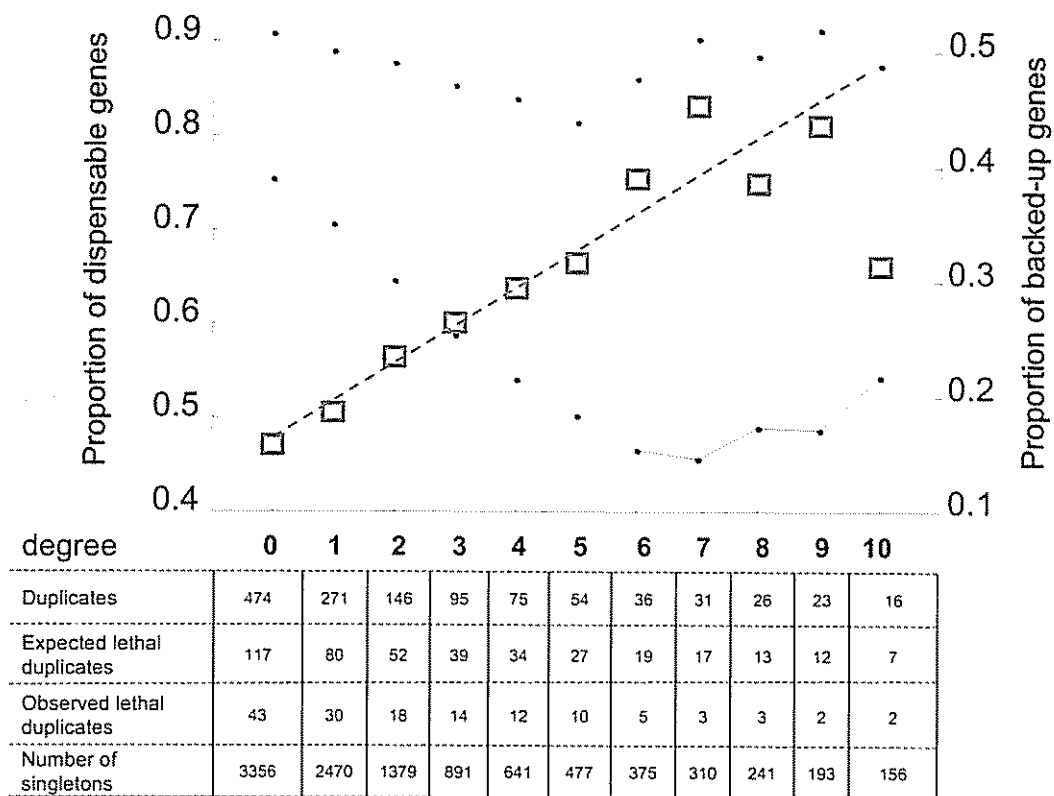
1. Z. Gu *et al.*, *Nature* **421**, 63 (Jan 2, 2003).
2. G. C. Conant, A. Wagner, *Proc R Soc Lond B Biol Sci* **271**, 89 (Jan 7, 2004).
3. J. L. t. Hartman, B. Garvik, L. Hartwell, *Science* **291**, 1001 (Feb 9, 2001).
4. B. Papp, C. Pal, L. D. Hurst, *Nature* **429**, 661 (Jun 10, 2004).
5. A. Wagner, *Bioessays* **27**, 176 (Feb, 2005).
6. H. Jeong, S. P. Mason, A. L. Barabasi, Z. N. Oltvai, *Nature* **411**, 41 (May 3, 2001).
7. J. D. Han *et al.*, *Nature* **430**, 88 (Jul 1, 2004).
8. J. C. Davis, D. A. Petrov, *PLoS Biol* **2**, E55 (Mar, 2004).
9. I. K. Jordan, Y. I. Wolf, E. V. Koonin, *BMC Evol Biol* **4**, 22 (Jul 6, 2004).
10. R. Kafri, A. Bar-Even, Y. Pilpel, *Nat Genet* **37**, 295 (Mar, 2005).
11. B. J. Breitkreutz, C. Stark, M. Tyers, *Genome Biol* **4**, R23 (2003).

**Figure 1A**



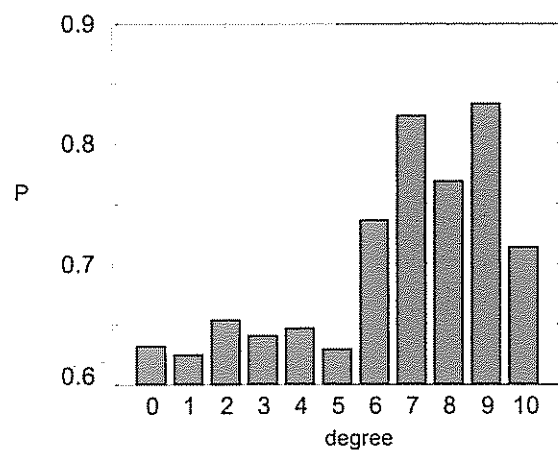


**Figure 1B**

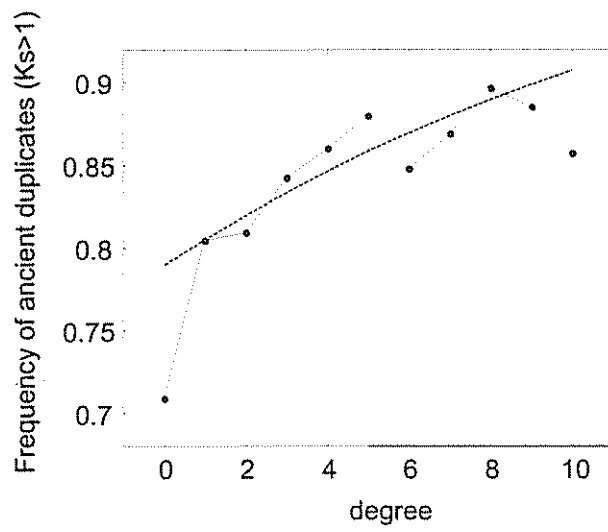




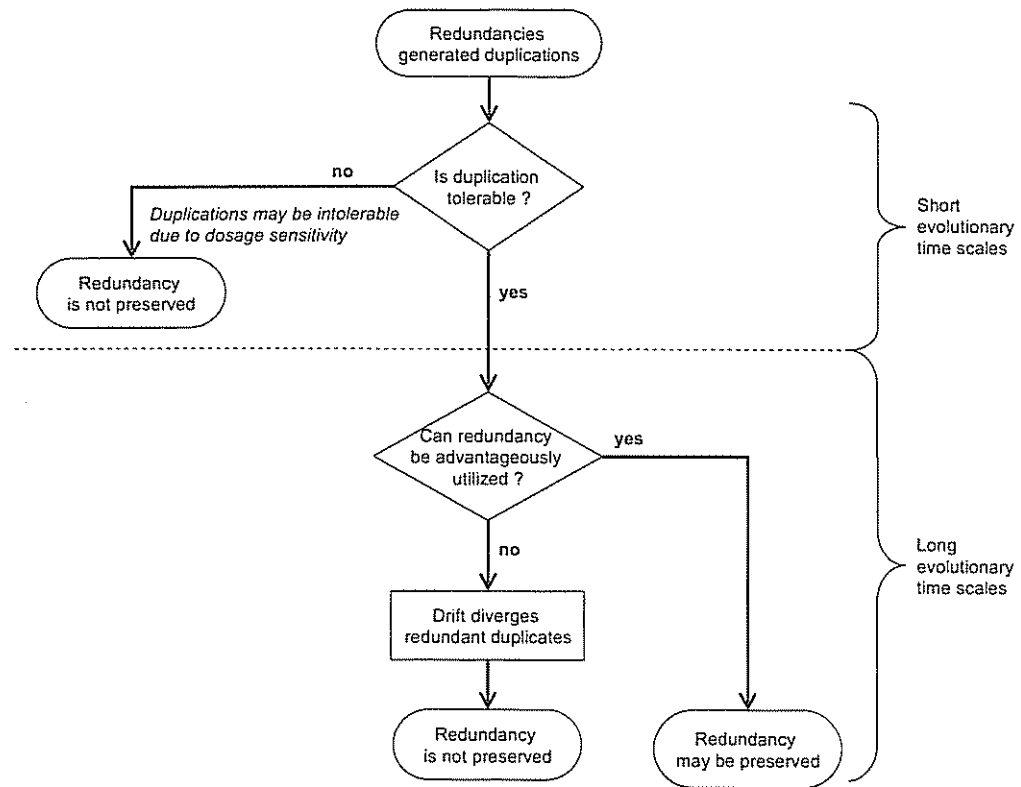
**Figure 1C**



**Figure 1D**



**Figure 1E**



### **The regulatory utilization of redundancy<sup>3</sup>.**

The central question raised in the article evolved from recent studies (Davis and Petrov 2004; Jordan, Wolf et al. 2004; Kafri, Bar-Even et al. 2005; Kafri and Pilpel 2006), performed in our lab and others, that surprisingly suggested an evolutionary selection and functional utilization of genomic redundancies. Functionally redundant pairs of genes, typified by examples such as Pearce et al. (Pearce, Senis et al. 2004) or Enns et al. (Enns, Kanaoka et al. 2005) and many more, are highly widespread and have been acknowledged in all fields of cellular biology. Remarkably, despite its tremendous prevalence, redundancy itself has rarely entered the spotlight of attention, except in the context of evolutionary biology where it has been suggested as the fuel for evolutionary change (Ohno 1970) and thus consequently, pre-destined for divergence (Lynch and Conery 2000). In fact, apart from buffering mutations of duplicated genes, redundancies are not expected have much of a role.

In attempt to answer these questions we systematically review a plethora of individual, well characterized, examples of redundant gene duplicates. We consider two lines of evidence to indicate a function's direct benefit from existing redundancy; first is the evolutionary conservation of the functional overlap and second is a non-trivial regulatory design that utilizes it. To maintain generality, we did not limit our search to any particular organism or gene functional class and compiled a large list of gene pairs answering our criteria. By inspecting our collected examples we came across a number of regulatory principles that systematically recur in many duplicate pairs and that we believe may highlight common biological functionalities associated with their redundancy. For example, one interesting common denominator of redundant duplicates describes a feedback regulation that allows one duplicate partner to sense and respond to the intactness of its partner. Based on this and other such "design principles" that we found to recur in many independent examples we formulate what seems to be the common mode of regulation of redundant duplicates, which we term a "*responsive backup circuit*". Interestingly, redundancies regulated by responsive circuits are not unique to any single species with examples ranging from *e. coli* to mouse and human, and regulate a wide variety of processes ranging from metabolism and transport to developmental pathways. Yet, in spite of this

---

<sup>3</sup> Manuscript under referee evaluation in *P.N.A.S.*

variability, we argue here that the unification of these examples into one generalized theme is rewarding as it has lead us to suggest regulatory principles that may have not been recognized otherwise.

We conclude by hypothesizing about the functional attributes of responsive backup circuits and the advantageous functionalities allowing their evolutionary preservation. For this purpose we classify the different responsive circuits in our compiled list according to functional class and regulatory design. We present the hypothesis that responsive backup circuits function as uni-directional noise filters removing accumulated noise from regulatory cascades. In other words, it has been recently demonstrated that the stochasticity in gene expression results in random variability in protein dosage that tends to accumulate for coupled interactions in pathways. The cellular system, in turn, can dampen the levels of these fluctuations by auto-repressors and negative feedbacks. In line with these we suggest responsive backup circuits as a highly efficient component for filtering out non-genetic noise by balancing fluctuations in one member of the redundant pair with reciprocal fluctuations of its partner. Furthermore, we rely on basic modeling techniques to compare the efficiency of responsive backups with that of auto-repressors and show clear advantages of the former. This hypothesis is also in line with a work that we have just recently sent for publication showing a preference of responsive circuits to be associated with high level regulators and hubs in the protein interaction network. By collectively inspecting a literature of seemingly unrelated examples of redundant gene partners we were able to recognize common denominators in the regulation of genetic redundancy. We further highlight a relevance of these denominators to aspects of developmental biology, metabolism, transport, transcriptional cascades and more. In doing so, we provide, for the first time, a “systems biology” approach for the aimed at understanding the implications of the so widely encountered genetic redundancies. Lastly, our conclusions challenge widely accepted notions of molecular evolution in proposing that redundancy itself may be selectable.

**The regulatory utilization of genetic redundancy  
through responsive backup circuits.**

Ran Kafri\$, Melissa Levy\$  
and Yitzhak Pilpel\*

Department of Molecular Genetics  
Weizmann Institute of Science  
Rehovot 76100, Israel

\$Both authors contributed equally

\*Corresponding author at: [pilpel@weizmann.ac.il](mailto:pilpel@weizmann.ac.il)

Fax: 972-8-934 4108

March 6, 2006



## Abstract

Functional redundancies, generated by gene duplications, are highly widespread throughout all known genomes. One consequence of this is a tremendous increase to the robustness of phenotypes, as mutational inactivation of many genes can be compensated for by functionally overlapping partners. Yet this very fact also renders these redundancies evolutionarily unstable and they are consequently predicted to have only transient life times.

Despite this, numerous reports describe instances of functional overlaps that have been conserved throughout extended evolutionary periods. Additionally, there are an increasing number of studies providing evidence for cross-regulatory interactions between redundant partners that result in an up-regulatory response of one gene partner in the case that the latter was mutationally inactivated. Thus the paradigm that emerges is that genes that are functionally redundant are often not independently controlled but rather they are regulated by a system that has evolved to both monitor and respond to their intactness. By manual inspection of the literature we have compiled a list of such '*responsive backup circuits*'. Reviewing these studies, we extract recurring principles that characterize their function and regulation. We then apply modeling approaches to further explore their dynamic properties. We demonstrate how such circuitries may generate various functionalities such as filtering non-genetic noise and fluctuations, thus providing phenotypic robustness. Furthermore, we classify different regulatory motifs based on their efficiency at exploiting and taking advantage of existing functional redundancies. Our conclusions, thus, challenge the view that such redundancies are simply leftovers of ancient duplications and suggest they are an additional component to the sophisticated machinery of cellular regulation. In this respect, we suggest that compensation for gene loss is merely a side effect of sophisticated design principles utilizing functional redundancy.

## Introduction

Duplicate genes and paralogous gene families have long been perceived as genomic sources of genetics robustness (Ohno 1970; Kirschner and Gerhart 1998; Wagner 2000; Hartman, Garvik et al. 2001; Krakauer and Plotkin 2002; Gu 2003; Gu, Steinmetz et al. 2003; Conant and Wagner 2004). The assumption is that a functional overlap of these genes acts to compensate against mutations, i.e., one paralog could substitute for the function of its partner in case the latter had suffered a mutation rendering it completely or partially non-functional. From an evolutionary perspective, redundancies serve to buffer phenotypes from genomic variations by reducing the phenotypic cost of mutations and, consequently, increasing the organism's' evolvability (Gerhart and Kirschner 1997; Kirschner and Gerhart 1998). But on the other hand, this very fact renders these redundancies instable on evolutionary time scales (Ohno 1970; Nowak, Boerlijst et al. 1997; Wolfe and Shields 1997; Lynch and Conery 2000; Brookfield 2003; Conant and Wagner 2003; Gu, Steinmetz et al. 2003; Makova and Li 2003). In other words, if a gene's function can be perfectly compensated for by a redundantly functioning partner, mutations in that gene would bear no consequence on the phenotype of that individual. Such mutations, therefore, could not be selected against and would tend to accumulate, leading to either loss of function of one of the duplicates (non-functionalization), or in the case of pleiotropy, the functions of the ancestral gene, prior to duplication, will be partitioned between the two duplicates (sub-functionalization) (Lynch and Conery 2000). A third and more rare possibility is that one or both of the duplicate partners will acquire novel functions not present in the ancestral gene (neo-functionalization).

Thus, redundancy generated by gene duplication is in most cases short-lived on evolutionary time scales and is rapidly eliminated by massive gene loss and specialization (Wolfe and Shields 1997; Lynch and Conery 2000; Kellis, Birren et al. 2004). Despite this being the general case, the numerous examples of paralogs retaining their functional overlap for extended evolutionary periods (for examples see (Hughes 1994; Maconochie, Nonchev et al. 1996; Nowak, Boerlijst et al. 1997; Schwarz, Alvarez-Bolado et al. 1997; Mansouri and Gruss 1998; Weiss, Stock et al. 1998)) suggest that, at least for a fraction of gene pairs, redundancies are conserved throughout evolution despite their predicted

instability. One such example is the pair of O-acyl-transferases isozymes, redundantly catalyzing the conjugation of sterols to fatty acids, for which functional overlap has been conserved all the way from yeast (Are1 and Are2) to mammals (ACAT1 and ACAT2) ((Yang, Bard et al. 1996; Yu, Kennedy et al. 1996; Cases, Novak et al. 1998)). In fact, although retention of redundancy is much less frequent than its loss, its widespread existence is non-trivial and cannot (Nowak, Boerlijst et al. 1997) be dismissed as leftovers of recent duplication events. In one study, modeling of evolutionary dynamics suggested that this conservation may be the result of asymmetries in the functional efficiencies or mutation rates between the redundant pair members (Krakauer and Nowak 1999). Alternatively, redundancies were suggested to be selected for their contribution to robustness and evolvability (Gerhart and Kirschner 1997; Kirschner and Gerhart 1998). In this review we wish to adapt the view that, at least in some pathways, redundancies are selected for based on some evolutionary advantage that they confer to the wild-type organism. In particular, we suggest the existence of regulatory designs that exploit redundancy to achieve functionalities such as control of noise in gene expression or extreme flexibility in gene regulation. In this respect, we suggest that compensation for gene loss is merely a side effect of sophisticated design principles utilizing functional redundancy.

Clues for regulatory designs controlling redundancy were first obtained in a recent study (Kafri, Bar-Even et al. 2005) that explored the dispensability of gene duplicates with various degrees of co-regulation. The underlying assumption of this study was that in order for one duplicate copy to compensate against the loss of its partner, both duplicates must not only perform the same function but do so at the same place and time. In other words, co-regulation of duplicates was perceived as a prerequisite for functional compensation (Gu, Steinmetz et al. 2003; Kafri, Bar-Even et al. 2005). In reality, however, similarly expressed paralogs were found to almost never back each other up, as evident from their high essentiality. In fact, tightly co-regulated gene duplicates were found more essential for viability than singleton genes. Functional redundancy and compensation were found to be most prevalent among gene duplicates that are regulated differently from one another (Kafri, Bar-Even et al. 2005). Further insight was provided by the observation that some differentially regulated duplicates maintain the ability to

become co-regulated under certain environmental conditions. Such *conditional co-regulation* (CCR), or *partial co-regulation* (PCoR) of these genes within the transcriptional network, was shown to be very strongly negatively correlated with the severity of the knockout-phenotypes of these genes. Thus the paradigm that has emerged is that genes that are functionally redundant are often not independently controlled, but rather they are regulated by a system that both monitors and responds to their intactness. In this study we will survey examples of such '*responsive backup circuits*' and draw a general outline for their function, design and evolution.

### **Responsive backup circuits.**

Two lines of evidence could indicate a function's direct benefit from existing redundancy; first is the evolutionary conservation of the functional overlap and second is a non-trivial regulatory design that utilizes it. Many well known examples meet both these criteria, one of which is that of the 1,3-beta-glucane synthase catalytic subunit in yeast that is encoded by the two alternative, functionally redundant and synthetically lethal, genes FKS1 and FKS2 (Zhao, Jung et al. 1998). The evolutionarily selectable advantage of this redundancy can be inferred from the fact that both isozymes are found as duplicates in all 12 sequenced yeast species except for the *Yarrowia lipolytica* (Leon, Sentandreu et al. 2002). Furthermore, in *S. cerevisiae*, these two genes obey a particular regulation whereby FKS2 transcriptionally responds to the intactness of FKS1 and is up-regulated upon FKS1 mutational inactivation (Garcia-Rodriguez, Trilla et al. 2000). Numerous other examples describing such '*responsive backup circuits* (RBCs) exist and cover a wide variety of organisms ranging from bacteria to mammals (see table 2 for more examples). In fact, the observed prevalence of this particular regulatory design for control of genetic redundancy raises questions as to the specific selectable functions that it performs. In this review we will highlight common denominators of known RBCs and use these to suggest principles that govern the utilization of redundancy and its evolution.

### **Conditional Co-Regulation and the maintenance of metabolic fluxes.**

A key requirement of the metabolic regulation is the maintenance of metabolite fluxes, despite the sometimes extreme changes in the external conditions and nutrient

availability. On evolutionary time scales, adaptation to extreme environmental changes is sometimes achieved by preservation of gene duplicates (Kondrashov, Rogozin et al. 2002). This is illustrated by the numerous observations of adaptive gene amplifications in response to antibiotics (Koch 1981; Romero and Palacios 1997), anticancer drug treatments (Stark and Wahl 1984; Schwab 1999; Montgomery, Price et al. 2001), nutrient limitations (Sonti and Roth 1989; Romero and Palacios 1997; Brown, Todd et al. 1998), and more (see (Kondrashov, Rogozin et al. 2002) for more examples). Also, it was suggested from a genome wide analysis that enzymes corresponding to reactions with higher metabolic fluxes are more likely to have duplicate partners (Papp, Pal et al. 2004). For a single cell, the ability to quickly and efficiently respond to fluctuating environments is crucial and offers an obvious evolutionary advantage. One avenue through which functional redundancy is utilized to facilitate this ability is by exploiting the differential efficiencies generated by divergence. For example, in yeast, the HXT gene family encodes a redundant set of membrane hexose transporters with varying affinities towards glucose and consequently different transport efficiencies (Ozcan and Johnston 1999). This variation together with glucose-tuned regulation enables the control of glucose flux. This is done by the expression of high affinity transporters when glucose is limited and low affinity transporters when glucose is abundant (Ozcan and Johnston 1999), thus allowing the cell to adapt to different external glucose availability. Other examples that fall under this principle include the transport of iron, copper, manganese, zinc and other metals in yeast (Marini, Soussi-Boudekou et al. 1997; Eide 1998). Interestingly, some of these transport systems are further regulated by RBCs (Nicholson, Sawamura et al. 1998; Ozcan 2002), which, as we will show later, may serve the beneficial role of providing robustness of the flux control to internal noise arising from genetic variation.

Within the metabolic network, fluxes are governed and regulated by the concentration of active enzymes catalyzing the different reactions. Although the detailed contribution of functional redundancy to this regulation is not fully established, such contribution is highly expected given the large amount of isozymes and other redundancies that exist within these networks (supplementary fig 1). Supplementary figure 1 shows that individual isozymes are less essential and produce less deleterious effects upon deletion than enzymes existing in single copy. An interesting twist to this

account comes from the fact that isozymes, although redundant and consequently dispensable, obey different regulatory programs and are transcribed at different times in response to environmental pressures (Gasch, Spellman et al. 2000; Ihmels, Levy et al. 2004; Kafri, Bar-Even et al. 2005). A recent finding supplying the compromise between these two seemingly opposing observations shows that many differentially regulated genes can be induced for co-expression given that particular environmental stress stimuli were applied (Kafri, Bar-Even et al. 2005). More so, gene pairs that maintain this capacity for conditional co-expression were shown to be the most likely candidates for compensating against deletion mutations (Kafri, Bar-Even et al. 2005). This conditional co-regulation (CCR) (referred to as PCoR in (Kafri, Bar-Even et al. 2005)) may provide essential clues for the function of these redundancies in the regulation of metabolic fluxes. The model that emerges is that while many isozymes are specialized for different environmental regimes, alarm signals induced by particular stress stimuli may call for their synergistic co-expression. Here, RBCs provide functional specialization together with extreme flexibility in gene control that could be activated when sufficient stress has been applied. For example, in yeast, glucose serves as a regulatory input for alternating between aerobic and anaerobic growth. Its presence is detected by two separate and independent signaling pathways, one probing intracellular glucose concentrations and the other probing extracellular concentrations (Ozcan 2002). This differential sensing enables some genes to be separately regulated by either intracellular or extracellular glucose. One consequence of this shows effect in the responsive backup circuit composed of Hxt1 and Hxt2. Here feedback is made possible by having Hxt2 controlled by two opposing signals. One is its induction by extracellular glucose and the second is its repression by intracellular glucose (Fig 2 & (Ozcan 2002)). The consequence of this is that while high glucose concentrations results in repression of Hxt2 expression, its induction could either be triggered by low environmental sugar, or alternatively, by mutations in genes responsible for glucose influx (Ozcan 2002). Other similar examples include the isocitrate dehydrogenases *idp2* and *Idh* where the glucose repression of *idp2* is reversed in the  $\Delta ihd$  mutant (McCammon and McAlister-Henn 2003) and for the pair *Acs1* and *Acs2* where *Acs2*'s expression is induced in the  $\Delta Acs1$  mutant (van den Berg, de Jong-Gubbels et al. 1996). In all these cases, the common denominator is that one of the two

duplicates is under repression in wild-type, and that that repression is relieved upon its partner's mutation.

### **Redundancies of developmental regulators**

The extent to which genomic functional redundancies have influenced the way we think about biology can be appreciated simply by inspecting the vast number of times the word 'redundancy' is specifically referred to in the biomedical literature (fig 1). Particularly interesting is the abundance with which it is addressed in studies of developmental biology (fig 1). In fact, it is here that concepts such as 'genetic buffering' and 'canalization' (Waddington 1942) had first been suggested. Furthermore, the robustness of the developmental phenotypes such as body morphologies and patterning have been repeatedly demonstrated (Gerhart and Kirschner 1997). So the question is, are these redundancies simply leftovers of ancient duplications or are they an additional component to the sophisticated machinery of cellular regulation?

In criticism one may argue that many of the reported redundancies do not actually represent functionally equivalent genes but rather reflect only partial functional overlap. In fact, knockout phenotypes have been described for a number of developmental genes that have redundant partners ((Qiu, Bulfone et al. 1995; Qiu, Bulfone et al. 1997; Enns, Kanaoka et al. 2005)

Although this may suggest that these redundancies have not evolved for the sake of buffering mutations, it has, in our opinion, little relevance to the question of whether or not they serve a functional role. The interesting question is then, can such a functional role for the duplicated state be inferred from the way the two genes are regulated?

For most cases of developmental redundancies, redundant partners are either temporally or spatially distinct in their expression patterns (table 2), however some level of expression overlap is usually observed. Cross-regulation of redundancies has only been tested for in a relatively small number of cases, yet from those, a few persuasive recurring themes do emerge. One of the better known cases of cross-regulated developmental regulators is that of the four master regulators of vertebrate skeletal muscle development: MyoD, Myf-5, myogenin and MRF4, collectively known as the MRF gene family (Sabourin and Rudnicki 2000). These four basic Helix-Loop-Helix

transcription factors specify and execute the process through which naïve mesoderm cells differentiate to form distinct skeletal muscles (for review see (Olson and Klein 1994; Molkenin and Olson 1996)) and are activated sequentially during myogenesis. The myogenic pathway consists of two separate phases. In the first phase, MyoD and Myf-5 specify the myogenic progenitors in the somites into myoblasts which are cells that are committed to become muscle fibers. In the second phase, myoblasts develop into myofibers, a process initiated by myogenin and MRF4 (Zhang, Behringer et al. 1995).

Sequence similarity between the myogenic transcription factors suggests that these have evolved through multiple gene duplication events early in the evolution of vertebrates, approximately with the appearance of fish (Krause, Fire et al. 1990; Michelson, Abmayr et al. 1990; Venuti, Goldberg et al. 1991; Holland, Holland et al. 1992; Atchley, Fitch et al. 1994). Interestingly, despite their long evolutionary separation, these regulators have largely conserved their functional redundancy. In fact, experiments on mice where MyoD was completely inactivated resulted in viable and fertile mice that exhibited phenotypically normal skeletal muscles (Rudnicki, Braun et al. 1992). In strong contrast, mice lacking both MyoD and Myf-5 lack skeletal muscle altogether and die soon after birth (Rudnicki, Schnegelsberg et al. 1993).

From the perspective of this review, myogenesis is a particularly interesting process as it harbors two responsive backup circuits. The first is manifested by the up-regulation of Myf-5 in response to mutations in MyoD (Rudnicki, Braun et al. 1992). The second is described by the induction of myogenin in response to mutations in MRF4 (Zhang, Behringer et al. 1995). An additional interesting feature of the MRF responsive backup circuits is by what we term '*dosage dependent linear response*'. By this we wish to account for the observation that the up-regulatory response induced by a heterozygote mutation is approximately half that of the homozygote one. In particular, for MyoD and Myf-5, mutations in one of the two MyoD alleles results in an 1.8-fold up-regulatory response of Myf-5, while disruption of both alleles results in a 3.5-fold response (Rudnicki, Braun et al. 1992). This type of linearity may hold clues as to the both function and regulation of these genetic circuits. One attractive possibility that may be suggested by this linearity is that the process carried out by these redundant regulators benefits from constancy of the sum of their protein concentrations. In other words, while



the concentration of MyoD may fluctuate due to noise in gene expression or false induction, the sum of MyoD + Myf-5 may have evolved to remain constant (see formal analysis below).

An additional example illustrating dosage dependent linear response constitutes the *Pax1* and *Pax9* regulators of sclerotome development. Here functional redundancy has been established at the phenotypic level from mutant mouse experiments showing that *Pax1* can fully rescue *Pax9* mutants and conversely *Pax9* can offset the *Pax1*-null phenotype to a substantial degree (Peters, Wilm et al. 1999). In line with what seems to be the general case for numerous examples of developmental redundancies (see table 1), *Pax1* and *Pax9* have partially overlapping expression domains during early development, particularly in the sclerotomes (Peters, Wilm et al. 1999). This overlap, though, decreases in the later stages of development. The responsive circuitry of these regulators was established using the lacZ/gal system to show an up-regulation and spatial expansion of *Pax9* expression in the sclerotomes of the *Pax1* mutant (Peters, Wilm et al. 1999). Thus *Pax9* expression in the *Pax1* mutants was observed in cells that in wild-type exhibit only *Pax1* expression. Dosage-dependency was observed by comparing phenotypes of combinations of wild-type, heterozygous and homozygous mutants of *Pax1* and *Pax9* (Peters, Wilm et al. 1999). It is worth noting that functional redundancy was also suggested for other members of the Pax gene family, in particular for the two pairs *Pax2/Pax5* (Schwarz, Alvarez-Bolado et al. 1997) and *Pax3/Pax7* (Mansouri and Gruss 1998). All nine member family of the Pax transcription factors carry roles in the genetic control of mammalian organogenesis (for review see (Dahl, Koseki et al. 1997)).

Other examples of responsive circuits of redundant developmental regulators include the closely related homeobox gene pair *Gsh1* and *Gsh2* for which mutational inactivation of *Gsh1* resulted in a pronounced expansion of *Gsh2* expression in the cerebral cortex and olfactory bulb of mice with an apparently normal phenotype (Toresson and Campbell 2001). The vertebrate *Distal-less* related regulators *dlx3* and *dlx7* where morpholino-induced inactivation of *dlx3* resulted in a strong induction of *dlx7* mRNA expression in zebrafish embryos (Solomon and Fritz 2002). The two functionally overlapping E3 ligases, Smurf-1 and Smurf-2 (Yamashita, Ying et al. 2005) for which knockouts of Smurf-1 were shown to result in an upregulatory response of Smurf-2

(Kavsak, Rasmussen et al. 2000; Lin, Liang et al. 2000; Zhang, Chang et al. 2001). The midkine and pleiotrophin cytokines for which functional redundancy was observed and a strong up-regulatory response of pleiotrophin was shown to result from double knockout of the midkine gene (Herradon, Ezquerro et al. 2005). And other examples as listed in table 2.

The abundance of redundancies occurring in genes related to developmental processes and their functional role as master regulators (fig 1) may be taken to suggest their utilization in either the flexibility or robustness of regulatory control. In fact, redundancies among high-level regulators have often been reported outside the context of developmental pathways (fig 1 and table 2). Although for most examples the regulation that they confer on one another was not assessed, some have been specifically identified as displaying negative cross-regulatory inhibitions (see table 2). One such example existing in *E. coli* is that of the pair *stpA*/*HN-S* that regulate genome-scale transcriptional response to DNA damage (Dorman 2004). This pair of regulators displays an additional complexity where its regulation is induced by pairwise associations to either form homodimers composed of either of the pair members or heterodimers containing both (Dorman 2004). Nevertheless, mutational inactivation of *HN-S* induced an up regulatory response of its partner, *stpA* with only marginal effect on phenotype (Zhang, Rimsky et al. 1996; Free and Dorman 1997). A more recent example indicates that the multidrug resistance phenomenon in *S. cerevisiae* is also regulated by an RBC encoding for the up-regulation of the transcription factor *YRR1* in response to the deletion of its partner, *YRM1* (Lucau-Danila, Delaveau et al. 2003; Onda, Ota et al. 2004).

### **Recurring regulatory patterns**

Two architectures of cross-regulated redundancies may exist. According to the first, inactivation of each of the redundant genes from a given pair would result in the induction of the other (fig 3A) and according to the second, only one of the pair members is responsive (fig 3B & 3C). We therefore suggest the terminology *bidirectional* and *unidirectional* RBCs. This distinction is important as from the current literature review, all but two of the examples seem to fall into the unidirectional category. In light of this,

we further suggest, for unidirectional RBCs, the distinction between the *responsive* gene and the *controller* gene.

The above is but one of several asymmetries and regulatory patterns that systematically recur throughout the literature. An additional example is the classification of redundant pair members into a ubiquitously expressed gene partner and a sporadically expressed one (see table 2). One of the most profound and insightful of these recurring regulatory themes is that, while both genes are capable of some functional compensation, disruption of the responder produces a significantly less deleterious phenotype than disruption of the controller (table 2). An insightful example illustrating this entails the pair of genes *Fks1* and *Fks2* redundantly encoding the catalytic subunit of the yeast 1,3- $\beta$ -glucan synthase (Douglas, Foor et al. 1994; Inoue, Takewaki et al. 1995). This enzyme is responsible for the generation of cross-links within the 1-3- $\beta$ -glucan matrix comprising the major structural component of the yeast cell wall. Being such, this process requires very tight regulation with cell wall degradation and cell cycle to enable budding and isotropic cell growth. This fact is also suggested from the numerous associations of *FKS1* within the genetic interaction network, illustrating its linkage to processes such as cell cycle control, environmental stress responses and mating (Lesage, Sdicu et al. 2004). The surprising aspect of this story is that despite this high connectivity of *FKS1*, *FKS2* is only sparsely connected (Lesage, Sdicu et al. 2004). More so, while deletion of *FKS1* induces an up-regulatory response of *FKS2* with mild phenotypic effects, deletion of *FKS2* induces no regulatory response of *FKS1* but also no detectable effect on the phenotype (Douglas, Foor et al. 1994). This result may seem counter-intuitive as it is *FKS2* that is up-regulated to rescue against deletion of *FKS1* and not vice-versa, yet *FKS2* is the more dispensable gene within this pair. (see table 2 for more examples). A simple potential interpretation may suggest that while the controller is the key player performing some essential biological role, the responder is merely a less efficient substitute. Yet, accepting the notion that redundancy could not have evolved for the sake of buffering mutations, this interpretation is still severely lacking.

A different, and more biologically reasonable hypothesis accounting these asymmetries is that one of the functions of the responder is to buffer dosage fluctuations of the controller. This buffering capacity requires a functional overlap that also manifests

itself in compensations against the more rare event of gene loss. Other models accounting for this are further discussed in this manuscript but our main point of argument is that this complex regulation of functionally redundant, yet evolutionarily conserved genes, strongly indicates utilization of redundancy.

### Regulatory designs

What regulatory design could account for a gene sensing and responding to its redundant partner's intactness? From the most general perspective there are three possible regulatory schemes that could answer this question. Scheme A (fig 4A) entails a direct negative regulation of a gene by its functionally redundant partner. Scheme B (fig 4B) utilizes the substrate abundance as a proxy for its partner's activity. In other words, over-accumulation of substrate, potentially caused by reduced or abolished efficiency of one of the RBC pair members, signals for over-production of the second member. Scheme C employs end-product inhibition. Assuming that an end-product may inhibit both redundant partners, the lack of function of one of the partners would result in absence of the product and hence relief of repression from the second partner. Conceptually, schemes B and C are symmetric.

One instance of an RBC that relies on a direct regulatory interaction between redundant partners without involving either substrate or end-product regulation constitutes the two vertebrate Distal-less related regulators, *dlx3* and *dlx7* (fig 5). These paralogous transcriptional regulators are both expressed in embryonic development and are involved in the development of auditory and olfactory placodes (Ekker, Akimenko et al. 1992; Akimenko, Ekker et al. 1994; Ellies, Stock et al. 1997). By injecting anti-*dlx3* and anti-*dlx7* morpholino oligonucleotides (MO) in zebrafish it was showed that whereas the simultaneous inhibition of both genes (*dlx3+7*-MO) resulted in embryos having severe defects in the auditory and olfactory placodes, *dlx7* loss-of-function embryos appeared phenotypically normal and *dlx3* MO embryos exhibited only smaller auditory placodes and inner ear structures than normal (or wild-type) embryos (Solomon and Fritz 2002). Increase in *dlx7* mRNA was observed in the *dlx3* MO embryos (Solomon and Fritz 2002).

The regulatory relationships between the *dlx3* and *dlx7* were tested by measuring the mRNA content of the different MO treated embryos (Solomon and Fritz 2002) and are summarized in a network diagram (Fig. 5) featuring both cross and auto regulation. Thus, the lesson is that, for this case, redundancy is embedded within a more complex interaction network that includes a unidirectional responsive circuit in which the controller (*dlx3*) also represses its own transcription while the responder (*dlx7*) is a positive auto-regulator.

Another interesting example for which the regulatory pathway leading to induction was well characterized constitutes the unidirectional RBC of Fks1 and Fks2, in yeast (discussed above in a different context). Here, the responder (Fks2), in addition to being activated in the *fks1*  $\Delta$  mutant, is also induced by heat-shock, cell wall damage, pheromone and  $\text{Ca}^{+2}$  (Garcia-Rodriguez, Trilla et al. 2000). The intricate design of this circuitry is realized by the fact that there are two different, alternative, signaling pathways that operate synergistically to control Fks2 expression. While response to Fks1 deletion is activated through a calcineurin/ $\text{Ca}^{+2}$  dependent pathway, response to heat shock and cell wall damage is induced by both the former and the Rho1 dependent cell integrity pathway (Zhao, Jung et al. 1998). Even more interestingly, it was found that these two pathways induce different and complementary dynamics of the Fks2 response (Zhao, Jung et al. 1998). Specifically, while the calcineurin -dependent pathway induces a rapid but transient response, the Rho1-dependent pathway induces a delayed response that is sustained for longer time scales.

An end-product activated feedback mechanism is demonstrated by the hexose transporters Htx1 and Hxt2 in yeast (Fig 2) where the expression of both genes is repressed by the level of intracellular glucose. Thus, once the flux of glucose from the environment to the yeast's cytoplasm decreases an additional glucose pump is induced for expression.

### **Functional consequence of cross-regulated redundancies.**

The up-regulation of the responsive gene in RBCs could be encoded for in a number of different ways, a few of which are schematically depicted in table 1 (first column). These different regulations, though, are not equivalent and result in differences

in both dynamics and steady states of the response. Inspired by a recent approach for the analysis of network motifs (Shen-Orr, Milo et al. 2002; Mangan, Zaslaver et al. 2003; Tyson, Chen et al. 2003) we quantified the different RBCs shown in table 1 for their steady state characteristics (see supplementary material for detailed calculations). Our basic hypothesis was that if there are biological functions that exploit redundancies between RBC pair members, such functions, to a first approximation, would be proportional to either the sum (*independent functions*) or the product (*synergistic functions*) of the concentrations of the two redundant proteins (fig 6 which for the sake of illustration shows a specific transcription factor oriented example). Examples of the first include reactions that are catalyzed by two independently functioning isozymes. In such cases the total rate of product production catalyzed by the pair of isozymes would be equal to the production rate contributed by the first isozyme plus that of the second. Examples of the second, i.e. synergistic exploitation of redundancies, could be understood by reactions that exploit cooperativity between RBC pair members. Examples are biological functions that are carried out more efficiently by heterodimers of the partially redundant proteins (see HN-S and StpA for an illustrative example).

Table 1 shows the steady state solution for differential equations that describe the dynamics of three different RBC motifs. We decided to limit the scope of the formal analysis only to these three motifs as we wish to address only sub-networks that maintain a negative feedback and that can be modeled with no more than two free variables. Given an RBC composed of two genes,  $G_1$  and  $G_2$ , we examined the capacity of each of the analyzed motifs to sustain constant steady state concentrations of  $G_1 + G_2$  or  $G_1 \times G_2$  with respect to variations in the concentration of  $G_1$ . In other words we computed the steady state susceptibility (Paulsson 2004),  $H_{ss}$ , of both synergistic and additive functions of RBCs to variations in the controller  $G_1$ . By doing this we asked whether an RBC may serve to filter the downstream processes from variation and fluctuations arising from non-genetic noise (Elowitz, Levine et al. 2002), genetic variability of  $G_1$  or genetic variability effecting  $G_1$ 's regulation. An extreme case of such variations entail the consequence of a deleterious mutation (or knockout/down) of the controller,  $G_1$ , as in the experiments described throughout this review.

The three RBC motifs that we examined include: Motif I (simple repression), described by  $G_1$  exerting a negative regulation on  $G_2$ ; Motif II (damped controller) where  $G_1$  exerts negative regulation on itself and on  $G_2$ ; and Motif III (cycled feedback) where  $G_1$  exerts negative regulation on  $G_2$ , and  $G_2$  in turn, positively regulates  $G_1$ .

The first observation that is immediately apparent from examination of the results in table 1 is that the product of the concentrations of the two redundant proteins is insensitive to variations of the controller,  $G_1$  ( $G_1 \times G_2$  is not dependent on  $v_1$  - the maximal production rate of  $G_1$ ). One important aspect of this result is its generality as it extends beyond the scope of RBCs and may hold true for complexes or heterodimers given that one monomer with these complexes negatively regulates the rest (as may be the case for some heterodimers (see HN-S/StpA, table 2)). An interesting question stemming from this is what are the functional advantages that are associated with this property? One attractive possibility is that this regulatory design masks the function from dosage fluctuations of the controller. Thus while in the extreme case of  $G_1$ 's complete deletion the circuit will be highly robust to variations in  $G_1$ 's level.

Since down-regulation of one member of an RBC leads to an up-regulation of its partner, it may be expected that the total sum of both proteins will stay relatively constant despite variations in  $G_1$ 's concentration. In other words, for a given RBC, a fluctuation of the controller,  $G_1$ , is expected to be counteracted by a *reciprocal effect* of the responsive partner,  $G_2$ . To quantify the relationship between these counteracting effects we computed the steady state susceptibility,  $H_{ss}$ , (Paulsson 2004) of  $G_1+G_2$  with respect to  $v_1$  for the different RBC motifs (Fig. 7). Since our analyses were strictly focused on long time scale fluctuations, we computed  $H$  by considering the effects of the fluctuations on the steady state levels rather than on changes in reaction rates as defined in (Paulsson 2004). Results of these analyses are presented in table 1 and shown in figure 7 for particular parameter values. Unlike the almost absolute buffering capacity RBCs offer for synergistic functions proportional to  $G_1 \times G_2$ , here the counteracting effect of the responder is inversely proportional to the square of the induction level. The consequence of this is that this counteracting effect becomes less effective at high production levels of the controller. This result is clearly demonstrated in figure 7 displaying the dependency of  $G_1+G_2$  on  $v_1$  for motifs I-III.

From figure 7 it can be clearly seen that RBC motif II provides an effective buffering for the broadest range of controller production levels. Interestingly, this specific responsive circuit was the one found to exist for the pair of regulators Dlx3/Dlx7, where Dlx3 is the controller and Dlx7 is the responder (Solomon and Fritz 2002). The buffering efficiency of this motif is due to both the complementing effect of the responsive gene and the negative auto-regulatory effect of the controller. Although the precise range of effectiveness of each of these motifs is very much parameter dependent, motif II has some fundamental advantages over the others. First, the strength of the restoration response, but not the inductive response, can be fine tuned by the level of induction of the responsive gene. This can be seen by the fact that while  $c_2$  is a function of  $v_2$  (see table 1),  $c_1$  is constant. Although this is also true for motif I, motif II has the additional advantage of the negative auto-regulation of the controller.

### **Prevalence of RBC in literature and in nature**

One of the central open questions that must still be addressed is just how common are RBCs in biology and how frequently do they appear in the different genomes. Also, we can ask whether this frequency, be it low or high, can be faithfully estimated from the amount of times RBCs are reported in the literature. In other words, do RBCs represent a 'genome wide' phenomenon or a collection of rare incidents? It is unfortunate that at this point no conclusive answer can be provided to that question, mainly because of the very limited number of studies that have specifically probed for cross-regulation among redundant protein pairs. Since absence of phenotype in a knockout experiment (a potential consequence of RBCs) is often looked upon by researchers as an 'insignificant result', analyses of the 'minor' regulatory consequences of such knockouts may have not seen the light of publication. One approach to overcome this is to collect pairs of genes for which the double knockout is lethal while the separate gene disruptions are not, i.e., synthetic lethals. In some respect these interactions could be regarded as proxies for functional compensatory interactions between these genes and thus some form of functional redundancy. The problem is that a close inspection reveals that the vast majority of synthetic lethal pairs are far from being even remotely functionally equivalent. For example, the list of synthetic lethals of FKS1 gene in yeast (Lesage, Sdicu et al.



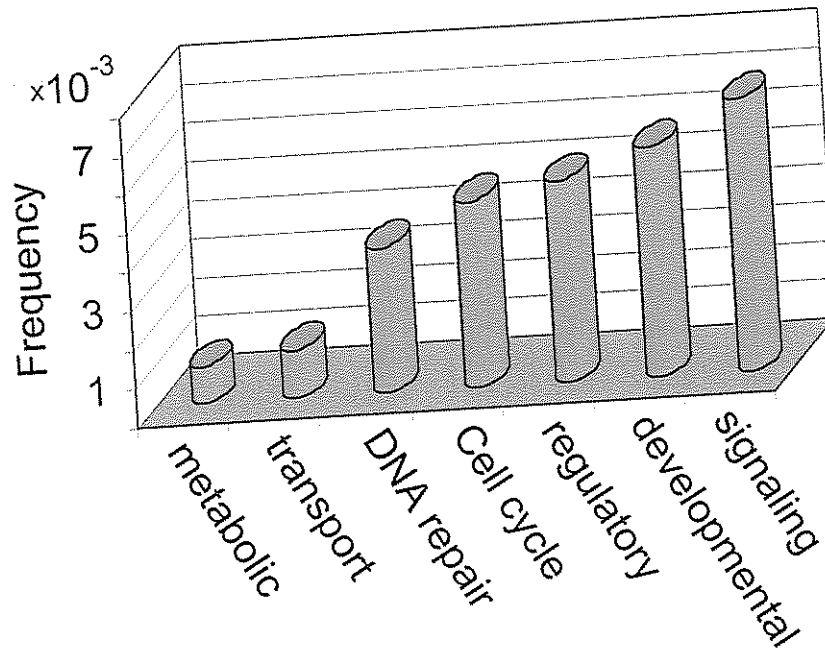
2004), include the gene FEN1. This gene is an enzyme that lies on the pathway of sphingolipid biosynthesis. Its genetic interaction with FKS1 is thought to result from an accumulation of its substrate, phytosphingosine, in the plasma membrane of the FEN1 deletion mutant. Phytosphingosine, in turn, is thought to repress the interaction between Fks1p and Rho1p, possibly by forming a microdomain around Fks1p and physically preventing its association with its regulatory subunit, Rho (Abe, Nishida et al. 2001). Thus, despite the genetic interaction, no redundancy can be implied.

Furthermore, we argue while the question of prevalence of RBCs is important, one can not judge their biological significance solely based on that criteria. This fact is emphasized, for example, by the pair of genes utrophin and dystrophin associated with structural components of muscle fiber. This pair of genes constitutes an RBC in that utrophin was found to be up regulated in the absence of its homolog dystrophin (Porter, Rafael et al. 1998). Particular attention has been attracted to this gene pair as dystrophin mutations were found causative of the Duchenne muscular dystrophy condition in human patients. Yet, in mice, utrophin has shown remarkable ability to compensate for dystrophin knockouts and it is estimated that partial compensation also occurs in humans (Deconinck, Rafael et al. 1997). Inspired by the efficient compensatory effect this RBC has in mouse studies have been suggested to artificially induce in human patients by means of gene therapy (Deconinck, Rafael et al. 1997; Porter, Rafael et al. 1998; Dowling, Culligan et al. 2002). Although these ambitions are not yet been realized they point a fruitful possibility.

One of the challenges of this work was to survey the literature for already documented cases of RBCs. The difficulty of this lies in the fact that there is, currently, no consistent ontology to describe backup circuits. As a result there is no simple pubmed search that could retrieve all such cases. Instead, we employed here a manual laborious search strategy in which papers whose abstracts contained key words such as "redundancy", "functional overlap", "paralogous", etc. were carefully examined. Future publications of knockout experiments with even very little phenotypic effects, using the RBC terminology coined here may facilitate further study of recurring principles that govern the regulation of genetic redundancies.

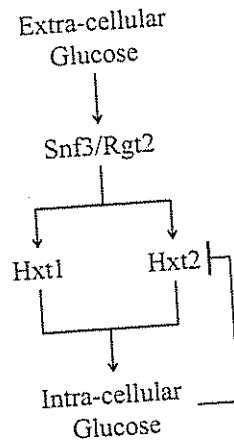
## Figures

Figure 1



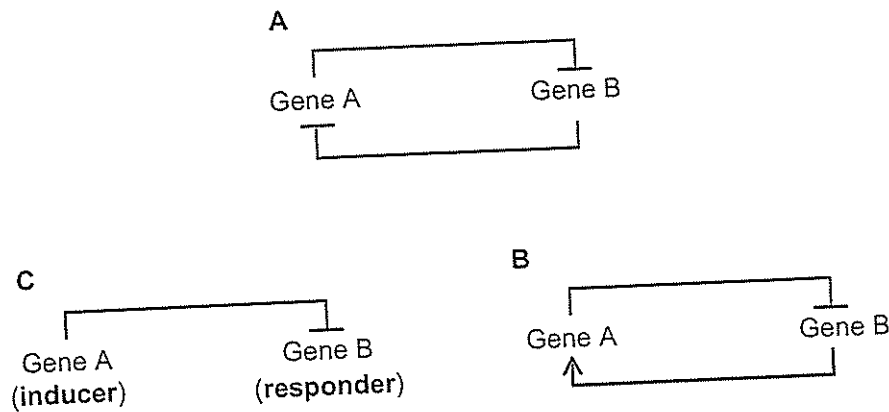
**Proportion of reported redundancies in different functional categories.** Proportions were estimated by the number of times a key word, such as “cell cycle”, appeared in Pubmed together with the term “functional redundancy” divided by the total number of times that that key word appeared.

Figure 2



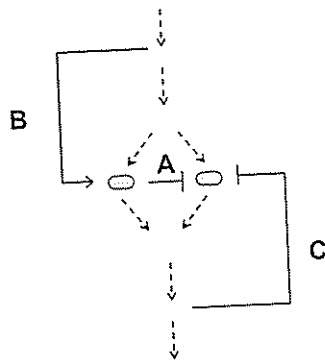
**The Hxt1/Hxt2 responsive backup circuit.** Extra-cellular glucose concentration is sensed by two membrane receptors on the outer yeast membrane, Rgt2 and Snf3. These, once activated by glucose, initiate a signal cascade that induces the transcription of the Hxt gene family of hexose transporters encoding membrane channels for glucose intake. The flux of incoming glucose generates an increase in intracellular glucose concentration, which, in turn, repress the transcription of Hxt2.

Figure 3



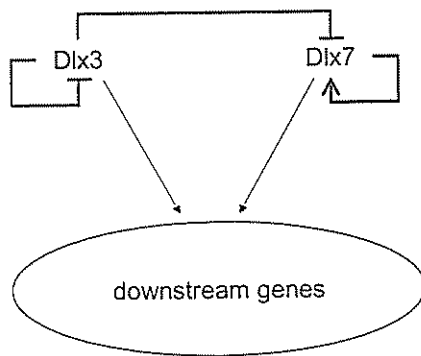
Bidirectional (A) versus unidirectional (B, C) responsive circuits.

Figure 4



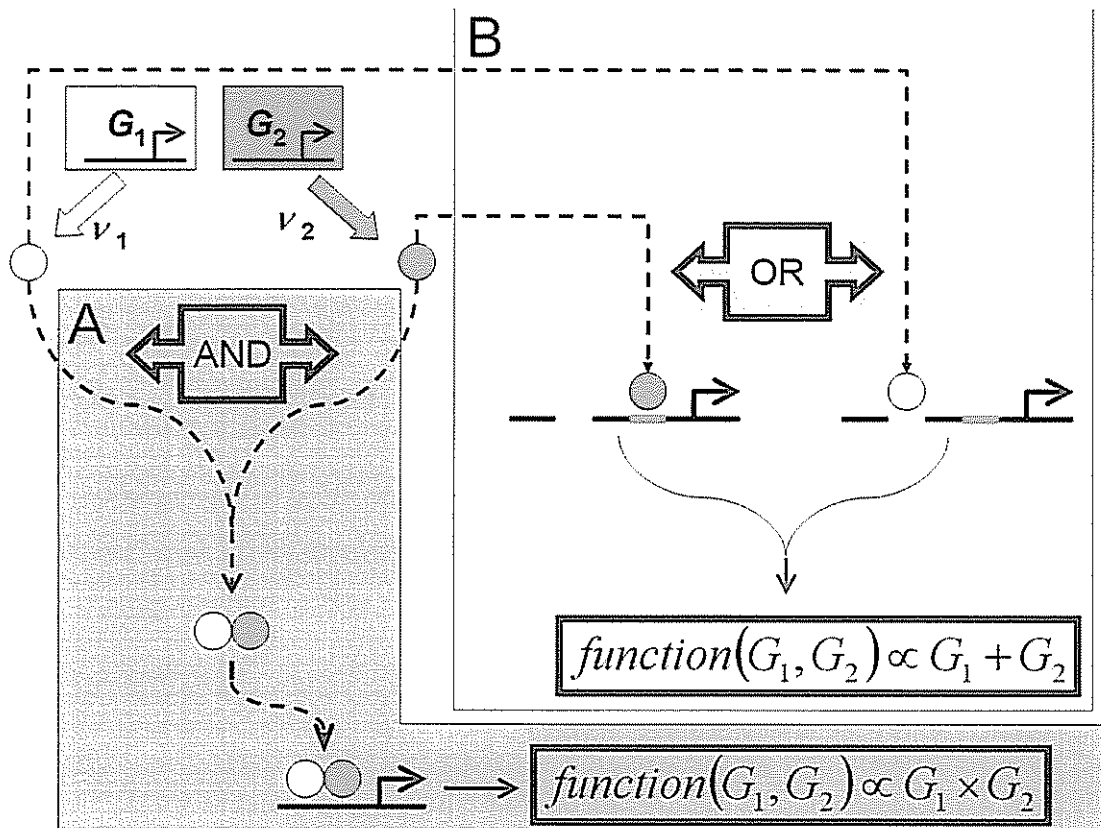
**Three possibilities for feedback in responsive backup circuits.** For one duplicate gene to sense and respond to its partners' intactness feedback mechanisms must be at play. In this diagram duplicates are represented as circles that lie embedded within a reaction pathway illustrated by the consecutive arrows. Lines **A**, **B** and **C** represent the three feedback possibilities, namely, simple negative regulation (**A**), substrate induction (**B**) and end-product regulation (**C**).

Figure 5



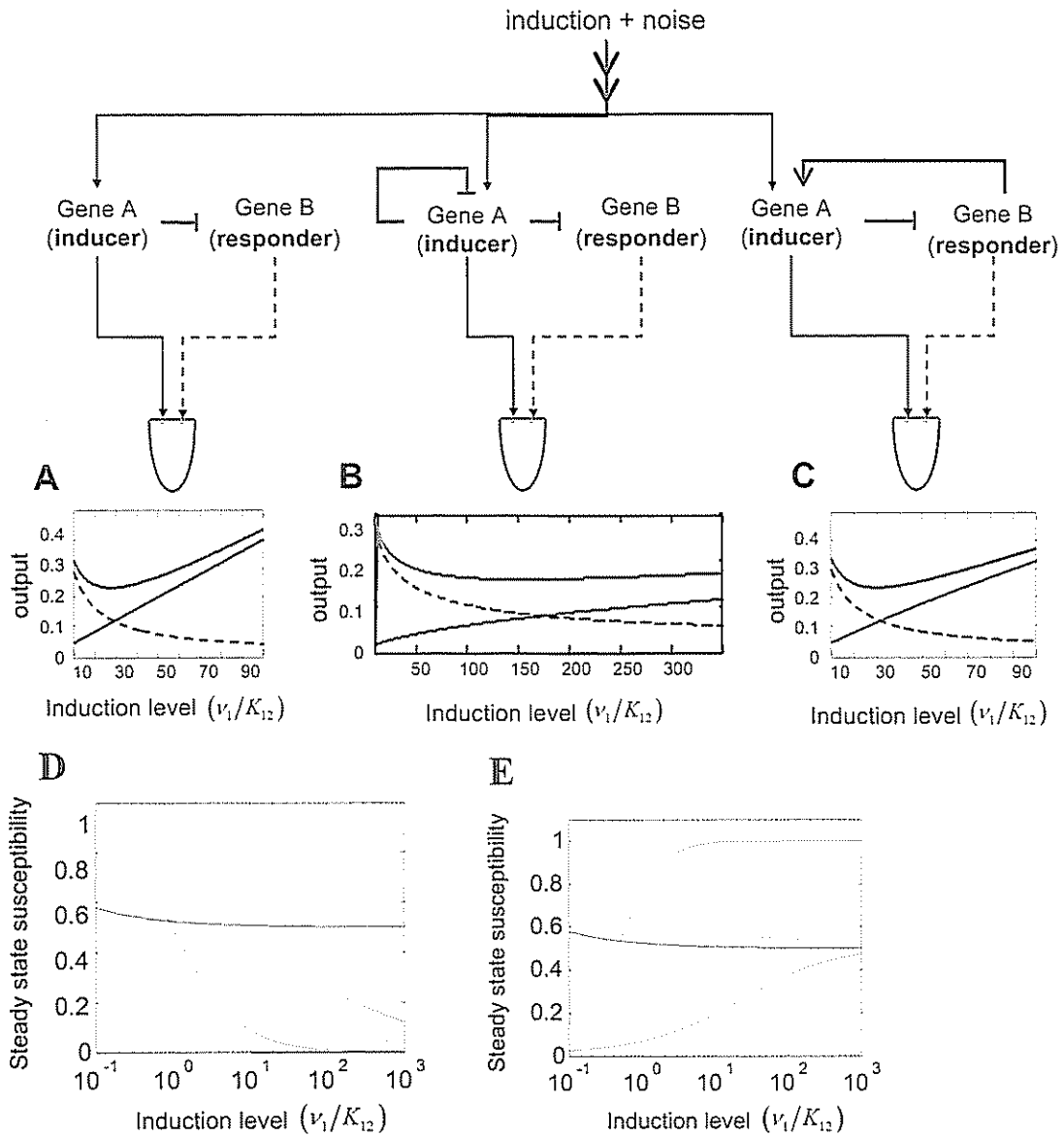
The regulatory wiring for the two distal-less developmental regulators Dlx3 and Dlx7 as deduced from morpholino antisense translation inhibitions (Solomon and Fritz 2002).

Figure 6



Example illustrative reactions proportional to the product (A, blue) or the sum (B, purple) of the two redundant gene products. Genes  $G_1$  and  $G_2$  encode proteins  $G_1$  and  $G_2$  (green and yellow circles) which act cooperatively to activate one function gene (blue) and alternatively to activate another (purple).

Figure 7

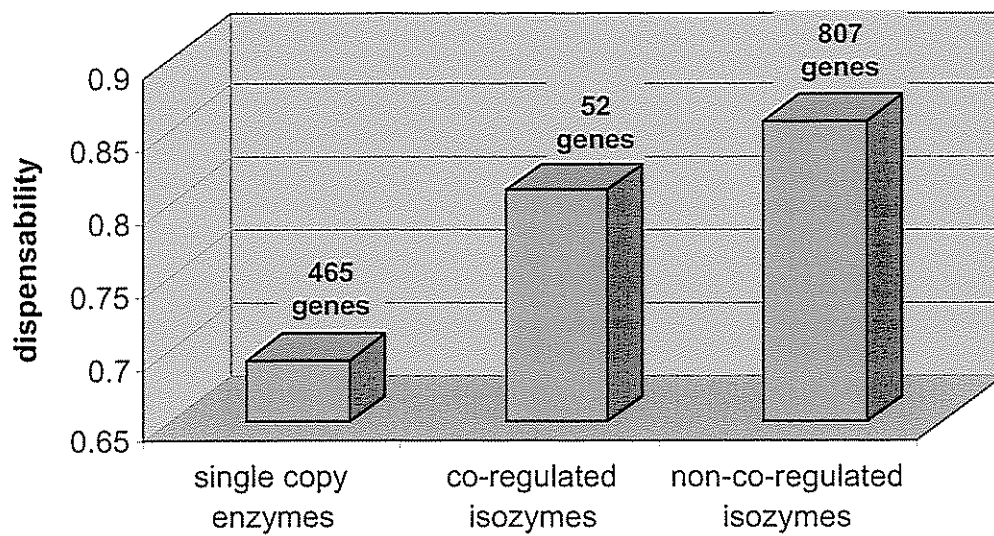


**Sustainability of output with respect to changes in input level for different RBC of genes with independent functions.** The plots show the dependency of the controller (solid black), the responder (broken black) and their sum,  $f_2 = G_1 + G_2$ , (green) on the induction level of  $G_1$ . Induction level is shown in units of  $K_{12}$ , such that an induction level of 1 corresponds 50% saturation of the  $K_{12}$  promoter element. The purple vertical line in **B** corresponds to  $v_1/K_{12} = 100$  to help comparison. **D** and **E** show, respectively,

the susceptibility,  $H = \frac{\partial \ln(f)}{\partial \ln(v_1)}$ , of  $f_2$  and  $f_1$  to fluctuations in  $v_1$ , i.e. the induction level of

$G_1$ . Susceptibilities were calculated separately for in the context of motif I (blue), motif II (red) and motif III (green).

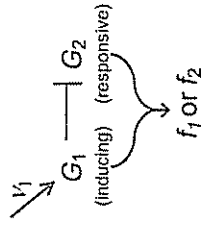
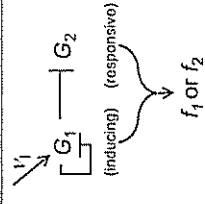
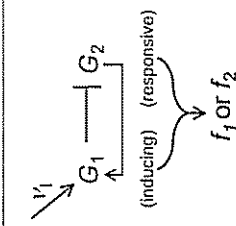
Supplementary figure 1



Proportion of genes with a viable deletion-phenotype (“dispensable genes”) among single copy enzymes, enzymes that are co-regulated with their paralogous partners and enzymes that are non-coregulated with their paralogous partners. Above each bar is the number of enzymes in that bin.



**Table 1: Properties of Responsive Backup Circuits**

	Model	Function I $f_1 = G_{1st} \times G_{2st}$	Function II $f_2 = G_{1st} + G_{2st}$	Susceptibility, $\left(\frac{v_1}{f_2} \frac{\partial f_2}{\partial v_1}\right)^*$ * The susceptibility to fluctuations in the production rate of $G_1$	Constants
<b>Motif I simple repression</b>		$v_2 K_{12}$	$v_1 c_1 + \frac{1}{v_1} c_2$	$\frac{v_1^2}{v_1^2 + c_2} - \frac{c_2}{v_1^2 + c_2}$	$c_1 = 1$ $c_2 = v_2 K_{12}$
<b>Motif II Dampened controller</b>		$v_2 K_{12}$	$\sqrt{v_1} c_1 + \frac{1}{\sqrt{v_1}} c_2$	$\frac{1}{2} \left( \frac{v_1}{v_1 + c_2} - \frac{c_2^*}{v_1 + c_2^*} \right)$	$c_1 = \sqrt{K_{11}}, c_2 = v_2 K_{12}$ $c_2^* = \frac{c_2}{c_1}$
<b>Motif III Cycled feedback</b>		$v_2 K_{12}$	$\sqrt{v_1} c_1 + \frac{1}{\sqrt{v_1}} c_2$	$\frac{v_1}{v_1 + K} - \frac{K}{v_1 + K}$ * Simplifications: 1) $K_{12} = K_{21} = K$ 2) $G_1 G_2 > K$	$c_1 = \sqrt{v_2} \sqrt{\frac{K_{12}}{K_{21}}}$ $c_2 = \sqrt{v_2} \sqrt{K_{12} K_{21}}$

\* All formulas in the table were calculated based on the simplifying assumption:  $K_{12} \ll G_1$ . For detailed calculation of solutions, with and without this simplification, see supplementary text.

$G_1$	Controller gene.	$\alpha$	Degradation rate
$G_2$	Responsive gene	$K_{IL}$	Thermodynamic dissociation constant between protein 'I' and its cis-regulatory motif on gene 'L'
$\beta$	Transcription rate	$\nu \equiv \frac{\beta}{\alpha}$	maximal steady state concentration

**Table 2 : Example Responsive Backup Circuits**

Genes	organism	directionality	Linear		essential		reference
			response	gene	gene	function	
HXT1/HXT2	yeast	unidirectional	Quantitatively responds to its own ploidity (Galitski, Saldanha et al. 1999)	HXT2	HXT1	Membrane transport	(Ozcan 2002)
Idp2 / idh	yeast	unidirectional	n.d.	Idp2	IHD	enzymes	(McCammon and McAlister-Henn 2003)
Fks1 / Fks2	yeast	unidirectional	n.d.	Fks2	Fks1	Enzymes	(Doughas, Foor et al. 1994)
YRR1 / YRM1	yeast	n.d.	n.d.	YRR1	YRM1	Transcriptional regulators	(Lucanu-Danila, Delaveau et al. 2003; Onda, Ota

Genes	organism	directionality	Linear	Responsive	essential		reference
			response	gene	gene	function	
							et al. 2004)
Acs1 / Acs2	yeast	unidirectional	n.d.	Acs2	Acs2	enzymes	(van den Berg, de Jong-Gubbels et al. 1996)
HN-S / StpA	e. coli	bidirectional	Homozygote organism	stpA	HN-S	Transcriptional regulators	(Zhang, Rinsky et al. 1996; Free and Dorman 1997)
MyoD / Myf5	Mouse / human	unidirectional	Yes	Myf-5	n.d.	Developmental regulators (myogenesis)	(Rudnicki, Braun et al. 1992)
MRF4 / myogenin	Mouse / human	unidirectional	Yes	MRF4	myogenin	Developmental regulators (myogenesis)	(Zhong, Behringer et al. 1995)
Dlx3 / Dlx7	Zebrafish	unidirectional		Dlx7	Dlx3	Developmental regulators	(Solomon and Fritz 2002)
Pax1 / Pax9	mouse	unidirectional	Yes	Pax9	Pax1 (in the context of axial)	Developmental regulators	(Peters, Wilm et al. 1999)

Genes	organism	directionality	Linear		Responsive		essential		reference
			response	gene	gene	gene	function		
						skeleton development)			
Gsh1 / Gsh2	mouse	Unidirectional	n.d.	Gsh2	Gsh2		Developmental regulators (myogenesis)	(Torsson and Campbell 2001)	
Dystrophin / A-utrophin	Mouse / human	n.d.	n.d.	n.d.	Dystrophin		Muscle structural components	(Porter, Rafael et al. 1998)	
SNRPB/SNRPN	human	n.d.	n.d.	SNRPB	SNRPN		Small Nuclear Ribonucleoprotein	(Gray, Smithwick et al. 1999)	
smurf1/smurf2	mouse	unidirectional	n.d.	Smurf2	n.d.		E3 ubiquitin ligases	(Yamashita, Ying et al. 2005)	
OSP/claudin-11 and PLP	mice	bidirectional	n.d.	both	OSP		Myelin structural genes	(Chow, Mottahedeh et al. 2005)	
runx1/runx3	human	Unidirectional	n.d.	Runx1			Transcriptional regulators ( also tumor sup. genes)	(Spender, Whiteman et al. 2005)	

Genes	organism	directionality	Linear		Responsive		essential		reference
			response	gene	gene	gene	gene	function	
Midkine / pleiotrophin	mouse	Unidirectional	n.d.	Pleiotrophin	Midkine			growth factors	(Herradon, Ezquerro et al. 2005)
nNOS/ eNOS	mouse	Unidirectional	n.d.	eNOS				Nitric oxide synthase	(Burnett, Nelson et al. 1996)
cat1/cat2/cat3	mice	unidirectional	n.d.	Cat2 & cat3	Cat1			Arg transporters	(Nicholson, Sawamura et al. 1998)

Genes	Organism	Mutation Phenotypes
HXT1/HXT2	Yeast	$\Delta Hxt2$ null mutant has higher growth rate than the $\Delta Hxt2$ under all tested growth conditions in a high-throughput analysis (Steinmetz, Scharfe et al. 2002)
Idp2 / idh	Yeast	$\Delta Idp1$ null mutant has higher growth rate than $\Delta Idh$ under all tested growth conditions in a high-throughput analysis (Steinmetz, Scharfe et al. 2002)
Fks1 / Fks2	Yeast	$\Delta Fks2$ null mutant has higher growth rate than $\Delta Fks1$ under all tested growth conditions in a high-throughput analysis (Steinmetz, Scharfe et al. 2002)
YRR1 / YRM1	Yeast	$\Delta Yrr1$ null mutant has higher growth rate than $\Delta Yrm1$ under all tested growth conditions in a high-throughput analysis (Steinmetz, Scharfe et al. 2002)
Acs1 / Acs2	Yeast	<ul style="list-style-type: none"> <li>○ Acs1 - Null mutant is viable and grows on ethanol or glucose (but not acetate) as sole carbon source (but with long lag-phase);</li> <li>○ Acs2 - Null mutant is viable, and grows on ethanol or acetate as sole carbon source, but is unable to grow on glucose as sole carbon source;</li> </ul> <p>* Phenotypes taken from <a href="http://www.yeastgenome.org/">http://www.yeastgenome.org/</a></p>
HN-S / StpA	E. Coli	<ul style="list-style-type: none"> <li>○ HN-S - altered frequencies of transposition, chromosomal deletions and site-specific recombination events . Mutations in hns are also highly pleiotropic, affecting the expression of a number of genes. The genes affected by H-NS are dispersed throughout the chromosome and are often regulated in response to</li> </ul>

		<ul style="list-style-type: none"> <li>○ changes in environmental conditions (Sonden and Uhlin 1996).</li> <li>○ <i>StpA</i> – no apparent phenotype to the <i>stpA</i> mutants (Sonden and Uhlin 1996).</li> </ul>
MyoD / Myf5	Mouse	<ul style="list-style-type: none"> <li>○ Myf5 - Homozygotes for targeted null mutations exhibit delayed appearance of myotomal cells in somites, and lack the distal portion of ribs resulting in inability to breathe and lethality at birth. Other mutants lack the rib phenotype (Eppig, Blake et al. 2002; Blake, Richardson et al. 2003).</li> </ul>
MRF4 / myogenin	Mouse	<ul style="list-style-type: none"> <li>○ MFR4 - Homozygotes for targeted mutations exhibit variable rib abnormalities, abnormal intercostal muscle morphology, reduced expression of Myf5, and postnatal mortality proportional to the severity of the rib defect (Eppig, Blake et al. 2002; Blake, Richardson et al. 2003).</li> <li>○ Myogenin - Homozygotes for targeted null mutations exhibit a severe reduction in muscle mass associated with delayed primary myogenesis and very little secondary myofiber formation, defects of the thoracic skeleton, and perinatal death (Eppig, Blake et al. 2002; Blake, Richardson et al. 2003).</li> </ul> <p>* Phenotypes taken from <a href="http://www.informatics.jax.org/">http://www.informatics.jax.org/</a></p>
Dlx3 / Dlx7	Zebrafish	<ul style="list-style-type: none"> <li>○ Embryos injected with <i>dlx7</i> MO do not exhibit any morphologically detectable defects and are indistinguishable from wild type (Solomon and Fritz 2002).</li> <li>○ In ~98% (79/81) of <i>dlx3</i> MO-injected embryos, otic vesicles are smaller than wild-type. By 24 h, the otic vesicles of these embryos exhibit smaller lumens and typically only one otolith (Solomon and Fritz 2002).</li> </ul>

		❖ In Human, mutation in <i>Dlx3</i> is thought to cause Trich-dento-osseous syndrome (TDO) characterized by abnormal hair, teeth and bone (Price, Bowden et al. 1998). The main clinical manifestations of TDO include taurodontism, enamel hypoplasia, kinky, curly hair at birth and increased thickness and density of the cranial bones.
Pax1 / Pax9	Mouse	<p>Although mutations in Pax9 are more deleterious than those in Pax1, this is not true in the context of axial skeleton development. In this context:</p> <ul style="list-style-type: none"> <li>○ Mutant mice of a defined Pax1 null allele are viable and exhibit morphological abnormalities of vertebrae and intervertebral discs similar to those seen in the recessive alleles <i>un</i> and <i>un<sub>ex</sub></i> (Wilm, Dahlt et al. 1998) .</li> <li>○ Homozygous Pax9 mutants have skeletal defects in the limbs and in the skull, but exhibit no obvious defects in the axial skeleton (Peters, Neubuser et al. 1998) .</li> </ul>
Gsh1 / Gsh2	Mouse	<ul style="list-style-type: none"> <li>○ Gsh1 –Homozygous mutants exhibit extreme dwarfism, sexual infantilism and significant perinatal mortality. The mutant pituitary is small in size and hypocellular, with severely reduced numbers of growth hormone- and prolactin-producing cells. Moreover, the pituitary content of a subset of pituitary hormones, including growth hormone, prolactin and luteinizing hormone, is significantly decreased. The hypothalamus, although morphologically normal, is also perturbed in mutants. The <i>gsh-1</i> gene is shown to be essential for growth hormone-releasing hormone (GHRH) gene expression in the arcuate nucleus of the hypothalamus. (Li, Zeitler et al. 1996)</li> </ul>



		<ul style="list-style-type: none"> <li>○ Gsh2 – Homozygous mutants uniformly failed to survive more than 1 day following birth. At the physiologic level the mutants experienced apnea and reduced levels of hemoglobin oxygenation. Histologically, the mutant brains had striking alterations of discrete components. In the forebrain the lateral ganglionic eminence was reduced in size. In the hindbrain, the area postrema, an important cardiorespiratory chemosensory center, was absent. The contiguous nucleus tractus solitarius, involved in integrating sensory input to maintain homeostasis, was also severely malformed in mutants (Szucsik, Witte et al. 1997). From the molecular level mutants have a significant decrease in the expression of numerous genes that mark early development of the lateral ganglionic eminence, the striatal anlage (Corbin, Gaiano et al. 2000). Accompanying this early loss of patterning genes is an initial expansion of dorsal telencephalic markers across the cortical-striatal boundary into the lateral ganglionic eminence. Interestingly, as development proceeds, there is compensation for this early loss of markers that is coincident with a molecular reestablishment of the cortical-striatal boundary that probably is related to compensation by Gsh1 (Toresson and Campbell 2001).</li> </ul>
Dystrophin / A-trophin	Mouse	<ul style="list-style-type: none"> <li>○ Dystrophin - Hemizygous males and homozygous mutant females show progressive muscle atrophy, increased K<sup>+</sup> and Ca<sup>2+</sup> levels in muscle with age, increased levels of blood pyruvate kinase, and other Duchenne- and Becker-like muscular dystrophy symptoms. Alleles vary in their phenotypic severity and may include cardiovascular and digestive system effects (Eppig, Blake et al. 2002; Blake, Richardson et al. 2003).</li> <li>○ A-trophin - Homozygous null mutants have reduced density of acetylcholine</li> </ul>

		<p>receptors and reduced number of junctional folds at neuromuscular junctions. Mice homozygous for utrophin and dystrophin knockouts die prematurely with severe, progressive muscular dystrophy (Eppig, Blake et al. 2002; Blake, Richardson et al. 2003).</p> <p>* Phenotypes taken from <a href="http://www.informatics.jax.org/">http://www.informatics.jax.org/</a></p>
SNRPB/SNRPN	Human	<ul style="list-style-type: none"> <li>○ SNRPB – No diseases are known to be associated with this gene.</li> <li>○ SNRPN – Mutations in this gene are known to be associated with the following diseases: prader-willi syndrome, angelman syndrome, lupus erythematosus systemic, beckwith-wiedemann syndrome, idiocy, teratoma</li> </ul> <p>*Data taken from GeneCards (<a href="http://www.genecards.org/">http://www.genecards.org/</a>)</p>
smurf1/smurf2	Mouse	<ul style="list-style-type: none"> <li>○ <i>Smurf1</i>-deficient mice are perinatally normal but exhibit an age-dependent increase of bone mass due to enhanced osteoblast activity (Eppig, Blake et al. 2002; Blake, Richardson et al. 2003)</li> </ul>
OSP/claudin-11 and PLP	Mouse	<ul style="list-style-type: none"> <li>○ OSP - Homozygous null mice exhibit tremors, impaired coordination, hindlimb weakness, abnormal myelination of the cranial nerves, increased auditory thresholds, and abnormal stria vascularis. Mutant males have small testes, abnormal seminiferous tubules, and sperm abnormalities resulting in infertility (Eppig, Blake et al. 2002; Blake, Richardson et al. 2003).</li> <li>○ PLP - Males hemizygous for X-linked missense and partially deleted mutations typically exhibit central nervous system demyelination, loss of oligodendrocytes, tremors, convulsions, and lethality, but targeted null mutants are essentially normal and fertile (Eppig, Blake et al. 2002; Blake, Richardson et al. 2003).</li> </ul>

		<p>* Phenotypes taken from <a href="http://www.informatics.jax.org/">http://www.informatics.jax.org/</a></p>
runx1/runx3	Human	<ul style="list-style-type: none"> <li>○ RUNX3 is required for the development of CD8-lineage T cells (Taniuchi, Osato et al. 2002) and TrkC-dependent dorsal root ganglion neurons (Inoue, Ozaki et al. 2002). RUNX3 also functions as a gastric tumor suppressor. About 60% of primary gastric cancer specimens express significantly lower levels of RUNX3 because of a combination of hemizygous deletion of the gene and hypermethylation of the RUNX3 promoter region (Jin, Jeon et al. 2004).</li> <li>○ The RUNX1 gene regulates hematopoietic cell-specific genes and is required for definitive hematopoiesis (Okuda, van Deursen et al. 1996). RUNX1 is the most frequent target for chromosomal translocation in leukemia and is responsible for about 30% of the cases of human acute leukemia (Look 1997). Haplo-insufficiency of RUNX1 causes an autosomal dominant congenital platelet defect that predisposes affected individuals to leukemia (Jin, Jeon et al. 2004)</li> </ul> <p>* Phenotypes taken from <a href="http://www.informatics.jax.org/">http://www.informatics.jax.org/</a></p>
Midkine / pleiotrophin	Mouse	<ul style="list-style-type: none"> <li>○ No significant defect in development has been reported in mice deficient in MK or PTN, except for a delay in postnatal hippocampal development in MK-deficient mice (Nakamura, Kadomatsu et al. 1998; Amel, Lauri et al. 2001). However, closer analysis has revealed that when heterozygotes are crossed, the numbers of <i>Mdk</i> (<math>-/-</math>) and <i>PTN</i> (<math>-/-</math>) mice born are significantly smaller than wild-type born (Maruyama, S. and Muramatsu, H., unpublished).</li> </ul>
nNOS	Mouse	<ul style="list-style-type: none"> <li>○ nNOS - Homozygous hypomorphic mice exhibit enlarged stomachs, abnormal pyloric and lower esophageal sphincters, age-related cardiac hypertrophy, altered</li> </ul>

		<p>alcohol consumption and responses, decreased ovulation and reduced REM sleep. Homozygous null mice display increased neurogenesis in the adult brain (Eppig, Blake et al. 2002; Blake, Richardson et al. 2003).</p> <ul style="list-style-type: none"> <li>o eNOS - Homozygotes for targeted null mutations exhibit reduced survival, hypertension, inhibited basal vasodilation, insulin resistance, fewer mitochondria, reduced heart rate, impaired ovulation and, in some, shortened limbs (Eppig, Blake et al. 2002; Blake, Richardson et al. 2003).</li> </ul> <p>* Phenotypes taken from <a href="http://www.informatics.jax.org/">http://www.informatics.jax.org/</a></p>
cat1/cat2/cat3	Mouse	<ul style="list-style-type: none"> <li>o Homozygous CAT-1 knockout mice die on day 1 after birth, are 25% smaller than their wild-type littermates and suffer from severe anaemia, while the heterozygous mice exhibit no phenotypic abnormalities(Perkins, Mar et al. 1997). The relatively normal development of most tissues in the homozygous CAT-1 knockout mice up to birth is probably due to the expression of CAT-3 during embryogenesis and fetal development(Nicholson, Sawamura et al. 1998). Both wild type and knockout cells transport arginine with comparable Km and Vmax. However, the apparent affinity for lysine transport was 2.4 times lower in <i>Cat1<sup>-/-</sup></i> cells when compared with wild type cells(Nicholson, Sawamura et al. 1998).</li> <li>o Homozygous CAT-2<sup>-/-</sup> mice show no apparent phenotypic abnormalities(Nicholson, Manner et al. 2001). However, the sustained NO production in peritoneal macrophages from these mice is almost abolished, underlining the important role of CAT-2B for the substrate supply of iNOS. Neither the expression of iNOS nor the intracellular L-arginine concentrations are reduced in these cells.</li> </ul>

			* Phenotypes taken from <a href="http://www.informatics.jax.org/">http://www.informatics.jax.org/</a>
--	--	--	---

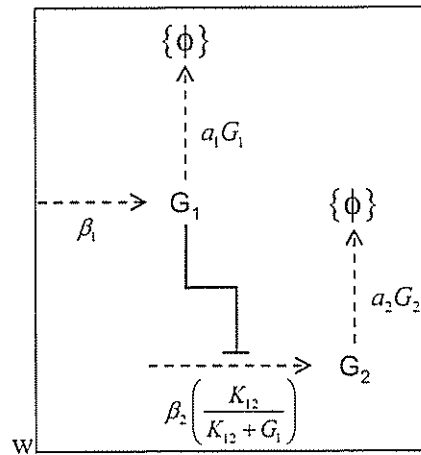
## Supplementary Text – Modeling steady states

In this section I will model how the steady state concentrations of the two functionally redundant paralogs change with respect to changes in the input to the controller of a RBC. Some of the calculations were performed twice, the second version having more simplifying assumptions than the first. The more simplified calculations are “boxed”.

Notation:

$G_1$	Controller gene.
$G_2$	Responsive gene.
$\beta$	Transcription rate.
$\alpha$	Degradation rate.
$K_{IL}$	Thermodynamic dissociation constant between protein ‘I’ and its cis-regulatory motif on gene ‘L’.
$\nu \equiv \frac{\beta}{\alpha}$	maximal steady state concentration

### Model I: simple repression:



$$1. \quad \frac{dG_1}{dt} = \beta_1 - a_1 G_1$$

$$2. \quad \frac{dG_2}{dt} = \beta_2 \left( \frac{K_{12}}{K_{12} + G_1} \right) - a_2 G_2$$

Solving for steady states:

$$3. \quad G_{1st} = \frac{\beta_1}{\alpha_1} \equiv v_1$$

$$4. \quad G_{2st} = v_2 \left( \frac{K_{12}}{K_{12} + v_1} \right) = v_2 \frac{K_{12}}{v_1} \left( \frac{1}{\frac{K_{12}}{v_1} + 1} \right)$$

Note: the term  $\left( \frac{1}{\frac{K_{12}}{v_1} + 1} \right)$  goes to 1 in the realistic realm of  $v_1 \gg K_{12}$

Solving for both product and sum of the steady state concentrations.

$$5. \quad f_p = G_{1st} \times G_{2st} = v_2 K_{12} \left( \frac{1}{\frac{K_{12}}{v_1} + 1} \right) \quad \text{output function 1 (f}_1\text{)}$$

Simplified version - Applying the simplification  $v_1 \gg K_{12}$

$$\diamond \quad G_{1st} \times G_{2st} = (v_2 K_{12}) = c_2 \quad f_1 - \text{simplified version}$$

$$\bullet \quad c_1 = 1$$

$$\bullet \quad c_2 = v_2 K_{12}$$

$$6. \quad f_s = G_{1st} + G_{2st} = v_1 + \frac{1}{v_1} (v_2 K_{12}) \left( \frac{1}{\frac{K_{12}}{v_1} + 1} \right) \quad \text{output function 2 (f}_2\text{)}$$

$$7. \quad \left( \frac{v_1}{G_{1st} + G_{2st}} \right) \frac{d(G_{1st} + G_{2st})}{dv_1} = \left( \frac{v_1}{v_1 + \frac{v_2 K_{12}}{v_1 + K_{12}}} \right) \left( 1 - \frac{v_2 K_{12}}{(K_{12} + v_1)^2} \right)$$

Simplified version - Applying the simplification  $v_1 \gg K_{12}$

$$\diamond \quad G_{1st} + G_{2st} = v_1 + \frac{1}{v_1} (v_2 K_{12}) = v_1 c_1 + \frac{1}{v_1} c_2 \quad f_2 - \text{simplified version}$$

$$\bullet \quad c_1 = 1$$

$$\bullet \quad c_2 = v_2 K_{12}$$

$$7.2 \quad \frac{d(G_{1st} + G_{2st})}{dv_1} = c_1 - \frac{1}{v_1^2} c_2$$

$$7.2 \quad \left( \frac{v_1}{G_{1st} + G_{2st}} \right) \frac{d(G_{1st} + G_{2st})}{dv_1} = \left( \frac{\frac{c_1}{c_2} v_1^2}{\frac{c_1}{c_2} v_1^2 + 1} \right) - \left( \frac{1}{\frac{c_1}{c_2} v_1^2 + 1} \right) = \frac{\frac{c_1}{c_2} v_1^2 - 1}{\frac{c_1}{c_2} v_1^2 + 1}$$

The effects of a haploid  $\rightarrow$  diploid fluctuation.

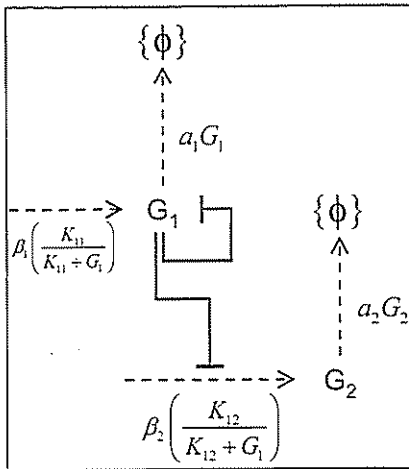
$$8. \quad \frac{f_s(2v_1) - f_s(v_1)}{f_s(v_1)} = \frac{v_1 + v_2 \left( \frac{K_{12}}{K_{12} + 2v_1} + \frac{K_{12}}{K_{12} + v_1} \right)}{v_1 + v_2 \left( \frac{K_{12}}{K_{12} + v_1} \right)} = \frac{\frac{v_1}{K_{12}} + \frac{v_2}{K_{12}} \left( \frac{1}{1 + 2\frac{v_1}{K_{12}}} + \frac{1}{1 + \frac{v_1}{K_{12}}} \right)}{\frac{v_1}{K_{12}} + \frac{v_2}{K_{12}} \left( \frac{1}{1 + \frac{v_1}{K_{12}}} \right)}$$

Substituting  $\bar{v}$  for  $\frac{v}{K_{12}}$  we obtain



$$= \frac{\overline{\nu_1 + \nu_2} \left( \frac{1}{1 + 2\nu_1} + \frac{1}{1 + \nu_1} \right)}{\overline{\nu_1 + \nu_2} \left( \frac{1}{1 + \nu_1} \right)}$$

### Model II: dampened controller



$$9. \quad \frac{dG_1}{dt} = \beta_1 \left( \frac{K_{11}}{K_{11} + G_1} \right) - a_1 G_1$$

$$10. \quad \frac{dG_2}{dt} = \beta_2 \left( \frac{K_{12}}{K_{12} + G_1} \right) - a_2 G_2$$

Simplified version - Solving for steady states ( $\frac{dG_1}{dt} = 0$ ) by applying the simplifying assumptions:  $G_1 \gg K_{11}$  and  $G_1 \gg K_{12}$

$$\diamond \quad \beta_1 \left( \frac{K_{11}}{G_1} \right) \approx a_1 G_1$$

$$\diamond \quad \beta_2 \left( \frac{K_{12}}{G_1} \right) \approx a_2 G_2$$

$$\diamond \quad G_{1st} = \sqrt{\nu_1} \sqrt{K_{11}}$$

$$\diamond \quad G_{2st} = \frac{1}{\sqrt{\nu_1}} \left( \nu_2 K_{12} \frac{1}{\sqrt{K_{11}}} \right)$$

Solving for steady states ( $\frac{dG_1}{dt} = 0$ ) without the above simplifications.

$$11. \quad \alpha_1 G_1^2 + \alpha_1 K_{11} G_1 - \beta_1 K_{11} = 0$$

Solving for the roots of the equation:

$$12. \quad G_{1st} = \frac{\sqrt{\alpha_1^2 K_{11}^2 + 4\alpha_1 \beta_1 K_{11}} - \alpha_1 K_{11}}{2\alpha_1}$$

Applying Taylor transformation and assuming  $K_{11} \ll 4v_1$

$$13. \quad G_{1st} = \sqrt{v_1 K_{11}} + \frac{K_{11}^2}{8\sqrt{v_1 K_{11}}} - \frac{K_{11}^4}{16\sqrt{v_1 K_{11}}^3} - \frac{K_{11}}{2}$$

Solving for steady state ( $\frac{dG_2}{dt} = 0$ )

$$14. \quad 0 = \beta_2 \left( \frac{K_{12}}{K_{12} + G_{1st}} \right) - a_2 G_{2st}$$

$$15. \quad G_{2st} = v_2 \frac{K_{12}}{G_{1st}} \left( \frac{1}{\frac{K_{12}}{G_{1st}} + 1} \right)$$

$$16. \quad G_{1st} \times G_{2st} = v_2 K_{12} \left( \frac{1}{\frac{K_{12}}{G_{1st}} + 1} \right) \quad \text{output function 1 (f}_1\text{)}$$

Simplified version - With the simplifying assumptions:  $G_1 \gg K_{11}$  and  $G_1 \gg K_{12}$

$$\diamond \quad G_{1st} \times G_{2st} = v_2 K_{12} \quad f_1 - \text{simplified version}$$

$$\bullet \quad c_1 = \sqrt{K_{11}}, \quad c_2 = v_2 K_{12}$$

$$17. \quad G_{1st} + G_{2st} = G_{1st} + \nu_2 \frac{K_{12}}{G_{1st}} \left( \frac{1}{\frac{K_{12}}{G_{1st}} + 1} \right) \quad \text{output function 2 } (f_2)$$

Simplified version - With the simplifying assumptions:  $G_1 \gg K_{11}$  and  $G_1 \gg K_{12}$

$$\diamond \quad G_{1st} + G_{2st} = \sqrt{\nu_1} \sqrt{K_{11}} + \frac{1}{\sqrt{\nu_1}} \frac{\nu_2 K_{12}}{\sqrt{K_{11}}}$$

$$= \sqrt{\nu_1} c_1 + \frac{1}{\sqrt{\nu_1}} c_2 \quad f_2 - \text{simplified version}$$

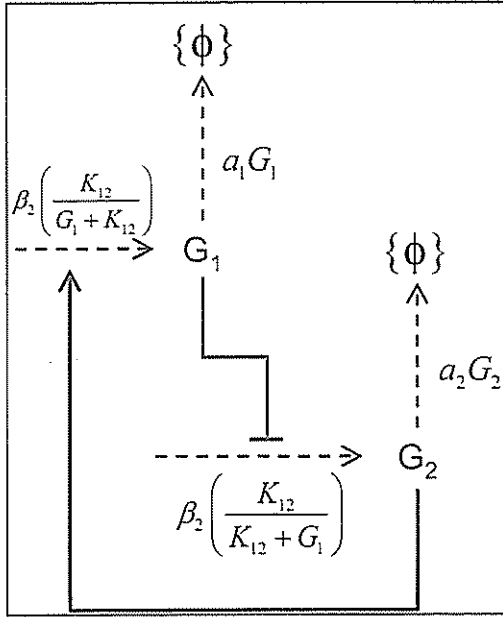
$$\bullet \quad c_1 = \sqrt{K_{11}}$$

$$\bullet \quad c_2 = \frac{\nu_2 K_{12}}{\sqrt{K_{11}}}$$

$$\diamond \quad \frac{d(G_{1st} + G_{2st})}{d\nu_1} = \frac{1}{2\sqrt{\nu_1}} c_1 - \frac{1}{2\sqrt{\nu_1^3}} c_2$$

$$\diamond \quad \left( \frac{\nu_1}{G_{1st} + G_{2st}} \right) \frac{d(G_{1st} + G_{2st})}{d\nu_1} = \frac{1}{2} \left( \frac{\frac{c_1}{c_2} \nu_1}{\frac{c_1}{c_2} \nu_1 + 1} \right) - \frac{1}{2} \left( \frac{1}{\frac{c_1}{c_2} \nu_1 + 1} \right) = \frac{1}{2} \left( \frac{\frac{c_1}{c_2} \nu_1 - 1}{\frac{c_1}{c_2} \nu_1 + 1} \right)$$

Model III: Two-way scaling negative feedback



$$18. \quad \frac{dG_1}{dt} = \beta_1 \left( \frac{G_2}{G_2 + K_{21}} \right) - a_1 G_1$$

$$19. \quad \frac{dG_2}{dt} = \beta_2 \left( \frac{K_{12}}{G_1 + K_{12}} \right) - a_2 G_2$$

Solving for steady state ( $\frac{dG_2}{dt} = 0$ ):

$$20. \quad 0 = \beta_2 \left( \frac{K_{12}}{G_{1st} + K_{12}} \right) - a_2 G_{2st}$$

$$21. \quad G_{2st} = \frac{\beta_2 K_{12}}{G_{1st} + K_{12}}$$

Solving for steady state ( $\frac{dG_1}{dt} = 0$ ):

$$22. \quad 0 = \beta_1 \left( \frac{G_{2st}}{G_{2st} + K_{21}} \right) - a_1 G_{1st}$$

$$23. \quad G_{1st} = \frac{\frac{v_2 K_{12}}{G_{1st} + K_{12}} v_1}{\frac{v_2 K_{12}}{G_{1st} + K_{12}} + K_{12}} = \frac{v_1 v_2 K_{12}}{v_2 K_{12} + K_{12} (G_{1st} + K_{12})}$$

$$24. \quad 0 = G_{1st}^2 K_{21} + G_{1st} (v_2 K_{12} + K_{12} K_{21}) - v_1 v_2 K_{12}$$

$$25. \quad G_{1st} = \frac{\sqrt{(v_2 K_{12} + K_{12} K_{21})^2 + 4 K_{12} K_{21} v_1 v_2} - (v_2 K_{12} + K_{12} K_{21})}{2 K_{21}}$$

$$26. \quad G_{2st} = v_2 \frac{K_{12}}{G_{1st}} \left( \frac{1}{\frac{K_{12}}{G_{1st}} + 1} \right)$$

Simplified version - Applying simplifying assumptions:  $G_1 \gg K_{12}$  and  $G_2 \gg K_{21}$  and resolving

$$\diamond \quad G_{1st} = \sqrt{v_1} \sqrt{v_2} \sqrt{\frac{K_{12}}{K_{21}}}$$

$$\diamond \quad G_{2st} = \frac{1}{\sqrt{v_1}} \sqrt{v_2} \sqrt{K_{12} K_{21}}$$

$$27. \quad G_{1st} \times G_{2st} = v_2 K_{12} \left( \frac{1}{\frac{K_{12}}{G_{1st}} + 1} \right) \quad \text{output function 1 (f}_1\text{)}$$

Simplified version - Applying simplifying assumptions:  $G_1 \gg K_{12}$  and  $G_2 \gg K_{21}$  and resolving

$$\diamond \quad G_{1st} \times G_{2st} = v_2 K_{12} \quad f_1 - \text{simplified version}$$

28.  $G_{1st} + G_{2st} = G_{1st} + v_2 \frac{K_{12}}{G_{1st}} \left( \frac{1}{\frac{K_{12}}{G_{1st}} + 1} \right)$  output function 2 ( $f_2$ )

Simplified version - Applying simplifying assumptions:  $G_1 \gg K_{12}$  and  $G_2 \ll K_{21}$  and resolving

♦  $G_{1st} + G_{2st} = \sqrt{v_1} c_1 + \frac{1}{\sqrt{v_1}} c_2$   $f_2$  – simplified version

♦  $c_1 = \sqrt{v_2} \sqrt{\frac{K_{12}}{K_{21}}}$

♦  $c_2 = \sqrt{v_2} \sqrt{K_{12} K_{21}}$

♦  $\frac{d(G_{1st} + G_{2st})}{dv_1} = \frac{1}{2\sqrt{v_1}} c_1 - \frac{1}{2\sqrt{v_1^3}} c_2$

## References

- Abe, M., I. Nishida, et al. (2001). "Yeast 1,3-beta-glucan synthase activity is inhibited by phytosphingosine localized to the endoplasmic reticulum." *J Biol Chem* **276**(29): 26923-30.
- Akimenko, M. A., M. Ekker, et al. (1994). "Combinatorial expression of three zebrafish genes related to distal-less: part of a homeobox gene code for the head." *J Neurosci* **14**(6): 3475-86.
- Amet, L. E., S. E. Lauri, et al. (2001). "Enhanced hippocampal long-term potentiation in mice lacking heparin-binding growth-associated molecule." *Mol Cell Neurosci* **17**(6): 1014-24.
- Atchley, W. R., W. M. Fitch, et al. (1994). "Molecular evolution of the MyoD family of transcription factors." *Proc Natl Acad Sci U S A* **91**(24): 11522-6.
- Blake, J. A., J. E. Richardson, et al. (2003). "MGD: the Mouse Genome Database." *Nucleic Acids Res* **31**(1): 193-5.
- Brookfield, J. F. (2003). "Gene duplications: the gradual evolution of functional divergence." *Curr Biol* **13**(6): R229-30.
- Brown, C. J., K. M. Todd, et al. (1998). "Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment." *Mol Biol Evol* **15**(8): 931-42.
- Burnett, A. L., R. J. Nelson, et al. (1996). "Nitric oxide-dependent penile erection in mice lacking neuronal nitric oxide synthase." *Mol Med* **2**(3): 288-96.
- Cases, S., S. Novak, et al. (1998). "ACAT-2, a second mammalian acyl-CoA:cholesterol acyltransferase. Its cloning, expression, and characterization." *J Biol Chem* **273**(41): 26755-64.
- Chow, E., J. Mottahedeh, et al. (2005). "Disrupted compaction of CNS myelin in an OSP/Claudin-11 and PLP/DM20 double knockout mouse." *Mol Cell Neurosci* **29**(3): 405-13.
- Conant, G. C. and A. Wagner (2003). "Asymmetric sequence divergence of duplicate genes." *Genome Res* **13**(9): 2052-8.
- Conant, G. C. and A. Wagner (2004). "Duplicate genes and robustness to transient gene knock-downs in *Caenorhabditis elegans*." *Proc R Soc Lond B Biol Sci* **271**(1534): 89-96.
- Corbin, J. G., N. Gaiano, et al. (2000). "The Gsh2 homeodomain gene controls multiple aspects of telencephalic development." *Development* **127**(23): 5007-20.
- Dahl, E., H. Koseki, et al. (1997). "Pax genes and organogenesis." *Bioessays* **19**(9): 755-65.
- Deconinck, A. E., J. A. Rafael, et al. (1997). "Utrophin-dystrophin-deficient mice as a model for Duchenne muscular dystrophy." *Cell* **90**(4): 717-27.
- Dorman, C. J. (2004). "H-NS: a universal regulator for a dynamic genome." *Nat Rev Microbiol* **2**(5): 391-400.
- Douglas, C. M., F. Foor, et al. (1994). "The *Saccharomyces cerevisiae* FKS1 (ETG1) gene encodes an integral membrane protein which is a subunit of 1,3-beta-D-glucan synthase." *Proc Natl Acad Sci U S A* **91**(26): 12907-11.
- Dowling, P., K. Culligan, et al. (2002). "Distal mdx muscle groups exhibiting up-regulation of utrophin and rescue of dystrophin-associated glycoproteins



- exemplify a protected phenotype in muscular dystrophy." Naturwissenschaften **89**(2): 75-8.
- Eide, D. J. (1998). "The molecular biology of metal ion transport in *Saccharomyces cerevisiae*." Annu Rev Nutr **18**: 441-69.
- Ekker, M., M. A. Akimenko, et al. (1992). "Regional expression of three homeobox transcripts in the inner ear of zebrafish embryos." Neuron **9**(1): 27-35.
- Ellies, D. L., D. W. Stock, et al. (1997). "Relationship between the genomic organization and the overlapping embryonic expression patterns of the zebrafish *dlx* genes." Genomics **45**(3): 580-90.
- Elowitz, M. B., A. J. Levine, et al. (2002). "Stochastic gene expression in a single cell." Science **297**(5584): 1183-6.
- Enns, L. C., M. M. Kanaoka, et al. (2005). "Two callose synthases, *GSL1* and *GSL5*, play an essential and redundant role in plant and pollen development and in fertility." Plant Mol Biol **58**(3): 333-49.
- Eppig, J. T., J. A. Blake, et al. (2002). "Corralling conditional mutations: a unified resource for mouse phenotypes." Genesis **32**(2): 63-5.
- Free, A. and C. J. Dorman (1997). "The *Escherichia coli* *stpA* gene is transiently expressed during growth in rich medium and is induced in minimal medium and by stress conditions." J Bacteriol **179**(3): 909-18.
- Galitski, T., A. J. Saldanha, et al. (1999). "Ploidy regulation of gene expression." Science **285**(5425): 251-4.
- Garcia-Rodriguez, L. J., J. A. Trilla, et al. (2000). "Characterization of the chitin biosynthesis process as a compensatory mechanism in the *fls1* mutant of *Saccharomyces cerevisiae*." FEBS Lett **478**(1-2): 84-8.
- Gasch, A. P., P. T. Spellman, et al. (2000). "Genomic expression programs in the response of yeast cells to environmental changes." Mol Biol Cell **11**(12): 4241-57.
- Gerhart, J. and M. Kirschner (1997). Cells, embryos, and evolution : toward a cellular and developmental understanding of phenotypic variation and evolutionary adaptability. Malden, Mass. ; Oxford, Blackwell Science.
- Gray, T. A., M. J. Smithwick, et al. (1999). "Concerted regulation and molecular evolution of the duplicated *SNRPN*'*B* and *SNRPN* loci." Nucleic Acids Res **27**(23): 4577-84.
- Gu, X. (2003). "Evolution of duplicate genes versus genetic robustness against null mutations." Trends in Genetics **19**(7): 354-356.
- Gu, Z., L. M. Steinmetz, et al. (2003). "Role of duplicate genes in genetic robustness against null mutations." Nature **421**(6918): 63-6.
- Hartman, J. L. t., B. Garvik, et al. (2001). "Principles for the buffering of genetic variation." Science **291**(5506): 1001-4.
- Herradon, G., L. Ezquerra, et al. (2005). "Midkine regulates pleiotrophin organ-specific gene expression: evidence for transcriptional regulation and functional redundancy within the pleiotrophin/midkine developmental gene family." Biochem Biophys Res Commun **333**(3): 714-21.
- Holland, P. W., L. Z. Holland, et al. (1992). "An amphioxus homeobox gene: sequence conservation, spatial expression during development and insights into vertebrate evolution." Development **116**(3): 653-61.

- Hughes, A. L. (1994). "The evolution of functionally novel proteins after gene duplication." *Proc Biol Sci* **256**(1346): 119-24.
- Ihmels, J., R. Levy, et al. (2004). "Principles of transcriptional control in the metabolic network of *Saccharomyces cerevisiae*." *Nat Biotechnol* **22**(1): 86-92.
- Inoue, K., S. Ozaki, et al. (2002). "Runx3 controls the axonal projection of proprioceptive dorsal root ganglion neurons." *Nat Neurosci* **5**(10): 946-54.
- Inoue, S. B., N. Takewaki, et al. (1995). "Characterization and gene cloning of 1,3-beta-D-glucan synthase from *Saccharomyces cerevisiae*." *Eur J Biochem* **231**(3): 845-54.
- Jin, Y. H., E. J. Jeon, et al. (2004). "Transforming growth factor-beta stimulates p300-dependent RUNX3 acetylation, which inhibits ubiquitination-mediated degradation." *J Biol Chem* **279**(28): 29409-17.
- Kafri, R., A. Bar-Even, et al. (2005). "Transcription control reprogramming in genetic backup circuits." *Nat Genet* **37**(3): 295-9.
- Kavsak, P., R. K. Rasmussen, et al. (2000). "Smad7 binds to Smurf2 to form an E3 ubiquitin ligase that targets the TGF beta receptor for degradation." *Mol Cell* **6**(6): 1365-75.
- Kellis, M., B. W. Birren, et al. (2004). "Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*." *Nature* **428**(6983): 617-24.
- Kirschner, M. and J. Gerhart (1998). "Evolvability." *Proc Natl Acad Sci U S A* **95**(15): 8420-7.
- Koch, A. L. (1981). "Evolution of antibiotic resistance gene function." *Microbiol Rev* **45**(2): 355-78.
- Kondrashov, F. A., I. B. Rogozin, et al. (2002). "Selection in the evolution of gene duplications." *Genome Biol* **3**(2): RESEARCH0008.
- Krakauer, D. C. and M. A. Nowak (1999). "Evolutionary preservation of redundant duplicated genes." *Semin Cell Dev Biol* **10**(5): 555-9.
- Krakauer, D. C. and J. B. Plotkin (2002). "Redundancy, antiredundancy, and the robustness of genomes." *Proc Natl Acad Sci U S A* **99**(3): 1405-9.
- Krause, M., A. Fire, et al. (1990). "CeMyoD accumulation defines the body wall muscle cell fate during *C. elegans* embryogenesis." *Cell* **63**(5): 907-19.
- Leon, M., R. Sentandreu, et al. (2002). "A single FKS homologue in *Yarrowia lipolytica* is essential for viability." *Yeast* **19**(12): 1003-14.
- Lesage, G., A. M. Sdicu, et al. (2004). "Analysis of beta-1,3-glucan assembly in *Saccharomyces cerevisiae* using a synthetic interaction network and altered sensitivity to caspofungin." *Genetics* **167**(1): 35-49.
- Li, H., P. S. Zeitler, et al. (1996). "Gsh-1, an orphan Hox gene, is required for normal pituitary development." *Embo J* **15**(4): 714-24.
- Lin, X., M. Liang, et al. (2000). "Smurf2 is a ubiquitin E3 ligase mediating proteasome-dependent degradation of Smad2 in transforming growth factor-beta signaling." *J Biol Chem* **275**(47): 36818-22.
- Look, A. T. (1997). "Oncogenic transcription factors in the human acute leukemias." *Science* **278**(5340): 1059-64.

- Lucau-Danila, A., T. Delaveau, et al. (2003). "Competitive promoter occupancy by two yeast paralogous transcription factors controlling the multidrug resistance phenomenon." *J Biol Chem* **278**(52): 52641-50.
- Lynch, M. and J. S. Conery (2000). "The evolutionary fate and consequences of duplicate genes." *Science* **290**(5494): 1151-5.
- Maconochie, M., S. Nonchev, et al. (1996). "Paralogous Hox genes: function and regulation." *Annu Rev Genet* **30**: 529-56.
- Makova, K. D. and W. H. Li (2003). "Divergence in the spatial pattern of gene expression between human duplicate genes." *Genome Res* **13**(7): 1638-45.
- Mangan, S., A. Zaslaver, et al. (2003). "The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks." *J Mol Biol* **334**(2): 197-204.
- Mansouri, A. and P. Gruss (1998). "Pax3 and Pax7 are expressed in commissural neurons and restrict ventral neuronal identity in the spinal cord." *Mech Dev* **78**(1-2): 171-8.
- Marini, A. M., S. Soussi-Boudekou, et al. (1997). "A family of ammonium transporters in *Saccharomyces cerevisiae*." *Mol Cell Biol* **17**(8): 4282-93.
- McCammon, M. T. and L. McAlister-Henn (2003). "Multiple cellular consequences of isocitrate dehydrogenase isozyme dysfunction." *Arch Biochem Biophys* **419**(2): 222-33.
- Michelson, A. M., S. M. Abmayr, et al. (1990). "Expression of a MyoD family member prefigures muscle pattern in *Drosophila* embryos." *Genes Dev* **4**(12A): 2086-97.
- Montgomery, J. S., D. K. Price, et al. (2001). "The androgen receptor gene and its influence on the development and progression of prostate cancer." *J Pathol* **195**(2): 138-46.
- Nakamura, E., K. Kadomatsu, et al. (1998). "Disruption of the midkine gene (Mdk) resulted in altered expression of a calcium binding protein in the hippocampus of infant mice and their abnormal behaviour." *Genes Cells* **3**(12): 811-22.
- Nicholson, B., C. K. Manner, et al. (2001). "Sustained nitric oxide production in macrophages requires the arginine transporter CAT2." *J Biol Chem* **276**(19): 15881-5.
- Nicholson, B., T. Sawamura, et al. (1998). "Increased Cat3-mediated cationic amino acid transport functionally compensates in Cat1 knockout cell lines." *J Biol Chem* **273**(24): 14663-6.
- Nowak, M. A., M. C. Boerlijst, et al. (1997). "Evolution of genetic redundancy." *Nature* **388**(6638): 167-71.
- Ohno, S. (1970). *Evolution by Gene and Genome Duplication*, Springer.
- Okuda, T., J. van Deursen, et al. (1996). "AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis." *Cell* **84**(2): 321-30.
- Onda, M., K. Ota, et al. (2004). "Analysis of gene network regulating yeast multidrug resistance by artificial activation of transcription factors: involvement of Pdr3 in salt tolerance." *Gene* **332**: 51-9.
- Ozcan, S. (2002). "Two different signals regulate repression and induction of gene expression by glucose." *J Biol Chem* **277**(49): 46993-7.
- Ozcan, S. and M. Johnston (1999). "Function and regulation of yeast hexose transporters." *Microbiol Mol Biol Rev* **63**(3): 554-69.

- Papp, B., C. Pal, et al. (2004). "Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast." *Nature* **429**(6992): 661-4.
- Paulsson, J. (2004). "Summing up the noise in gene networks." *Nature* **427**(6973): 415-8.
- Perkins, C. P., V. Mar, et al. (1997). "Anemia and perinatal death result from loss of the murine ecotropic retrovirus receptor mCAT-1." *Genes Dev* **11**(7): 914-25.
- Peters, H., A. Neubuser, et al. (1998). "Pax9-deficient mice lack pharyngeal pouch derivatives and teeth and exhibit craniofacial and limb abnormalities." *Genes Dev* **12**(17): 2735-47.
- Peters, H., B. Wilm, et al. (1999). "Pax1 and Pax9 synergistically regulate vertebral column development." *Development* **126**(23): 5399-408.
- Porter, J. D., J. A. Rafael, et al. (1998). "The sparing of extraocular muscle in dystrophinopathy is lost in mice lacking utrophin and dystrophin." *J Cell Sci* **111** (Pt 13): 1801-11.
- Price, J. A., D. W. Bowden, et al. (1998). "Identification of a mutation in DLX3 associated with tricho-dento-osseous (TDO) syndrome." *Hum Mol Genet* **7**(3): 563-9.
- Qiu, M., A. Bulfone, et al. (1997). "Role of the Dlx homeobox genes in proximodistal patterning of the branchial arches: mutations of Dlx-1, Dlx-2, and Dlx-1 and -2 alter morphogenesis of proximal skeletal and soft tissue structures derived from the first and second arches." *Dev Biol* **185**(2): 165-84.
- Qiu, M., A. Bulfone, et al. (1995). "Null mutation of Dlx-2 results in abnormal morphogenesis of proximal first and second branchial arch derivatives and abnormal differentiation in the forebrain." *Genes Dev* **9**(20): 2523-38.
- Romero, D. and R. Palacios (1997). "Gene amplification and genomic plasticity in prokaryotes." *Annu Rev Genet* **31**: 91-111.
- Rudnicki, M. A., T. Braun, et al. (1992). "Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene Myf-5 and results in apparently normal muscle development." *Cell* **71**(3): 383-90.
- Rudnicki, M. A., P. N. Schnegelsberg, et al. (1993). "MyoD or Myf-5 is required for the formation of skeletal muscle." *Cell* **75**(7): 1351-9.
- Sabourin, L. A. and M. A. Rudnicki (2000). "The molecular regulation of myogenesis." *Clin Genet* **57**(1): 16-25.
- Schwab, M. (1999). "Oncogene amplification in solid tumors." *Semin Cancer Biol* **9**(4): 319-25.
- Schwarz, M., G. Alvarez-Bolado, et al. (1997). "Conserved biological function between Pax-2 and Pax-5 in midbrain and cerebellum development: evidence from targeted mutations." *Proc Natl Acad Sci U S A* **94**(26): 14518-23.
- Shen-Orr, S. S., R. Milo, et al. (2002). "Network motifs in the transcriptional regulation network of Escherichia coli." *Nat Genet* **31**(1): 64-8.
- Solomon, K. S. and A. Fritz (2002). "Concerted action of two dlx paralogs in sensory placode formation." *Development* **129**(13): 3127-36.
- Sonden, B. and B. E. Uhlin (1996). "Coordinated and differential expression of histone-like proteins in Escherichia coli: regulation and function of the H-NS analog StpA." *Embo J* **15**(18): 4970-80.

- Sonti, R. V. and J. R. Roth (1989). "Role of gene duplications in the adaptation of *Salmonella typhimurium* to growth on limiting carbon sources." Genetics **123**(1): 19-28.
- Spender, L. C., H. J. Whiteman, et al. (2005). "Transcriptional cross-regulation of RUNX1 by RUNX3 in human B cells." Oncogene **24**(11): 1873-81.
- Stark, G. R. and G. M. Wahl (1984). "Gene amplification." Annu Rev Biochem **53**: 447-91.
- Steinmetz, L. M., C. Scharfe, et al. (2002). "Systematic screen for human disease genes in yeast." Nat Genet **31**(4): 400-4.
- Szucsik, J. C., D. P. Witte, et al. (1997). "Altered forebrain and hindbrain development in mice mutant for the Gsh-2 homeobox gene." Dev Biol **191**(2): 230-42.
- Taniuchi, I., M. Osato, et al. (2002). "Differential requirements for Runx proteins in CD4 repression and epigenetic silencing during T lymphocyte development." Cell **111**(5): 621-33.
- Toresson, H. and K. Campbell (2001). "A role for Gsh1 in the developing striatum and olfactory bulb of Gsh2 mutant mice." Development **128**(23): 4769-80.
- Tyson, J. J., K. C. Chen, et al. (2003). "Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell." Curr Opin Cell Biol **15**(2): 221-31.
- van den Berg, M. A., P. de Jong-Gubbels, et al. (1996). "The two acetyl-coenzyme A synthetases of *Saccharomyces cerevisiae* differ with respect to kinetic properties and transcriptional regulation." J Biol Chem **271**(46): 28953-9.
- Venuti, J. M., L. Goldberg, et al. (1991). "A myogenic factor from sea urchin embryos capable of programming muscle differentiation in mammalian cells." Proc Natl Acad Sci U S A **88**(14): 6219-23.
- Waddington, C. H. (1942). Nature **150**: 563-565.
- Wagner, A. (2000). "Robustness against mutations in genetic networks of yeast." Nat Genet **24**(4): 355-61.
- Weiss, K., D. Stock, et al. (1998). "Perspectives on genetic aspects of dental patterning." Eur J Oral Sci **106 Suppl 1**: 55-63.
- Wilm, B., E. Dahl, et al. (1998). "Targeted disruption of Pax1 defines its null phenotype and proves haploinsufficiency." Proc Natl Acad Sci U S A **95**(15): 8692-7.
- Wolfe, K. H. and D. C. Shields (1997). "Molecular evidence for an ancient duplication of the entire yeast genome." Nature **387**(6634): 708-13.
- Yamashita, M., S. X. Ying, et al. (2005). "Ubiquitin ligase Smurf1 controls osteoblast activity and bone homeostasis by targeting MEKK2 for degradation." Cell **121**(1): 101-13.
- Yang, H., M. Bard, et al. (1996). "Sterol esterification in yeast: a two-gene process." Science **272**(5266): 1353-6.
- Yu, C., N. J. Kennedy, et al. (1996). "Molecular cloning and characterization of two isoforms of *Saccharomyces cerevisiae* acyl-CoA:sterol acyltransferase." J Biol Chem **271**(39): 24157-63.
- Zhang, A., S. Rimsky, et al. (1996). "Escherichia coli protein analogs StpA and H-NS: regulatory loops, similar and disparate effects on nucleic acid dynamics." Embo J **15**(6): 1340-9.

- Zhang, W., R. R. Behringer, et al. (1995). "Inactivation of the myogenic bHLH gene MRF4 results in up-regulation of myogenin and rib anomalies." Genes Dev **9**(11): 1388-99.
- Zhang, Y., C. Chang, et al. (2001). "Regulation of Smad degradation and activity by Smurf2, an E3 ubiquitin ligase." Proc Natl Acad Sci U S A **98**(3): 974-9.
- Zhao, C., U. S. Jung, et al. (1998). "Temperature-induced expression of yeast FKS2 is under the dual control of protein kinase C and calcineurin." Mol Cell Biol **18**(2): 1013-22.

#### **Statistics of biological and non-biological molecular associations<sup>4</sup>**

Biological complexity is, in part, reflected by the dense and crowded intracellular environment (Minton 2006). Continuous collisions and associations between a variety of chemical entities ranging from proteins and nucleic acids to sugars and sterols collectively define the biochemical entity. Physical associations within the cell range from tight “key-lock” type of interactions typified by protein complexes to weak and labile associations governing membrane domains, enzyme channeling and intracellular dynamics (better refs xxx) (Lancet, Sadovsky et al. 1993; Rosenwald, Kafri et al. 2002). In principle, any compound can be described by the distribution of affinities it has with all other cellular constituents (Lancet, Sadovsky et al. 1993). This affinity distribution could, therefore, be perceived as a fingerprint of its evolved function (Kauvar, Higgins et al. 1995). This notion is captured, to some extent, by the protein interaction network where instead of continuous affinity measures proteins’ are assigned with and identity of the partners with which they interact with an affinity above a particular threshold value (Bader, Heilbut et al. 2003). The distribution of the number of partners proteins share within this network was shown to follow a power-law decay, the origins of which were interpreted on basis of evolutionary basis (Yook, Oltvai et al. 2004). A more precise attempt to characterize biological affinity distributions include the Sips distribution (Sips 1948) describing the range specific to non-specific antibody-antigen affinities. Other more general models have also been described, including our own (Goldstein 1975; Lancet, Sadovsky et al. 1993; Detours, Sulzer et al. 1996; Rosenwald, Kafri et al. 2002; Kafri and Lancet 2004).

A question pertinent to this is that asking what would be the distribution of affinities within a random collection of molecules that has not evolved to constitute the cellular machinery. This type of distribution is expected to be governed by the statistical chemical properties of the ensemble, such as molecular size distribution and chemical heterogeneity. Such a distribution likely describes the “raw material” from which new associations have evolved. Motivated by this ambition, we searched for databases containing affinity values of random molecular encounters from which we could reconstruct the statistics of random associations. To this end, we relied on a commercially available dataset describing column separations of chiral compounds

---

<sup>4</sup> Kafri, R. and D. Lancet (2004). "Probability rule for chiral recognition." *Chirality* 16(6): 369-78.

(Koppenhoefer, Graf et al. 1994). By employing previously derived methodologies (Chaiken 1987; Winzor 2001) we were able to extract from the elution rates, the mobile to stationary phase affinities. From this we showed that that distribution of affinities between such random chemical moieties follows a lognormal distribution (Rosenwald, Kafri et al. 2002; Kafri and Lancet 2004). Furthermore, the number of interactions (defined by an affinity threshold) with which a single molecule is involved is binomial. Specifically considering chirality of compounds, we showed that the distribution of enantiomeric separation factors is exponential (Kafri and Lancet 2004). These results are collectively described our work published journal of chirality. Despite the initial motivation, final conclusions have somewhat diverged from their initial ambition of shedding light on the statistics of protein interactions. We nevertheless argue that considering the generality of the data, its relevance to biochemical ensembles should not be dismissed.



# Probability Rule for Chiral Recognition

RAN KAFRI AND DORON LANCET\*

*Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel*

**ABSTRACT** Molecular Chirality is of central interest in biological studies because enantiomeric compounds, while indistinguishable by most inanimate systems, show profoundly different properties in biochemical environments. Enantioselective separation methods, based on the differential recognition of two optical isomers by a chiral selector, have been amply documented. Also, great effort has been directed towards a theoretical understanding of the fundamental mechanisms underlying the chiral recognition process. Here we report a comprehensive data examination of enantioseparation measurements for over 72,000 chiral selector-selectand pairs from the chiral selection compendium CHIRBASE. The distribution of  $\alpha = k'_D/k'_L$  values was found to follow a power law, equivalent to an exponential decay for chiral differential free energies. This observation is experimentally relevant in terms of the number of different individual or combinatorial selectors that need to be screened in order to observe  $\alpha$  values higher than a preset minimum. A string model for enantioselective recognition (SMED) formalism is proposed to account for this observation on the basis of an extended Ogston three-point interaction model. Partially overlapping molecular interaction domains are analyzed in terms of a string complementarity model for ligand-receptor complementarity. The results suggest that chiral selection statistics may be interpreted in terms of more general concepts related to biomolecular recognition. *Chirality* 16:369–378, 2004. © 2004 Wiley-Liss, Inc.

**KEY WORDS:** chiral recognition; chiral discrimination; receptor affinity distribution (RAD); string model; SMED; CHIRBASE; chiral statistics; binding statistics

The field of chiral recognition and separation of chiral compounds has received considerable attention in the past few decades. This is due to its pharmaceutical importance<sup>1</sup> and to its relevance to the more general aspects of molecular recognition.<sup>2–4</sup> To date, the most applicable tool for obtaining pure enantiomers remains column chromatography on chiral stationary phases (CSPs).<sup>5–9</sup> Here, enantiomers (selectands) are separated based on their differential recognition by an immobilized asymmetric molecule (selector), in either the gas or liquid phase. The detailed mechanisms by which such recognition takes place have been characterized for a number of model systems.<sup>10–13</sup> It has been shown, for example, that the discrimination free-energy differences are typically 2–3 orders of magnitude lower than the binding energies of the ligand and stationary phase.<sup>3</sup> Also, it was shown that in most cases the molecular determinant displaying maximum binding affinity towards the separated compounds is that which is most highly enantioselective.<sup>3,14</sup>

An extensive body of theoretical descriptions of the selective process has been established describing the thermodynamics and kinetics of the chromatographic process.<sup>2,9,15–21</sup> One issue at the center of many of these studies is the distinction between selective vs. nonselective interactions taking place between the stationary phase and the separated ligand.<sup>14,17</sup> It has been established that

the measured retention factors,  $\alpha$ , are the sum of both these contributions.<sup>15,17,22</sup>

Formalisms and models describing chromatographic selection, and in particular enantioselection, have been amply documented.<sup>4,9,18,23–30</sup> One such model is the Ogston three-point attachment model.<sup>28,31–34</sup> This model, however, is largely qualitative, providing only a limited understanding of the quantitative and energetic parameters of the stereoselective process.<sup>35</sup> More detailed extensions with a higher number of attachment points have been described.<sup>18,26,29,36</sup>

While chirality is usually considered a discrete property, molecules being either chiral or nonchiral, Avnir and colleagues<sup>37–39</sup> have initiated a novel approach where the chiral content of structures is quantified by a continuous chirality measure (CCM). Continuous chirality is computed from the molecular structure by averaging over the distances from the asymmetric center to the geometric

\*Correspondence to: D. Lancet, Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel. E-mail: doron.lancet@weizmann.ac.il

Received for publication 8 October 2003; Accepted 5 March 2004

DOI: 10.1002/chir.20049

Published online in Wiley InterScience (www.interscience.wiley.com).

trace of the molecule, and is thus a purely geometric measure, insensitive to chemical descriptors.<sup>38</sup> A major achievement of this approach has been the successful prediction of the magnitude of discrimination and other chemical properties of chiral compounds.<sup>37,39</sup>

Chiral recognition reactions may be considered a subset pertaining to the wider realm of molecular recognition. Thus, chiral recognition stems from the difference,  $\Delta\Delta G$ , in the Gibbs free energy of binding between the two enantiomers of a given chiral compound.<sup>40</sup> Such recognition reactions may be further characterized by enthalpic ( $\Delta\Delta H$ ) and entropic ( $-T \Delta\Delta S$ ) contributions, whose relative magnitude determines the temperature dependence of the discrimination.<sup>3,40-47</sup> It is even possible that reversal of enantioselectivity would occur as a function of temperature.<sup>47</sup>

An understanding of the more general field of molecular recognition has previously benefited from studies employing phenomenological models and statistical analyses of reported binding data.<sup>48-50</sup> Several authors have utilized this approach to account for the affinity distributions that describe the binding of members of large repertoires or chemical libraries towards specified targets.<sup>45,51,52</sup> In this realm, a series of studies used string complementarity to model antigen-antibody binding in the immune system<sup>53-55</sup> and odorant-receptor recognition in the olfactory system.<sup>56</sup> In such string representations, complementarity between two surfaces is defined by a set of simple rules and the degree of complementarity can be quantified and used as a phenomenological measure of affinity between the two molecules. To assess their validity, some of the relevant models have been tested based on large binding datasets derived from ligand library screens.<sup>49,57</sup>

We conjectured that an analogous insight could be derived for chiral recognition by analyzing data obtained from a large number of chiral separation experiments. CHIRBASE<sup>58,59</sup> is perhaps the largest and most comprehensive chiral separation database, listing details for more than 70,000 chiral separation experiments. Although the database is constructed as a tool for deciding on separation procedures for particular analytes, it may also be used for broad statistical analyses of chiral recognition. One such analysis<sup>60</sup> was aimed at linking various chemical descriptors of both selector and selectands to separation efficiencies and at clustering separation parameters according to these descriptors. Here we attempt the task of mapping the distribution of separation constants and related free-energy differences for the entire CHIRBASE dataset. We demonstrate that this approach reveals new facets of chiral recognition.

## MATERIALS AND METHODS

### *Chiral Chromatographic Separation Data*

Chromatographic enantiomer separation data were retrieved from the CHIRBASE database (1997-2000 ENSS-PICAM, Marseille, France).<sup>61,62</sup> At the time of the study the data consisted of 72,076 chiral separation records for

selectors with pairs of enantiomeric selectands. Each record includes molecular structures, molecular mass, chromatographic data, and experimental conditions. The numerical value presented for all triads was the separation factor:

$$\alpha = k'_1/k'_2 \quad (1)$$

where  $k'$  are the capacity factors for the two optical isomers, relative to a given chiral stationary phase.

**Binding constant and free-energy computations.** The absolute values of the free-energy differences,  $\Delta\Delta G$ , associated with discriminating a pair of enantiomers (D and L) were calculated by use of:

$$\Delta\Delta G_i = |RT \ln(K_{Li}/K_{Di})| = |RT \ln(\alpha_i)| \quad (2)$$

where  $R$  is the universal gas constant and  $T$  is the absolute temperature. Equation 2 is based on the theory of Quantitative Affinity Chromatography<sup>2,9,21,23,25,40,42</sup> and has been previously used as a measure of chiral recognition free-energy differences.<sup>22,40,46,63</sup> Despite this, its validity as such has been argued,<sup>15,17</sup> stating that the real free-energy differences should be higher owing to the contributions of the nonselective associations. In fact, Jung et al.<sup>16</sup> have shown that, in the case of gas chromatography,  $\alpha$  is slightly dependent on stationary phase concentrations and cannot be used directly to calculate the absolute interaction free energy. Additionally, there are reports indicating such effects also influence liquid chromatographic separations and suggest a distinction between an apparent, measured separation factor,  $\alpha_{app}$ , and the true separation factor,  $\alpha_{tr}$ .<sup>17,64</sup> Despite this, we argue that since our goal is concerned with a statistical insight associated with the distribution of free energies, rather than a strict computation of a particular association energy, Eq. 2 will be sufficient for our purposes. We are aware, though, of the caution required in pursuing such an approach and therefore provide a detailed discussion and analysis of its implications (see Discussion). Note that throughout this article we use the notation  $K_L$  and  $K_D$  to index the relationship between enantiomers and not specify their absolute configuration.

Analysis was performed by subjecting the discrimination data to binning whereby normalized histograms were constructed describing the frequencies of the different free-energy values or energy differences. This and all other calculations (except where otherwise specified) were performed with MATLAB 6.5 (MathWorks, Natick, MA).

**Fit of experimental data to the string model.** We conceptualize the enantiomeric structures through strings of varying length constructed by the following set of rules: a Bernoulli trial is employed with a particular success probability value ( $p_1$ ) for each digit of a string whereby the first success is interpreted as an asymmetric center and the second defines the length of the string analogous to

molecular size. Accordingly, both molecule length and asymmetric center location are sampled from a geometric distribution. An interaction intensity is calculated for each of the string domains defined by the "asymmetric digit," (Fig. 4) through the use of an algorithm as described for the receptor affinity distribution (RAD) formalism.<sup>56</sup> Each digit is assigned a probability for successful interaction ( $p_2$ ) and the number of such Bernoulli successes is taken as proportional to the total binding free energy, resulting in a binomial distribution.<sup>56</sup> In each such experiment the string domain having the minimal affinity towards

the selector represents the free arm giving rise to the stereoselection (Fig. 4). The computation is repeated  $N$  times and a simulated distribution of  $\Delta\Delta G$  is obtained.

## RESULTS

### Distribution of Chiral Recognition Parameters

We extracted from CHIRBASE elution separation factors,  $\alpha$ , 72,076 chiral column experiments and constructed a normalized binned frequency plot for the  $\alpha$  values (Fig. 1). The enantiomer discrimination parameters,  $\alpha$ , spanned a range from 1.01–8.5 and showed a distinct functional distribution, whereby lower values were much more prevalent, while the highest values were very rare. As an example, 90% of all  $\alpha$  values were lower than 2.2, while only 1% were higher than 5.0. The absence of values lower than 1 reflects a consistent convention by which the higher capacity factors were always assigned to the  $k'_2$  parameter regardless of absolute configuration. Intriguingly, the chiral discrimination factors were found to obey a distinct power law  $P(\alpha) = \lambda\alpha^{-\lambda}$ , with a best fit value of  $\lambda = 3.76$ . The observation that there are many more cases of poor chiral recognition as compared to the number of resolutions with high  $\alpha$  values agrees with common-sense prediction. However, the specific functional shape, namely, that the distribution of  $\alpha$  values follow a power law as opposed to exponential, absolute value Gaussian or t-density function, as well as the particular values of the fit parameters are novel.

The identified trends are statistical and are subject to inherent inaccuracies associated with a public dataset containing numerous measurements, including heterogeneity of experimental parameters such as solvent or matrix, as well as human error. Still, because such deviations and errors are often random, they are less likely to generate systematic artifacts. This is demonstrated in Figure 1B. Furthermore, the database used is artificially depleted in the range of  $\alpha$  values very near to unity, corresponding to unsuccessful separations (see also legend to Fig. 3A). This is manifested in the first few points of the graph in Figure 1A, which deviate from the linear trend. Thus, the observed rule may not be applicable for small  $\alpha$  values.

Because  $\Delta\Delta G$  has both enthalpic and entropic contributions,<sup>21,42,64</sup> it is important to consider the temperature dependence of the observed distribution. Figure 2A shows an analysis of such dependence. It may be seen that somewhat different distributions are obtained at different temperatures, but the general power law trend (overall linearity of the double logarithmic plot) is observed for all temperature bins. Interestingly, at higher temperatures the probability of high values of  $\alpha$  diminishes, which may be taken as an indication that, on average, negative  $\Delta\Delta H$  values might govern the observed population behavior.

We next asked whether the distribution of  $\alpha$  values, as inferred from the presently analyzed data, depends on molecular size of the selector. For this, we recalculated the frequency distribution for selector/selectand triads, sorted according to the molecular masses of the selector

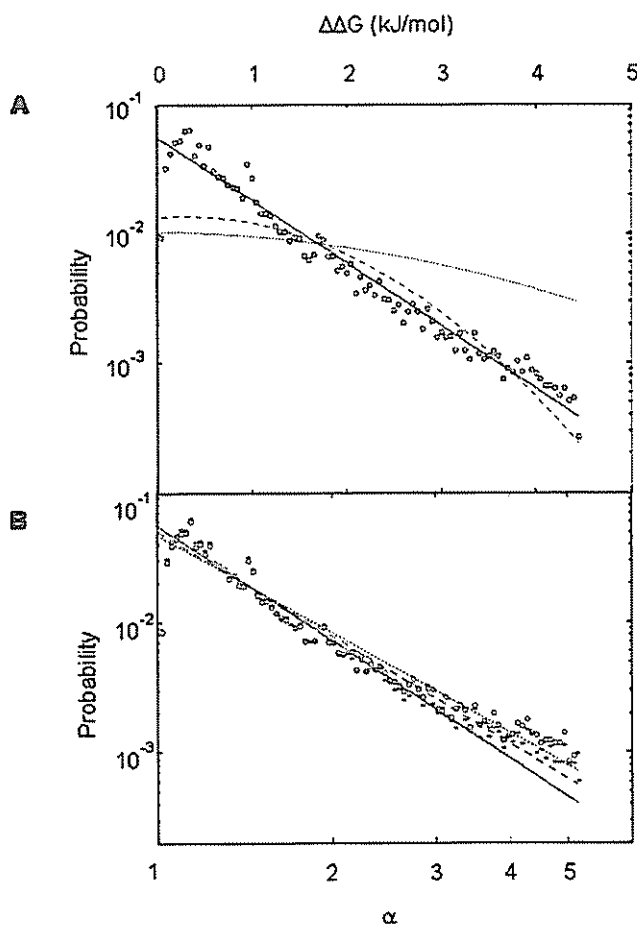
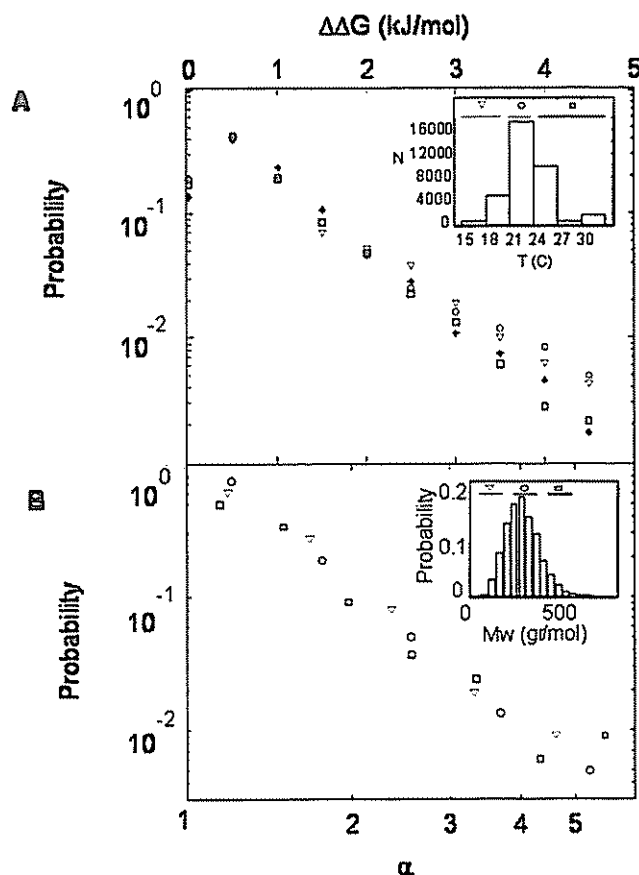
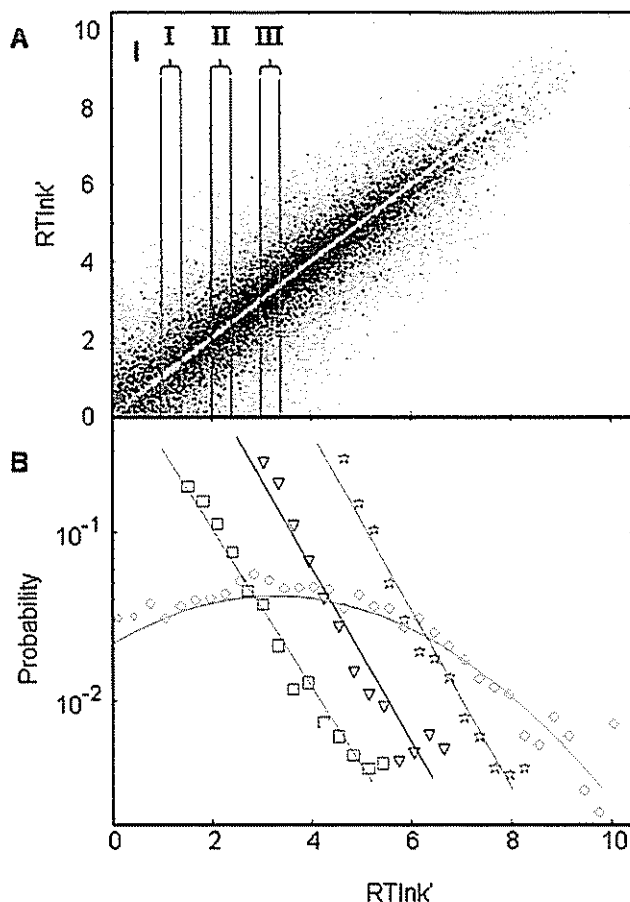


Fig. 1. A: The frequency histogram (O) for enantiomer discrimination factors,  $\alpha$ , fit to a power law ( $\lambda = 3.76$ ) (solid line) and log-normal distribution ( $\mu = 3.0$ ,  $\sigma = 0.74$ ) (dashed curve). The two horizontal axes are equivalent representations based on Eq. 2 by assuming room temperature for all separations. It was verified that the resulting temperature constrained distribution is significantly independent of this assumption. Also plotted (dotted curve) is a distribution of discrimination free-energy differences computed from randomly paired capacity factors illustrating the expected statistics if enantiomer binding affinities were noncorrelated. Only discrimination values greater than 0.05 kJ/mol were taken for the analysis, resulting in 57,650 data values which were binned into 297 bins. B: The effect of random errors on the distribution of chiral recognition free-energy change,  $\Delta\Delta G$ . Errors were introduced by randomly shuffling all  $k'_1$  values with respect to their  $k'_2$  partners and then computing the new ratios ( $\alpha_{ERR}$  values). We then replaced a randomly sampled fraction (5%, dashed curve; 10%, dotted curve) of the  $\alpha$  values with the respective  $\alpha_{ERR}$  values and computed the resulting  $\Delta\Delta G$  density functions. [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]



**Fig. 2. A:** The dependence of the distribution of  $\ln\alpha$  on the temperature at which experiment was performed. The full dataset was divided into groups according to experimental temperature (inset). Frequency distributions were calculated separately for all separations performed at a temperature lower than 294 K (triangle), between 294 and 297 K (circle), and greater than 297 K (square). Additionally, we show the distribution from experiments for which a temperature was not recorded (diamonds). The four distribution plots were calculated from 25,136 data values with a bin size of 0.5. The lower horizontal axis represents an approximate scale computed through Eq. 2 assuming room temperature. **B:** The dependence of the distribution  $P(\Delta\Delta G)$ , the enantiomer discriminating free-energy difference distribution, on the molecular weight of the chiral selectors. The full dataset was divided into groups according to the molecular mass of the selector (inset), and a frequency histogram was calculated for each such size group accordingly 0–200 gr/mol (triangle), 200–400 gr/mol (circle), and 400–600 gr/mol (square). The three distribution plots (0–200, 200–400, 400–600) were calculated from 1,527 data values with a bin size of 0.9, 3,450 values with a bin size of 1.0, and 333 values with a bin size of 0.7, respectively. [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

(Fig. 2B). Interestingly, the  $P(\alpha)$  distribution was found to be practically independent of molecular size. Previously, a molecular size-dependence was computationally suggested for fractal diffusion-limited aggregates.<sup>65</sup> While these were suggested as prototypes of chiral selectors, our data suggest that molecular size-dependence might be true only under a limited set of assumptions. It should be noted that our analyses were done only for selector up to a size of 600 Da, above which the data were too sparse for significant analysis. Antibiotics and cyclodextrins, which are important chiral selectors,<sup>66,67</sup> typically have molecular masses above 500 Da. Future focused



**Fig. 3. A:** Correlation of the  $RT\ln(k')$  values for the paired enantiomers, drawn for all 72,076 pairs. The red lines illustrate the intervals for the computations in B. A reduced concentration of data points on the diagonal reflects the fact that only successful chiral resolutions were included in the analysis, depleting very small values of  $\Delta\Delta G$ . **B:** Conditional affinity distributions for different values of  $k'$ . Frequency histograms of the Y axis  $RT\ln k'$  values of A were calculated for three subsets of  $RT\ln k'$  values on the X axis of the same figure as follows: Interval I, 1–1.4; Interval II, 2–2.4; and Interval III, 3–3.4, as indicated in A. Best fit straight lines were obtained by optimizing the slope and intercept parameters  $a$  and  $b$  in Eq. 4, and they were, respectively,  $a = 3.09$ ,  $b = 1.08$  (square);  $a = 24.6$ ,  $b = 1.19$  (triangle); and  $a = 167.47$ ,  $b = 1.21$  (star). All frequency plots were calculated with a bin size of 0.3. Best fit to  $P(\Delta\Delta G)$  using a subset of the same size as that for the red line gave the parameters  $a = 1.07$ ,  $b = 1.03$ . Also shown is a distribution of  $RT\ln k'$  values (diamond) within a random molecular set of the same size (~1,000) with a best fit for a Gaussian approximation ( $\mu = 3.28$ ,  $\sigma = 2.85$ ). [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

analyses done with such compounds could help determine the statistical behavior of  $\alpha$  in this size class.

#### Interpreting $P(\alpha)$ Through Affinity Distributions

In an attempt to interpret the above power law result, we addressed the discrimination parameters as ratios of values sampled from an underlying affinity distribution.<sup>45,48,49,56,68,69</sup> We examined the distribution that would emerge based on the null hypothesis that  $k'_L$  and  $k'_D$  for every compound are completely uncorrelated. For this, random pairs were drawn out of the collection of experimental  $k'_i$  values, and the probability distribution  $P(k'_i/k'_j)$

was plotted (Fig. 1, red dashed line). Obviously, the resultant curve deviates strongly from the actually observed  $P(\alpha)$  distribution. This discrepancy is not unexpected, however, because there are previous indications that the  $k_i$  for enantiomer pairs are correlated.<sup>10</sup> Indeed, plotting a correlation diagram between  $k'_L$  and  $k'_D$  values for all enantiomer pairs (Fig. 3A) shows a correlation coefficient of 0.83. This high value relative to the null hypothesis of nonexistent correlation between  $k'_L$  and  $k'_D$  is statistically significant at  $P < 10^{-9}$  and suggests that such correlation underlies the actually observed  $P(\Delta\Delta G)$  distribution. Such a correlation value, however, is necessary, but not sufficient for generating the observed statistical behavior, as this would also depend on the specific set of conditional affinity distributions defined by the enantiomer/selector interaction.

To find out the exact nature of this correlation, we generated a set of conditional density curves for the experimental  $k'_{Li}$  values related to compounds for which  $k'_{Di}$  was within a specified magnitude range (Fig. 3B). The resultant curves showed an exponential decline:

$$P(X = q_{Li} | q_{Di}) = ae^{-bx} \quad (3)$$

where  $q_{Li} = RT \ln k'_{Li}$  and  $q_{Di} = RT \ln k'_{Di}$ . The best fit parameters  $a$  and  $b$  were found to follow a linear dependence according to:

$$P(q_{Li} | q_{Di}) = (a_1 + a_2 q_{Li})e^{-q_{Di}(b_1 + b_2 q_{Li})} \quad (4)$$

Based on Eq. 4 and the discovered power law, a probability distribution for  $\alpha$  may be computed analytically based on a standard formula<sup>70</sup> (p. 36):

$$P(A) = \sum_{i=1}^n P(A | B_i)P(B_i) \quad (5)$$

where  $B_i$  represents the  $q_{Di}$  range specifying the condition. The resulting exponential dependence (Fig. 5A) provides a measure of rationalization to the presently discovered density function of chiral recognition parameters.

Figure 3 addresses another question, namely, whether the separation factors  $\alpha$  are correlated to the values of the capacity factors  $k'$ . In other words, we ask whether a strong molecular association between selector and selectand is the prerequisite to high enantioselectivity. Figure 3A shows that by and large such correlation does not exist. This is seen by the fact that off-diagonal points, which signify large (or small)  $k'_L/k'_D$  values, hence large  $\alpha$  values, are not found considerably more often at any particular  $k'_L$  or  $k'_D$  values range.

#### String Complementarity Model of Enantioselection

We examined the possibility that a toy-model may be constructed that will rationalize the distribution of  $\alpha$  values. We considered an L-shaped string of length  $B$  as a model for an asymmetric selector (Fig. 4). The two antipodal selectands are represented by similar strings of length  $B$ , which are mirror images of each other, not

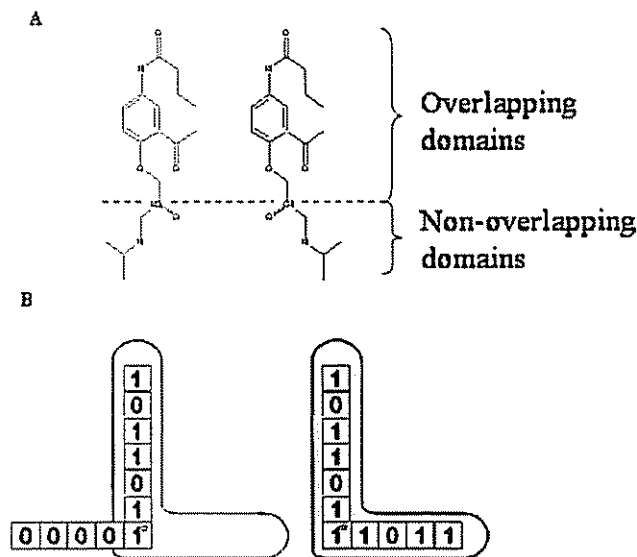


Fig. 4. The string model for chiral recognition (SMED). A: A realistic representation of the difference in binding site recognition between enantiomers. B: String representations of two chiral selectands and a selector binding surface (gray) representing the underlying selector related string. [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

superimposable in two dimensions (see Materials and Methods). This model merges features of previously explored string-based models for molecular recognition, with concepts stemming from the three-point attachment model for chiral recognition or discrimination. The latter model for stereoselectivity invokes three-point interactions between receptor and substrate. Accordingly, one enantiomer binds to a selector entity simultaneously at three sites, while the opposite enantiomer cannot bind to the same three sites. Later treatises have pointed out the significance of conformational flexibility, steric factors, and even repulsive interactions for the recognition of chiral isomers.<sup>71</sup> However, the basic concepts of a three-point interaction continues to be accepted.<sup>26,33,34,72-74</sup> In our model, the two attachment points to the selector that preserve their interaction in both selectand enantiomers are represented by a long arm of length  $M$ , delimited by the chiral center. The third attachment point, which can form an interaction only in one of the isomers, is represented by the shorter, B-M long arm. This model harbors a quantitative manifestation of the partial correlation between the binding energies of the two enantiomers:  $M$  elementary interactions are identical between the antipodes, while another B-M such interactions are only present in one of the isomers. Notably, similar analogies may be drawn between our model and other more elaborate molecular descriptions of chiral recognition.<sup>18,28,35,50</sup> An assumption of our model, similar to that which underlies other  $n$ -point attachment models,<sup>27,29</sup> is that the two enantiomers bind to the partially overlapping recognition sites. It is obvious that this simplification does not always hold, and alternative models could better address such a scenario.

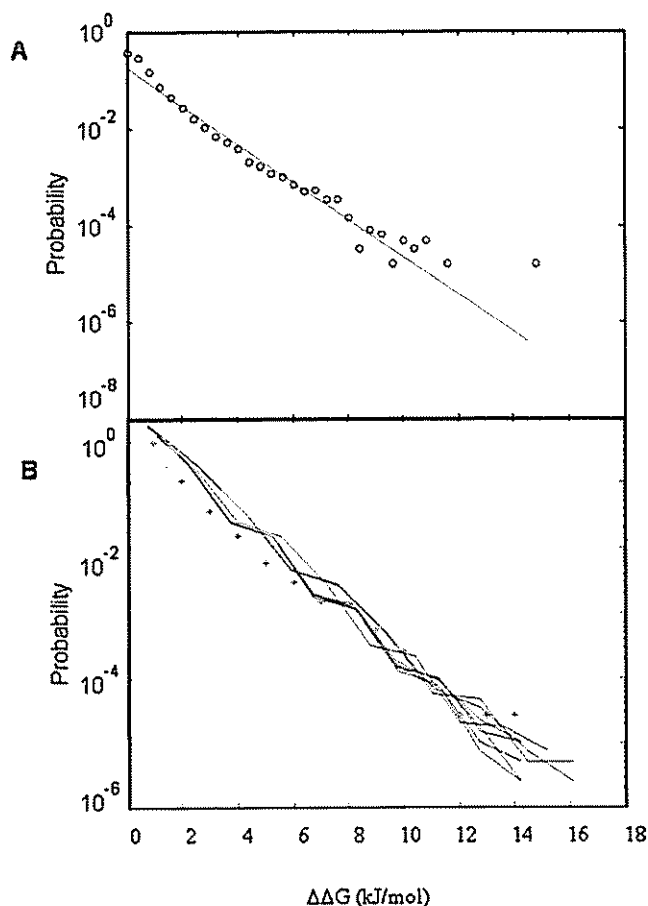


Fig. 5. A: An analytically calculated distribution of enantiomer separation factors. The computation was based on Eq. 4 and the power law distribution by use of the standard formula given in Eq. 5. B: Model results from eight different iterations of a probability distribution computation, each with 10,000  $\Delta\Delta G$  values, plotted along with experimental discrimination frequencies.  $P_1 = 0.11$ ,  $P_2 = 0.20$ . [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

The model representation for the free energy of binding is computed by a string complementarity rule,<sup>52,55</sup> employing a complete geometrical overlap for one of the isomers, and overlap restricted to the long arm for the other isomer (see Materials and Methods). A more elaborate procedure that may have been employed is that in which grid elements are assigned different weights<sup>75,76</sup>; such elaborations may be considered to enhance the power of future analyses. The employed procedure allowed us to generate a large number of  $\Delta\Delta G$  values in a Monte Carlo simulation, using a set of two parameters. From these, a  $P(\Delta\Delta G)$  function is computed and an iterative fit procedure subsequently employed to obtain the parameters showing best fit to the experimental points. Figure 5B shows that the invoked string model for chiral recognition (SMED) may adequately predict the functional dependence of measured  $P(\Delta\Delta G)$  data.

#### DISCUSSION

The main ideas derived from the present analysis are the particular probability distribution form that governs

chiral recognition and the potential mechanistic rationalization for such quantitative behavior. We suggest that this distribution may be applicable to practical questions in the field of chiral recognition and separation, as discussed below.

A-posteriori analyses of cumulative results within public databases form the basis for many discoveries in the emerging field of bioinformatics. A similar approach is applied here by analyzing a large dataset of chiral separation results from numerous laboratories. An inherent weakness of such wide-ranging analysis is the lack of control over experimental variables associated with data accumulation in a single laboratory. These variables may include heterogeneity of chromatography dead times, buffer types, column densities, as well as human experimental and database generation errors. Therefore, the capacity of random errors to annihilate a specific conclusion should be carefully assessed.<sup>77–79</sup> This was done here (Fig. 1B) showing that randomization of as many as 20% of all data points does not appreciably affect the conclusion drawn about the statistical behavior of the discrimination factors.

Although the interpretation of  $\alpha$  values as free-energy differences by use of Eq. 2 have been employed throughout the literature,<sup>5,11,12,22,40,44,46,63,80</sup> there is substantial evidence showing that this interpretation may be inaccurate, due to deviations resulting from nonselective binding contributions.<sup>15–17,21,41,42,81,82</sup> Such deviations have been reported to be more pronounced in gas chromatography as compared to liquid chromatography.<sup>15,16</sup> Schurig and colleagues<sup>15</sup> provided a comparison between  $\Delta\Delta G$  values computed with Eq. 2 and those computed more accurately by taking nonselective interactions into account. This analysis was performed on gas chromatographic columns and showed ligand concentration dependence. In that report, the deviation between the two energy terms is on the order of 0.3 kJ/mol, a rather small difference relative to the range where most of the analyzed  $\Delta\Delta G$  values reside, between 1–5 kJ/mol (Fig. 1). Furthermore, because all our analyses were restricted to liquid rather than gas chromatography, we believe that the general shape of the  $\Delta\Delta G$  distribution is valid.

Temperature has a significant effect on retention, chiral recognition, resolution, and column efficiency for chromatographic separation. Thus, variable temperature has been frequently used as an optimization parameter for chromatographic separations of enantiomers.<sup>83–86</sup> Analyses based on van't Hoff plots are often used to derive thermodynamic functions of enantioselective adsorption, which may be interpreted in terms of mechanistic aspects of chiral recognition. With some assumptions, the temperature dependence on the retention of a given analyte can be expressed by a derivative of the Van't Hoff equation as applied to  $\alpha$  and  $\Delta\Delta H$ .<sup>87</sup> Although the temperature dependence of the measured separation factors have proven important for mechanistic determination in many studies,<sup>41,46,47,83</sup> their relevance is decreased in the case of statistical studies, where influences of particular temperature dependencies may be canceled out due to the heterogeneity of the ensemble. Our results show that the observation of a linear

dependence of the logarithm of the probability of  $\alpha$  on the logarithm of  $\alpha$  only has a weak temperature dependence. Future studies could indicate the detailed contribution of enthalpy and entropy to this general behavior.

Simple phenomenological models based on string complementarity have proven effective in predicting the functional form of affinity distributions.<sup>49,51,52,55,56</sup> We therefore attempted to develop an analogous approach for predicting the conditional probability phenomena that underlie the stereoselective processes which allow chiral separations. The analysis of a large dataset of chiral recognition values led to a hitherto unsuspected relationship, whereby the propensity of chiral discrimination magnitude diminishes as a power law. The simplicity of this behavior was perceived as indicating a possible explanation in terms of the elementary interactions that underlie enantioselection.

As a basis for constructing our quantitative string model for chiral recognition, SMED, we used one of the most widely accepted qualitative rationalizations, the three-point attachment model. More specifically, we invoke the notion that the interaction subsites shared by both enantiomers contribute a large free energy change as compared to the nonshared subsites. This echoes the notion that the two strongest binding interactions within the selector-selectand pair contribute to the retention of the enantiomers, whereas the weakest interaction determines the discrimination selectivity.<sup>18,36</sup>

The fit of data to the described string complementarity model could be influenced by structural heterogeneity. For example, in GC reversals of elution order were observed in response to minute changes of the structure of the selector or selectand.<sup>44,88</sup> Impurities of the stationary phase selector obviously harbors similar risks. Despite this, due to the success of the string model approach in the description of affinity distributions<sup>45,49,51,54-56,68,69,89</sup> for which similar structural complexities are relevant, we were encouraged to try this approach in the field of chiral recognition.

The choice of a model for chiral recognition may also be influenced by potential knowledge of the entropy and enthalpy contributions. The binding of ligand to a receptor, and similarly of a selectand to a selector, often includes favorable enthalpic contributions due to hydrogen bonds, dipole and ionic interactions. In parallel, favorable entropic contributions could come from hydrophobic interactions. In contrast, unfavorable contributions are often related to the fact that the actual formation of the complex involves an increase in order, and hence negative entropy change. In this realm, the negative (unfavorable) entropy contribution would become stronger as the complex is more structured and involves more subsite interactions. In the case of the three-point attachment model, the combination of three binding contacts could be favored by enthalpy but, as this combination represents a more ordered structure, it is at the same time disfavored by entropy. The individual figures for  $\Delta\Delta H$  and  $\Delta\Delta S$  and the value of the temperature will determine whether chiral recognition is mainly enthalpy- or entropy-controlled. Thus, as more knowledge is accumulated about the above thermodynamic param-

eters the choice of a model to explain chiral recognition could be refined.

Criticism has been directed against the underlying simplicity of the three-point approach,<sup>28-30,35</sup> and more elaborate models have been proposed, including a four-point attachment model.<sup>29,35</sup> However, our model constitutes sufficient departure from the basic three-point attachment view to be minimally affected by such less favorable appraisal. Specifically, we invoke the three-point concept only in a broad outline, and our model brings into play a large number of elementary interactions, thus being a more realistic representation of the molecular contacts that might mediate chiral recognition.

Our analysis attempts to make conjectural inferences on the properties of molecular interactions between chiral selectands and stationary phase selectors based on the statistics of  $\Delta\Delta G$  values. This process has a limited validity because of the lack of detailed thermodynamic information regarding the individual measurements. Thus, a complete knowledge of binding site configurations, temperature dependences, selector concentration, and mobile phase parameters are largely lacking. One has to bear in mind that every  $\Delta\Delta G$  value is a compound physical measure that includes enthalpy and entropy terms. For all these reasons, the observations and the mechanistic inferences should be regarded as tentative, and as a basis for future scrutiny. However, the statistical inference reported here for chiral recognition may provide clues for future research, when additional data may become available.

A significant contribution to an overall understanding of enantiomer recognition is the quantitative chirality measure approach, based on purely geometric considerations.<sup>38</sup> We perceive that an extension merging the SMED approach with such a geometrical analysis could help enhance the utility of both models. Because comparisons based on the quantitative chirality measure apply mainly to series compounds that share functional attributes, this could be realized by performing restricted statistical analysis within compound libraries of defined chemical attributes, such as the helicene family.<sup>37</sup>

In the SMED formalism considered here we assume that for the low-affinity enantiomer zero free-energy contribution occurs by the molecular aspect that underlies discrimination. This parsimonious assumption is based on precedence, whereby other authors have performed analyses based on such a similar simplification.<sup>90</sup> Our model, however, could be elaborated in the future by assuming that the zero-energy moiety may also mediate repulsion, contributing a positive free-energy term, as proposed.<sup>29</sup>

An important outcome of the present study is the observation of a clear difference between the functional form of the chiral recognition probability density function and that of the binding free-energy density functions that were previously described.<sup>45,49,56</sup> This discrepancy is shown to be the result of a correlation of the binding free energies within enantiomer pairs, consistent with both enantiomers partially binding to the same structural elements on the selector molecule. This inference is



statistical in nature, and does not necessarily apply to every CSP/ligand. In the future, more detailed studies, performed at well-defined specific experimental conditions, could highlight the composite relationships between the statistical behaviors of  $\Delta G$  and  $\Delta\Delta G$  in chiral recognition.

Chiral separation is often a "trial and error" process,<sup>91,92</sup> which requires the application of diverse techniques as well as the utilization of numerous selector types in order to attain high  $\alpha$  values.<sup>7,9,93</sup> Combinatorial libraries have been used to address this problem.<sup>94–96</sup> Consequently, the number of different selector columns to be scanned, or the size of the combinatorial library to be employed, are important decisions for an experimenter in the field of chiral recognition. In this context, our derivation of a specific statistical rule that predicts the rarity of  $\alpha$  values higher than a preset minimum may generate a useful guide for method optimization.

#### LITERATURE CITED

- Eichelbaum M, Gross AS. Stereochemical aspects of drug action and disposition. *Adv Drug Res* 1996;28:1–64.
- Ringo MC, Evans CE. Liquid chromatography as a measurement tool for chiral interactions. *Anal Chem News Features* 1998; 315A–321.
- Lipkowitz KB, Pearl G, Coner B, Peterson MA. Explanation of where and how enantioselective binding takes place on permethylated beta-cyclodextrin, a chiral stationary phase used in gas chromatography. *J Am Chem Soc* 1997;119:600–610.
- Lipkowitz KB. Atomistic modeling of enantioselective binding. *Acc Chem Res* 2000;33:555–562.
- Hare PE, Gil-av E. Separation of D and L amino acids by liquid chromatography: use of chiral eluants. *Science* 1979;204:1226–1228.
- Thomas E, Beesley RPWS. *Chiral chromatography*. New York: John Wiley & Sons; 1999.
- Lough WJ. *Chiral liquid chromatography*. London: Blackie Academic & Professional; 2002.
- Lochmuller CH, Souter RW. Chromatographic resolution of enantiomers selective review. *J Chromatogr* 1975;113:283–302.
- Guiochon G. Preparative liquid chromatography. *J Chromatogr A* 2002;965:129–161.
- Lipkowitz KB, Coner B, Peterson MA, Morreale A. Enantioselective binding in gas chromatography: a computational study of chiral selection by permethyl-beta-cyclodextrin. *J Phys Organ Chem* 1997;10: 311–322.
- Schefzick S, Lindner W, Lipkowitz KB, Jalaie M. Enantiodiscrimination by a quinine-based chiral stationary phase: a computational study. *Chirality* 2000;12:7–15.
- Maier NM, Schefzick S, Lombardo GM, Feliz M, Rissanen K, Lindner W, Lipkowitz KB. Elucidation of the chiral recognition mechanism of cinchona alkaloid carbamate-type receptors for 3,5-dinitrobenzoyl amino acids. *J Am Chem Soc* 2002;124:8611–8629.
- Suzuki T, Timofei S, Iuoras BE, Uray G, Verdino P, Fabian WM. Quantitative structure-enantioselective retention relationships for chromatographic separation of arylalkylcarbinols on Pirkle type chiral stationary phases. *J Chromatogr A* 2001;922:13–23.
- Lipkowitz KB, Coner R, Peterson MA. Locating regions of maximum chiral discrimination: A computational study of enantioselection on a popular chiral stationary phase used in chromatography. *J Am Chem Soc* 1997;119:11269–11276.
- Spanik I, Krupcik J, Schurig V. Comparison of two methods for the gas chromatographic determination of thermodynamic parameters of enantioselectivity. *J Chromatogr A* 1999;843:123–128.
- Jung M, Schmalzing D, Schurig V. Theoretical approach to the gas-chromatographic separation of enantiomers on dissolved cyclodextrin derivatives. *J Chromatogr* 1991;552:43–57.
- Gotmar G, Fornstedt T, Guiochon G. Apparent and true enantioselectivity in enantioseparations. *Chirality* 2000;12:558–564.
- Davankov VA, Meyer VR, Rais M. A vivid model illustrating chiral recognition induced by achiral structures. *Chirality* 1990;2:208–210.
- Zhou D, Kaczmarek K, Cavazzini A, Liu X, Guiochon G. Modeling of the separation of two enantiomers using a microbore column. *J Chromatogr A* 2003;1020:199–217.
- Skavrada M, Jandera P, Cherrak DE, Aced A, Guiochon G. Adsorption isotherms and retention behavior of 1,1'-bis(2-naphthol) on CHIRIS AD1 and CHIRIS AD2 columns. *J Chromatogr A* 2003;1016: 143–154.
- Fornstedt T, Guiochon G. Nonlinear effects in LC and chiral LC. *Anal Chem* 2001;73:608A–617A.
- Maier NM, Franco P, Lindner W. Separation of enantiomers: needs, challenges, perspectives. *J Chromatogr A* 2001;906:3–33.
- Winzor DJ. Quantitative affinity chromatography. *J Biochem Biophys Methods* 2001;49:99–121.
- Dunn BM, Chaiken IM. Quantitative affinity chromatography. Determination of binding constants by elution with competitive inhibitors. *Proc Natl Acad Sci USA* 1974;71:2382–2385.
- Chaiken IM. Analytical affinity chromatography in studies of molecular recognition in biology: a review. *J Chromatogr* 1986;376: 11–32.
- Sundaresan V, Abrol R. Towards a general model for protein-substrate stereoselectivity. *Protein Sci* 2002;11:1330–1339.
- Topiol S. A general criterion for molecular recognition: implications for chiral interactions. *Chirality* 1989;1:69–79.
- Meyer VR, Rais M. A vivid model of chiral recognition. *Chirality* 1989; 1:167–169.
- Mesecar AD, Koshland DE. Sites of binding and orientation in a four-location model for protein stereospecificity. *Jubmb Life* 2000;49: 457–466.
- Booth TD, Wahnon D, Wainer IW. Is chiral recognition a three-point process? *Chirality* 1997;9:96–98.
- Labrie J, Mercier JF, Slater GW. An exactly solvable Ogston model of gel electrophoresis. V. Attractive gel-analyte interactions and their effects on the Ferguson plot. *Electrophoresis* 2000;21:823–833.
- Nederkoorn PH, van Lenthe JH, van der Goot H, Donne-Op den Kelder GM, Timmerman H. The agonistic binding site at the histamine H2 receptor. I. Theoretical investigations of histamine binding to an oligopeptide mimicking a part of the fifth transmembrane alpha-helix. *J Comput Aid Mol Des* 1996;10:461–478.
- Nederkoorn PH, van Gelder EM, Donne-Op den Kelder GM, Timmerman H. The agonistic binding site at the histamine H2 receptor. II. Theoretical investigations of histamine binding to receptor models of the seven alpha-helical transmembrane domain. *J Comput Aid Mol Des* 1996;10:479–489.
- Ahn S, Ramirez J, Grigorean G, Lebrilla CB. Chiral recognition in gas-phase cyclodextrin: amino acid complexes—is the three point interaction still valid in the gas phase? *J Am Soc Mass Spectrom* 2001; 12:278–287.
- Bentley R. Three-point attachment: past, present, but no future. *Trans NY Acad Sci* 1983;41:5–24.
- Davankov VA. The nature of chiral recognition: is it a three point interaction? *Chirality* 1997;9:99–102.
- Katzenelson O, Edelstein J, Avnir D. Quantitative chirality of helixes. *Tetrahedron Asymmetry* 2000;11:2695–2704.
- Zabrodsky H, Peleg S, Avnir D. Continuous symmetry measures. *J Am Chem Soc* 1992;114:7843–7851.
- Lipkowitz KB, Schefzick S, Avnir D. Enhancement of enantiomeric excess by ligand distortion. *J Am Chem Soc* 2001;123:6710–6711.
- Feibush B, Gil-av E. Interaction between asymmetric solutes and solvents. Peptide derivatives as stationary phases in gas liquid partition chromatography. *Tetrahedron* 1970;26:1361–1368.



41. Miyabet K, Guiochon G. Extrathermodynamic relationships in reversed-phase liquid chromatography. *Anal Chem* 2002;74: 5754–5465.
42. Miyabet K, Guiochon G. Thermodynamic interpretation of retention equilibrium in reversed-phase liquid chromatography based on enthalpy-entropy compensation. *Anal Chem* 2002;74:5982–5992.
43. Peter A, Vekes E, Armstrong DW. Effects of temperature on retention of chiral compounds on a ristocetin A chiral stationary phase. *J Chromatogr A* 2002;958:89–107.
44. Wang X, Ching CB. Liquid chromatographic retention behavior and enantiomeric separation of three chiral center beta-blocker drug (nadolol) using heptakis (6-azido-6-deoxy-2, 3-di-O-phenylcarbamoylated) beta-cyclodextrin bonded chiral stationary phase. *Chirality* 2002;14: 798–805.
45. Yee E. Reconstruction of the antibody affinity distribution from experimental binding data by a minimum cross-entropy procedure. *J Theor Biol* 1991;153:205–227.
46. Dietrich A, Maas B, Mosandl A. Diluted modified cyclodextrins as chiral stationary phases influence of the polysiloxane solvent — heptakis(2,3-di-O-acetyl-6-O-tert-butyltrimethylsilyl)-beta-cyclodextrin. *J High Resol Chromatogr* 1995;18:152–156.
47. Karlsson A, Skoog A, Ohlen K. Effect of temperature on the reversal in the retention order of the enantiomers of mosapride on Chiral-AGP. *J Biochem Biophys Methods* 2002;54:347–356.
48. Janin J. Quantifying biological specificity: the statistical mechanics of molecular recognition. *Proteins* 1996;25:438–445.
49. Rosenwald S, Kafri R, Lancet D. Test of a statistical model for molecular recognition in biological repertoires. *J Theor Biol* 2002;216: 327–336.
50. Topiol S. A general criterion for molecular recognition — implications for chiral interactions. *Chirality* 1989;1:69–79.
51. Perelson AS. Immune network theory. *Immunol Rev* 1989;110:5–36.
52. Lancet D, Horovitz A, Katchalski-Katzir E. Molecular recognition in biology: models for analysis of protein-ligand interactions. In: Behr JP, editor. *The lock-and-key principle*, vol. 2. New York: John Wiley & Sons; 1994. p 25–71.
53. Detours V, Perelson AS. Explaining high alloreactivity as a quantitative consequence of affinity-driven thymocyte selection. *Proc Natl Acad Sci USA* 1999;96:5153–5158.
54. Detours V, Perelson AS. The paradox of alloreactivity and self MHC restriction: quantitative analysis and statistics. *Proc Natl Acad Sci USA* 2000;97:8479–8483.
55. Perelson AS, Weisbuch G. Immunology for physicists. *Rev Mod Phys* 1997;69:1219–1267.
56. Lancet D, Sadovsky E, Seidemann E. Probability model for molecular recognition in biological receptor repertoires: significance to the olfactory system. *Proc Natl Acad Sci USA* 1993;90:3715–3719.
57. Inman JK, Barnett AL. Affinities of antibodies for diverse ligands—theoretical and practical aspects. In: Hutchens TW, editor. *Protein recognition of immobilized ligand; UCLA symposia on molecular and cellular biology*, vol. 80. New York: Liss A.R.; 1988. p 35–44.
58. Roussel C, Pierrot-Sanders J, Heitmann I, Piras P. Database current status and derived research applications using molecular similarity, decision tree and 3D enantiophore search. In: Subramanian G, editor. *Weinheim: Wiley-VCH*; 2001. p 95–125.
59. Piras P, Roussel C, Pierrot-Sanders J. Reviewing mobile phases used on Chiralcel OD through an application of data mining tools to CHIRBASE database. *J Chromatogr A* 2001;906:443–458.
60. Roussel C, Piras P, Heitmann I. An approach to discriminating 25 commercial chiral stationary phases from structural data sets extracted from a molecular database. *Biomed Chromatogr* 1997;11: 311–316.
61. Koppenhoefer B, Graf R, Holzschuh H, Nothdurft A, Trettin U, Piras P, Roussel C. CHIRBASE, a molecular database for the separation of enantiomers by chromatography. *J Chromatogr A* 1994;666: 557–563.
62. Koppenhoefer B, Nothdurft A, Pierrot-Sanders J, Piras P, Popescu C, Roussel C, Stiebler M, Trettin U. CHIRBASE, a graphical molecular database on the separation of enantiomers by liquid-, supercritical fluid-, and gas chromatography. *Chirality* 1993;5:213–219.
63. Ching CB, Lim BG, Lee EJ, Ng SC. Chromatographic resolution of the chiral isomers of several beta-blockers over cellulose tris(3,5-dimethylphenylcarbamate) chiral stationary phase. *Chirality* 1992;4: 174–177.
64. Oberleitner WR, Maier NM, Lindner W. Enantioseparation of various amino acid derivatives on a quinine based chiral anion-exchange selector at variable temperature conditions. Influence of structural parameters of the analytes on the apparent retention and enantioseparation characteristics. *J Chromatogr A* 2002;960:97–108.
65. Katzenelson O, Avnir D. Quantitative chirality/enantioselectivity relations in large random supramolecular structures. *Chemistry* 2000; 6:1346–1354.
66. Jandera P, Skavrada M, Klemmova K, Backovska V, Guiochon G. Effect of the mobile phase on the retention behaviour of optical isomers of carboxylic acids and amino acids in liquid chromatography on bonded Teicoplanin columns. *J Chromatogr A* 2001;917: 123–133.
67. Bicchi C, D'Amato A, Rubiolo P. Cyclodextrin derivatives as chiral selectors for direct gas chromatographic separation of enantiomers in the essential oil, aroma and flavour fields. *J Chromatogr A* 1999; 843:99–121.
68. Goldstein B. Theory of hapten binding to IgM: the question of repulsive interactions between binding sites. *Biophys Chem* 1975;3: 363–367.
69. Detours V, Sulzer B, Perelson AS. Size and connectivity of the idiotypic network are independent of the discreteness of the affinity distribution. *J Theor Biol* 1996;183:409–416.
70. Nelson R. Probability, stochastic processes, and queueing theory. *The mathematics of computer performance modeling*. Berlin: Springer; 1995.
71. Koshland DE Jr, Nemethy G, Filmer D. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* 1966;5:365–385.
72. Vidyasankar S, Ru M, Arnold FH. Molecularly imprinted ligand-exchange adsorbents for the chiral separation of underivatized amino acids. *J Chromatogr A* 1997;775:51–63.
73. Fersht A. *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*. New York: W.H. Freeman; 1999.
74. Copeland R. *Enzymes: a practical introduction to the structure, mechanisms and data analysis*. New York: John Wiley & Sons; 2000.
75. Heifetz A, Katchalski-Katzir E, Eisenstein M. Electrostatics in protein-protein docking. *Protein Sci* 2002;11:571–587.
76. Strynadka NC, Eisenstein M, Katchalski-Katzir E, Shoichet BK, Kuntz ID, Abagyan R, Totrov M, Janin J, Cherfils J, Zimmerman F, et al. Molecular docking programs successfully predict the binding of a beta-lactamase inhibitory protein to TEM-1 beta-lactamase. *Nat Struct Biol* 1996;3:233–239.
77. Siemer M, Joermann J. Power and measures of effect size in analysis of variance with fixed versus random nested factors. *Psychol Methods* 2003;8:497–517.
78. Abecasis GR, Cherny SS, Cardon LR. The impact of genotyping error on family-based analysis of quantitative traits. *Eur J Hum Genet* 2001;9:130–134.
79. Abecasis GR, Cardon LR, Cookson WO, Sham PC, Cherny SS. Association analysis in a variance components framework. *Genet Epidemiol* 2001;21(Suppl 1):S341–346.
80. Agranat I, Caner H, Caldwell J. Putting chirality to work: the strategy of chiral switches. *Nat Rev Drug Discov* 2002;1:753–768.
81. Gotmar G, Fornstedt T, Andersson M, Guiochon G. Influence of the solute hydrophobicity on the enantioselective adsorption of beta-blockers on a cellulase protein used as the chiral selector. *J Chromatogr A* 2001;905:3–17.
82. Fornstedt T, Zhong G, Bensetiü Z, Guiochon G. Experimental and theoretical study of the adsorption behavior and mass transfer kinetics of propranolol enantiomers on cellulase protein as the selector. *Anal Chem* 1996;68:2370–2378.

83. Kazusaki M, Shoda T, Kawabata H. Enthalpy-entropy compensation for enantio-separation on cellulose and amylose tris (3, 5-dimethyl-phenylcarbamate) derivatives as stationary phases. *Chromatography* 2003;24:121–126.
84. Kusters E, Loux V, Schmid E, Floersheim P. Enantiomeric separation of chiral sulfoxides — screening of cellulose-based sorbents with particular reference to cellulose tribenzoate. *J Chromatogr A* 1994; 666:421–432.
85. Schurig V. Separation of enantiomers by gas chromatography. *J Chromatogr A* 2001;906:275–299.
86. Kirkland KM, McCombs DA. Changes in chiral selectivity with temperature for an ovomucoid protein-based column. *J Chromatogr A* 1994;666:211–219.
87. Eisenberg D, Crothers D. *Physical chemistry with applications to the life sciences*. Upper Saddle River, NJ: Pearson Education; 1979.
88. Okamoto M. Reversal of elution order during the chiral separation in high performance liquid chromatography. *J Pharm Biomed Anal* 2002;27:401–407.
89. Inman JK. A study of multispecific interactions by quantitative affinity chromatography. In: Chaiken IM, Wilcheck M, Parkih I, editors. *Affinity chromatography and biological recognition*. Orlando: Academic Press; 1983.
90. Katzenelson O, HelOr HZ, Avnir D. Chirality of large random supra-molecular structures. *Chem Eur J* 1996;2:174–181.
91. Bereznitski Y, Thompson R, O'Neill E, Grinberg N. Thin-layer chromatography—a useful technique for the separation of enantiomers. *J AOAC Int* 2001;84:1242–1251.
92. Wang F, Khaledi MG. Chiral separations by nonaqueous capillary electrophoresis. *Anal Chem* 1996;68:3460–3467.
93. Davankov VA. Enantioselective ligand exchange in modern separation techniques. *J Chromatogr A* 2003;1000:891–915.
94. Wang Y, Bluhm LH, Li T. Identification of chiral selectors from a 200-member parallel combinatorial library. *Anal Chem* 2000;72:5459–5465.
95. Zhong W, Yeung ES. Combinatorial enantiomeric separation of diverse compounds using capillary array electrophoresis. *Electrophoresis* 2002;23:2996–3005.
96. Brahmachary E, Ling FH, Svec F, Frechet JM. Chiral recognition: design and preparation of chiral stationary phases using selectors derived from ugi multicomponent condensation reactions and a combinatorial approach. *J Comb Chem* 2003;5:441–450.

## Conclusions

From the collection of studies described (Kafri, Bar-Even et al. 2005; Kafri, Levy et al. 2006; Kafri and Pilpel 2006) it is apparent that genes which are redundant in function are often not independently controlled but rather they are regulated by a system that has evolved to both monitor and respond to their intactness. Furthermore we show that such ‘*responsive backup circuits*’ (RBCs) are preferentially associated with protein network hubs (Kafri and Pilpel 2006) and have been evolutionarily conserved in many different organisms (Kafri, Levy et al. 2006). Together, these results suggest that genetic networks have evolved to utilize the functional overlap of many particular gene pairs. Nevertheless, the nature of this utilization and the evolutionarily selectable advantageous functionalities associated with it still remain unclear. To speculate in this direction we performed a mathematical analysis characterizing the residual contribution of redundancies to the dynamics and steady states of gene functions (Kafri, Levy et al. 2006). Using this approach we demonstrate that responsive backup circuits may function as increasingly efficient genetic devices that filter non-genetic noise from transcriptional pathways. This functionality is achieved by having dosage fluctuations of one duplicate counteracted with reciprocal fluctuations of the other. Although still speculative, this model predicts that responsive backup circuits are preferentially associated with genes having higher levels of stochasticity or noise (Kafri, Levy et al. 2006).

A currently exiting opportunity to test the association of responsive backup circuits with non-genetic noise has just recently become available by data that has been experimentally generated in the Barkai Lab in collaboration with Bar-Even et al. (Bar-Even et al., *Nature Genetics* in press). In this work the noise of individual yeast genes was quantified by measuring the distribution of the expression levels GFP-tagged yeast proteins. It is, thus, now possible to test for a preferential association between responsive backup circuits and noisy genes. By collecting duplicates for which redundancy has been established one could specifically examine whether they are more noisy or of lower protein dosage than the genome average.

One of the central open questions that must still be addressed is just how common are RBCs in biology and how frequently do they appear in the different genomes. Also, we can ask whether this frequency, be it low or high, can be faithfully estimated from the amount of times RBCs are reported in the literature. In other words, do RBCs represent a ‘genome wide’ phenomenon or a collection of rare

incidents? It is unfortunate that at this point no conclusive answer can be provided to that question, mainly because of the very limited number of studies that have specifically probed for cross-regulation among redundant protein pairs or for the lack of a systematic nomenclature describing it. Since absence of phenotype in a knockout experiment (a potential consequence of RBCs) is often looked upon by researchers as an 'insignificant result', analyses of the 'minor' regulatory consequences of such knockouts may have not seen the light of publication. One approach to overcome this is to collect pairs of genes for which the double knockout is lethal while the separate gene disruptions are not, i.e., synthetic lethals. In some respect these interactions could be regarded as proxies for functional compensatory interactions between these genes and thus some form of functional redundancy. The problem is that a close inspection reveals that the vast majority of synthetic lethal pairs are far from being even remotely functionally related. For example, the list of synthetic lethals of FKS1 gene in yeast (Lesage, Sdicu et al. 2004), include the gene FEN1. FEN1 is an enzyme that lies on the pathway of sphingolipid biosynthesis. Its genetic interaction with FKS1 is thought to result from an accumulation of its substrate, phytosphingosine, in the plasma membrane of the FEN1 deletion mutant. Phytosphingosine, in turn, is thought to repress the interaction between Fks1p and Rho1p, possibly by forming a microdomain around Fks1p and physically preventing its association with its regulatory subunit, Rho (Abe, Nishida et al. 2001). Thus, despite the genetic interaction, no redundancy can be implied.

Furthermore, while the question of prevalence of RBCs is important, one can not judge their biological significance solely based on that criterion. This fact is emphasized, for example, by the pair of genes utrophin and dystrophin associated with structural components of muscle fiber. This pair of genes constitutes an RBC in that utrophin was found to be up regulated in the absence of its homolog dystrophin (Porter, Rafael et al. 1998). Particular attention has been attracted to this gene pair as dystrophin mutations were found causative of the Duchenne muscular dystrophy condition in human patients. Yet, in mice, utrophin has shown remarkable ability to compensate for dystrophin knockouts and it is estimated that partial compensation also occurs in humans (Deconinck, Rafael et al. 1997). Inspired by the efficient compensatory effect this RBC has in mouse studies have been suggested to artificially induce in human patients by means of gene therapy (Deconinck, Rafael et al. 1997;

Porter, Rafael et al. 1998; Dowling, Culligan et al. 2002). Although these ambitions are not yet been realized they point a fruitful possibility.

One of the challenges of this work was to survey the literature for already documented cases of RBCs. The difficulty of this lies in the fact that there is, currently, no consistent ontology to describe backup circuits. As a result there is no simple pubmed search that could retrieve all such cases. Instead, we employed here a manual laborious search strategy in which papers whose abstracts contained key words such as "redundancy", "functional overlap", "paralogous", etc. were carefully examined. Future publications of knockout experiments with even very little phenotypic effects, using the RBC terminology coined here may facilitate further study of recurring principles that govern the regulation of genetic redundancies.

## References

- Abe, M., I. Nishida, et al. (2001). "Yeast 1,3-beta-glucan synthase activity is inhibited by phytosphingosine localized to the endoplasmic reticulum." J Biol Chem 276(29): 26923-30.
- Bader, G. D., A. Heilbut, et al. (2003). "Functional genomics and proteomics: charting a multidimensional map of the yeast cell." Trends Cell Biol 13(7): 344-56.
- Braun, T. and H. H. Arnold (1995). "Inactivation of Myf-6 and Myf-5 genes in mice leads to alterations in skeletal muscle development." Embo J 14(6): 1176-86.
- Brookfield, J. F. (2003). "Gene duplications: the gradual evolution of functional divergence." Curr Biol 13(6): R229-30.
- Chaiken, I. M. (1987). Analytical Affinity Chromatography, Boca Raton, FL : CRC Press, 1987.
- Conant, G. C. and A. Wagner (2003). "Asymmetric sequence divergence of duplicate genes." Genome Res 13(9): 2052-8.
- Conant, G. C. and A. Wagner (2004). "Duplicate genes and robustness to transient gene knock-downs in *Caenorhabditis elegans*." Proc R Soc Lond B Biol Sci 271(1534): 89-96.
- Davis, J. C. and D. A. Petrov (2004). "Preferential duplication of conserved proteins in eukaryotic genomes." PLoS Biol 2(3): E55.
- Deconinck, A. E., J. A. Rafael, et al. (1997). "Utrophin-dystrophin-deficient mice as a model for Duchenne muscular dystrophy." Cell 90(4): 717-27.
- Detours, V., B. Sulzer, et al. (1996). "Size and connectivity of the idiotypic network are independent of the discreteness of the affinity distribution." J Theor Biol 183(4): 409-16.
- Dowling, P., K. Culligan, et al. (2002). "Distal mdx muscle groups exhibiting up-regulation of utrophin and rescue of dystrophin-associated glycoproteins exemplify a protected phenotype in muscular dystrophy." Naturwissenschaften 89(2): 75-8.
- Elowitz, M. B., A. J. Levine, et al. (2002). "Stochastic gene expression in a single cell." Science 297(5584): 1183-6.
- Enns, L. C., M. M. Kanaoka, et al. (2005). "Two callose synthases, GSL1 and GSL5, play an essential and redundant role in plant and pollen development and in fertility." Plant Mol Biol 58(3): 333-49.
- Ferris, S. D. and G. S. Whitt (1979). "Evolution of the differential regulation of duplicate genes after polyploidization." J Mol Evol 12(4): 267-317.

- Foster, D. V., S. A. Kauffman, et al. (2006). "Network growth models and genetic regulatory networks." Phys Rev E Stat Nonlin Soft Matter Phys 73(3 Pt 1): 031912.
- Fraser, H. B., A. E. Hirsh, et al. (2004). "Noise minimization in eukaryotic gene expression." PLoS Biol 2(6): e137.
- Gasch, A. P., P. T. Spellman, et al. (2000). "Genomic expression programs in the response of yeast cells to environmental changes." Mol Biol Cell 11(12): 4241-57.
- Gerhart, J. and M. Kirschner (1997). Cells, embryos, and evolution : toward a cellular and developmental understanding of phenotypic variation and evolutionary adaptability. Malden, Mass. ; Oxford, Blackwell Science.
- Goldstein, B. (1975). "Theory of hapten binding to IgM: the question of repulsive interactions between binding sites." Biophys Chem 3(4): 363-7.
- Gu, X. (2003). "Evolution of duplicate genes versus genetic robustness against null mutations." Trends in Genetics 19(7): 354-356.
- Gu, Z., D. Nicolae, et al. (2002). "Rapid divergence in expression between duplicate genes inferred from microarray data." Trends Genet 18(12): 609-13.
- Gu, Z., L. M. Steinmetz, et al. (2003). "Role of duplicate genes in genetic robustness against null mutations." Nature 421(6918): 63-6.
- Hartman, J. L. t., B. Garvik, et al. (2001). "Principles for the buffering of genetic variation." Science 291(5506): 1001-4.
- He, X. and J. Zhang (2006). "Higher duplicability of less important genes in yeast genomes." Mol Biol Evol 23(1): 144-51.
- Jeong, H., S. P. Mason, et al. (2001). "Lethality and centrality in protein networks." Nature 411(6833): 41-2.
- Jordan, I. K., Y. I. Wolf, et al. (2004). "Duplicated genes evolve slower than singletons despite the initial rate increase." BMC Evol Biol 4: 22.
- Kafri, R., A. Bar-Even, et al. (2005). "Transcription control reprogramming in genetic backup circuits." Nat Genet 37(3): 295-9.
- Kafri, R. and D. Lancet (2004). "Probability rule for chiral recognition." Chirality 16(6): 369-78.
- Kafri, R., M. Levy, et al. (2006). The utilization of genetic redundancy by responsive backup circuits.
- Kafri, R. and Y. Pilpel (2006). Evidence for preferential genetic backup of hubs in the protein interaction network.

- Kauvar, L. M., D. L. Higgins, et al. (1995). "Predicting ligand binding to proteins by affinity fingerprinting." Chem Biol 2(2): 107-18.
- Kellis, M., B. W. Birren, et al. (2004). "Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*." Nature 428(6983): 617-24.
- Kirschner, M. and J. Gerhart (1998). "Evolvability." Proc Natl Acad Sci U S A 95(15): 8420-7.
- Kondrashov, F. A., I. B. Rogozin, et al. (2002). "Selection in the evolution of gene duplications." Genome Biol 3(2): RESEARCH0008.
- Koppenhoefer, B., R. Graf, et al. (1994). "CHIRBASE, a molecular database for the separation of enantiomers by chromatography." J Chromatogr A 666(1-2): 557-63.
- Krakauer, D. C. and M. A. Nowak (1999). "Evolutionary preservation of redundant duplicated genes." Semin Cell Dev Biol 10(5): 555-9.
- Lancet, D., E. Sadovsky, et al. (1993). "Probability model for molecular recognition in biological receptor repertoires: significance to the olfactory system." Proc Natl Acad Sci U S A 90(8): 3715-9.
- Lesage, G., A. M. Sdicu, et al. (2004). "Analysis of beta-1,3-glucan assembly in *Saccharomyces cerevisiae* using a synthetic interaction network and altered sensitivity to caspofungin." Genetics 167(1): 35-49.
- Lynch, M. and J. S. Conery (2000). "The evolutionary fate and consequences of duplicate genes." Science 290(5494): 1151-5.
- Makova, K. D. and W. H. Li (2003). "Divergence in the spatial pattern of gene expression between human duplicate genes." Genome Res 13(7): 1638-45.
- Minton, A. P. (2006). "Macromolecular crowding." Curr Biol 16(8): R269-71.
- Nowak, M. A., M. C. Boerlijst, et al. (1997). "Evolution of genetic redundancy." Nature 388(6638): 167-71.
- Ohno, S. (1970). Evolution by Gene and Genome Duplication, Springer.
- Papp, B., C. Pal, et al. (2003). "Dosage sensitivity and the evolution of gene families in yeast." Nature 424(6945): 194-7.
- Papp, B., C. Pal, et al. (2003). "Evolution of cis-regulatory elements in duplicated genes of yeast." Trends Genet 19(8): 417-22.
- Papp, B., C. Pal, et al. (2004). "Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast." Nature 429(6992): 661-4.



- Pearce, A. C., Y. A. Senis, et al. (2004). "Vav1 and vav3 have critical but redundant roles in mediating platelet activation by collagen." J Biol Chem 279(52): 53955-62.
- Porter, J. D., J. A. Rafael, et al. (1998). "The sparing of extraocular muscle in dystrophinopathy is lost in mice lacking utrophin and dystrophin." J Cell Sci 111 ( Pt 13): 1801-11.
- Pyne, S., S. Skiena, et al. (2005). "Copy correction and concerted evolution in the conservation of yeast genes." Genetics 170(4): 1501-13.
- Raser, J. M. and E. K. O'Shea (2004). "Control of stochasticity in eukaryotic gene expression." Science 304(5678): 1811-4.
- Rosenwald, S., R. Kafri, et al. (2002). "Test of a statistical model for molecular recognition in biological repertoires." J Theor Biol 216(3): 327-36.
- Rudnicki, M. A., T. Braun, et al. (1992). "Inactivation of MyoD in mice leads to up-regulation of the myogenic HLH gene Myf-5 and results in apparently normal muscle development." Cell 71(3): 383-90.
- Schwarz, M., G. Alvarez-Bolado, et al. (1997). "Conserved biological function between Pax-2 and Pax-5 in midbrain and cerebellum development: evidence from targeted mutations." Proc Natl Acad Sci U S A 94(26): 14518-23.
- Sips, R. J. (1948). "On the structure of a catalyst surface." J. Chem. Phys. 16(5): 490-495.
- Swain, P. S., M. B. Elowitz, et al. (2002). "Intrinsic and extrinsic contributions to stochasticity in gene expression." Proc Natl Acad Sci U S A 99(20): 12795-800.
- Taylor, J. S. and J. Raes (2004). "Duplication and divergence: the evolution of new genes and old ideas." Annu Rev Genet 38: 615-43.
- van den Berg, M. A., P. de Jong-Gubbels, et al. (1996). "The two acetyl-coenzyme A synthetases of *Saccharomyces cerevisiae* differ with respect to kinetic properties and transcriptional regulation." J Biol Chem 271(46): 28953-9.
- Veitia, R. A. (2002). "Exploring the etiology of haploinsufficiency." Bioessays 24(2): 175-84.
- Veitia, R. A. (2004). "Gene dosage balance in cellular pathways: implications for dominance and gene duplicability." Genetics 168(1): 569-74.
- Vilar, J. M., H. Y. Kueh, et al. (2002). "Mechanisms of noise-resistance in genetic oscillators." Proc Natl Acad Sci U S A 99(9): 5988-92.
- Wagner, A. (2000). "Robustness against mutations in genetic networks of yeast." Nat Genet 24(4): 355-61.

- Wagner, A. (2001). "The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes." Mol Biol Evol 18(7): 1283-92.
- Winzor, D. J. (2001). "Quantitative affinity chromatography." J Biochem Biophys Methods 49(1-3): 99-121.
- Wolfe, K. H. and D. C. Shields (1997). "Molecular evidence for an ancient duplication of the entire yeast genome." Nature 387(6634): 708-13.
- Yook, S. H., Z. N. Oltvai, et al. (2004). "Functional and topological characterization of protein interaction networks." Proteomics 4(4): 928-42.
- Zhang, Y., M. J. Tessaro, et al. (2003). "Knockout analysis of Arabidopsis transcription factors TGA2, TGA5, and TGA6 reveals their redundant and essential roles in systemic acquired resistance." Plant Cell 15(11): 2647-53.