Clipboard

Genes and proteins: Dogmas in decline

A long-standing assumption of modern biology is the fidelity of the genotype-phenotype relationship. Embryonic development is often characterized as the process by which the one-dimensional information in a genome is decoded to construct a three-dimensional organism, while the neo-Darwinian model of evolutionary change is dependent on a close mapping between the genes of an organism and its outward form and behaviour. But to the extent that phenotypes are not entirely predictable expressions of genotypes, causal agents beyond the gene must be invoked to explain how organisms transmit their characteristics across generations, and how organisms become different from one another over time. Of course, the inherent noisiness of development (Pritchard *et al* 2006), a feature of any complex dynamical process (Kaneko 2007), will necessarily undermine strict determination of the phenotype by the genotype. However, such random effects do not represent a real challenge to standard developmental or evolutionary models, since the phenotypic norm is a populational concept. As long as a given genotype in a given environment is associated with a narrow range of phenotypes, with no discontinuity of outcome due to other modes of causation, all is well with the "genetic program" notion in developmental biology and gene-centered neo-Darwinian evolutionary theory.

A recent report by Kimchi-Sarfaty *et al* (2007) has now called into question a basic tenet of these generally accepted developmental and evolutionary models. According to what has been termed "Anfinsen's dogma" (e.g. Chatani and Goto 2005), a protein's native (i.e. three-dimensional, functional) structure is determined only by its amino acid sequence. The new report provides an exception to this rule, and in doing so demonstrates that the influence of genotype over phenotype even at the most fundamental cellular level – the production of proteins of defined function – is dependent on contextual factors.

Studying P-gp [the product of the Multidrug Resistance 1 (*MDR*1) gene], an ATP-driven efflux pump of the plasma membrane involved in the multidrug resistance of cancer cells, the investigators made a puzzling finding: HeLa cells with a particular single-nucleotide polymorphism (SNP) that was synonymous (i.e. did not produce an altered coding sequence) with the most common, or wild-type, allelic sequence of *MDR*1 exhibited reduced transport functionality of P-gp. There were several possible explanations for this, including linkage disequilibrium (i.e. nonrandom association) with other, function-impairing, non-synonymous SNPs in the *MDR*1 gene, and alteration in mRNA secondary structure leading to variations in splicing or translational efficiency. The authors put each of these to rest, however, by assaying for P-gp function as part of different haplotypes and in several different cell lines. They found that it was the SNP in question, C3435T (which substitutes a rarer codon for isoleucine than that of the wild-type gene), that was indeed responsible for the impaired protein. Although the impairment was only seen in constructs in which C3435T was present along with certain other *MDR*1 SNPs (a genetic status they term "haplotype C3435T"), the same haplotype contexts had no effect on P-gp specified by genes with the wild-type sequence at position 3435, and so linkage disequilibrium was not involved. They also found that haplotype C3435T was not misspliced, expressed at aberrantly low levels, or in truncated form, and that cells expressing haplotype C3435T produced the same amount of P-gp protein as ones expressing its wild-type counterpart.

How could a synonymous substitution, a supposedly "silent" polymorphism, change a protein's function? Using a conformation-sensitive antibody against P-gp, and a trypsin digestion assay, Kimchi-Sarfaty and coworkers determined that P-gp specified by haplotype C3435T is folded differently from P-gp specified by wild-type P-gp, despite the two proteins having identical amino acid sequence. In the presence of verapamil, one of P-gp's substrates, the protein specified by haplotype C3435T assumed a conformation similar or identical to that of the wild-type-specified protein suggesting that the two conformations are interconvertible.

Keywords. Synonymous substitution; protein folding

This conclusive violation of Anfinsen's dogma, which is itself based on the assumption that proteins fold to assume a unique state of minimum free energy consistent with the cytoplasmic microenvironment (Sela *et al* 1957), requires a mechanistic account for its deviation from the thermodynamic expectation. The authors suggest, based on earlier studies, that the explanation resides in the fact that use of rare codons appears to influence translation rate, which in turn may affect protein folding (Purvis *et al* 1987; Komar *et al* 1999). In other words, if translation of haplotype C3435T-specified P-gp is slowed down or paused due to its use of a rare codon (a phenomenon that would be exacerbated by rate-slowing depletion of tRNA species associated with the compromising haplotype contexts), this may cause portions of it to fold before other portions get made. The polypeptide chain would then be trapped in a state representing a local, rather than the global, free energy minimum. The result would be a protein different from the wild-type one, if "protein" is defined by shape and function and not simply by amino acid sequence. The translation-rate mechanism for the effect of the synonymous substitution remains a speculation, but a reasonable one.

The existence and prevalence of alternative splicing has long dispelled the notion that a given RNA transcript encodes a unique protein (reviewed in Blencowe 2006). There is also nothing surprising in the possibility that synonymous SNPs, which may alter translation rates or splice choice, have phenotypic consequences that can lead to their being selected for or against (Chamary *et al* 2006). Finally, the existence of prions (Prusiner 1998) has made familiar the idea that a given polypeptide can have more than one conformation in a cellular microenvironment, though the alternative, atypical, conformation is usually pathogenic (Caughey 2001).

The study of Kimchi-Sarfaty *et al* (2007), however, can be seen as going beyond these earlier findings in undermining exclusively gene-centered models of development and evolution. Although the authors describe their results in terms of the different effects of a particular "wild-type" and "mutant" codon of one gene, it is also reasonable to consider the implications of the described phenomenon for the synthesis of typical cellular proteins under different conditions. One possible inference is that, apart from the relative effects of rare codons and haplotype context, *any of a cell's mRNAs, if subjected to translational braking, can potentially generate a protein with an alternative conformation.* [We can call this the "translation-dependent folding" (TDF) hypothesis.] Since the percentage of proteins that actually fold, per Anfinsen's dogma, into the thermodynamically most stable conformation is completely unknown, many "standard" protein conformations could, in fact, be metastable ones, dependent on submaximal rates of translation. It would thus appear that any variation in the cellular microenvironment that affects translation rate, generally or selectively, in a positive or negative fashion, could qualitatively influence the array of proteins a cell produces.

The findings of Kimchi-Sarfaty *et al* (2007), considered in the light of well-characterized mechanisms of microenvironmental control of translation elongation rate in eukaryotic cells (Proud 2007), could undermine the importance of a precept of modern molecular biology even more celebrated than Anfinsen's dogma: the "central dogma" propounded by Francis Crick (Crick 1958). This principle, whereby sequence information flows from DNA to RNA to protein, but not in the reverse direction, is at the core of the idea that the causality and logic of development is embodied in networks of transcription factor-promoter interactions (Davidson 2006). In this widely accepted model, differential gene expression, reflected in a particular array of transcripts, defines the developed phenotype. But if (by the TDF hypothesis) the set of proteins a cell produces is not uniquely determined by the population of mRNAs it contains, the central dogma, while not being thereby disconfirmed, becomes less dispositive in the determination of phenotype. One could infer, for example that two embryonic cells with identical programs of gene expression, but subject to different microenvironments, could make different sets of proteins and follow different developmental trajectories.

The possible implications for neo-Darwinism are even more unsettling. Two subpopulations of organisms with identical genomes, encountering different environments, could produce different arrays of proteins, becoming phenotypically different and reproductively isolated in a single generation! Although the provocative findings of Kimchi-Sarfaty *et al* (2007) may turn out to be a biological exception, and the TDF hypothesis and the scenarios that flow from it flights of fancy, they must now nevertheless be admitted into the realm of biological possibility.

References

Blencowe B J 2006 Alternative splicing: new insights from global analyses; *Cell* **126** 37–47

Caughey B 2001 Prion protein interconversions; *Philos. Trans. R. Soc. London B Biol. Sci.* **356** 197–200 (discussion 200–202)

Chamary J V, Parmley J L and Hurst L D 2006 Hearing silence: non-neutral evolution at synonymous sites in mammals; *Nat. Rev. Genet.* **7** 98–108

Chatani E and Goto Y 2005 Structural stability of amyloid fibrils of beta(2)-microglobulin in comparison with its native fold; *Biochim. Biophys. Acta* **1753** 64–75

Crick F H C 1958 On protein synthesis; *Symp. Soc. Exp. Biol.* **12** 138–163

Davidson E H 2006 *The regulatory genome: gene regulatory networks in development and evolution* (Amsterdam, London: Elsevier/Academic Press)

Kaneko K 2007 Evolution of robustness to noise and mutation in gene expression dynamics; *PLoS ONE* **2** e434

Kimchi-Sarfaty C, Oh J M, Kim I W, Sauna Z E, Calcagno A M, Ambudkar S V and Gottesman M M 2007 A "silent" polymorphism in the MDR1 gene changes substrate specificity; *Science* **315** 525–528

Komar A A, Lesnik T and Reiss C 1999 Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation; *FEBS Lett.* **462** 387–391

Sela M, White F H, Jr. and Anfinsen C B 1957 Reductive cleavage of disulfide bridges in ribonuclease; *Science* **125** 691–692

Pritchard C, Coil D, Hawley S, Hsu L and Nelson P S 2006 The contributions of normal variation and genetic background to mammalian gene expression; *Genome Biol.* **7** R26

Proud CG 2007 Signalling to translation: how signal transduction pathways control the protein synthetic machinery; *Biochem. J.* **403** 217–234

Prusiner S B 1998 Prions; *Proc. Natl. Acad. Sci. USA* **95** 13363–13383

Purvis I J, Bettany A J, Santiago T C, Coggins J R, Duncan K, Eason R and Brown A J 1987 The efficiency of folding of some proteins is increased by controlled rates of translation *in vivo*. A hypothesis; *J. Mol. Biol.* **193** 413–417

Stuart A Newman* and Ramray Bhat
Department of Cell Biology and Anatomy, Basic Science Building, New York Medical College, Valhalla, New York 10595, USA
**(Email: newman@nymc.edu)*

*e*Publication: 30 July 2007