Messy biology and the origins of evolutionary innovations

Dan S Tawfik

Biological messiness relates to infidelity, heterogeneity, stochastic noise and variation—both genetic and phenotypic—at all levels, from single proteins to organisms. Messiness comes from the complexity and evolutionary history of biological systems and from the high cost of accuracy. For better or for worse, messiness is inherent to biology. It also provides the raw material for physiological and evolutionary adaptations to new challenges.

Explaining why something happens in a given way and not in any other way is an intellectual tendency that dominates human thought in general and the biological sciences in particular. In this vein, biological systems, be they individual proteins or whole organisms, are studied with the aim of showing how their structure, function and evolution are purposely directed and optimally designed. However, although the hallmarks of optimal functionality can be seen throughout biology, there are also clear-cut indications that biology is messy and, at times, even sloppy. In other words, the accuracy and fidelity of biological systems are limited.

The term 'messiness' has negative connotations, and indeed, biological systems make errors that can be manifested, for example, in diseases. However, there is a 'method to the madness', and there are systematic trends in biological messiness. I discuss here some general trends of messiness and their evolutionary manifestations. The wealth of phenomena addressed here fall into three categories: genetic, phenotypic and environmental (**Box 1**, **Table 1**). A key general feature of messiness is that genetic, phenotypic and environmental variances tend to correlate. Thus, as far as generalizations apply to biology, at all levels—from single molecules to pathways and whole cells—infidelity, inaccuracy, heterogeneity and noise compose the most accessible source of physiological as well as evolutionary adaptations. Finally, I discuss the origins of messiness and argue that although evolution capitalizes on messy traits, it does not explicitly promote messiness.

Biology is messy

Natural systems have not been optimized in the engineering sense. They have evolved

Box 1 | Definitions: varying and noisy traits

Genetic variance: Mutations, or variations in the genetic content of individuals within populations, including single-nucleotide polymorphisms (SNPs), insertions, deletions and variations in gene copy numbers. Mutations may fix owing to a selective advantage (due to a necessity (positive selection)) but may also fix stochastically, by chance (drift). The latter, although often ignored, plays a key role in evolution[1] and is a primary source of stochasticity by survival of those who are reasonably fit and particularly lucky.

Phenotypic noise and variability: Heterogeneity, or stochastic variation, in phenotypes (molecular, cellular and/or organismal traits) within the same genotype and under a single given condition or environment. Examples are shown in **Table 1**.

Environmental variance: Changes in phenotypes within the same genotype but under different environments—for example, variations in gene expression and metabolic content under changing environments.

Table 1 | Examples for phenotypic noise and variability

Component	Variable trait	Variance or noise	Exemplifying references
Individual proteins	Primary sequence	Transcriptional and/or translational errors (phenotypic mutations); read-through of stop codons and frameshifts	4,26,35
	Structure	Alternative protein conformations	18
	Function	Promiscuous protein functions	18
	Expression levels	Variable expression levels due to low mRNA abundance or stability, variability in epigenetics, etc.	8,15,16
Metabolic pathways and networks	Metabolites	Concentrations of individual metabolites	20
		Promiscuously synthesized metabolites; turning on of alternative pathways	20,21
Signaling, regulatory networks	'On'/'off' switching of a given cellular trait	Composition and concentration, sequence and conformation of the network's mediating proteins	7,11,12,24,27
Developmental networks	Triggering of a given developmental route	As above	9

stepwise, tier by tier, by diversifying existing functions and components. Many of the components have evolved to function in a particular way not only out of necessity but also by chance. Indeed, chance is a key feature of evolution, resulting in the haphazard dominance of certain traits, molecules and organisms. The role of stochastic events (survival of the luckiest) is sometimes hard to distinguish from the role of optimization (survival of the fittest), especially with traits that provide modest advantages within small populations[1]. Biological components, be they molecular (for example, enzymes), cellular or organismal, have remarkable properties, but they are far from textbook perfection. The specificity of biomolecules is inherently limited by interactions that are transient and mediated by weak forces. In cases in which high specificity is absolutely essential, mechanisms such as proof editing have evolved, although these are energetically costly and reduce processivity. Wherever the cost of inaccuracy proves bearable, biological systems produce marked heterogeneity. Further, although individual components and processes can be remarkably accurate and specific, combinations of multiple components and processes result in considerable infidelity, noise and heterogeneity. Take, as an illustration, a one-pot, ten-step chemical reaction in which each step proceeds with 99% product purity—impressive, by any measure. However, although the desired final product is obtained with very high yield (0.99^{10} = 90%), the remaining 10% of the products are a variable and heterogeneous mixture of byproducts. Similarly, all biological systems show messiness in the sense that certain of their traits are variable or fluctuating. In addition, if one of the steps involves a small number of molecules, the outcome of the process may vary from one run to another, resulting in stochastic variations or noise.

Variations and noise in expression levels are the most widely studied aspect of messiness[2]. However, wide variations, heterogeneity and noise are characteristic of all biological components and properties, as described in other contributions in this issue. Take, for example, a single gene encoding an enzyme. Within a given population, this gene may have various polymorphs (genetic mutations). Further, drift may result in the coincidental fixation of mutations that are neutral, or even partly deleterious, thus introducing a dominating stochastic element into genetic variability. In addition, errors in protein synthesis (phenotypic mutations) give rise to a range of enzyme molecules that deviate from those encoded by the original gene sequence. For a 350-residue enzyme (an average prokaryotic protein), one out of ten enzyme molecules will carry a phenotypic mutation[3]. Variations in length owing to read-through of stop codons and frameshifts are even more frequent[4]. As both the enzyme's sequence and expression levels vary, both the concentration and the composition of the metabolite(s) this enzyme uses and/or produces may vary. Variations in composition also result from byproducts produced by promiscuous activities of enzymes and by their genetic and phenotypic mutants. Variations in the metabolite(s) used and produced by an enzyme will also affect the down- and upstream enzymes, such that overall, significant compositional variations may be observed between one cell and another.

These stochastic variations in the concentration of individual proteins, metabolites and signaling molecules can result in different subpopulations of cells, each showing a different behavior (phenotype) with respect to a given trait[5–9]. Indeed, single-cell analyses reveal how widely distributed certain traits are (for example, the concentration of a given protein or metabolite)[5]. In some cases, the distribution of single-cell values has been found to be bimodal, and thus the average value per population does not even represent the values of most individual cells[5]. This phenomenon is most conspicuous in cases of phenotypic switches that are triggered by signaling molecules present at low copy numbers[10]. The lactose operon of *E. coli*, for example, is controlled by the LacI transcriptional repressor, which is present at ~10 molecules per cell. At low induction level, a bimodal distribution of cells can be observed in which expression of the *Lac* operon is either turned off or on[11]. A similar phenomenon is observed in antibiotic persistence: fluctuations in the concentration of an endogenous toxin give rise to the coexistence of normally growing cells alongside a small subpopulation of dormant, drug-resistant cells in which toxin levels are above a certain threshold[12].

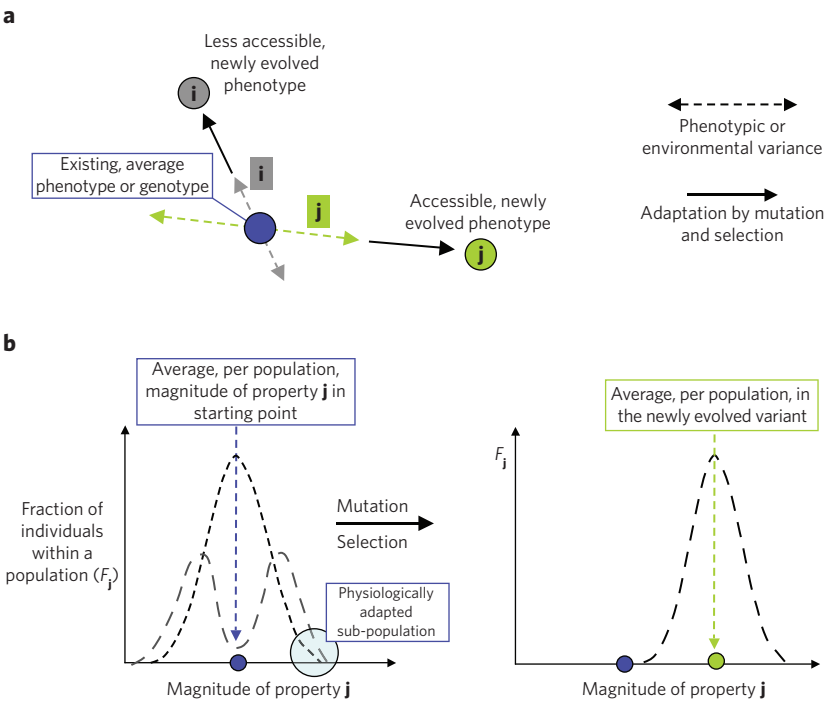

Figure 1 | The correlation of phenotypic and genetic variances and the origins of evolutionary innovations. (**a**) Biological components (for example, proteins) and systems (for example, regulatory networks) show phenotypic variations, or noise, along different traits, depicted here as vectors **i** and **j**. Genetic variance (mutations) shift the component and/or system along the same directions as the most pronounced phenotypic variances. Once a mutation is fixed, a newly evolved variant emerges with the advantageous trait. Thus, the space of possible new adaptations is described by the relative magnitudes of variances along various possible directions (**j** represents the most accessible adaptation in the depicted scheme). (**b**) An alternative representation describing one trait, **j**. Phenotypic variance results in a distribution of different values of **j** within single molecules or single cells—the distribution can be normal, or even bimodal, such that the average value does not represent the majority of single molecules or cells. If high values of **j** provide a selective advantage under certain circumstances, mutations that shift the distribution to higher values will be fixed, resulting in a newly evolved variant with the advantageous trait.

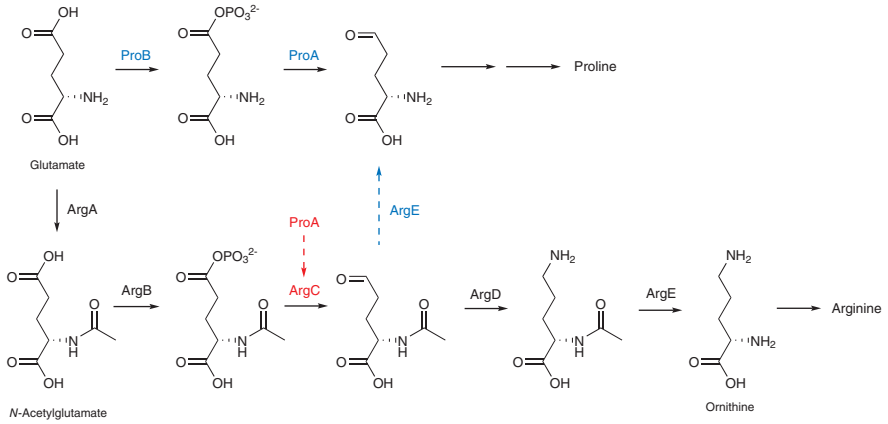

Figure 2 | Promiscuous enzyme functions (protein messiness) mediate physiological and evolutionary adaptations in the arginine and proline biosynthetic pathways. Perturbations in the proline biosynthetic pathway by deletion of *ProB* and *ProA* are bypassed owing to the promiscuous action of ArgE on the substrate of ArgD (annotated in blue). Similarly, an *ArgC* knockout strain is rescued by a mutation that increases the promiscuous activity of ProA toward the generation of the product of ArgC (annotated in red).

Messiness: the central hypotheses

Two generalizations or hypotheses have been made with respect to messiness and its role in the evolution of new traits (**Fig. 1**). (i) Phenotypic, environmental and genetic variances are often correlated, in both their direction and magnitude. (ii) The phenotypic and environmental variances indicate, and possibly dictate, the possible routes that evolution can take. Thus, for a given component or system, genetic variance (mutations) induces effects similar to those of phenotypic variance, and the latter thereby dictates the space of solutions available for evolutionary innovations. The origins of these ideas date back to Waddington (in the 1950s) and even earlier (for earlier references, see refs. 13,14). Waddington's ideas and the general hypotheses described here stem from the notion that environmental perturbations and mutations are all likely to affect organisms in similar ways and that the very same phenotypic effects can thus be seen either in one specific environment, or in a small fraction of molecules or cells.

Although the above hypotheses have been only sporadically explored, they are supported by several observations and theoretical analyses. Analysis of the yeast genome suggests that genes that confer robustness to environmental and/or stochastic fluctuations also buffer the effects of genetic changes[13]. Measurements of fluctuations in gene expression indicate that genes that stochastically change expression levels are more responsive to environmental perturbations and also vary in response to randomly accumulating mutations[15]. A similar overlap is seen between environmentally perturbed genes and genes that vary in their expression in closely related yeast species[16].

A theoretical framework has also been proposed that borrows from the fluctuation-dissipation theorem and indicates that the larger the amplitude of a phenotypically fluctuating trait, the more responsive it is to mutations[17]. However, it is not just the magnitude of the response that is correlated but also its directionality. That is, a particular phenotypic variance that mediates survival under a given environment (physiological adaptation) is also most likely to respond to mutations, and eventually, mutations that alter the very same variance will fix to produce the newly adapted variant (**Fig. 1**). Take, for example, the case of antibiotic persistence. Although all cells are genetically identical, an antibiotic-resistant subpopulation of cells has toxin levels above a given threshold. Accordingly, a selected mutant strain that has a thousand-fold higher frequency of resistant cells shows two-fold lower repression levels and higher free toxin levels, thus resulting in higher probability of persistence[12]. A similar mechanism has been hypothesized for the evolution of regulatory reproduction mechanisms in social insects—an initial stochastic decision that originates from phenotypic variability could eventually be fixed through mutations[9].

Protein noise and the evolution of new protein functions

An example taken from my own research concerns 'protein messiness'—coincidental, promiscuous protein functions and alternative protein conformations coexisting alongside the native function and conformation. Functional promiscuity, multispecificity and structural plasticity are inherent to proteins[18]. Mutations that accumulate as neutral with respect to the native function may have large effects on the promiscuous functions. As promiscuous activities are mostly latent (meaning they have no physiological role or effects), they appear and disappear at random[19]. Alternative conformations are also likely to vary in response to mutations, with little effect on the native conformation. Although this has not been directly demonstrated, promiscuous functions and alternative conformations are also likely to be more responsive to environmental perturbations (pH, salt, changes in metabolite concentrations, and so on) than the native function or conformation. Finally, when new protein functions are needed, promiscuous functions, as well as the alternative conformations mediating them, provide the most accessible solution. This can be seen in physiological adaptations (without any genetic changes) as well as in evolutionary adaptations. For example, the biosynthetic pathways of arginine and proline involve analogous reactions on different substrates (**Fig. 2**). Strains with deletions of the first two enzymes of proline synthesis (ProB, ProA) are viable because ArgE, which normally deacylates *N*-acetylornithine, promiscuously deacetylates *N*-acetylglutamate semialdehyde and thereby produces the product of ProAB[20]. Thus, physiological adaptation is mediated through the infidelity of an enzyme. The same principle governs the evolutionary adaptation of an *ArgC* (N-acetylglutamylphosphate reductase) knockout strain via a mutation that increased a latent, promiscuous activity of ProA (glutamylphosphate reductase) that leads to the product of ArgC[21]. Thus, at the single-protein level, phenotypic heterogeneity and genetic variability are often correlated, and evolution capitalizes on existing, latent routes that are initially manifested as phenotypic variability.

Robustness and evolvability

The hypothesis that phenotypic, environmental and genetic variances are correlated in direction as well as magnitude accounts for the robustness (genetic and environmental) of biological systems and also for their adaptability (or evolvability). As suggested by Waddington (see refs. 13,14,22 for original references), the robustness of phenotypes toward mutations (genetic robustness,

or canalization) evolved in response to the need to cope with environmental and stochastic perturbations. Native functions or traits that play a permanent physiological role and have therefore been constantly under selection have become robust with respect to all three variables—they show smaller phenotypic fluctuations and lower environmental variability and are also not easily perturbed by mutations[13,22]. In contrast, latent, promiscuous or coincidental traits are easily perturbed, stochastically, environmentally and genetically, and thereby provide the raw material for evolutionary novelty.

The origins of messiness

Finally, I wish to make a point regarding the origins of biological messiness and argue that although evolution exploits the messiness of biology, it does not explicitly promote it. First, messiness is an inevitable, rather than a desirable, feature of biology. There are numerous evidences for evolution of higher fidelity and minimal noise, including the proof editing of DNA replication and of protein translation. Enzymes have evolved toward high specificity, in particular with respect to alternative substrates that lead to undesirable products[18]. Gene expression is regulated by elaborate mechanisms that generally provide adequate control, and stochastic variations seem to have been evolutionary minimized[23]. Gene order within bacterial operons may have evolved to minimize stochastic variations in protein levels[24]. Evolution therefore acts to minimize noise and infidelity. However, there exist chemophysical constraints on specificity, and accuracy and fidelity are costly, thus resulting in tradeoffs between the benefits of accuracy and fidelity and their costs.

The notion that certain elements of messiness may have evolved *per se* is often prompted by the observation that under stress, noise increases and fidelity is reduced. This is seen at many levels, from genetic mutation rates[25] to translational errors[4,26] and gene regulation[15,16,27]. Under stress, populations become increasingly heterogeneous. Adaptation (physiological and evolutionary) may therefore be promoted by the increased likelihood of having a small subpopulation that can cope with the stress owing to either a phenotypic or a genetic variation. However, are stress-induced noise and infidelity traits purposely evolved with the aim of bet-hedging under extreme conditions[22,25,27,28]? Or are they a mere byproduct of stress that is opportunistically exploited? This issue is as yet unresolved. I surmise that the latter is the more common scenario. Only in rare cases, such as with RNA viruses, have high error rates evolved explicitly, and in this case increased fidelity has been shown to be disadvantageous[29]. Another example is the prion state [PSI+] in yeast under stress, which increases the rate of stop codon read-through, leading to extended proteins (see refs. 22,26 and references therein). However, by and large, messiness and heterogeneity are neither a desired nor a deliberate outcome of evolution. Heterogeneity and diversity compose the very basis of evolution, not only within genetically diverse populations but also within the same allele or genome. Thus, messiness is an inevitable feature of biological components and processes. Enzymes, for example, have not evolved to be promiscuous or conformationally flexible—it is simply that proteins are flexible polymers and that absolute specificity is unattainable.

Messiness is also a byproduct of the multifunctionality of biological systems. Multifunctionality is seen at all levels, from single proteins (many of which perform multiple, often very different functions[18]) to regulatory networks—the latter, for example, appear to contain more widely connected and more interconnected components than expected[30]. Multifunctionality results in an increased capacity to cope with diverse challenges[31], including ones that have never been encountered before[7,30]. However, multifunctionality also increases complexity and messiness, as is apparent in the difficulty of associating changes in phenotypes with changes in genes that are known to mediate these phenotypes[31].

The origins of messiness also lie in the opportunistic, tinkering nature of evolution. Gould's 'panda's thumb'[32] and Jacob's 'nature as a tinkerer'[33] have become icons of this fundamental principle that is observable at all levels, from the evolution of enzyme functions[18,34] to the emergence of new developmental patterns[9]. Biological systems, as Darwin put it, are machines made of "old

Box 2 | Some open questions

- The hypotheses described in the text are supported by various findings but have not been systematically explored, even at the single-protein level. Can, for example, alternative protein forms generated by synthesis errors (by read-through of stop codons under stress[4] or under prion states in yeast[26]) provide a physiological advantage? Do these alternative lengths become genetically fixed in later generations via mutations leading to a new stop codon? Or are promiscuous or moonlighting functions triggered upon certain environmental and/or genetic changes?

- What are the underlining features of exceptions to the above hypotheses? For example, physiological adaptations of hemoglobin occur primarily via changes in the half-saturation level for O_2, whereas evolutionary adaptations seem mostly to alter the cooperativity coefficient[14]. These parameters are mechanistically interlinked, but it remains unclear how the parameters' responses to physiological changes (such as in pH or O_2 pressure) compare to their responses to mutations.

- Which of the oft-observed stress-induced noises and infidelities are traits purposely evolved for enabling bet-hedging under extreme conditions and which ones are inevitable byproducts of stress that are capitalized upon?

- Stochastic phenotypic switches owing to low copy numbers of a regulatory protein are common (for example, LacI (ref. 11) or HipA toxin[12]), and stabilized switches have been shown to induce different physiological and environmental responses (for example, see ref. 36). How would other systems behave under conditions of increased fidelity (for example, see ref. 29) or reduced noise? What are, for example, the physiological and evolutionary implications of higher translational fidelity, increased enzyme specificity and lower variability in the levels of metabolic enzymes?

- The idea that phenotypic and environmental variances pave the road for evolutionary adaptations is attractive. However, the time gap between physiological adaptations and evolutionary ones is wide (same generation versus dozens or even thousands of generations). How this gap is bridged remains unclear. One may speculate, however, that for microorganisms, the transfer of biochemical content (such as proteins or metabolites) may mediate 'inheritance' of physiologically adaptive states[11], whereas in multicellular organisms, epigenetics could play a bridging role.

wheels, springs and pulleys, only slightly altered"[32]. Thus, behind the façade of perfection and optimality lies messy biology that originates from evolution and provides the basis for the evolution of all living forms. Further research in this area will reveal more examples for the principles outlined here and will also shed light on the mechanisms that govern the evolutionary exploitation of messiness (**Box 2**).

Dan S. Tawfik is in the Department of Biological Chemistry, Weizmann Institute of Science, Rehovot, Israel.
e-mail: tawfik@weizmann.ac.il

References

1. Lynch, M. *Nat. Rev. Genet.* **8**, 803–813 (2007).
2. Elowitz, M.B., Levine, A.J., Siggia, E.D. & Swain, P.S. *Science* **297**, 1183–1186 (2002).
3. Willensdorfer, M., Burger, R. & Nowak, M.A. *PLOS Comput. Biol.* **3**, e203 (2007).
4. Meyerovich, M., Mamou, G. & Ben-Yehuda, S. *Proc. Natl. Acad. Sci. USA* **107**, 11543–11548.
5. Wang, D. & Bodovitz, S. *Trends Biotechnol.* **28**, 281–290 (2010).
6. Avery, S.V. *Nat. Rev. Microbiol.* **4**, 577–587 (2006).
7. Stern, S., Dror, T., Stolovicki, E., Brenner, N. & Braun, E. *Mol. Syst. Biol.* **3**, 106 (2007).
8. Cohen, A.A. *et al. Science* **322**, 1511–1516 (2008).
9. Kilfoil, M.L., Lasko, P. & Abouheif, E. *HFSP J.* **3**, 379–385 (2009).
10. Raj, A. & van Oudenaarden, A. *Cell* **135**, 216–226 (2008).
11. Robert, L. *et al. Mol. Syst. Biol.* **6**, 357–368 (2010).
12. Rotem, E. *et al. Proc. Natl. Acad. Sci. USA* **107**, 12541–12546.
13. Lehner, B. *PLoS ONE* **5**, e9035 (2010).
14. Milo, R., Hou, J.H., Springer, M., Brenner, M.P. & Kirschner, M.W. *Proc. Natl. Acad. Sci. USA* **104**, 16998–17003 (2007).
15. Landry, C.R., Lemos, B., Rifkin, S.A., Dickinson, W.J. & Hartl, D.L. *Science* **317**, 118–121 (2007).
16. Tirosh, I., Weinberger, A., Carmi, M. & Barkai, N. *Nat. Genet.* **38**, 830–834 (2006).
17. Kaneko, K. & Furusawa, C. J. *Theor. Biol.* **240**, 78–86 (2006).
18. Khersonsky, O. & Tawfik, D.S. *Annu. Rev. Biochem.* **79**, 471–505 (2010).
19. Amitai, G., Gupta, R.D. & Tawfik, D.S. *HFSP J.* **1**, 67–78 (2007).
20. D'Ari, R. & Casadesus, J. *Bioessays* **20**, 181–186 (1998).
21. McLoughlin, S.Y. & Copley, S.D. *Proc. Natl. Acad. Sci. USA* **105**, 13497–13502 (2008).
22. Masel, J. & Trotter, M.V. *Trends Genet.* **26**, 406–414 (2010).
23. Lehner, B. *Mol. Syst. Biol.* 4, 170–175 (2008).
24. Kovács, K., Hurst, L.D. & Papp, B. *PLoS Biol.* **7**, e1000115 (2009).
25. Galhardo, R.S., Hastings, P.J. & Rosenberg, S.M. *Crit. Rev. Biochem. Mol. Biol.* **42**, 399–435 (2007).
26. Masel, J. & Bergman, *A. Evolution* **57**, 1498–1512 (2003).
27. Fraser, D. & Kaern, M. *Mol. Microbiol.* **71**, 1333–1340 (2009).
28. Radman, M., Matic, I. & Taddei, F. *Ann. NY Acad. Sci.* **870**, 146–155 (1999).
29. Vignuzzi, M., Stone, J.K., Arnold, J.J., Cameron, C.E. & Andino, R. *Nature* **439**, 344–348 (2006).
30. Zhang, Z. & Zhang, J. *PLoS ONE* **4**, e5686 (2009).
31. Greenspan, R.J. *Nat. Rev. Genet.* **2**, 383–387 (2001).
32. Gould, S.J. *The Panda's Thumb: More Reflections in Natural History* (W. W. Norton & Company, 1980).
33. Jacob, F. *Science* **196**, 1161–1166 (1977).
34. Austin, M.B., O'Maille, P.E. & Noel, J.P. *Nat. Chem. Biol.* **4**, 217–222 (2008).
35. Goldsmith, M. & Tawfik, D.S. *Proc. Natl. Acad. Sci. USA* **106**, 6197–6202 (2009).
36. Süel, G.M., Kulkarni, R.P., Dworkin, J., Garcia-Ojalvo, J. & Elowitz, M.B. *Science* **315**, 1716–1719 (2007).

Acknowledgments

The author is grateful to J. Noel for inspirational discussions and for a comment he made in the opening of his seminar at the Weizmann Institute— "Biology is messy"—used in this article. He is also grateful to R. Milo, E. Braun, N. Balaban, B. Lehner and S. Stern for insightful and enjoyable discussions and valuable comments. Financial support from the Israel Science Foundation and the Sasson and Marjorie Peress Foundation are gratefully acknowledged. The author is the Nella and Leon Benoziyo Professor of Biochemistry.

Competing financial interests

The author declares no competing financial interests.