# The causes of protein evolutionary rate variation

## James O. McInerney

Department of Biology, National University of Ireland, Maynooth, Co. Kildare, Ireland

**The rate of protein evolution varies more than 1000-fold and, for the past 30 years, it was thought that the rate was determined by protein function. Drummond and co-workers have now shown that a single factor underlying mRNA expression, protein abundance and synonymous codon usage is the chief causal agent of protein evolutionary rate in yeast. It will be interesting to see whether this is shown to be a universal rule for all biological systems.**

## Functional constraint and protein evolutionary rate

The strength and extent of natural selection on individual amino acids in a protein greatly influences the evolutionary rate of that protein. Strong purifying selection (see Glossary) on a large number of residues leads to a reduced overall evolutionary rate and, conversely, strong positive selection on large numbers of residues leads to a rapid rate of evolution. This has been our paradigm for 30 years: functional constraints dictate protein evolutionary rates. However, we have also known of many other correlates with evolutionary rate that are not connected with purifying and positive selection and this has led to difficulties in separating extrinsic from causal correlations. Now, in a new paper, Drummond and co-workers provide evidence that, in yeast, a single factor is responsible for the rate of protein evolution.

## Correlations with protein evolutionary rate

Working with yeast data, Pál *et al.* [1] used pairwise $dN$ values and information about mRNA transcript abundance gathered by Holstedge *et al.* [2] to show a strong negative correlation ($r = -0.584, p < 10^{-6}$) between amino acid substitution rate and the number of mRNA transcripts in the cell. Although Pál *et al.* concluded that expression level affects protein evolution, other correlates were observed, including the correlation between protein evolutionary rate and the codon adaptation index (CAI) [3].

Hirsch and Fraser [4] revealed a correlation between evolutionary rate and dispensability, which is the fitness effect of deleting a gene. Fraser *et al.* [5] have shown that protein evolutionary rate inversely correlates with the number of protein–protein interactions, that is, the greater the number of interactions a protein has with other proteins, the slower it is likely to evolve. Wall *et al.* [6] found significant independent correlations between

protein dispensability, mRNA expression level and evolutionary rate. Recently, Wyckoff *et al.* [7] reported an unexpected correlation between the probability of fixation of nonsynonymous substitutions in human–mouse orthologues and the rate of silent site evolution. They showed that, at elevated rates of $dS$, the $dN$:$dS$ ratio was also elevated.

In some organisms, there is a strong correlation between protein abundance and synonymous codon usage, a phenomenon known as translational selection [8]. Those codons that correspond to the most abundant isoaccepting tRNA are used more frequently in highly expressed genes. This is a coevolutionary process that facilitates the rapid and accurate translation of highly expressed genes. Akashi [9] has further shown that, in yeast, there is a stronger correlation between tRNA concentration and corresponding amino acid content in highly expressed genes than is seen in genes with low expression levels. Additionally, synonymous codon usage can be influenced by whether the gene is found on the leading- or lagging-strand of replication [10]; a correlation has also been observed between codon usage and gene length [11]. So clearly, several additional intercorrelates with protein evolutionary rate exist.

## Which correlates are most important?

Recently, Drummond *et al.* [12] used multivariate analysis to explore the relative contributions of seven commonly used predictors of protein evolutionary rate for proteins of the yeast *Saccharomyces cerevisiae*. These predictors included mRNA expression, CAI, protein length, protein dispensability, degree (the number of protein–protein interactions), the centrality of a protein in an interaction network and protein abundance. Using Principal Component Regression (PCR) analysis [13], the authors identified a single

---

## Glossary

**CAI:** codon adaptation index.
***dN***: number of nonsynonymous substitutions per nonsynonymous site.
***dS***: number of synonymous substitutions per synonymous site.
**Isoaccepting tRNA:** transfer RNA species that are charged by the same specific amino acid.
**Positive selection:** selection for new variant nucleotides or amino acids.
**Protein dispensability:** fitness effect of deleting a gene.
**Purifying selection:** conservation of existing residues.
**Silent sites:** sites within a codon where a nucleotide substitution does not change the encoded amino acid.
**Synonymous codon usage:** pattern of usage of all the synonymous codons for a particular amino acid.

*Corresponding author:* McInerney, J.O. (james.o.mcinerney@nuim.ie).

principal component that accounted for 43% of the variance in amino acid evolutionary rate. Approximately 90% of this single dominant component was made up of three predictors: mRNA expression level, protein abundance and CAI. In addition, these predictors also dominated the principal component governing the rate of synonymous substitution. In total, 61.9% of the variation in synonymous substitution rate was explained by the seven predictors. The first PCR axis explained 50.8% of all variation and 87% of the total explained variation was governed by mRNA expression level, protein abundance and CAI. One of their most interesting findings was that the quality of the result depended strongly on using the most appropriate multivariate analysis method. Their choice of PCR is based on its ability to identify when a single underlying variable, for instance expression level, influences two or more other variables, say, evolutionary rate and dispensability. There is a real need in the meta-analysis of large data sets to use the appropriate methods and, in their paper, Drummond *et al.* demonstrate that an erroneous choice of method can lead to spurious results, such as the finding of correlations where none exist [12].

### What causes the correlation?
The correlation between mRNA expression level, protein abundance, CAI (i.e. the three principals) and evolutionary rate begs the question of causality. It is unlikely that evolutionary rate is causing variation in the three principals. However, it is relatively easy to argue that the three principals can and do influence evolutionary rate.

In an organism, small selectively beneficial effects can accrue from the substitution of a codon for a synonymous alternative that can be translated more quickly. Translation rate for a codon is largely determined by the abundance of the isoaccepting tRNA in the cytoplasm and, in most organisms, there are <61 tRNA species. Therefore, not all codons are optimal for rapid and accurate translation. Given that all codons can be translated, the selective advantage of substituting one codon for another is small. It is only when two other conditions prevail that this small selective difference can be seen: first, if the gene is highly expressed, then the effect of codon substitution will be stronger; second, if the long-term effective population size of the organism in which this substitution occurs is large enough, then random genetic drift is reduced. Therefore, selection for preferred codons will only overcome random genetic drift in highly expressed genes in organisms with large long-term effective population sizes [8]. This explains the close link between mRNA expression level and CAI in the analysis. Protein abundance is a consequence of the roughly constant number of proteins produced from each mRNA transcript [14].

From their observation of CAI, mRNA expression and protein abundance being the most important correlates, Drummond *et al.* [15] hypothesise that selection for translational robustness is the underlying selection pressure. In this hypothesis, they speculate that the purifying selection pressure that affects proteins most is selection to minimise misfolding of proteins and that this effect is stronger in more abundant proteins.

## The universality of this finding
The observations of Drummond *et al.* are restricted both taxonomically and to limited experimental conditions (i.e. the expression data is from exponentially growing cells). Therefore, it remains to be seen how universal these correlations are, although Drummond *et al.* do make the prediction that the same underlying determinant influences protein evolutionary rate across all of life. In species with large, long-term effective population sizes, such as yeast, many prokaryotes and *Drosophila*, translational selection for optimal codons can be found in highly expressed genes. Organisms with small effective population sizes, such as mammals, do not exhibit such a pattern. Future work will determine whether Drummond *et al.*. are correct in their assertion that selection for translational robustness is universally affecting protein evolutionary rate. Nonetheless, it will be interesting to see how these results affect our views on the effects of functional constraint on amino acid substitution.

## The solution to a conundrum?
But how can we explain the unusual finding made by Wyckoff *et al.* [7] that, when a high rate of silent substitution is observed, then a high rate of replacement to silent substitution is also observed (i.e. there is a correlation between $dS$ and the $dN{:}dS$ ratio)? Currently, no theory covers this observation, although Wyckoff *et al.* offer a possible explanation based on differences in mutation rates in different genes. However, Drummond *et al.* have demonstrated that the same factors influence the evolution of both silent and replacement substitution rates. It seems possible therefore that selection, rather than mutation, might be responsible. It will be intriguing to see whether our new insights into molecular evolutionary rate variation provide the answer to this unusual correlation.

### References
1 Pál, C. *et al.* (2001) Highly expressed genes in yeast evolve slowly. *Genetics* 158, 927–931
2 Holstege, F.C. *et al.* (1998) Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 95, 717–728
3 Sharp, P.M. and Li, W.H. (1987) The codon adaptation index: a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15, 1281–1295
4 Hirsh, A.E. and Fraser, H.B. (2001) Protein dispensability and rate of evolution. *Nature* 411, 1046–1049
5 Fraser, H.B. *et al.* (2002) Evolutionary rate in the protein interaction network. *Science* 296, 750–752
6 Wall, D.P. *et al.* (2005) Functional genomic analysis of the rates of protein evolution. *Proc. Natl. Acad. Sci. U. S. A.* 102, 5483–5488
7 Wyckoff, G.J. *et al.* (2005) A highly unexpected strong correlation between fixation probability of nonsynonymous mutations and mutation rate. *Trends Genet.* 21, 381–385
8 Gouy, M. and Gautier, C. (1982) Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res.* 10, 7055–7074

9 Akashi, H. (2003) Translational selection and yeast proteome evolution. *Genetics* 164, 1291–1303

10 McInerney, J.O. (1998) Replicational and transcriptional selection on codon usage in *Borrelia burgdorferi*. *Proc. Natl. Acad. Sci. U. S. A.* 95, 10698–10703

11 Marais, G. and Duret, L. (2001) Synonymous codon usage, accuracy of translation, and gene length in *Caenorhabditis elegans*. *J Mol. Evol.* 52, 275–280

12 Drummond, D.A. *et al.* A single determinant dominates the rate of yeast protein evolution. *Mol. Biol. Evol.* (in press)

13 Mandel, J. (1982) Use of the singular value decomposition in regression analysis. *Am. Stat.* 36, 15–24

14 Ghaemmaghami, S. *et al.* (2003) Global analysis of protein expression in yeast. *Nature* 425, 737–741

15 Drummond, D.A. *et al.* (2005) Why highly expressed proteins evolve slowly. *Proc. Natl. Acad. Sci. U. S. A.* 102, 14338–14343
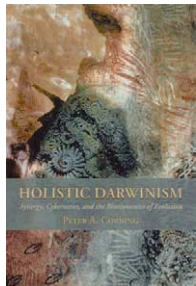
---

Book Reviews

# The collective survival enterprise

**Holistic Darwinism. Synergy, Cybernetics, and the Bioeconomics of Evolution** by Peter A. Croning. University of Chicago Press, 2006. £44.50/£18.50 hbk/pbk (504 pages) ISBN 0226116131/0226116166

## Eva Jablonka

The Cohn institute for the History and Philosophy of Science and Ideas, Tel-Aviv University, Tel Aviv 69978, Israel

In *Holistic Darwinism*, Peter Croning takes the work of biologists, psychologists, social scientists, system theorists, economists and others who have taken an evolutionary approach to their subject matter, and tries to weld their findings together within a Darwinian framework based on functional interactions. He believes that the best approach to evolutionary problems is to focus on the relations between parts and processes within teleonomic (goal-directed) systems. Such systems exhibit what he calls 'functional synergies', that is all kinds of combined or cooperative functional effects. According to Croning, the progressive aspect of the evolution of biological and social–cultural life is fueled by new synergies.

In some ways, Croning is reviving Herbert Spencer's 19th-century project to give a general evolutionary interpretation to all aspects of existence, or at least of life. Indeed, in a refreshing and honest way, Croning acknowledges his debt to Spencer: he has studied Spencer's work, and does not repeat disparaging clichés based on tertiary sources. Croning argues that biological and social systems are unified at all levels of organization because they share a basic enterprise, that of survival. All organisms have to earn a living to survive and reproduce. When organisms are part of a social system that is more than a mere aggregate, the survival enterprise is individual and collective, and the two are intimately linked. Very often, Croning argues, there is no conflict between the interests of the individual and that of the collective, so the problem of free-riders has been overstated. In most cases, he maintains, the collective enterprise is the winning strategy – those individuals that form functionally integrated collectives have a selective advantage compared with those that do not. Moreover, the cooperative strategy is often the one that leads to an increase in functional complexity, division of labor, the range of functions covered, and so on.

Despite the similarity in the ambition and scope of their projects, I do not think that Croning manages to do for the 21st century what Spencer did for the 19th. Croning creates such a large conceptual umbrella that every evolutionary biologist, from Dawkins to Sober and Wilson, can feel comfortable under it, and every point of view can be accommodated. Moreover, his analysis of 'wholes' is too general and he does not distinguish between different types of wholes, different types of functional interactions and different constituent parts with different degrees of autonomy. For example, there is no explicit discussion of the demographic paradox that characterizes the richer parts of the world.

A second problem is that it is not clear for whom the book is intended: it often reads as if Croning is writing for well-read experts, although his intention seems to have been to produce a book for the general reader. This ambiguity is reflected in the uneven writing style. It is also not clear to me why part III, which consists of three chapters about thermoeconomics (Croning's term for the economy of energy in biological and social systems) was included. The book would have been better without it as it would have been more coherent and focused. The important philosophical issue of the nature of functional and semantic information could have been summarized concisely, without going into long critiques (however justified) of the various misuses of the concept of information in biology, but with more attention given to current biological–philosophical discussions.

Some of the other problems are the result of it being a collection of edited and more or less updated papers and lectures. Despite the author's intention to create a cohesive monograph and to connect the various issues, the seams are obvious. The end of the last chapter, for example, is a series of lengthy comments about some of the topics discussed in that chapter (primers on capitalism,

---

*Corresponding author:* Jablonka, E. (jablonka@post.tau.ac.il).
Available online 10 January 2006