REVIEW: CELL BIOLOGY

Principles for the Buffering of Genetic Variation

John L. Hartman IV, Barbara Garvik, Lee Hartwell*

Most genetic research has used inbred organisms and has not explored the complexity of natural genetic variation present in outbred populations. The translation of genotype to phenotype is complicated by gene interactions observed as epistasis, canalization, robustness, or buffering. Analysis of double mutations in inbred experimental organisms suggests some principles for gene interaction that may apply to natural variation as well. The buffering of variation in one gene is most often due to a small number of other genes that function in the same biochemical process. However, buffering can also result from genes functioning in processes extrinsic to that of the primary gene.

The importance of genetic variation for understanding human disease is increasingly appreciated, as exemplified by the large-scale public and private initiatives aiming to identify hundreds of thousands of SNPs (single-nucleotide polymorphisms) along the human genome. SNPs are defined as chromosomal positions where two or more variant bases exist, each with 1% or greater prevalence within a population. SNPs serve as genetic markers and potentially identify variant alleles that contribute to phenotypic traits (*1, 2*). Human genomes contain approximately one SNP per 1500 bases of DNA sequence. This means that most genes have several polymorphic sites distributed throughout their coding and regulatory regions. Genetic variation is abundant in all natural species, and most is expected to be neutral or nearly neutral with respect to fitness (*3*). This variation can be seen as quantitative differences for nearly every phenotype, and artificial selection experiments in experimental organisms demonstrate that variation for most traits is heritable. Human diseases, whether they are inherited in Mendelian fashion or not, are quantitative traits (*4, 5*) for which the same allele can present different phenotypes in different individuals (*6–8*). The genetic component of human phenotypic variation is of great interest because of its impact on the quality of human life even though much of it may have little consequence for fitness (*5*).

One of the ultimate goals of the SNP project is to refine the assignment of genotype to phenotype in individuals. Many factors will make achieving this goal difficult (*9, 10*), including the large number of polymorphisms, the possibility that many polymorphisms contribute small effects to a single phenotype, unrecognized population admixture (*11*), and the fact that phenotypic expression of variant alleles might be influenced differentially by environment (*12*), stochastic events (*13, 14*), and interactions with multiple other genetic loci. If only 1 in 10,000 of the polymorphisms present in the human population had some quantitative phenotypic effect, then there would be more than enough unique combinations of these polymorphisms to assure that every human being (with the exception of identical twins) should have a unique phenotype, in agreement with our anecdotal experience.

Our progress in understanding human disease genes owes much to research in experimental organisms. One of the most important outcomes of the accumulation of knowledge from model organisms is the realization that orthologous genes are present ubiquitously in living organisms. This knowledge has been extremely important in human genetics, where the function of many of the 1000 identified human disease genes has been surmised from functional information about its ortholog in model organisms (*15*). Although the phenotypic effects of genetic variation are most evident in humans, because of the relative intensity with which human phenotypic richness has been investigated and characterized, investigators studying quantitative traits in humans have exhorted caution in extending techniques such as linkage disequilibrium mapping, that have been used to identify rare alleles with major effects inherited in Mendelian fashion, to the analysis of complex traits (*5, 11, 16*). It is hardly a leap of faith to assume that, just as model organisms have been instrumental in defining the roles of genes and the structure of genetic pathways that are important for human disease, they will be equally useful in defining the principles of gene interaction. Orthologs of disease-modifying genes are also likely to be ubiquitous. Moreover, experimental organisms may be even more useful for discovering gene interactions than for the characterization of the functions of individual genes, because the power resulting from genetic tractability will be compounded in studies of gene interaction.

Gene Interactions Underlie Buffering

Given the growing recognition of both the importance of genetic variation and the usefulness of model organisms for understanding human biology and disease, it is an appropriate time for geneticists to go beyond studying single genes in model organisms and attempt to derive principles about gene interactions. Currently, geneticists use inbred, often clonal populations of experimental organisms, limiting the genotype-phenotype relationship to the allele of interest in order to obtain a uniform result. It is a common experience to find that a specific allele produces somewhat different phenotypes in different inbred strains of the same species (*17–21*). The diversity of phenotypes produced by identical mutations in different strain backgrounds has been attributed to suppressors, enhancers, and modifiers. These concepts imply that a fundamentally singular phenotype is expected of a given allele and that this expectation is somehow obscured by the influences of other single-gene variants. However, whenever the effects of genetic background have been investigated, they always reveal much greater complexity than these concepts indicate (*17–19*). In nature, there is no wild type, rather, all phenotypes are quantitative traits; "disease" merely lies beyond some arbitrarily defined point along a spectrum. Although we are beginning to understand the functional consequences of mutations in individual genes, we have very limited understanding of how genetic variation in different genes influences one another.

Living systems maintain phenotypic stability in the face of a great variety of perturbations arising from environmental changes, stochastic events, and genetic variation. Although this universal biological feature was appreciated long ago, our understanding of how robustness is attained at the cellular and molecular level remains quite limited. C. H. Waddington was among the first biologists to appreciate the role that homeostatic mechanisms play in the mo-

Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109–1024, USA.

*To whom correspondence should be addressed. E-mail: lhartwel@fhcrc.org

lecular reactions underlying development (*22*). He wrote: ". . . developmental reactions, as they occur in organisms submitted to natural selection, are in general canalized. That is to say, they are adjusted so as to bring about one definite end result regardless of minor variations in conditions during the course of the reaction." He was brought to this conclusion by the emergence of distinct tissue types adjacent to one another during development and the absence of intermediate forms. His metaphor was of a number of well-defined canals down which development could flow. He continues, "the constancy of the wild type must be taken as evidence of the buffering of the genotype against minor variations not only in the environment in which the animals developed but also in its genetic make-up."

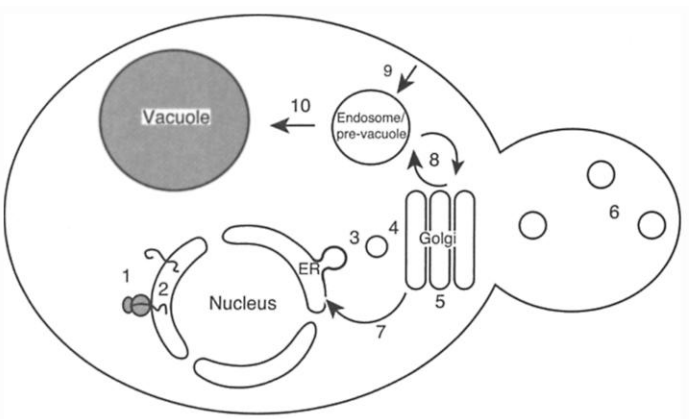
Waddington and others have pointed out that buffering occasionally breaks down, and heritable differences in the buffering capacity for environmental and genetic perturbations have been demonstrated (*23, 24*). A recent example showed that loss of buffering due to mutation of the HSP90 gene permits phenotypic expression of "cryptic" genetic variation (*25*). Additionally, the finding that fluctuating asymmetry (quantitative differences between anatomical structures with bilateral symmetry) increases with inbreeding points to the common existence of genes that buffer developmental processes (*26, 27*).

Systematic Studies of Gene Interaction

Our aim is to point out that some principles regarding the molecular mechanisms that buffer the phenotypic consequences of genetic variation can be gleaned from the literature on gene interactions in experimental organisms, particularly yeast. Some types of gene interaction have been extensively studied in inbred model organisms. These studies usually involve comparing the phenotypes of single mutants with those of double mutants in an otherwise isogenic genetic background and are usually done for the purpose of identifying genes that function in related processes or to order genes in a pathway. We suggest that the results of these studies are informative about potential buffering relationships between naturally occurring variant alleles of genes in natural populations. Whereas the primary focus of this report is on the buffering of genetic variation, it is likely that mechanistic overlap exists between buffering of genetic, environmental, and stochastic perturbations to the organism. An understanding of the molecular basis for the buffering of genetic variation should enhance our appreciation for the principles of molecular circuit design, clarify our understanding of the relationship of genotype to phenotype in outbred populations, aid in identifying alleles that act as human disease modifiers, and may be relevant to how genetic variation is accumulated and expressed during evolution. An excellent discussion of the relationship between genetic buffering and evolutionary questions has appeared recently (*28*).

How can genes that buffer variation in other genes be identified? If the product of gene A buffers the phenotypic consequences of variations (whether genetically, environmentally, or stochastically induced) in gene B, then there may be alleles of gene A that lose that buffering capacity. Typically, geneticists have screened with a mutant allele of a known gene, with function greater or less than the wild-type allele, for mutations in other genes which either "suppress" or "enhance" the original mutant phenotype. In yeast, where growth is a convenient phenotype, mutations in two different genes are said to be "synthetically lethal" if either mutation is viable in an otherwise wild-type background, but the combination of both alleles prevents growth. Synthetic lethality defines a relationship where the presence of one gene (A) allows the organism to tolerate genetic variation (b) in another gene (B) that would be lethal in the absence of the first gene (a) (Fig. 1). Synthetic lethality has been used extensively to study genes of the secretion pathway in yeast, which are divided by genetic and biochemical analysis into 10 different steps, each defined by specific biochemical markers (*29*). The relationships between the roles of proteins in the biochemistry of secretion and the interactions of mutations in their genes to produce synthetic lethality are summarized in Fig. 2. Examples of screens for synthetic gene interactions in other organisms include the analysis of the RAS pathways involved in eye development in *Drosophila* (*30*) and vulva development in

Fig. 1. Synthetic lethality identifies buffering relationships. If an organism is viable with full function of either one of two genes (A or B) and no function of the other, then the organism will be viable with full function of either gene and partial function of the other. Thus, one gene buffers variation in the other.



			Number of synthetic lethal interactions		
Step	Function	Number of genes	Same step	Different step	Not secretion
1	Translocation to Golgi	23	6	0	4
2	Maturation in ER	16	10	0	2
3	Vesicle budding	12	11	3	2
4	Vesicle fusion	18	11	13	1
5	Transport to Golgi	11	2	9	3
6	Fusion w/ plasma membrane	30	43	13	12
7	Retrieval	20	15	9	1
8	Vacuolar targeting	16	6	6	4
9	Endocytosis	20	12	12	22
10	Endosome to vacuole	7	0	3	2
Total		173	116	68	53

Fig. 2. Synthetic lethal relationships in the secretion pathway of *S. cerevisiae*. The secretion pathway of yeast has been divided into 10 different biochemical steps. Synthetic lethal interactions have been found for many genes in the pathway, with nearly half of all interactions occurring between genes acting at the same step, approximately one-quarter occurring between genes in different steps, and the remaining quarter occurring between genes not known to be involved in secretion. A synthetic lethal interaction within a step involves two members of that step, but is counted as only one interaction. A few genes act in two steps (e.g., KAR2 in steps 1 and 2 and ACT1 in steps 6 and 9). Interactions between these genes and any other gene in either step were allotted to the step of the second gene (*34*). [The diagram has been reprinted from (*29*) with permission.]

Caenorhabditis elegans (*31*). Many complex biochemical pathways have been elucidated by identifying mutations in genes that either enhance or suppress the phenotype of mutation in a known gene of that pathway. It seems likely that such genetic interactions identify relationships relevant to the buffering of natural variation and to the expression of quantitative traits in natural populations (*32*).

Three generalizations can be drawn from extensive work on synthetic gene interactions in yeast. First, the majority of synthetic lethal relationships occur among genes acting in a single process or pathway, providing a biochemical logic to the compartmentalization of buffering (Fig. 2). We term this "intrinsic" buffering. Second, although the majority of buffering relationships appear to be intrinsic, some synthetic relationships occur between genes located in biochemically distinct circuits ["extrinsic" buffering (Fig. 3)]. The latter type of gene interactions identify processes that are functionally, but not necessarily biochemically, redundant; e.g., DNA repair pathways function to compensate for defects in DNA replication (*33*). Third, the number of synthetic lethal interactions for a particular gene is generally small. For example, we found five studies that saturated the yeast genome for synthetic lethal interactions by searching until more than one mutant allele was recovered from each synthetic lethal locus. The number of synthetic lethal interactions discovered in each study ranged from three to eight (Table 1). Examination of the synthetic lethal relationships reported in the Yeast Proteome Database reveal that a defect in a single gene is usually synthetically lethal with less than 10 other genes and at most 26 other genes (Fig. 4). The small number of synthetic lethal interactions for any gene has implications for the amount of genetic variation that can be accommodated in a population, suggesting that variation in any given gene is insulated from the effects of variation in the vast majority of other genes in the genome.

Although these generalizations appear to have broad experimental support, there are several caveats that limit one's confidence in them. First, the purpose of the studies from which we have drawn our conclusions was not to identify the buffering relationships between genes, but rather, was usually to identify new genes that function in the same or a related pathway or to order known genes within a pathway. Consequently, in some cases only selected genes were tested for synthetic lethal interactions (e.g., by candidate crosses). In other cases, unbiased searches were made (i.e., by mutagenesis of the entire genome), but the searches either were not comprehensive (e.g., not "saturating") or only selected results were reported. Another caveat of extrapolating from results of genetic screens not designed to investigate buffering per se is that only extreme phenotypes are scored. Extreme phenotypes identify the genes that are least well buffered, and the genes most relevant for their buffering, but may miss more subtle effects. What we view as ideal for the purpose of defining buffering relationships would be a comprehensive search of all gene combinations together with measurements that allow quantitative comparison of the degree of buffering. The results of such a study would provide a catalog of buffering relationships. New opportunities

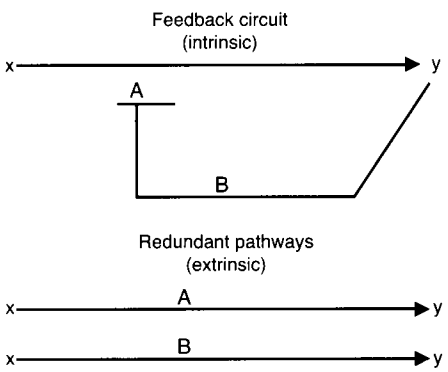


Fig. 3. Intrinsic and extrinsic buffering. In an intrinsic relationship, the properties of a circuit foster the buffering of variation, as in a feedback control circuit. In an extrinsic relationship, buffering arises from a partially redundant or compensatory process.

Table 1. Comprehensive synthetic lethal screens.

Mutant gene	Number of interacting genes	Reference
gle1	4	York *et al.* (*46*)
abp1	3	Holtzman *et al.* (*47*)
orc2	8	Hardy (*48*)
pom152	4	Atchinson *et al.* (*49*)
msb1	3	Pringle, Bender (*50*)

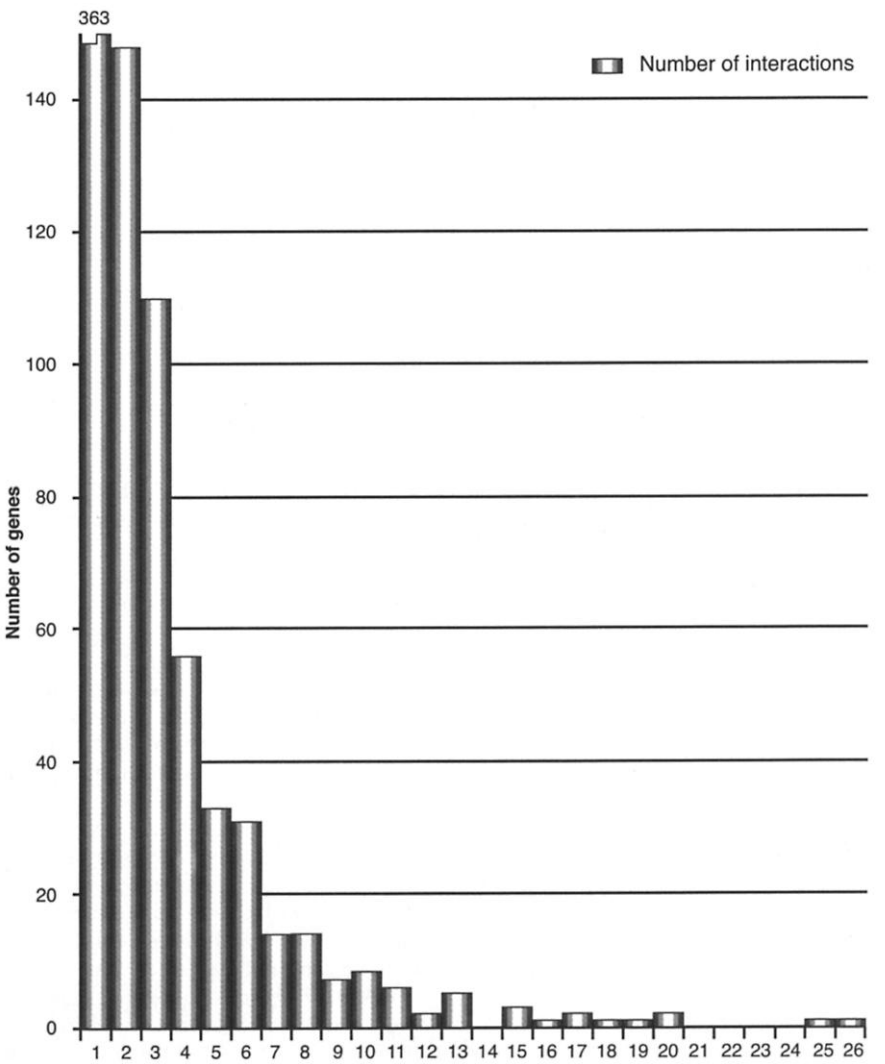


Fig. 4. The distribution of synthetic lethal relationships in the proteome database. For most genes, only one synthetic lethal relationship has been reported, and only a few genes have 10 or more such interactions. Most of this data was derived from studies that were not comprehensive.

for studying the subtleties of gene interactions are enabled by entire genome sequences, microarray technologies for parallel analysis of nucleic acids and proteins, and computational methods that increase the power to analyze large amounts of data. In yeast, fewer than 30% of the 6000 suspected genes are without known or suspected function (*34*), and a complete set of individual gene deletion strains for *Saccharomyces cerevisiae* is nearing completion.

Mechanisms of Buffering

In addition to a comprehensive catalog of gene interactions, it would be desirable to understand the principles and mechanisms that underlie buffering relationships. Although it makes intuitive sense that, if a process is weakened, then further inactivation of that process (or of a compensatory process) would bring its activity below some debilitating threshold, we lack rigorous understanding of these relationships. Which components of the circuit interact to affect circuit output? What is the output threshold? How much weakening of the output can be tolerated? Why do some genes in the same pathway display buffering interactions, but not others? Are there common design principles that account for buffering in different pathways? Are most biochemical pathways backed up by compensating pathways, and will their identification lead to an understanding of how molecular circuits are insulated from and integrated with one another?

Although we are far from a sophisticated understanding of the mechanisms of buffering, some insights have been obtained. The simplest, most obvious, and best appreciated mechanism for buffering genetic variation is redundancy (*35, 36*). For example in humans, all autosomal genes are redundant due to the presence of two alleles, one on each chromosome, thus explaining why most mutations and genetic diseases are recessive. Functional genetic redundancy may result from gene duplication events with residual, overlapping function of the evolutionarily divergent duplicated genes. The family of adhesins of yeast are an example of a set of proteins that can substitute for one another when inappropriately expressed, but that normally play distinct physiological roles as a result of differential regulation and compartmentalization (*37*). There are other well-documented examples of partial redundancy (*38*), however, the generality of this mechanism for buffering is much less clear than in the case of diploidy. Although ~13% of the yeast genome is a relic of an ancient duplication of the entire genome (*39*), duplicated genes do not seem to make a disproportionate contribution to buffering (*38*).

Kirschner and Gerhart have discussed a variety of molecular and cellular mechanisms that permit the accumulation of genetic variation by reducing the potential of mutations to be lethal (*40*). These mechanisms act by reducing the number, specificity, and interdependence of molecular and cellular events. Examples include flexible, versatile proteins with broad substrate specificity, weak linkage of protein interactions in signaling pathways, transcriptional regulatory pathways, molecular systems that generate a variety of states from which the optimal state is chosen, and compartmentation that reduces the interdependence of events. Kirschner and Gerhart point out that these mechanisms have been selected to provide robustness and versatility of molecular and cellular processes in the face of stochastic and environmental variability, with the consequence that such processes would be buffered from the lethal effects of mutation. They suggest that evolvability is a by-product of flexible, robust processes that facilitate phenotypic variation. Evolutionary plasticity is evident for genetic modules controlling development (*41*).

Buffering can also result from the distributed properties of systems. Negative feedback regulation is ubiquitous in biology and acts to optimize the activity of a circuit in the presence of alleles with altered activities. For example, large alterations in the amounts of proteins that comprise the signaling system for bacterial chemotaxis have negligible effect on the ability of the system to exhibit adaptation following perturbation (*42*), a property that has been attributed to the system's feedback regulation (*43*). Certain types of metabolic pathways are relatively insensitive to changes in enzyme concentration. If all of the enzymes of a pathway follow Michaelis-Menton kinetics and none are saturated for their substrate, then changes in pathway flux resulting from reduction of a single enzyme is distributed throughout the pathway (*44, 45*).

The fact that mechanisms exist, as described above, to diminish the phenotypic consequences of mutant alleles has important implications. First, identifying the relationships that allow one gene to buffer the consequences of mutation in another reveals gene redundancy, compensatory pathways, and robust properties of molecular circuits. Second, the genes that buffer disease-producing genes are strong candidates for disease gene modifiers that influence the penetrance of a disease-causing allele (*27*). Finally, buffering relationships provide a force to maintain genetic variation in an unexpressed state in some genotypes, but allow it to be expressed in other genotypes. Mutations that reduce buffering could, therefore, be important in evolution by fostering phenotypic expression of previously unexpressed alleles. For example, Rutherford and Lindquist have demonstrated that mutant alleles of the chaperonin, HSP90, can reveal cryptic alleles in a *Drosophila* population that produce developmental abnormalities (*25*).

References and Notes

1. M. Cargill *et al.*, *Nature Genet.* **22**, 231 (1999).
2. D. G. Wang *et al.*, *Science* **280**, 1077 (1998).
3. M. Kimura, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 5969 (1991).
4. K. Dipple, E. McCabe, *Am. J. Hum. Genet.* **66**, 1729 (2000).
5. K. M. Weiss, *Genetic Variation and Human Disease* (Cambridge Univ. Press, New York, 1993).
6. J. Zielenski *et al.*, *Nature Genet.* **22**, 128 (1999).
7. J. Zielenski, L. Tsui, *Annu. Rev. Genet.* **29**, 777 (1995).
8. M. Steinberg, *Hemoglobin* **20**, 1 (1996).
9. E. Lander, N. Schork, *Science* **265**, 2037 (1994).
10. J. Nadeau, W. Frankel, *Nature Genet.* **25**, 381 (2000).
11. J. Terwilliger, K. Weiss, *Curr. Opin. Biotechnol.* **9**, 578 (1998).
12. A. Imasheva, D. Bosenko, O. Bubli, *Heredity* **82**, 187 (1999).
13. K. Gartner, *Lab. Anim.* **24**, 71 (1990).
14. H. McAdams, A. Arkin, *Trends Genet.* **15**, 65 (1999).
15. S. Antonarakis, V. McKusick, *Nature Genet.* **25**, 11 (2000).
16. L. Kruglyak, *Nature Genet.* **22**, 139 (1999).
17. W. Frankel, N. Schork, *Nature Genet.* **14**, 371 (1996).
18. R. Fijneman *et al.*, *Nature Genet.* **14**, 465 (1996).
19. T. van Wezel *et al.*, *Nature Genet.* **14**, 468 (1996).
20. G. Gibson, M. Wemple, S. van Helden, *Genetics* **151**, 1081 (1999).
21. D. Threadgill *et al.*, *Science* **269**, 230 (1995).
22. C. H. Waddington, *Nature* **150**, 563 (1942).
23. T. Dobzhansky, B. Spassky, *Genetics* **29**, 270 (1944).
24. R. Dun, A. Fraser, *Nature* **181**, 1018 (1958).
25. S. Rutherford, S. Lindquist, *Nature* **396**, 336 (1998).
26. J. Phelan, S. Austad, *J. Gerontol.* **49**, B1 (1994).
27. R. Thornhill, A. Moller, *Biol. Rev.* **72**, 497 (1997).
28. S. Rutherford, *BioEssays* **22**, 1095 (2000).
29. C. Kaiser, R. Gimeno, D. Shaywitz, in *The Molecular and Cellular Biology of the Yeast* Saccharomyces, J. Pringle, J. Broach, E. Jones, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1997), pp. 91–227.
30. I. Rebay *et al.*, *Genetics* **154**, 695 (2000).
31. P. Sternberg, M. Han, *Trends Genet.* **14**, 466 (1998).
32. A. Long *et al.*, *Genetics* **139**, 1273 (1995).
33. E. Friedberg, G. Walker, W. Siede, *DNA Repair and Mutagenesis* (ASM Press, Washington, DC, 1995).
34. M. C. Costanzo *et al.*, *The Yeast Proteome Database (YPD) and* Caenorhabditis elegans *Proteome Database (WormPD): Comprehensive Resources for the Organization and Comparison of Model Organism Protein Information* [*Nucleic Acids Res.* **28**, 73 (2000) and www.proteome.com/databases/index.html]
35. A. Wilkins, *BioEssays* **19**, 257 (1997).
36. D. Tautz, *BioEssays* **14**, 263 (1992).
37. B. Guo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 12158 (2000).
38. A. Wagner, *Nature Genet.* **24**, 355 (2000).
39. K. Wolfe, D. Shields, *Nature* **387**, 708 (1997).
40. M. Kirschner, J. Gerhart, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8420 (1998).
41. A. Minelli, *Mol. Phylogenet. Evol.* **9**, 340 (1998).
42. U. Alon *et al.*, *Nature* **397**, 168 (1999).
43. T. M. Yi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 4649 (2000).
44. H. Kacser, J. Burns, *Genetics* **97**, 639 (1981).
45. D. Dykhuizen, A. Dean, D. Hartl, *Genetics* **115**, 25 (1987).
46. J. D. York *et al.*, *Science* **285**, 96 (1999).
47. D. Holtzman, S. Yang, D. Drubin, *J. Cell Biol.* **122**, 635 (1993).
48. C. Hardy, *Mol. Cell. Biol.* **16**, 1832 (1996).
49. J. Aitchison *et al.*, *J. Cell Biol.* **131**, 1133 (1995).
50. J. Pringle, A. Bender, *Mol. Cell. Biol.* **11**, 1295 (1991).

51. We thank A. Emili, P. Paddison, and J. Appelbaum for discussions. The authors were supported in part by grant GM17709 from the National Institutes of Health, and J.L.H. IV is supported by a fellowship from the Howard Hughes Medical Institute.