

# Why there is more to protein evolution than protein function: splicing, nucleosomes and dual-coding sequence

Tobias Warnecke, Claudia C. Weber and Laurence D. Hurst<sup>1</sup>

Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, U.K.

## Abstract

There is considerable variation in the rate at which different proteins evolve. Why is this? Classically, it has been considered that the density of functionally important sites must predict rates of protein evolution. Likewise, amino acid choice is usually assumed to reflect optimal protein function. In the present article, we briefly review evidence suggesting that this protein function-centred view is too simplistic. In particular, we concentrate on how selection acting during the protein's production history can also affect protein evolutionary rates and amino acid choice. Exploring the role of selection at the DNA and RNA level, we specifically address how the need (i) to specify exonic splice enhancer motifs in pre-mRNA, and (ii) to ensure nucleosome positioning on DNA have an impact on amino acid choice and rates of evolution. For both, we review evidence that sequence affected by more than one coding demand is particularly constrained. Strikingly, in mammals, splicing-related constraints are quantitatively as important as expression parameters in predicting rates of protein evolution. These results indicate that there is substantially more to protein evolution than protein functional constraints.

## Introduction

What determines a protein's rate of evolution and its amino acid content? The simplest model supposes that both of these features reflect the protein's biochemical functions. Integral inner membrane proteins, for example, comprise a high proportion of hydrophobic amino acids, diagnostic of their role in the cell [1]. More generally, signature amino acid arrangements can be recognized across proteins (domain architecture), permitting prediction of protein function from sequence alone. One might expect that proteins with a high density of such functional motifs evolve more slowly than those with a low density [2,3]. This model is consistent with the fact that a given protein tends to have a characteristic rate of evolution, such that the amount of protein divergence between, for example, mice and rats can be predicted with some confidence from the rate of evolution of the same protein measured between, say, humans and cows [4]. But does this imply that for any one protein its composition and evolution are solely, or even predominantly, a reflection of what the protein does? In the present article, we briefly review evidence that suggests that the protein function-centred view of gene and protein evolution is too simplistic. We particularly highlight some recent evidence indicating that all through the chain of events, from transcription initiation through splicing to translation, mechanisms necessary for the manufacture of proteins can

and do have an impact on the resulting protein. We start by considering evidence that more general mutational pressures on nucleotide content might also affect amino acid choice.

## Forces that influence genomic nucleotide content can skew amino acid content

Although biochemical function is evidently critical, its power to explain biases in amino acid choice is also strikingly limited. Figure 1 illustrates how amino acid usage across genomes varies, often quite drastically, as a function of intergenic GC content. By implication, genome-wide mutational dynamics or selection on nucleotide content independent of protein function have a considerable impact on protein composition [5,6]. Some of the variation in nucleotide content might be attributable to ecological factors. Unrelated bacteria derived from the same metagenomic sample tend to have more similar nucleotide content than expected by chance [7]. In part, this may reflect the fact that bacteria that reside in oxygen-rich environments have GC-rich genomes [8]. That GC-related forces can skew amino acid usage is also evident within genomes. In mammals, a gene's position within the isochore landscape affects the amino acid usage of the corresponding protein [9], reflected in a correlation between the GC usage at silent third sites in codons and at the first two codon sites (Figure 2).

## Gene expression and rates of evolution

The above results do not challenge the notion that when selection acts directly on a protein's amino acid content it does so to optimize protein function. More provocatively, evidence

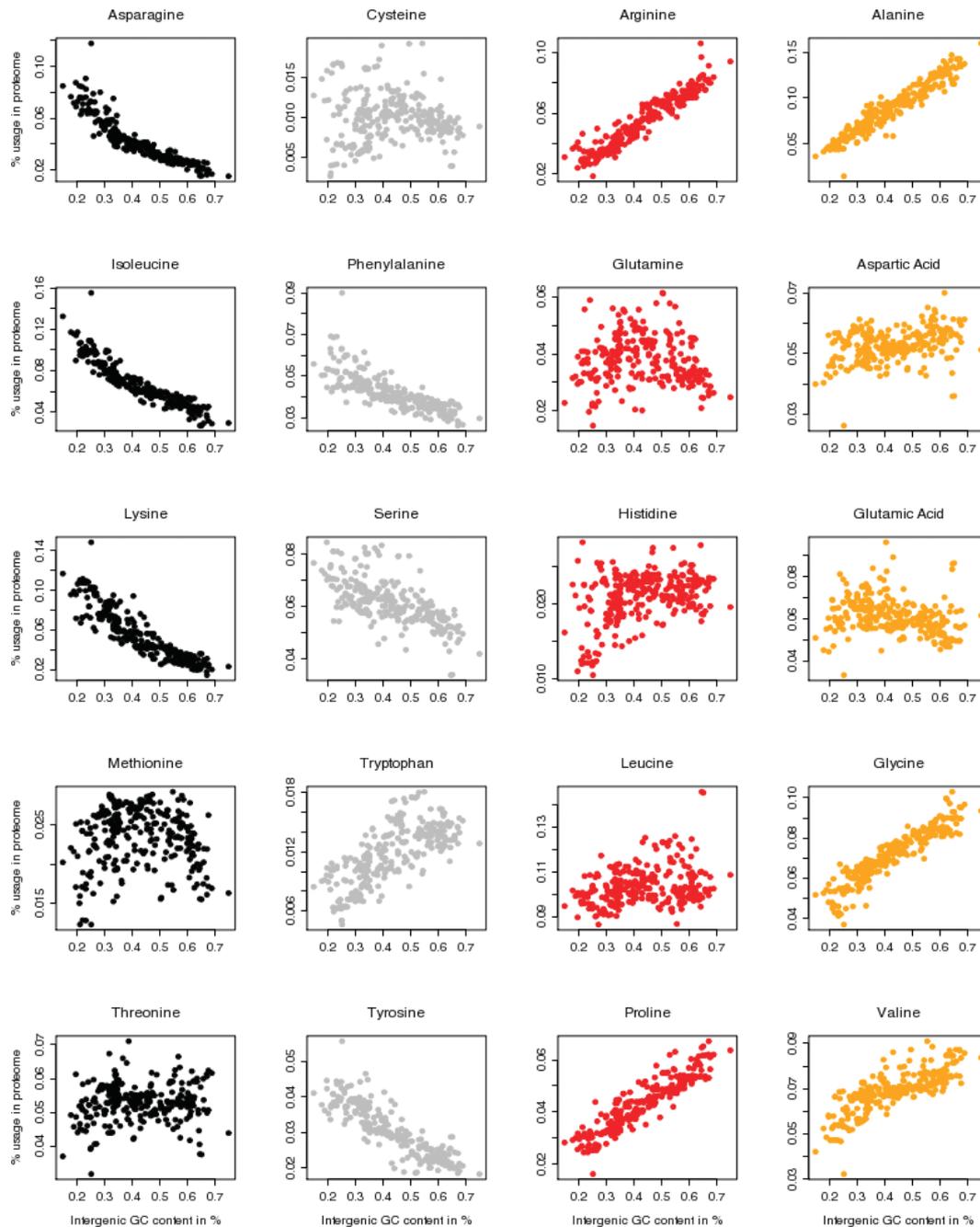
**Key words:** dual-coding sequence, exonic splicing enhancer (ESE), nucleosome, protein evolution, protein function, splicing.

**Abbreviations used:** ESE, exonic splicing enhancer; SNP, single-nucleotide polymorphism.

<sup>1</sup>To whom correspondence should be addressed (email l.d.hurst@bath.ac.uk)

**Figure 1 | Background nucleotide content affects amino acid usage between bacterial genomes**

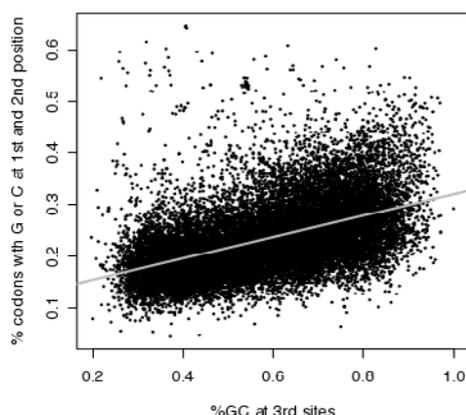
Mean GC usage in intergenic regions is specified on the x-axis for 304 bacterial genomes. Amino acids are grouped according to first codon position: A-- (black), T-- (grey), C-- (red), G-- (orange). Note that amino acids specified by GC-rich codons are highly abundant in GC-rich genomes.



is steadily accumulating that amino acid choice is shaped to a significant extent by selection unrelated to protein function. For example, if protein function was the sole concern to be satisfied, amino acid choice should be agnostic to the fact that amino acids vary substantially in the energy required to assemble them. Yet, in *Escherichia coli* and other organisms, highly expressed genes tend to be encoded by metabolically cheap amino acids [10,11]. Expression parameters have indeed repeatedly been singled out as principal correlates

of protein evolution [3,12]. Certainly, in fast-growing single-celled organisms, they can account for much greater compositional and especially rate variation than protein characteristics more intuitively associated with protein function, such as the protein's dispensability or its centrality in interaction networks [3]. Indeed, the observation that proteins with multiple heterotopic binding partners evolve slowly [13] appears to be fully explained by co-variance with rates of expression [14]. That proteins have characteristic

**Figure 2** | Genes in GC-rich isochores specify amino acids that are specified by GC-rich codons in humans



rates of evolution is then likely to be partly accounted for by proteins having characteristic expression rates.

Selection to avoid translational error and subsequent misfolding as well as minimization of ribosome dwelling times are currently the major contenders to provide a mechanistic explanation for this correlation [15], although this remains an area of active debate. Similarly, the quantitative importance of selection mediated by expression rate compared with selection owing to protein functionality is very much an open issue. In an elegant test, Wolf et al. [16] recently offered a first quantification of the relative contributions that these forces make to shaping rate variation among human proteins. The authors tested whether the evolutionary rate of a focal domain corresponds more closely to that of other domains in the same protein (i.e. with the same translation history) or the same domain in different proteins (putatively shared function) [16]. Their results suggest that both factors make a strong independent contribution of approximately equal weight [16].

### Coding sequence as a complex information carrier

The above observations suggest that selection on amino acid content is not simply dictated by what is best for protein function. They also highlight that, to gain an understanding of protein evolution, we require a solid understanding of the protein's production history. In the case of translational parameters, selection acts on mRNA rather than the protein itself. That mRNA frequently harbours biologically relevant information beyond specifying a string of amino acids is, of course, not a new insight. Selection for favourable mRNA secondary structure, which may affect mRNA half-life or translational properties and therefore have fitness consequences, provides one well-known example [17–19]. Signatures of selection for such added information are typically detected at third sites [19]. Here, the degeneracy of the genetic code affords obvious flexibility in nucleotide choice. Whether, for example, the codon is CCA, CCC, CCG or CCT makes no difference

to the amino acid specified (proline), but may generate secondary structures of differential fitness and hence call selection into play. It is important to realize, however, that mRNA secondary structure concerns ought to be agnostic to reading frame. Third sites, rather than being intrinsically linked to this layer of selection, are simply the loci where, unopposed by selection for a specific amino acid, selection for secondary structure is most easily diagnosed. As a consequence, selection on protein precursors unrelated to protein function is frequently discussed in terms of codon choice, distracting from the fact that there is no *a priori* reason that selection could not act on amino acid content as well. The implicit model here is that information is added into empty slots (third sites) in the substrate that carries protein-coding information. We call this the 'opportunistic model' of multicoding. This model appears reasonable as long as the underlying assumption holds that specifying a certain amino acid is the overriding selective concern. But need this always be the case? Here, we consider two processes that suggest not.

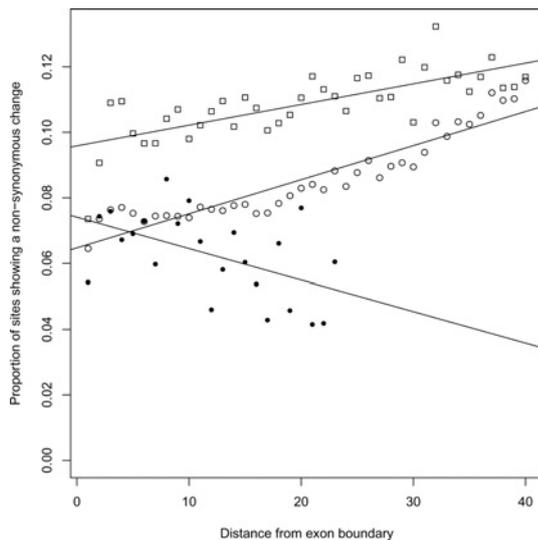
### Selection on pre-mRNA: the case of splicing

Over the last few years, protein composition has turned out to be surprisingly plastic and evolutionary rate sometimes very sensitive to selection acting on parts of the expression process before formation of the mature protein. Particularly startling has been the extent to which splicing-related concerns can drive patterns of protein evolution.

Although classically it was supposed that information for intron removal is almost exclusively intronic, efficient intron removal, particularly in species where introns are long and numerous, frequently requires more information than is provided by the canonical splicing motifs in the intron [5' and 3' splice sites, branch point, poly(Y)-tract]. Enhancer motifs located near splice junctions can attract splice-auxiliary proteins that help the spliceosome to home in on the correct sites and facilitate communication between splice sites [20]. Where introns exceed a certain length, such communication typically occurs across the exon (exon definition) rather than the intron [21]. Consequently, enhancer motifs located in exons [ESEs (exonic splicing enhancers)] are a regular, and critical, feature of intron-containing transcripts in humans and other species [22]. Although they catalyse splicing more efficiently the closer they are to the splice junction, ESEs can act at a variable distance from the intron–exon boundary [23]. In conjunction with the fact that ESE motifs are short and degenerate [22], one might suppose that ESEs are prime candidates to adhere to the opportunistic coding model, i.e. selection may alter third site identity, but largely fits around a given amino acid structure.

As expected, we find a strong bias in codon choice near splice junctions, with codons favoured in ESEs found near junctions [24]. More surprisingly, however, Parmley et al. [25] also reported that the relative abundance of the majority of amino acids changes markedly as one approaches the exon–intron boundary: 4-fold and 2-fold blocks of the same 6-fold degenerate amino acids show divergent behaviour,

**Figure 3** | The proportion of informative sites in intron-containing genes (○, all exons; □, exons >80 codons) and functional retrogenes (●) that show non-synonymous change (human and mouse comparison) as a function of distance from the current (or, for retrogenes, former) intron–exon boundary



demonstrating that the effect is founded at the nucleotide level, but strong enough to affect amino acid choice. A sizeable proportion of amino acids exhibit skewed relative usage near splice junctions and, rather than being confined to mammals, such biases appear to be a general property of intron-rich genomes [26]. Selection operating on splice control elements near intron–exon boundaries is further supported by a rarity of SNPs (single-nucleotide polymorphisms) in such domains [27,28].

Importantly, these effects are not modest in quantitative terms. The proportion of exonic sequence in close vicinity of introns turns out to be as strong a predictor of evolutionary rate as expression breadth, identified previously as a major predictor of evolutionary rates in mammals [25,29]. Accordingly, non-synonymous SNPs are rare in genes with an abundance of sequence near exon–intron junctions [30] and exon cores evolve faster than exon edges (Figure 3), this also being true when comparing parts of exons from the same gene.

### Selection on DNA: chromatin organization affects coding sequence evolution

More recently, we and others have noted that the DNA template of a protein can also be subject to constraints that affect codon and amino acid choices, as well as rates of evolution. In its chromosomal context, most DNA, including protein-coding sequence, is intimately associated with histone proteins, which not only provide a scaffold for organization into chromatin, but also play a vital role in regulating gene expression [31]. Notably, nucleosomes, the fundamental repeat units of chromatin, exhibit a stereotypical architecture at the 5′-end of genes in many species, thereby controlling promoter

accessibility [32–36]. At the same time, nucleosomes show sequence preferences, governed by the differential bending and twisting attributes of the underlying nucleotide string [37]. Relative affinities vary by several orders of magnitude [38]. Although nucleosomes can be coerced on to rigid DNA by remodelling factors, this incurs an energy cost. Where nucleosome position should be stable or readily re-induced, as is the case for the promoter-associated nucleosomes, sequence might evolve to assist functional positioning.

Where promoter regions are close to the coding portion of the gene, as tends to be the case for yeast protein-coding genes, nucleosome positioning concerns might then be expected to affect coding sequence evolution. Consistent with this hypothesis, we found coding sequence bound to and free from nucleosomes to evolve at different rates [39]. Interestingly, nucleosome-free (‘linker’) sequence shows greater constraint. This may, in part, result from linker sequence being more accessible to DNA repair machineries [40], i.e. a neutral effect. However, the effect is most pronounced at the 5′-end of genes, where not only synonymous, but also non-synonymous, rates of evolution vary with nucleosome occupancy. This suggests that an exclusively mutational explanation is incomplete where coding sequence is concerned. It should be noted that the effect on overall rates of evolution are much more modest than those that we observe for splice-related functions.

Codon and amino acid composition also vary between linker and nucleosome-bound sequences [39,41]. This is, of course, what one would predict given the inherent sequence preferences of nucleosomes. However, the above result, that constraint differs as a function of intragenic position, suggests that compositional differences do not merely reflect a passive alignment to the underlying DNA, but imply evolution towards functionally favourable nucleotide composition. These recent findings once again emphasize the importance of understanding a protein’s ontology in the cell, starting from the ultimate DNA template, and provide a timely caveat that the catalogue of factors co-determining coding sequence evolution is unlikely to have yet opened all of its chapters.

### Opportunism and conflict

Although it is now well-documented that selection for functions other than those immediately related to its biochemical role influence protein composition, one might argue that this is insufficient to invalidate the opportunistic model of multiple coding. That amino acid content is responsive to splicing and other forces is equally consistent with the notion that, for a substantial proportion of residues, amino acid choice, like codon choice, is functionally degenerate, i.e. more than one option exists that satisfactorily does the same job. Changes in amino acid composition could hence be ‘functionally/structurally opportunistic’. This is consistent with the observation that high levels of structural homology can be achieved at low levels of sequence identity [42–44]. In addition, individual amino acid replacements appear, on average, to have very little or no selectable fitness

consequence, especially for species with a small effective population size [45].

To challenge the primacy of biochemical function as a coding concern, one needs to demonstrate that an added coding demand, such as specifying an ESE, has led to a change in amino acid composition that has compromised protein function: the 'conflict model' of multiple coding. As regards codon choice (for a given amino acid), one case history strongly supports a conflict view. In *Drosophila*, a set of codons that maximize translation rates and accuracy of translation is well described [46,47]. Importantly, these codons are unlikely to participate in ESEs. Towards the ends of exons in *Drosophila*, we see a striking decrease in the usage of translationally optimal codons and a rise in usage of ESE-related codons: you cannot both splice accurately and be translationally optimal [48].

Perhaps the most suggestive evidence of a similar effect at the amino acid level comes from analysis of retrogenes. The rate of evolution in retrogenes is higher close to where the intron–exon boundary used to be [25] (Figure 3). Similarly, retrogenes that are derived from genes in which a high proportion of the sequence was near exon boundaries (genes with predominantly small exons) tend to have higher overall rates of evolution. The results are consistent with selection towards a more favourable amino acid content after release from splicing-related selection.

Splicing-related selection more generally is an ideal candidate to test whether hypothesized pleiotropic trade-offs actually exist. First, selection on splice-regulatory motifs can, in principle, be sufficiently strong to override protein-functional concerns. Simply speaking, selection might tolerate a protein 'performance hit' if this ensures correct splicing. Secondly, the selective *raison d'être* of ESEs is easily enough abolished: removing the associated intron immediately makes the ESE obsolete. Subsequent tinkering with amino acid identity might then reveal whether an increase in protein performance can be achieved by using amino acids whose nucleotide content cannot be reconciled with ESE function.

Establishing the extent to which protein-coding sequence composition is a product of coding conflict rather than opportunistic accommodation is not only important for a general understanding of protein evolution, but also has direct implications for using real genes as templates to develop proteins for industrial applications and gene therapy, where optimizing protein performance is key. Providing a comprehensive description of the catalogue of coding demands affecting protein-coding sequence is a first crucial step towards this goal. The more complete our understanding of the diverse demands on coding sequence, the better our position to devise critical tests to establish whether trade-offs between these demands, at both the codon and the amino acid level, do exist and how they are negotiated in biological systems.

## Acknowledgements

L.D.H. is a Royal Society Wolfson Research Merit Award Holder.

## Funding

T.W. is funded by the Medical Research Council, and C.C.W. is funded by the University of Bath.

## References

- Nishikawa, K. and Ooi, T. (1982) Correlation of the amino acid composition of a protein to its structural and biological characters. *J. Biochem.* **91**, 1821–1824
- Zuckerandl, E. (1976) Evolutionary processes and evolutionary noise at the molecular level. II. A selectionist model for random fixations in proteins. *J. Mol. Evol.* **7**, 269–311
- Pal, C., Papp, B. and Lercher, M.J. (2006) An integrated view of protein evolution. *Nat. Rev. Genet.* **7**, 337–348
- Williams, E.J. and Hurst, L.D. (2002) Is the synonymous substitution rate in mammals gene-specific? *Mol. Biol. Evol.* **19**, 1395–1398
- Lobry, J.R. and Gautier, C. (1994) Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 *Escherichia coli* chromosome-encoded genes. *Nucleic Acids Res.* **22**, 3174–3180
- Pascal, G., Medigue, C. and Danchin, A. (2005) Universal biases in protein composition of model prokaryotes. *Proteins* **60**, 27–35
- Foerster, K.U., von Mering, C., Hooper, S.D. and Bork, P. (2005) Environments shape the nucleotide composition of genomes. *EMBO Rep.* **6**, 1208–1213
- Naya, H., Romero, H., Zavala, A., Alvarez, B. and Musto, H. (2002) Aerobiosis increases the genomic guanine plus cytosine content (GC%) in prokaryotes. *J. Mol. Evol.* **55**, 260–264
- Sabbia, V., Piovani, R., Naya, H., Rodriguez-Maseda, H., Romero, H. and Musto, H. (2007) Trends of amino acid usage in the proteins from the human genome. *J. Biomol. Struct. Dyn.* **25**, 55–59
- Akashi, H. and Gojobori, T. (2002) Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3695–3700
- Swire, J. (2007) Selection on synthesis cost affects interprotein amino acid usage in all three domains of life. *J. Mol. Evol.* **64**, 558–571
- Drummond, D.A., Raval, A. and Wilke, C.O. (2006) A single determinant dominates the rate of yeast protein evolution. *Mol. Biol. Evol.* **23**, 327–337
- Kim, P.M., Lu, L.J., Xia, Y. and Gerstein, M.B. (2006) Relating three-dimensional structures to protein networks provides evolutionary insights. *Science* **314**, 1938–1941
- Batada, N.N., Reguly, T., Breitkreutz, A., Boucher, L., Breitkreutz, B.J., Hurst, L.D. and Tyers, M. (2007) Still stratus not altocumulus: further evidence against the date/party hub distinction. *PLoS Biol.* **5**, e154
- Drummond, D.A. and Wilke, C.O. (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* **134**, 341–352
- Wolf, M.Y., Wolf, Y.I. and Koonin, E.V. (2008) Comparable contributions of structural-functional constraints and expression level to the rate of protein sequence evolution. *Biol. Direct* **3**, 40
- Chamary, J.V. and Hurst, L.D. (2005) Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. *Genome Biol.* **6**, R75
- Carlini, D.B. (2005) Context-dependent codon bias and messenger RNA longevity in the yeast transcriptome. *Mol. Biol. Evol.* **22**, 1403–1411
- Shabalina, S.A., Ogurtsov, A.Y. and Spiridonov, N.A. (2006) A periodic pattern of mRNA secondary structure created by the genetic code. *Nucleic Acids Res.* **34**, 2428–2437
- Berget, S.M. (1995) Exon recognition in vertebrate splicing. *J. Biol. Chem.* **270**, 2411–2414
- Fox-Walsh, K.L., Dou, Y., Lam, B.J., Hung, S.P., Baldi, P.F. and Hertel, K.J. (2005) The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16176–16181
- Blencowe, B.J. (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem. Sci.* **25**, 106–110
- Graveley, B.R., Hertel, K.J. and Maniatis, T. (1998) A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. *EMBO J.* **17**, 6747–6756
- Parmley, J.L. and Hurst, L.D. (2007) Exonic splicing regulatory elements skew synonymous codon usage near intron–exon boundaries in mammals. *Mol. Biol. Evol.* **24**, 1600–1603

- 25 Parmley, J.L., Urrutia, A.O., Potrzebowski, L., Kaessmann, H. and Hurst, L.D. (2007) Splicing and the evolution of proteins in mammals. *PLoS Biol.* **5**, e14
- 26 Warnecke, T., Parmley, J.L. and Hurst, L.D. (2008) Finding exonic islands in a sea of non-coding sequence: splicing related constraints on protein composition and evolution are common in intron-rich genomes. *Genome Biol.* **9**, R29
- 27 Fairbrother, W.G., Holste, D., Burge, C.B. and Sharp, P.A. (2004) Single nucleotide polymorphism-based validation of exonic splicing enhancers. *PLoS Biol.* **2**, E268
- 28 Carlini, D.B. and Genot, J.E. (2006) Synonymous SNPs provide evidence for selective constraint on human exonic splicing enhancers. *J. Mol. Evol.* **62**, 89–98
- 29 Duret, L. and Mouchiroud, D. (2000) Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. *Mol. Biol. Evol.* **17**, 68–74
- 30 Liu, J., Zhang, Y., Lei, X. and Zhang, Z. (2008) Natural selection of protein structural and functional properties: a single nucleotide polymorphism perspective. *Genome Biol.* **9**, R69
- 31 Henikoff, S. (2008) Nucleosome destabilization in the epigenetic regulation of gene expression. *Nat. Rev. Genet.* **9**, 15–26
- 32 Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R. and Nislow, C. (2007) A high-resolution atlas of nucleosome occupancy in yeast. *Nat. Genet.* **39**, 1235–1244
- 33 Shivaswamy, S., Bhinge, A., Zhao, Y., Jones, S., Hirst, M. and Iyer, V.R. (2008) Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS Biol.* **6**, e65
- 34 Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C. et al. (2008) Nucleosome organization in the *Drosophila* genome. *Nature* **453**, 358–362
- 35 Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., Zeng, K., Malek, J.A., Costa, G., McKernan, K. et al. (2008) A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res.* **18**, 1051–1063
- 36 Schones, D.E., Cui, K., Cuddapah, S., Roh, T.Y., Barski, A., Wang, Z., Wei, G. and Zhao, K. (2008) Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887–898
- 37 Widom, J. (2001) Role of DNA sequence in nucleosome stability and dynamics. *Q. Rev. Biophys.* **34**, 269–324
- 38 Davey, C., Pennings, S., Meersseman, G., Wess, T.J. and Allan, J. (1995) Periodicity of strong nucleosome positioning sites around the chicken adult  $\beta$ -globin gene may encode regularly spaced chromatin. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11210–11214
- 39 Warnecke, T., Batada, N.N. and Hurst, L.D. (2008) The impact of the nucleosome code on protein-coding sequence evolution in yeast. *PLoS Genet.* **4**, e1000250
- 40 Washietl, S., Machne, R. and Goldman, N. (2008) Evolutionary footprints of nucleosome positions in yeast. *Trends Genet.* **24**, 583–587
- 41 Chen, K., Meng, Q., Ma, L., Liu, Q., Tang, P., Chiu, C., Hu, S. and Yu, J. (2008) A novel DNA sequence periodicity decodes nucleosome positioning. *Nucleic Acids Res.* **36**, 6228–6236
- 42 Murzin, A.G., Brenner, S.E., Hubbard, T. and Chothia, C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540
- 43 Bauer, F., Schweimer, K., Kluver, E., Conejo-Garcia, J.R., Forssmann, W.G., Rosch, P., Adermann, K. and Sticht, H. (2001) Structure determination of human and murine  $\beta$ -defensins reveals structural conservation in the absence of significant sequence similarity. *Protein Sci.* **10**, 2470–2479
- 44 Hill, E.E., Morea, V. and Chothia, C. (2002) Sequence conservation in families whose members have little or no sequence similarity: the four-helical cytokines and cytochromes. *J. Mol. Biol.* **322**, 205–233
- 45 Eyre-Walker, A. and Keightley, P.D. (2007) The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* **8**, 610–618
- 46 Akashi, H. (1994) Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. *Genetics* **136**, 927–935
- 47 Vicario, S., Moriyama, E.N. and Powell, J.R. (2007) Codon usage in twelve species of *Drosophila*. *BMC Evol. Biol.* **7**, 226
- 48 Warnecke, T. and Hurst, L.D. (2007) Evidence for a trade-off between translational efficiency and splicing regulation in determining synonymous codon usage in *Drosophila melanogaster*. *Mol. Biol. Evol.* **24**, 2755–2762

Received 5 January 2009  
doi:10.1042/BST0370756