Many molecules that control genetic regulatory circuits act at extremely low intracellular concentrations. Resultant fluctuations (noise) in reaction rates cause large random variation in rates of development, morphology and the instantaneous concentration of each molecular species in each cell. To achieve regulatory reliability in spite of this noise, cells use redundancy in genes as well as redundancy and extensive feedback loops and other features to produce the phenotypic outcomes. To achieve the switching and randomize phenotypic outcomes. To achieve the switching and randomize phenotypic outcomes.

Ericson, J. et al. (1998) Integrated FGF and BMP signaling controls the progression of progenitor cells in epithelial differentiation and the emergence of pattern in the embryonic amnion plating. Development 125, 3303-3311


This review focuses on the cell to cell variations in the concentrations of regulatory molecules that arise from internal cellular processes rather than from differing environments. These variations are commonly observed as irreducible cell to cell concentration differences in well-stirred cultures of single-cell organisms. (In tissue cultures, uniform extracellular environments are virtually impossible to achieve.) There are several internal sources of regulatory noise. For example, there are inevitable statistical variations in the random partitioning of small numbers of regulatory molecules between daughter cells when cells divide. Many regulatory molecules are present in bacterial cells at extremely low concentrations – anywhere from a few tens to a few hundred molecules per cell. Thus, in random partitioning of, say, 50 molecules between equal-sized daughters, 6% of the daughters will
TABLE 1. Competitive reactions controlling the expression of alternative genes

<table>
<thead>
<tr>
<th>Organism</th>
<th>Mechanism</th>
<th>Function</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli Pap system</td>
<td>Differential methylation of alternative Lrp</td>
<td>Phase variation in pili expression, affecting virulence</td>
<td>36</td>
</tr>
<tr>
<td>E. coli Fim system</td>
<td>Invertible DNA segments</td>
<td>Phase variation in type 1 pili, affecting virulence</td>
<td>37</td>
</tr>
<tr>
<td>Phage Mu</td>
<td>Invertible DNA segments</td>
<td>Phase variation in type 1 pili, affecting virulence</td>
<td>38</td>
</tr>
<tr>
<td>Salmonella typhimurium Hin system</td>
<td>Invertible DNA segments</td>
<td>Phase variation in pilin alters antigen response</td>
<td>39</td>
</tr>
<tr>
<td>Moraxella bovis</td>
<td>Invertible DNA segments</td>
<td>Phase variation in pilin alters antigen response</td>
<td>39</td>
</tr>
</tbody>
</table>

Reviews

Genetic regulatory circuitry

A less obvious but more important cause of cellular variation is the distinctive statistical properties of regulatory chemical reactions that involve a small number of reaction centers and slow reaction rates. Regulation of bacterial gene transcription, for example, predominantly involves reactions of small intracellular populations of one to three regulatory species. These bind to the promoter region of a given gene, and there are generally two or less copies of the gene in a growing bacterial cell. Even at fully activated bacterial promoters, the average time between transcript initiations at each promoter can be many seconds and the distribution of intertranscript times is highly skewed around the average. Detailed consideration of the statistical properties of transcript initiation and translation suggests that proteins are ultimately produced from an activated promoter as short bursts of variable numbers of proteins, and that the bursts occur at random time intervals, both in bacteria and eukaryotic cells. Stochastic gene expression has been observed directly in eukaryotic cells.

Protein production from eukaryotic genes is erratic and bursty as in prokaryotes, but with longer average intervals between bursts.

Stochastic outcomes at regulatory switch points

A particularly interesting case occurs when two independently produced regulatory proteins are involved in the competitive control of a developmental switch that selects between alternative pathways. Because the independent, stochastic temporal patterns of production of each regulatory protein can vary widely from cell to cell, the pathway selection by the competitively regulated switch can be random. The probabilities of selecting each pathway will depend on the stochastic properties of the gene expression mechanisms and the design of the switch circuit. Cells can take advantage of stochastic expression of the regulatory proteins to randomize the regulatory outcome — the pathway choice — using appropriately designed regulatory circuits. The simple cross-repressive configuration in Fig. 1 illustrates how this phenomenon can produce subpopulations expressing alternative phenotypes, even in genetically homogeneous populations in identical environments. In the example shown, competitive autoregulating feedback loops can lock the cell into one or another pathway with some fraction of the cells, by chance, taking each path. In such systems, environmental signals can act on the parameters of the regulatory circuit to bias the probabilities of path choice under different conditions. Organisms exploit this mechanism to achieve diversity and increase the likelihood of species survival over a wide range of environments. Two examples of such a mechanism are the phage λ lyso-lysogeny decision circuit (see below) and the networks controlling Bacillus subtilis commitment to competence and sporulation.

Another stochastic bistable genetic regulatory mechanism is the random inversion of DNA segments used in many organisms to produce subpopulations of distinct phenotypes. Table 1 shows a small sample of well-known cases of bistable regulatory mechanisms used in genetic circuits that produce stochastic phenotype outcomes. These stochastic bistable switching mechanisms are common virulence mechanisms in pathogenic organisms. For example, random alteration of proteins on the bacterial surface or in external features such as flagella can aid in avoidance of the host’s immune response.

Common features of the dynamics of these stochastic regulatory switches that randomly select among several alternative pathways include: (1) transient, low-level expression of key regulatory proteins; (2) stochastic progress toward pathway commitment as concentrations of the controlling proteins in each cell change from moment to moment, so that there is a transient period of partial (i.e. reversible) commitment before a definitive choice eventually emerges; and (3) multiple feedback loops that reinforce the activation of the selected path and repression of rejected alternative(s).

Kinetics of stochastic regulatory circuits

Conventional kinetics does not model statistics of regulatory systems that produce probabilistic outcomes, such as the
stochastic switching mechanisms discussed above, and might not even describe the average behavior of such systems correctly\(^\text{15,16}\). For these cases, a stochastic kinetic analysis\(^\text{14}\) can be used to predict the behavior of systems that are probabilistically regulated and might permit improved exploitation of information in the statistics of phenotypic outcomes. Analytical resolution of the resulting systems of stochastic reaction equations is only practical for simple reaction systems. The so-called ‘Langevin approach’ for approximation of the effect of fluctuations has been used to model the microscopic kinetics of stochastic regulatory systems, but this practice is theoretically unsound\(^\text{15}\) and can yield invalid predictions for bistable systems\(^\text{2,16}\). However, the Monte Carlo simulation algorithm described by Gillespie\(^\text{14}\) does provide valid numerical solutions for complex systems of coupled stochastic reactions. Stochastic switching in the phage λ lysis–lysogeny decision circuit (Fig. 2) has been analyzed using stochastic kinetics and the Gillespie algorithm\(^\text{14}\) to show in detail how initially homogeneous cell populations can partition randomly into distinct phenotypic subpopulations\(^\text{2}\). Host-cell hunger and higher numbers of phage particles infecting the cell bias the decision circuit to produce a higher percentage of lysogens.

**How do cells achieve regulatory determinism?**

In spite of the randomness in basic regulatory mechanisms discussed above, many regulatory pathways in cells have highly predictable outcomes. The strategies that cells use to ensure that critical proteins are expressed when needed, in spite of infrequent and stochastic gene expression, include: (1) population transcriptional cooperation (i.e. it is not necessary for every cell in a population to make all the gene products\(^\text{2}\)); (2) checkpoints to assure that cascaded events are adequately synchronized\(^\text{2}\) and (3) widespread redundancy in genes\(^\text{20–22}\) and in regulatory pathways\(^\text{23,24}\).

Even in uniform conditions, normal fluctuations in protein production can be large, relative to the regulatory thresholds that control the expression of downstream genes. One consequence is wide variations from cell to cell in the ‘switching time’ for the controlling protein to activate the genes it controls. Without a coordinating mechanism, these timing variations will cause errors in synchronization of cellular functions when complex networked signal pathways control the sequencing of cellular functions. One mechanism to provide coordination is provided by regulatory checkpoints that halt regulatory cascades until conditions for further progress (e.g. availability of essential nutrients, external environmental signals or completion of precursor cellular events) are satisfied\(^\text{1,15}\). Checkpoints assure the orderly execution of cellular activities, but the time required to execute cascaded functions can still vary widely between cells. Thus, checkpoint yield certainty in outcome, but not certainty in the timing of regulatory events. A common example is the random distribution of generation times of cells in growing cell cultures that causes progressive desynchronization of initially synchronized cell populations\(^\text{25,26}\).

Development of large metaorganisms from egg to adult requires the highly reliable execution of very large numbers of developmental processes, with correct timing, sequencing and spatial positioning. The regulatory processes controlling the development must act predictably, in spite of large fluctuations in the function of elemental regulatory mechanisms and fluctuations in environmental conditions. The reliability requirement for individual somatic developmental processes depends on the function’s criticality for production of a successful adult. For example, regulatory processes early in embryonic development that are prerequisites for extensive downstream cell lineages, and processes whose failure might allow dangerously uncontrolled cellular proliferation, have to be particularly reliable\(^\text{27}\). Thus, regulatory circuit designs and the molecular details that determine kinetic parameters must be under selective pressure for reliable and robust operation (including robustness to large variations in the organism’s normal external environment).

Several complementary strategies can be combined to construct a reliable regulatory system from noisy biochemical elements and inherently mutable genes. Dynamic stability in the regulatory circuit designs results principally from the exploitation of redundancy and feedback\(^\text{28,29}\). Redundancy is applied both at the level of individual components (i.e. genes\(^\text{30}\)) and through parallelism and interlinking in the control pathways, so that regulatory networks are more reliable than their parts\(^\text{28}\). A common genetic criterion for functional redundancy between two genes is that single gene mutations have little phenotypic effect while mutation of all paralogues produces a strong effect\(^\text{13,28}\). This test does not imply that redundant genes are necessarily genetic duplicates. Indeed, genes with redundant or overlapping function but unrelated sequences are well known\(^\text{31}\).
Reliability through redundancy

In the 1940s and 1950s, the notion that there are genetically specified, self-stabilizing capabilities in developing organisms was recognized and characterized as ‘canalization’.

In the 1950s, redundant genes were suggested as a possible mechanism for this genetic capacity to buffer development pathways against mutational or environmental perturbations.

Recently, analysis of the connection between genetic redundancy and reliability has focussed largely on conditions for the evolutionary stability of redundant genes, with general agreement that the redundancy provides some sort of ‘back up’ for somatic development functions.

The dramatic increases in chromosome size at the phylogenetic transition to vertebrates and the interrelated and interlinked evolutionary boundaries are attributed to chromosome duplications that provided the opportunity to create the redundant genetic network designs necessary for reliable regulatory operation of pathways against mutational or environmental perturbations.

A highly simplified reliability analysis shows that even simple redundant configurations provide high payoff to regulatory link reliability. (a) A simple, genetically coupled link where effector B (controlled by A) controls the downstream gene c. Assume the statistics of operation of the link are such that a single gene is capable of producing an effective signal (i.e., the link is operational) with probability $P = 0.90$. (b) A link of two genes A and B in series is only operational if both genes are operational, so the reliability is $0.90^2 = 0.81$. (c) A parallel configuration will only fail if both genes fail and so has reliability $0.19$. (d) With the same statistics, two redundant, homoeologous duplicated genes are operational if one out of the four genes is operational, so the reliability is $0.99^4 = 0.9999$. (e) The redundant configuration also decreases sensitivity to individual mutations. Here, two mutations producing a configuration AaBb in series with configuration RbRb still has a reliability of $0.998$. (f) The duplication can provide a fitness advantage owing to increased regulatory reliability and, hence, be preserved.

Addition of other independent and parallel regulatory pathways involving different gene products can further increase the reliability of a regulatory network’s performance.

Figure 3 illustrates the dramatic increases in overall link reliability from parallel redundancy. Although the simple examples in Fig. 3 demonstrate the benefits of redundancy, quantitative analysis of particular genetic networks is complex because the statistical characteristics of gene expression are determined by the stochastic properties of the molecular mechanisms controlling the expression of each gene. Regardless of the statistical details, however, networks with redundant elements will perform more predictably and with less variance in outcome than non-redundant networks of the same elements owing to the statistical independence of variations in the different chemical reactions.

In the small number of regulatory networks where molecular mechanisms are relatively well known, the statistics of operation of each mechanism can be estimated and the effects of redundant components on the robustness of genetic network performance can be analyzed. Given the complexity of even small genetic networks, the use of simulation techniques is necessary for such an analysis.

Redundancy affords resilience in genetic network performance both to gene mutations and to the short-term, transient regulatory failures caused by erratic protein production. Figure 3(e) illustrates how mutations in one or more of the genes in a redundant regulatory network can increase the probability of failure of the network. This phenomenon results from the statistical character of the robustness afforded by redundancy against ‘outages’ in a regulatory link: the $n$-fold redundancy will always provide less reliability than $n$-fold redundancy. For low levels of redundancy, mutations can increase the probability of failure of the regulatory network significantly during development, so that a fraction of the population exhibits a mutant phenotype; that is, there is partial penetrance of the mutant phenotype.
The dynamic stability of genetic networks arises in part from feedback loops that are formed by the regulatory interlocking of components that are based on negative and positive feedback loops. The stability of the network depends on the presence of both types of feedback loops and the degree of interconnection between them. In some cases, the network may exhibit bistability, where the system can exist in one of two stable states, depending on the initial conditions.

References