



Efficient inefficiency: Biochemical “junk” may represent molecular bridesmaids awaiting emergent function as a buffer against environmental fluctuation

Anthony J. Yun ^{a,*}, Patrick Y. Lee ^a, John D. Doux ^b

^a *Stanford University, Radiology, 470 University Avenue, Palo Alto, CA 94301, United States*

^b *Palo Alto Institute, Palo Alto, CA 94301, United States*

Received 3 January 2006; accepted 1 February 2006

Summary The biochemical function of many parts of the genome, transcriptome, proteome, and interactome remain largely unknown. We propose that portions of these fundamental building blocks of life have no current biochemical function per se. Rather, sections of these “omes” may contribute to an inventory of biochemical parts and circuits that participate in the development of emergent functions. Low fidelity deoxyribonucleic acid replication, transcription, translation, and post-translational modification all represent potential mechanisms to produce an inventory of parts. Stochastic processes that influence the conformations of ribonucleic acid molecules and proteins may also contribute to potential biochemical inventory. Some components of the biochemical inventory may enable future adaptations, some may produce disease, and some may remain useless. The function of many of these components await discovery, not by science, but by evolution. While carrying such purposeless biochemical units may appear to dilute fitness by exacting a thermodynamic cost, we argue that net fitness becomes enhanced when considering the value for potential future innovations. One can envision components that intermingle, interact, and act out mock pathways, but in most cases remain molecular bridesmaids. Given sufficiently low thermodynamic cost, such stochastic cycling may persist until a markedly advantageous or cataclysmically disadvantageous trait emerges. Maladaptive screening and utilization of inventory content can lead to disease phenotypes, a process buffered and regulated in part by the heat shock protein and stress response network. Whereas failure of the ubiquitin pathway to recycle misfolded proteins has become increasingly recognized as a source of disease, protein misfolding may itself represent one step in a process that maximizes functional innovation through increasing proteomic diversity. Fractal correlates of these processes occur at the organizational level of cells and organisms. That the abnormal accumulation of units induces local collapse may serve to limit the extension of damage to the greater system at large. The immune and cognitive systems that selectively sample and prune environmental content may serve as additional portals for innovation.

© 2006 Elsevier Ltd. All rights reserved.

* Corresponding author. Tel.: +1 650 387 6667; fax: +1 650 325 5028.
E-mail address: ayun@stanford.edu (A.J. Yun).

Hypothesis

Energetic efficiency constitutes an important driver of evolution. We propose that evolution also selects for designs that exhibit robust performance of existing biochemical functions and the potential for adaptive generation of novel functions. Such design features provide a buffer against environmental fluctuations and allow organisms to adaptively fill environmental niches. This hypothesis implies that organisms may harbor a large content of biochemical inventory with no specified function. This biochemical inventory may be carried in the genome, transcriptome, proteome, or on a synthetic basis in the interactome. During the life span of an organism and across generations, novel functions may emerge within these broad categories that either enhance or reduce fitness. In humans, such emergent functioning of biochemical inventory may contribute to both health and disease.

Evidence

The genome: current junk as future innovations

The function of many parts of the genome currently remains unknown [1]. More than 90% of the human genome has been designated as “junk” deoxyribonucleic acid (DNA) whose function remains unknown [2–5]. With the notable exception of certain species of frogs and amoebae, most organisms possess a genome of smaller size than that found in humans. However, size does not necessarily connote actively expressed content: the puffer fish (*fugu*) genome constitutes one-tenth the size of the human genome, yet appears to have a relatively similar number of genes [6].

Given the competition for efficiency, the presence of useless molecular units in many higher eukaryotes in the form of DNA seems unusual, since each additional base pair of DNA exacts an obligate thermodynamic cost that would seem to impair efficiency and fitness. In our model, “junk” DNA would contribute to the inventory of current biochemical parts available to generate emergent functions. In terms of fitness, as long as the net present value of future utility of these parts exceeds their current carrying cost, organisms can justify carrying currently useless “junk”. This notion is supported by evidence that genome size inversely correlates with evolutionary rate [7]; in our model the extra content carried in larger genomes

would obviate the need for more dramatic evolutionary drift and speciation. Our model would predict that mature, stable environmental niches would favor organisms that are maximizing current efficiency and living in the moment over those that are carrying excess “junk” and not maximizing current utility. On the other hand, less mature unstable environmental niches with emergent stress or opportunity may favor those organisms which appear inefficient and which carry an excess inventory of “junk” parts over those that seemingly maximize current efficiency. In either case, the net present value of the future stream of fitness gains remains a useful metric for analysis.

Apparently non-functional regions of DNA can serve as repositories for content in several ways. The genetic sequence directly regulates DNA replication in part [8]. Although DNA replication generally occurs under exacting content and chronologic control [9], genetic shift and drift may also constitute regulated events. Dedicated sections of the genome may control the probability that micro and macro changes in the genome occur. Indeed, certain regions of the genome exhibit higher rates of sequence diversity than other regions [10], with immunoglobulin gene recombination and somatic mutation constituting an extreme example [11,12]. On a different time scale, speciation and chromosome rearrangement may represent a programmed phenomenon. In cancer, control of this process becomes subverted, thereby accelerating the pace of evolution. Therefore, the vast potential content carried by non-coding portions of the genome may contribute to the complexity of the cancer phenotype.

The genome also directly regulates RNA production. Genomic DNA serves as a docking station for chromatin, for transcription factor binding, and for the assembly of transcription complexes [13–15], all of which play central roles in controlling RNA production and processing. Sophisticated folding and looping machinery is also embedded in genomic sequence [16]. While certain binding sites and positioning elements play critical and non-redundant roles in regulating gene expression, organism viability may not necessarily require many features of gene expression. For example, variant histones are positioned at specific sites in yeast and higher organisms [17–19]; however, yeast strains harboring deletions in some histone subtypes remain viable, suggesting that the histone positional matrix as directed by genomic sequence carries dispensable information [20]. Organisms may use such optional material in times of environmental stress or in the development of new pathways that facilitate tailoring to niches. A recent

argument has posited that adaptive evolution in *Drosophila* may utilize non-coding DNA [21].

DNA can also function directly as an epigenetic substrate [22,23]. Although epigenetic modifications can influence gene expression and replication timing, to what extent this form of information has undergone pre-selection remains unknown. Environmental influences can alter methylation patterns in mice – which then pass on to their progeny [24–26]. External forces can also alter the methylation patterns of plant DNA [27,28]. DNA sequences thus may exert an emergent regulatory influence on actively expressed genes. As such, these sequences may contain unspecified information that awaits environmental instruction. Although each potentially methylated region of DNA may have withstood a high degree of selection to confer specific regulatory information in certain environments, novel environments that modify methylation patterns would clearly enable the DNA sequences involved to provide new, context dependent instructions to the cell. The intriguing possibility remains that methylation emerged as a hydrocarbon signature to endow an additional dimension of functional plasticity in response to changing energy conditions.

The transcriptome: a world of unrealized content

The ribonucleic acid (RNA) transcriptome contains an enormous amount of biologic content. Traditional considerations have presumed RNA to function predominantly as a template for protein production. However, increasing evidence suggests that RNA processing affords a major opportunity to generate additional biochemical function. Portions of unprocessed RNA molecules modulate splicing, RNA half-life, RNA–protein interactions, and RNA localization [29–33]. Each of these processes offers a substantial opportunity to diversify content.

Splicing may afford the most compelling example of this principle. The control of alternate splicing has become the subject of intense investigation [34,35]. For the majority of coding units, research has not fully explained the contribution of alternate splicing to pools of messenger RNA. In some cases, careful investigation of alternate splicing patterns for a given gene has led to identification of a surprising amount of alternate spliceoforms. In *Drosophila* the *dscam* locus can yield over thirty thousand alternate RNA moieties implicated in the generation of both immune and neuronal diversity [36–38]. Indeed, random alternate splicing from a gene with 17 exons could theoretically give rise to more than 100,000 distinct proteins.

Recent discoveries have shown that an enormous amount of RNA transcribed from DNA does not undergo subsequent translation into proteins [39–41]. The functioning of these so-called “non-coding” RNAs remains under study. Larger non-coding RNAs that undergo seemingly random transcription could potentially form functional structures that participate in defined biochemical reactions. Alternatively, they may contribute random biospace content that participates in expanding the interactome, the conceptual space which encompasses all reactions that occur within a system. Certain environmental circumstances may compel reinforcement of nodes within the interactome that utilize non-coding RNAs as adaptive molecular bridges. Epigenetic stabilization of the transcription of a given non-coding RNA moiety may also occur as part of a selective process.

Interest has recently developed in engineering RNA aptamers that recognize small molecules bound to proteins [42]. The feasibility of this approach as reverse engineering of the cataloging process for content suggests that non-coding RNAs may play a major role in responding to novel compounds encountered in the environment. Previous work has implied that non-coding RNA can function as a primitive immune system with recognition of nucleic acid sequences in viruses [43]. However, our model does not limit the scope of RNA based recognition to other nucleic acids. Processes such as reverse transcription may theoretically enable RNA-based molecular recognition to incorporate more durable paradigms of response into the genome itself [44].

The proteome: heterogeneous folding promotes diversity

Ample evidence supports our hypothesis with regards to the proteome. While certain proteins with well defined biochemical functions are highly conserved across many species, the specific roles of many other proteins remain unknown. Although in many cases purpose may eventually become apparent, large parts of the proteome may not possess a current biochemical function per se. As certain sequence polymorphisms promote alternate protein structures, folding variability has increasingly become considered as dysfunctional, as recognition of its role as a source of human disease increases [45–47]. Deleterious misfolds that escape surveillance may represent occasional maladaptive byproducts of an otherwise robust process that maximizes functional innovation through increasing proteomic diversity.

In addition to obligate thermodynamic and alternate sequence contributions to heterogeneous protein folding, in some cases variability in protein folding may undergo adaptation to provide a source for greater proteomic diversity. Recent data suggests an unexpectedly high α -galactosidase A misfolding rate among normal individuals. A recent clinical trial sought to assess the safety of a chaperone drug designed to help patients suffering from alpha-galactosidase A deficiency due to misfolding. Normal volunteers also showed marked increase in α -galactosidase A levels with use of the drug, suggesting the tantalizing possibility that a high level of baseline misfolding may exist in many if not all proteins of normal individuals. Misfolding may serve an adaptive function when the net present value of fitness associated with diversity-enabled future innovations exceeds the present carrying costs of the misfolds and any associated diseases.

Sequence diversity within exons can also manifest as protein diversity. Work by Lindquist and colleagues has shown that systems exist to buffer potentially emergent properties of mutations or sequence changes via the heat shock protein (Hsp) system [48]. Times of cellular stress provoke release of this silent additional content. Natural selection will then presumably promote retention of buffered mutations that confer survival advantage under certain conditions. Large capacitance within the buffer systems themselves would also prove advantageous under rapidly changing environmental conditions. In the context of an appropriate deployment system, this capacitance represents a proteomic form of content inventory awaiting function.

The potential for enzymatic modification of protein substrates may serve as another source of combinatorial proteomic diversity. Although certain protein modifications function as critical molecular switches in response to signals, these alterations may represent the exception rather than the rule. Indeed, although multiple sites for phosphorylation, sumoylation, glycosylation, acetylation, palmitoylation, and/or ubiquitination exist in many proteins, defined biologic function for these domains in many specific contexts remains unidentified [49,50]. These malleable domains may have yet to undergo recruitment into a functional biochemical activity. If random changes confer an advantage under specific environmental conditions, they may become linked to specific signaling events. Potential flexibility in protein domains thus provides additional biochemical inventory.

Proteins, misfolded and otherwise, undergo proteolysis by the ubiquitin-proteasome system (UPS), with recycling of the yielded components. As with

misfolding itself, failings of this recycling system may also participate in the pathogenesis of disease [51]. Whether proteolytic churning relies on any particular regime to select among proteins remains unclear. One might anticipate that an intelligent selection regime would confer significant selection advantage.

The interactome: a universe of functional inventory

An alternative conceptual paradigm for inventory involves considering biochemical function as a matrix known as the interactome that encompasses all interactions between biochemical units in a cell. That the interactome contains interactions that have no defined purpose represents a corollary of the notion that the genome, transcriptome and proteome harbor “extra” biochemical units. Although regulation of putative deleterious interactions may occur via modulation of constituent parts or additional control systems such as shuttles and degradation, random interactions could also modify cellular phenotype. Previous work has proposed that alternate splicing of RNA can serve as a network control element, whereby distinct interactomes can arise in response to a stimulus at the level of RNA processing [44]. A stochastic component to interactome selection at any level would potentially allow for adaptive evolution.

One potentially intriguing class of interactions involves an expansion of the concept of prions. A prion is a protein conformation that can induce a similar folding conformation in other identical proteins [52,53]. Thus, a prion suffices to catalyze a protein phenotype to pass between cells. Generalization of this concept leads to consideration of auto-regulatory loops that depend on the presence of a given protein in a given conformation. For example, suppose a protein X exists in two possible conformations A and B. Conformation A self-induces more expression of Protein X. Conformation A, which possesses energetically unfavorable features and occurs only in acidic conditions, can become catalytically induced in an emergent fashion at normal pH by the environmental presence of another molecule. In this model, simple exposure to conformation A will stably increase expression of Protein X. If the inductive molecule remains unprocessed and persists within the cell, unrestrained production of protein X may occur, with resultant deleterious effects. Cell death might then release the molecule; uptake by neighboring cells may then induce another round of unrestrained protein X production, albeit amplified.

Food additives and drugs may represent particular classes of agents which function in this fashion as inductive molecules.

For complex organisms to process proteins at such an inefficient rate, and for promiscuous interactions to occur between proteins within the biospace, incurs not only a large thermodynamic cost but also a risk of disease. We postulate that this arrangement represents a type of Faustian bargain – protein misfolding, along with other facets of DNA, RNA and protein production and processing that exhibit low fidelity, can then serve as a rich source of component diversity. The generation of variety may not yield substantial or immediate benefits in terms of new functions, and may even predispose to disease. However, as long as the net present value of future innovations exceeds the current carrying costs, such processes will persist and continue to create diversity. One can envision misfolded proteins as components that intermingle, interact, and act out mock pathways. In most cases, such random associations, driven by thermodynamic conditions, lead to inconsequential outcomes, and the components remain molecular bridesmaids. However, given the low cost of having such processes continue, such stochastic cycling may persist until a markedly advantageous or catastrophically disadvantageous trait emerges.

Implications

We propose that the sorting and recycling of extra content may both promote diversity and give rise to disease in an inherently stochastic process. This assertion carries three significant sets of implications: one with respect to the fractal consideration of the nature of information in systems at all levels, and two concerned specifically with the function of the immune and cognitive systems.

Fractals: diversity, pruning, escape, and systemic stress

Fractals are patterns that exhibit self-similarity at multiple scales [54]. The roster of processes that comprise the life of a protein may exhibit fractal counterparts at the scales of both cells and organisms, namely humans.

We have proposed that protein misfolding merely constitutes part of a larger drive for proteomic diversity. The UPS may serve to both prune this diversity of proteins and promote the ongoing recycling of its constituent parts. When proteins escape the certainty of proteolysis, they can accumulate

in the cell. Accumulating proteins can occupy space, overconsume resources at the expense of the cell, and cause systemic cellular stress as well as stress in the endoplasmic reticulum [55]. When combined with inappropriate distribution of proteins outside of the contexts in which they appropriately function, this stress can manifest in cellular dysfunction or demise, and cause disease. Examples include amyloid and Alzheimer's disease, in which beta pleated sheets accumulate in cells and induce neuronal degeneration. Indeed, aging may manifest as a stochastic dysfunction, as time allows content to adopt progressively dysfunctional forms. In that regard, if stress associated with mass production of a single protein promotes alternate folding, protein based therapeutics generated through cell based over-expression systems may contain unexpected diversity [56]. Idiopathic reactions to protein based therapies may occur through the action of or as a reaction to rare alternate protein isoforms.

Like protein folding, cellular differentiation generates tremendous diversity. Programmed cell death, or apoptosis, serves the pruning and recycling function assumed by the UPS at the protein level. In a similar manner to the phenomena seen with proteins, when cells evade apoptosis, they can accumulate, deplete resources at the expense of the host, and induce systemic stress. In this case, this progression describes the process of malignant transformation, also known as cancer. Compounds that induce mutation may do so indirectly by activating stress responses that cause infidelity in DNA replication, which can produce both adaptive and maladaptive results. Cancer can then proceed to disseminate and evade therapy through stochastic sorting via content inventory. As in the case of cancer, accelerating cycles of replication effectively increases the frequency of such sorting events.

Recent human history may follow a similar path, albeit at a larger scale. Growth and differentiation produces diversity among individuals. All individuals eventually succumb to death, described previously as an autocatalytic, self-terminating process [57]. Whereas the human life-cycle previously harbored close ties to nature, due to modern developments in science and technology, individuals may have begun to separate their existence from environmental cues and persist beyond what previously defined the limits of their biologic destiny. The delayed age of procreation among individuals who perceive a longer lifespan may effectively slow the rate of replacement and partially serve as a natural control system for population growth. Nonetheless, due to longevity and other factors,

humans have also begun to engage in behaviors of accumulation and overconsumption. As for dissemination, human civilization may have arisen in Mesopotamia, but has since assumed wide distribution in multiple forms and guises, not all of which necessarily promote systemic health. Indeed, as a result of this behavioral trajectory, the planet may have become subjected to increasing degrees of wholesale stress, a subject which has recently undergone closer scrutiny as a potential source of systemic collapse [58].

Systemic collapse at one level may ultimately protect the system at a higher level. The death of cells stressed by the accumulation of selfish proteins may protect the organism. Indeed, inhibiting the UPS through proteasome inhibitors, which would promote protein accumulation, appears to induce cell apoptosis by inhibiting Bcl-2, an anti-apoptotic protein [59]. Similarly, individual organisms that undergo stress through the accumulation of immortal cells also ultimately succumb to cancer, and in doing so temper population growth. The collapse of a system in response to stress engendered by lower order components may work as a control system to prevent the extension of stress to higher levels of organization.

The immune system: a sieve for innovation

Ample evidence exists that many of the key genes held in the inventory of the human genome, such as mitochondrial genes, came from other organisms, both directly and indirectly through vectors [60]. The genome may have also acquired many of the apparent non-coding sequences from microorganisms and viruses. Indeed, microorganisms may function as one of the most efficient sources of genetic innovation for the host. The relationship of the host to the microbial and viral world may prove more commensal and symbiotic than previously understood; the destruction of microbes by the host may only represent a specific pruning response when the relationship turns sour. Content recognition guides the most basic form of immune function in bacteria – namely, through restriction endonuclease mediated recognition specific sequences of DNA – and this action also leads to exchange and acquisition of content in the form of the resultant cleaved strands.

Traditionally viewed as primarily a system of defense, the immune system may thus merit more proper consideration as another sensory organ, one that functions to sample the environment. The primary function of the immune system may involve gathering external information and acting

as a selective filter to assimilate external innovations into the host genome and proteome, both through selective modulation and amplification of existing content, as well as incorporation of novel content. In accordance with how other systems fit into this conceptual framework, induction of tolerance may in fact represent the primary mandate of the immune system – activating specific elements in the inventory of content so as to generate a complementary diversity to fit the exigencies of a particular spectrum of externalities. The development of asthma in individuals raised in specific environments may reflect an incomplete generation of appropriate diversity so as to contend with future scenarios. The defense function of the immune system may simply represent a fortuitous manifestation of the pruning process, modulating incumbent content so as to definitively exclude potentially harmful external entities from infiltrating the catalog of parts. This line of thinking suggests that the use of immunomodulators may either promote or interfere with the sieving process for microbial innovations that could undergo integration into the host.

In the world of the virtual network, humans have chosen to implement anti-viral strategies that mimic the immune system model in many ways – specifically through recognition and subsequent eradication of certain stretches of code that suggest the presence of an unwanted virus. However, by only recognizing and emulating the defense function of the immune system, current anti-viral solutions fail to account for the potential value gained by sampling possible innovations residing in the viral code that the network might otherwise acquire and repurpose for its own benefit. In the same fashion that they have developed in the realm of organic biology, perhaps host operating systems and immune systems may eventually emerge in the realm of silicon-based life that will also selectively sample the content landscape of the network in search of useful innovations.

The cognitive system: new Frontier for innovation

The emergence of complex cognitive systems has enabled memes to become attractive alternatives to biomolecules as vehicles for innovative traits. Nowadays humans store memes both in their own brains as well as in external storage systems such as books and communication networks. The low thermodynamic costs of acquisition, storage, and processing associated with memes allow hosts to carry excess inventory to a far greater degree than

would be the case with conventional biomolecular traits. As long as the net present value of future benefits exceeds their current carrying costs, apparent “junk” ideas can be retained for possible future functions. Pursuing science for science’s sake, even without demonstrable value for immediate applications, exemplifies this phenomenon, as scientific investigation has generally progressed with the idea of maximizing the net present value of future utility as opposed to necessarily maximizing current utility.

Nevertheless, the dramatic proliferation of content due to technical innovations has also resulted in noise proliferation and information oversupply. Both information oversupply and illegitimate signaling can produce maladaptive host responses including obesity, metabolic dysfunction, behavioral disorders, and inflammatory conditions [61]. Thus, in the current environment of content proliferation, effective pruning is gaining in importance as a trait. In the brain, synaptic pruning begins at least as early as the neonatal period and continues throughout childhood [62]. Ubiquitin, which regulates memory deletion in neuronal synapses, may participate in the pruning of memes acquired from the environment [63]. On the other hand, dysfunctional accumulation of protein content can produce Alzheimer’s disease, which impairs meme storage and retrieval. Selective forgetting and ignoring may emerge as more valuable traits in the long run than the ability to perceive and remember.

Acknowledgement

Special thanks to Anthony N. Gerber, M.D., Ph.D. for helpful comments and suggestions.

References

- [1] Gardiner K. Human genome organization. *Curr Opin Genet Dev* 1995;5:315–22.
- [2] Zuckerkandl E. Revisiting junk DNA. *J Mol Evol* 1992;34:259–71.
- [3] Chiaromonte F, Weber RJ, Roskin KM, Diekhans M, Kent WJ, Haussler D. The share of human genomic DNA under selection estimated from human-mouse genomic alignments. *Cold Spring Harb Symp Quant Biol* 2003;68:245–54.
- [4] Waterston RH, Lindblad-Toh K, Birney E, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature* 2002;420:520–62.
- [5] Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921.
- [6] Elgar G. Quality not quantity: the pufferfish genome. *Hum Mol Genet* 1996;5:1437–42.
- [7] Knight CA, Molinari NA, Petrov DA. The large genome constraint hypothesis: evolution, ecology and phenotype. *Ann Bot (Lond)* 2005;95:177–90.
- [8] Zannis-Hadjopoulos M, Price GB. Eukaryotic DNA replication. *J Cell Biochem* 1999;73:1–14.
- [9] Aladjem MI. The mammalian beta globin origin of DNA replication. *Front Biosci* 2004;9:2540–7.
- [10] Raynard SJ, Baker MD. *cis*-acting regulatory sequences promote high-frequency gene conversion between repeated sequences in mammalian cells. *Nucleic Acids Res* 2004;32:5916–27.
- [11] Maizels N. Immunoglobulin gene diversification. *Annu Rev Genet* 2005;39:23–46.
- [12] Seki M, Gearhart PJ, Wood RD. DNA polymerases and somatic hypermutation of immunoglobulin genes. *EMBO Rep* 2005;6:1143–8.
- [13] Zaidi SK, Young DW, Choi JY, et al. The dynamic organization of gene-regulatory machinery in nuclear microenvironments. *EMBO Rep* 2005;6:128–33.
- [14] Svejstrup JQ. The RNA polymerase II transcription cycle: cycling through chromatin. *Biochim Biophys Acta* 2004;1677:64–73.
- [15] Remenyi A, Scholer HR, Wilmanns M. Combinatorial control of gene expression. *Nat Struct Mol Biol* 2004;11:812–5.
- [16] Vilar JM, Saiz L. DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Curr Opin Genet Dev* 2005;15:136–44.
- [17] Chow CM, Georgiou A, Szutorisz H, et al. Variant histone H3.3 marks promoters of transcriptionally active genes during mammalian cell division. *EMBO Rep* 2005;6:354–60.
- [18] Guillemette B, Bataille AR, Gevry N, et al. Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. *PLoS Biol* 2005;3:e384.
- [19] Raisner RM, Hartley PD, Meneghini MD, et al. Histone variant H2A.Z marks the 5’ ends of both active and inactive genes in euchromatin. *Cell* 2005;123:233–48.
- [20] Kolodrubetz D, Rykowski MC, Grunstein M. Histone H2A subtypes associate interchangeably in vivo with histone H2B subtypes. *Proc Natl Acad Sci USA* 1982;79:7814–8.
- [21] Andolfatto P. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* 2005;437:1149–52.
- [22] Holliday R. The inheritance of epigenetic defects. *Science* 1987;238:163–70.
- [23] Jablonka E, Lamb MJ. The inheritance of acquired epigenetic variations. *J Theor Biol* 1989;139:69–83.
- [24] Cooney CA, Dave AA, Wolff GL. Maternal methyl supplements in mice affect epigenetic variation and DNA methylation of offspring. *J Nutr* 2002;132:2393S–400S.
- [25] Rakyan VK, Chong S, Champ ME, et al. Transgenerational inheritance of epigenetic states at the murine *Axin* (*Fu*) allele occurs after maternal and paternal transmission. *Proc Natl Acad Sci USA* 2003;100:2538–43.
- [26] Morgan HD, Sutherland HG, Martin DI, Whitelaw E. Epigenetic inheritance at the *agouti* locus in the mouse. *Nat Genet* 1999;23:314–8.
- [27] Steward N, Ito M, Yamaguchi Y, Koizumi N, Sano H. Periodic DNA methylation in maize nucleosomes and demethylation by environmental stress. *J Biol Chem* 2002;277:37741–6.
- [28] Kakutani T. Epi-alleles in plants: inheritance of epigenetic information over generations. *Plant Cell Physiol* 2002;43:1106–11.
- [29] Bentley D. The mRNA assembly line: transcription and processing machines in the same factory. *Curr Opin Cell Biol* 2002;14:336–42.
- [30] Gott JM, Emeson RB. Functions and mechanisms of RNA editing. *Annu Rev Genet* 2000;34:499–531.

- [31] Zaphiropoulos PG. Mechanisms of pre-mRNA splicing: classical versus non-classical pathways. *Histol Histopathol* 1998;13:585–9.
- [32] VanDongen AM, VanDongen HM. Effects of mRNA untranslated regions on translational efficiency of NMDA receptor subunits. *Neurosignals* 2004;13:194–206.
- [33] Carmo-Fonseca M, Custodio N, Calado A. Intranuclear trafficking of messenger RNA. *Crit Rev Eukaryot Gene Expr* 1999;9:213–9.
- [34] Fededa JP, Petrillo E, Gelfand MS. A polar mechanism coordinates different regions of alternative splicing within a single gene. *Mol Cell* 2005;19:393–404.
- [35] Black DL. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 2003;72:291–336.
- [36] Schmucker D, Clemens JC, Shu H. Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* 2000;101:671–84.
- [37] Neves G, Chess A. Dscam-mediated self- versus non-self-recognition by individual neurons. *Cold Spring Harb Symp Quant Biol* 2004;69:485–8.
- [38] Watson FL, Puttmann-Holgado R, Thomas F. Extensive diversity of Ig-superfamily proteins in the immune system of insects. *Science* 2005;309:1874–8.
- [39] Kampa D, Cheng J, Kapranov P. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res* 2004;14:331–42.
- [40] Huttenhofer A, Schattner P, Polacek N. Non-coding RNAs: hope or hype. *Trends Genet* 2005;21:289–97.
- [41] Costa FF. Non-coding RNAs: new players in eukaryotic biology. *Gene* 2005;357:83–94.
- [42] Plummer KA, Carothers JM, Yoshimura M, Szostak JW, Verdine GL. In vitro selection of RNA aptamers against a composite small molecule–protein surface. *Nucleic Acids Res* 2005;33:5602–10.
- [43] Forsdyke DR, Madill CA, Smith SD. Immunity as a function of the unicellular state: implications of emerging genomic data. *Trends Immunol* 2002;23:575–9.
- [44] Herbert A, Rich A. RNA processing in evolution. The logic of soft-wired genomes. *Ann NY Acad Sci* 1999;870:119–32.
- [45] Gregersen N, Bolund L, Bross P. Protein misfolding, aggregation, and degradation in disease. *Mol Biotechnol* 2005;31:141–50.
- [46] Lee C, Yu MH. Protein folding and diseases. *J Biochem Mol Biol* 2005;38:275–80.
- [47] Selkoe DJ. Cell biology of protein misfolding: the examples of Alzheimer's and Parkinson's diseases. *Nat Cell Biol* 2004;6:1054–61.
- [48] Sangster TA, Lindquist S, Queitsch C. Under cover: causes, effects and implications of Hsp90-mediated genetic capacitance. *Bioessays* 2004;26:348–62.
- [49] Wohlschlegel JA, Johnson ES, Reed SI, Yates 3rd JR. Global analysis of protein sumoylation in *Saccharomyces cerevisiae*. *J Biol Chem* 2004;279:45662–8.
- [50] Fukuta K, Matsumoto K, Nakamura T. Multiple biological responses are induced by glycosylation-deficient hepatocyte growth factor. *Biochem J* 2005;388:555–62.
- [51] Zhang K, Kaufman RJ. The unfolded protein response. A stress signaling pathway critical for health and disease. *Neurology* 2006;66:S102–9.
- [52] Johnson RT. Prion diseases. *Lancet Neurol* 2005;4:635–42.
- [53] Coughlan CM, Brodsky JL. Use of yeast as a model system to investigate protein conformational diseases. *Mol Biotechnol* 2005;30:171–80.
- [54] Mandelbrot BB. The fractal geometry of nature. New York: WH Freeman; 1983.
- [55] Kim R, Emi M, Tanabe K, Murakami S. Role of the unfolded protein response in cell death. *Apoptosis* 2006;11:5–13.
- [56] Mattanovich D, Gasser B, Hohenblum H, Sauer M. Stress in recombinant protein producing yeasts. *J Biotechnol* 2004;113:121–35.
- [57] Yun AJ, Lee PY, Doux JD, Conley BR. A general theory of evolution based on energy efficiency: its implications for diseases. *Med Hypotheses* 2006;66:664–70.
- [58] Diamond J. Collapse. How societies choose to fail or succeed. New York: Viking; 2004.
- [59] Lara Jr PN, Bold RJ, Mack PC, Davies AM, Gumerlock PH, Gandara DR. Proteasome inhibition in small-cell lung cancer: preclinical rationale and clinical applications. *Clin Lung Cancer* 2005;7:567–71.
- [60] Margulis L, Sagan D. Acquiring genomes: a theory of the origins of species. New York: Basic Books; 2003.
- [61] Bazar KA, Yun AJ, Lee PY, Daniel SM, Doux JD. Obesity and ADHD may represent different manifestations of a common environmental oversampling syndrome: a model for revealing mechanistic overlap among cognitive, metabolic, and inflammatory disorders. *Med Hypotheses* 2006;66:263–9.
- [62] Kesler SR, Met LR, Vohr B, Pajot SK, Schneider KC, Katz KH, et al. Volumetric analysis of regional cerebral development in preterm children. *Pediatr Neurol* 2004;31:318–25.
- [63] Bingol B, Schuman EM. Synaptic protein degradation by the ubiquitin proteasome system. *Curr Opin Neurobiol* 2005;15:536–41.

Available online at www.sciencedirect.com

