Rethinking biology education in the light of single cell and single molecule studies and stochastic processes.

For Science's Education Forum.

M.W. Klymkowsky

Molecular, Cellular & Developmental Biology, University of Colorado Boulder

Boulder, Colorado 80309 U.S.A. email: klym@colorado.edu / cell: 720-425-8266

Summary: Most biology textbooks, and presumably the courses that use them, describe the molecular and cellular interactions that underlie biological behaviors from a deterministic perspective. Increasingly single cell (Levine et al., 2013; Symmons and Raj, 2016) and single molecule (Li and Xie, 2011; Yu, 2016) studies are providing important mechanistic insights into a range of biological processes, from evolutionary to pathogenic mechanisms. At the same time the "Framework for K12 science education" (Schweingruber et al., 2012) ignores stochastic processes altogether, while the Vision and Change in Undergraduate Biology Education" document (Science, 2011) calls for "incorporating stochasticity into biological models" (p. 17), but omits details of what this means in practice. People (even scholars) often have a difficult time developing an accurate understanding of stochastic processes (see Garvin-Doxas and Klymkowsky, 2008; Taleb, 2005). The failure to appreciate stochastic processes has been an obstacle to the acceptance of Darwinian evolution (Bowler, 1992). In this light, it seems well past time for a rethinking of the foundational roles of stochastic processes in biological (as well as chemical and physical) systems and how best to introduce such processes to students through coherent course narratives and supporting materials.

Stochastic processes are often presented in terms of random, that is unpredictable, events. This framing that obscures the reality that stochastic processes, while more or less unpredictable at the level of individual events, are well behaved at the population level. It also obscures the role of stochastic processes in a wide range of predictable phenomena; in atomic systems, for example, unknown factors determine the timing of the radioactive decay of a particular unstable isotope atom, at the same time when examining a large enough population of atoms, the rate of radioactive decay is highly predictable. Similarly, in the classical double slit experiment, the passage of a single photon, electron, or C60 molecule (see Nairz et al., 2003) is unpredictable while the behavior of a larger population is perfectly predictable. The macroscopic (population) predictability of the Brownian motion (a stochastic process) enable Einstein (Einstein and Infeld, 1938) to argue for the reality of atoms.

Similarly, the dissociation of a molecular complex or the occurrence of a chemical reaction, driven as they are by thermal collisions are stochastic processes, whereas dissociation constants and reaction rates are predictable. In fact this type of unpredictability at the individual level and predictability at the population level is the hallmark of stochastic, as compared to truly random, that is, unpredictable behaviors.

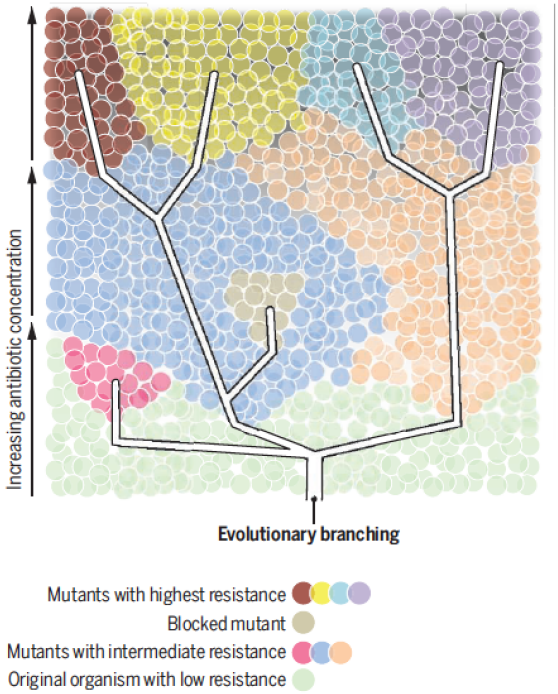

FIG.1 A schematic of the mutational origin of antibiotic resistance over time (and space) from McNally & Brown, 2016).

In a biological context the effects of stochastic events are complicated by the developing and adaptive nature of such systems and appears to influenced by the genetic background (Ansel et al., 2008). In some cases, homeostatic (feedback) mechanisms return the system to its original state. In others, the stochastic expression (or mutation) of a particular gene (or set of genes) leads to a cascade of downstream effects that change the system (McAdams and Arkin, 1997), such that subsequent events become more or less probable, a process nicely illustrated in recent real time studies of the evolution of antibiotic resistance in bacteria (**FIG. 1**) (from McNally & Brown, 2016; discussed in Baym et al., 2016). The stochastic (molecular clock) nature of mutation has recently been used with the EXAC system (Lek et al., 2016) to visualize the impact of selective and non-adaptive effects on human genes.

Pedagogical studies: A number of studies indicate that students call upon deterministic models to explain a range of stochastic processes (Andrews et al., 2012; Cooper and Klymkowsky, 2013b; Garvin-Doxas and Klymkowsky, 2008). The fact that all too often students are introduced to the behaviors of cellular and molecular level biological systems through depictions that are overtly deterministic does not help the situation. In the majority of instructional videos, for example, molecules appear to know were they are heading and move there with a purpose (see XVIVO, 2011). Similarly the folding of polypeptides is often depicted as a deterministic process[1] although the proliferation of model-based simulations offers a more realistic depiction (see below). Macromolecules are commonly depicted as rigid rather than as dynamic (Craveur et al., 2015); for example, the thermally driven opening and closing of the DNA double-helix (a consequence of the weakness of intermolecular

[1] Scientific animation: protein production and folding: https://youtu.be/2dV5s6v2v8Q

interactions) is rarely illustrated. Molecules recognize one another and (apparently) stay locked in their mutual embrace forever; the role of thermal collisions in driving molecular dissociation is rarely (apparently) considered in most textbooks, and presumably, in the classes that use these books. Moreover, the factors involved in inter-molecular interactions are often poorly understood, even after the completion of conventional university level chemistry courses (Williams et al., 2015). The energetic factors that determine enzyme specificity and reaction rates and the binding of transcription factors to their target DNA sequences, as well as the effects of mutations on these and other processes, often go uncommented on. It is not at all clear whether students appreciate that thermal collisions supply a reaction's activation energy or that they are responsible for the reversal of molecular interactions. Cells with the same genotype are implicitly expected to behave in identical ways (display the same phenotypes), a situation at odds with direct observation (see Elowitz et al., 2002; Symmons and Raj, 2016) and the general processes involved in cellular differentiation and social behaviors. Phenotypic penetrance and expressivity also involve stochastic behaviors (Raj et al., 2010). It certainly does not help when instructors introduce a stochastic process, such as genetic drift, in the context of the Hardy-Weinberg model, a situation in which genetic drift does not occur; in fact, such presentations appear to increase student confusion (Andrews et al., 2012).

It is our impression that the typical instructional approach is to present molecular level processes in terms of large populations of molecules that behave in a deterministic manner. Consider the bacteria *Escherichia coli*'s lac operon, a group of genes that has been a workhorse in modern molecular biology and a common context through which to present the regulation of gene expression. Expression of the lac operon results in the synthesis of two proteins (lactose permease and β-galactosidase) that enables lactose to enter the cell and converts lactose into the monosaccharides glucose and galactose (which are further metabolized) and allolactone, which inhibits the binding of the lac repressor protein to DNA , allowing the expression of the lac operon. When the bulk behavior of a bacterial culture is analyzed, the expression of the lac operon increases as a smooth function over time (in the absence of other energy sources)(**FIG. 2**). The result is that the expression of the proteins required for lactose

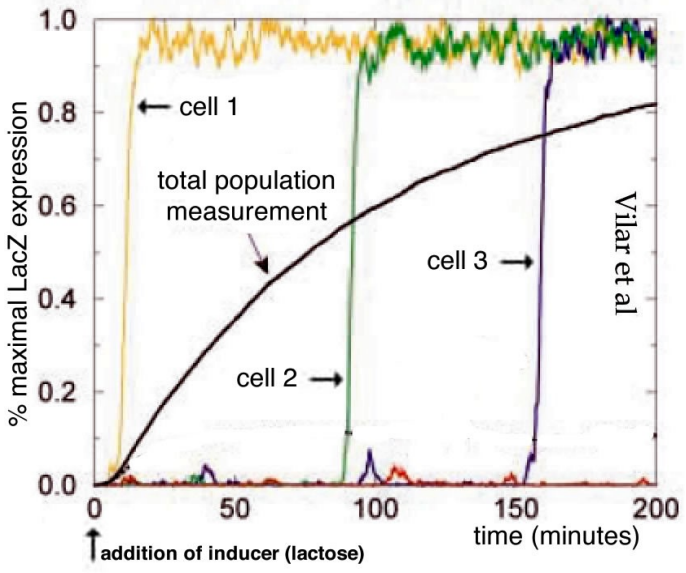

FIG. 2 A comparison between the expression of lacZ as measured in a bulk population and in individual cells (adapted from Vilar et al., 2003).

metabolism is restricted to situations in which lactose is present.

The mechanistic quandary, rarely if ever explicitly considered as far as we can discern, is how can the lac operon "turn on", that is, how can the lac operon go from repressed to expressed when the entry of lactose into the cell and the inactivation of the lac repressor both depend upon the expression of the lac operon? The situation becomes clear once we consider the behavior of individual cells; LacZ expression goes from off to fully on in a stochastic manner (**FIG. 2**)(Vilar et al., 2003).[2] Given that there are ~5 lac repressor molecules and one to two copies of the lac operon per cell, presumably the lac operon is expressed when the operon is free of bound repressor. If such a "noisy" event occurs when lactose is present in the media, expression of the lac operon allows lactose to enter the cell, the conversion of lactose into allolactone, the inactivation of the lac repressor, and stable expression of the lac operon. The stochastic behavior of the system works enables individual cells to sample their environment and respond when useful metabolites are present while minimizing unnecessary metabolic expense (the synthesis of irrelevant polypeptides) when they are not. A similar logic is involved in the social process of quorum sensing (Weber and Buceta, 2013) and a number of other processes, such as the emission of light (via the luciferase system)(Fuqua et al., 1994) and the regulation of the DNA uptake system (Johnston et al., 2014).

A diverse range of cell and molecular behaviors are the result of stochastic processes, including transcriptional (Corrigan et al., 2016; Fukaya et al., 2016; Pedraza and Paulsson, 2008; Suter et al., 2011) and translational bursting (Symmons and Raj, 2016). Various cell fate decisions involving stem cells (Buchholz et al., 2016; Rué and Arias, 2015), as well as molecular analyzes of enhancer syntax (i.e., the order, orientation, and spacing of transcription factor binding sites) underscore the fact that cellular and molecular systems must accommodate, and often exploit the inherent noisiness of molecular interactions. In the case of enhancers, organisms insure developmental robustness by exploiting regulatory redundancy and suboptimal transcription factor binding sites (Farley et al., 2015; Farley et al., 2016). Similarly, neural systems depend upon spontaneous (noisy) neural activity (Faisal et al., 2008) that arises from thermally driven association/dissociation interactions and transitions between active states. In microbial systems, molecular level noise can generate multiple phenotypes from a common genotype. Examples include altruistic programmed cell death, mutator phenotypes, and the generation of metabolically inactive "persister" cells (Eldar and Elowitz, 2010; Kussell et al., 2005; Uphoff et al., 2016; Veening et al., 2008). Similarly, the ubiquity of stochastic behaviors necessitates the presence of feedback and signaling systems as well as redundant systems (e.g. enhancers) that control these effects (Ansel et al., 2008; Farley et al., 2016; Sangster et al., 2008)

---

[2] http://bionumbers.hms.harvard.edu/bionumber.aspx?&id=102022&ver=6&trm=lac%20repressor

channeling the system to particular phenotypic outcomes – as illustrated by Waddington's landscape metaphor (Waddington, 1957).

Whats a biology educator to do? The question that faces the reflective educational designer and enlightened instructor is how should their course address the multiple roles of stochastic processes within biological systems? We have a short set of recommendations that we think both designers and instructors might want to consider; many have been incorporated into our ongoing efforts at course design (Klymkowsky et al., 2016b). First, and rather obviously, instructors should illustrate and articulate the role of stochastic processes in range of biological systems, from phenotypic variation and evolutionary events, including the effects of mutations and various non-adaptive processes (such as genetic drift) to gene expression, drug-target interactions, and reaction kinetics. Second, stochastic behaviors should be accurately and explicitly illustrated. This can range from simulations that present representations of molecular (and cellular) level processes (Wu et al., 2009).[3] Among currently available examples there are those that illustrate the movement of a water molecule through a membrane either through an aquaporin molecule or on its own,[4] as well as the PhET applet on gene expression that illustrates the Elowitz et al (Elowitz et al., 2002) study on GFP-expression in *E. coli* (and allows for student manipulation of key regulatory parameters).[5] A simulation of the nature of intermolecular interactions and the role of molecular collisions in their formation has been developed for use with the CLUE Chemistry curriculum (Cooper and Klymkowsky, 2013a).[6] An instructor can reveal (to themselves and their students) the understanding of stochastic

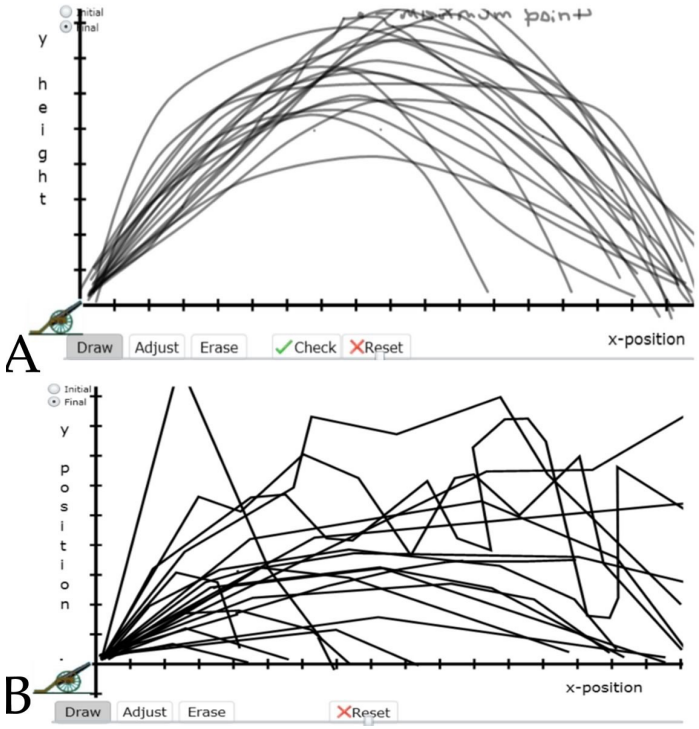

FIG.3 Students are asked to predict the trajectory of a cannon ball (A) or a molecule (B); few students recognize the transition from projectile to Brownian motion (adapted from Klymkowsky et al., 2016).

[3] See this recording of stem cells http://jcb.rupress.org/content/187/4/513/suppl/DC1

[4] Aquaporin: https://youtu.be/1Uw6u0fzNsE and a pure lipid membrane: https://youtu.be/ePGqRaQiBfc

[5] PhET gene expression basics https://phet.colorado.edu/en/simulation/gene-expression-basics - see panel 3

[6] http://virtuallaboratory.colorado.edu/LDF+binding-interactions/1.2-interactions-0.html

processes using various assessment tools (Garvin-Doxas and Klymkowsky, 2008; Klymkowsky et al., 2016a; Price et al., 2014). For example, students can be asked to draw a graph that reflects the movement of a macroscopic projectile (**FIG.3A**) versus a molecular (microscopic) object (**FIG.3B**); such a task can reveal whether students can make the transition from the well behaved (deterministic) to the stochastic behavior. Drawing (and explanation) has been used extensively in the analysis of student understanding with the context of the CLUE project (Cooper et al., 2010; Underwood et al., 2016; Williams et al., 2015). In a similar vein, network dynamics, including the cascade effects driving cell level divergence and the feedback and regulatory interactions involved in limiting the effects of noise and generating various outcomes (cellular differentiation) can be presented to students (Kellogg and Tay, 2015; Milo et al., 2002; Trujillo et al., 2012). One can consider the role of stochastic events within social systems, including responses to various aberrant behaviors (social cheating, cancer) and in terms of social feedback mechanisms (apoptosis, positive and negative feedback, lateral inhibition of cell differentiation) and in the context of the decisions involved in stem cells division and differentiation (Shahriyari and Komarova, 2013; Till et al., 1964). By introducing students to the various roles of stochastic processes in biological systems, we can help them considered the comprehensible and predictable but not completely deterministic nature of such systems.

Acknowledgements: We thank Melanie Cooper for many useful discussions and Ernst Hafen for comments on the manuscript. MWK's work on developing and assessing scientific understanding has been supported over the years through funds from the National Science Foundation (USA), mostly in collaboration with Melanie Cooper (Michigan State University).

Literature cited:

Andrews, T., Price, R., Mead, L., McElhinny, T., Thanukos, A., Perez, K., Herreid, C., Terry, D. and Lemons, P. (2012). Biology undergraduates' misconceptions about genetic drift. *CBE-Life Sciences Education* 11, 248-259.

Ansel, J., Bottin, H., Rodriguez-Beltran, C., Damon, C., Nagarajan, M., Fehrmann, S., François, J. and Yvert, G. (2008). Cell-to-cell stochastic variation in gene expression is a complex genetic trait. *PLoS genetics* 4, e1000049.

Bowler, P. J. (1992). The eclipse of Darwinism: Anti-Darwinian evolution theories in the decades around 1900. Baltimore: Johns Hopkins University Press.

Buchholz, V. R., Schumacher, T. N. and Busch, D. H. (2016). T cell fate at the single-cell level. *Annual review of immunology* 34, 65-92.

Cooper, M. M., Grove, N., Underwood, S. and Klymkowsky, M. W. (2010). Lost in Lewis Structures: an investigation of student difficulties in developing representational competence. *J. Chem. Educ.* **87**, 869–874.

Cooper, M. M. and Klymkowsky, M. W. (2013a). Chemistry, life, the universe, and everything: a new approach to general chemistry, and a model for curriculum reform. *Journal of Chemical Education* **90**, 1116-1122.

Cooper, M. M. and Klymkowsky, M. W. (2013b). The trouble with chemical energy: why understanding bond energies requires an interdisciplinary systems approach. *CBE Life Sci Educ* **12**, 306-312.

Corrigan, A. M., Tunnacliffe, E., Cannon, D. and Chubb, J. R. (2016). A continuum model of transcriptional bursting. *eLife* 5, e13051.

Craveur, P., Joseph, A. P., Esque, J., Narwani, T. J., Noël, F., Shinada, N., Goguet, M., Leonard, S., Poulain, P. and Bertrand, O. (2015). Protein flexibility in the light of structural alphabets. *Frontiers in molecular biosciences* 2.

Einstein, A. and Infeld, L. (1938). *The evolution of physics*. New York: Norton.

Eldar, A. and Elowitz, M. B. (2010). Functional roles for noise in genetic circuits. *Nature* **467**, 167-173.

Elowitz, M. B., Levine, A. J., Siggia, E. D. and Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science* 297, 1183-1186.

Faisal, A. A., Selen, L. P. and Wolpert, D. M. (2008). Noise in the nervous system. *Nature reviews neuroscience* 9, 292-303.

Farley, E. K., Olson, K. M., Zhang, W., Brandt, A. J., Rokhsar, D. S. and Levine, M. S. (2015). Suboptimization of developmental enhancers. *Science* 350, 325-328.

Farley, E. K., Olson, K. M., Zhang, W., Rokhsar, D. S. and Levine, M. S. (2016). Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. *Proceedings of the National Academy of Sciences* **113**, 6508-6513.

Fukaya, T., Lim, B. and Levine, M. (2016). Enhancer Control of Transcriptional Bursting. *Cell*.

Fuqua, W. C., Winans, S. C. and Greenberg, E. P. (1994). Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *Journal of bacteriology* **176**, 269.

Garvin-Doxas, K. and Klymkowsky, M. W. (2008). Understanding Randomness and its impact on Student Learning: Lessons from the Biology Concept Inventory (BCI). *Life Science Education* **7**, 227-233.

Johnston, C., Martin, B., Fichant, G., Polard, P. and Claverys, J.-P. (2014). Bacterial transformation: distribution, shared mechanisms and divergent control. *Nature Reviews Microbiology* **12**, 181-196.

Kellogg, R. A. and Tay, S. (2015). Noise facilitates transcriptional control under dynamic inputs. *Cell* **160**, 381-392.
Klymkowsky, M. W., Kohler, K. and Cooper, M. M. (2016a). Diagnostic assessments of student thinking about stochastic processes. In *bioArXiv: http://biorxiv.org/content/early/2016/05/20/053991*.
Klymkowsky, M. W., Rentsch, J. D., Begovic, E. and Cooper, M. M. (2016b). The design and transformation of Biofundamentals: a non-survey introductory evolutionary and molecular biology course. *LSE Cell Biol Edu, in press.* **in press.**
Kussell, E., Kishony, R., Balaban, N. Q. and Leibler, S. (2005). Bacterial persistence. *Genetics* 169, 1807-1814.
Lek, M., Karczewski, K., Minikel, E., Samocha, K., Banks, E., Fennell, T., O'Donnell-Luria, A., Ware, J., Hill, A. and Cummings, B. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *BioRxiv*, 030338.
Levine, J. H., Lin, Y. and Elowitz, M. B. (2013). Functional roles of pulsing in genetic circuits. *Science* **342**, 1193-1200.
Li, G.-W. and Xie, X. S. (2011). Central dogma at the single-molecule level in living cells. *Nature* **475**, 308-315.
McAdams, H. H. and Arkin, A. (1997). Stochastic mechanisms in gene expression. *Proc Natl Acad Sci U S A* **94**, 814-819.
Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science* 298, 824-827.
Nairz, O., Arndt, M. and Zeilinger, A. (2003). Quantum interference experiments with large molecules. *American Journal of Physics* **71**, 319-325.
Pedraza, J. M. and Paulsson, J. (2008). Effects of molecular memory and bursting on fluctuations in gene expression. *Science* **319**, 339-343.
Price, R. M., Andrews, T. C., McElhinny, T. L., Mead, L. S., Abraham, J. K., Thanukos, A. and Perez, K. E. (2014). The Genetic Drift Inventory: a tool for measuring what advanced undergraduates have mastered about genetic drift. *CBE-Life Sciences Education* **13**, 65-75.
Raj, A., Rifkin, S. A., Andersen, E. and van Oudenaarden, A. (2010). Variability in gene expression underlies incomplete penetrance. *Nature* **463**, 913-918.
Rué, P. and Arias, A. M. (2015). Cell dynamics and gene expression control in tissue homeostasis and development. *Molecular systems biology* **11**, 792.
Sangster, T. A., Salathia, N., Undurraga, S., Milo, R., Schellenberg, K., Lindquist, S. and Queitsch, C. (2008). HSP90 affects the expression of genetic variation and developmental stability in quantitative traits. *Proceedings of the National Academy of Sciences* **105**, 2963-2968.
Schweingruber, H., Keller, T. and Quinn, H. (2012). A Framework for K-12 Science Education:: Practices, Crosscutting Concepts, and Core Ideas: National Academies Press.
Science, A. A. f. t. A. o. (2011). Vision and change in undergraduate biology education: a call to action. *Washington, DC*.
Shahriyari, L. and Komarova, N. L. (2013). Symmetric vs. asymmetric stem cell divisions: an adaptation against cancer? *PLoS One* **8**, e76195.
Suter, D. M., Molina, N., Gatfield, D., Schneider, K., Schibler, U. and Naef, F. (2011). Mammalian genes are transcribed with widely different bursting kinetics. *Science* 332, 472-474.
Symmons, O. and Raj, A. (2016). What's Luck Got to Do with It: Single Cells, Multiple Fates, and Biological Nondeterminism. *Molecular cell* **62**, 788-802.

Taleb, N. N. (2005). Fooled by Randomness: The hidden role of chance in life and in the markets. (2nd edn). New York: Random House.
Till, J. E., McCulloch, E. A. and Siminovitch, L. (1964). A stochastic model of stem cell proliferation, based on the growth of spleen colony-forming cells. *Proceedings of the National Academy of Sciences* **51**, 29-36.
Trujillo, C., Cooper, M. M. and Klymkowsky, M. W. (2012). Using graph-based assessments within socratic tutorials to reveal and refine students' analytical thinking about molecular networks. *Biochem Mol Biol Educ* **40**, 100-107.
Underwood, S. M., Reyes-Gastelum, D. and Cooper, M. M. (2016). When do students recognize relationships between molecular structure and properties? A longitudinal comparison of the impact of traditional and transformed curricula. *Chemistry Education Research and Practice* **in press.**
Uphoff, S., Lord, N. D., Okumus, B., Potvin-Trottier, L., Sherratt, D. J. and Paulsson, J. (2016). Stochastic activation of a DNA damage response causes cell-to-cell mutation rate variation. *Science* **351**, 1094-1097.
Veening, J.-W., Smits, W. K. and Kuipers, O. P. (2008). Bistability, epigenetics, and bet-hedging in bacteria. *Annu. Rev. Microbiol.* **62**, 193-210.
Vilar, J. M., Guet, C. C. and Leibler, S. (2003). Modeling network dynamics: the lac operon, a case study. *J Cell Biol* **161**, 471-476.
Waddington, C. H. (1957). *The Strategy Of The Genes.* London: George Allen & Unwin.
Weber, M. and Buceta, J. (2013). Dynamics of the quorum sensing switch: stochastic and non-stationary effects. *BMC systems biology* **7**, 6.
Williams, L. C., Underwood, S. M., Klymkowsky, M. W. and Cooper, M. M. (2015). Are Noncovalent Interactions an Achilles Heel in Chemistry Education? A Comparison of Instructional Approaches. *Journal of Chemical Education* **92**, 1979–1987.
Wu, Z., Luby-Phelps, K., Bugde, A., Molyneux, L. A., Denard, B., Li, W.-H., Süel, G. M. and Garbers, D. L. (2009). Capacity for stochastic self-renewal and differentiation in mammalian spermatogonial stem cells. *The Journal of cell biology* 187, 513-524.
XVIVO (2011). The Inner Life of the Cell.
Yu, J. (2016). Single-Molecule Studies in Live Cells. *Annual review of physical chemistry* **67**, 565-585.