Science and Society

Evolution, language and analogy in functional genomics

Almost a century ago, Wittgenstein pointed out that theory in science is intricately connected to language. This connection is not a frequent topic in the genomics literature. But a case can be made that functional genomics is today hindered by the paradoxes that Wittgenstein identified. If this is true, until these paradoxes are recognized and addressed, functional genomics will continue to be limited in its ability to extrapolate information from genomic sequences.

Those who ask 'What is the function of my protein?' expect a linguistic answer[1], a sentence or two written in the language of the biologist. The answer might take, for example, the form: 'your protein is a leptin, which regulates the feeding behavior of mice. When the gene is mutated or deleted, the mouse becomes obese'[2].

How does one get such a linguistic construct from a genomic sequence, which is no more (and no less) than a chemical formula for an organic molecule? This question, central to contemporary FUNCTIONAL GENOMICS (see Glossary), is not easy to answer. The simpler task of predicting how the MOLECULAR BEHAVIOR (not function) of an organic molecule is determined by its structure remains one of the great unsolved problems in chemistry. In principle, we should be able to solve this problem. The First Law of Chemistry states that the behavior of all matter is determined by the behavior of its constituent molecules, even behavior that a biologist might observe and call a phenotype. However, the problem has not been solved convincingly for any but the simplest of molecules, and we are far from solving it for the general molecule, let alone a protein.

And even if we could do so, behavior would not necessarily lead to a statement about function. For example, it might become predictable that the benzodiazapene receptor binds tightly to Valium. But the implied statement, 'the purpose of this receptor is to bind to Valium', is transparently misderived because Valium is synthetic. To go from molecular behavior to organismal fitness, which is the darwinian definition of function, information is required about the entire organism and the entire ecosystem.

'Functional equivalency'

To obtain functional annotation, contemporary bioinformatics generally attempts to relate chemical sequence to biological fitness using a doctrine of functional EQUIVALENCY (e.g. Ref. 3). This doctrine seeks to write a linguistic construct for a new protein sequence by expropriating the linguistic construct from another sequence having a similar chemical structure, under the assumption that the two proteins with similar chemical structures have equivalent functions. A protein with unknown function is found in one genome. It is inferred, from its sequence similarity, to be homologous to a different protein found in a different organism. Homologous proteins are then assumed to have equivalent functions. The functional language assigned to the protein with the known function is then transferred to the new protein.

Long before the genomics revolution began, many cases were known where this doctrine failed[4]. Figure 1 illustrates just one example. Here, four proteins from microbial metabolism, adenylosuccinate lyase, argininosuccinate lyase, aspartase and fumarase clearly group into homologous pairs on the basis of sequence similarity, and are part of an evolutionary superfamily that includes all four proteins[5]. However, one protein is involved in nucleic acid biosynthesis, another is involved in amino acid biosynthesis, another is involved in amino acid degradation, and the last is involved in central metabolism. The biologist certainly does not regard the function of these proteins as equivalent.

But should they? All of these proteins use fumarate as a substrate. They all, in the language of the chemist, add the elements of H–X to fumarate using a Michael reaction, where the carboxylic acid functional group acts as an electron sink. This type of language is very close to that used by the Enzyme Commission when it assigns 'EC' numbers to enzymes. In the language of the chemist, all of these proteins have analogous function because they all catalyze an E2 addition reaction to fumarate. Evolutionary recruitment in this family presumably occurred because of this mechanistic similarity[6].

The point to be made here is not that one cannot infer function by homology alone. Neither do we wish to argue that the biologist's view of function is right, and the Enzyme Commission's view is wrong. Rather, the point is that the analysis of function is tied to the language used to describe it. The language used to describe the systems determines whether one sees 'equivalency' or 'non-equivalency'.

Orthologs as functional analogs?

Some attempts to alleviate these problems are based on identification of orthologous sequences. Here, the 'homology-implies-equivalency' assumption is restricted to a subset of homologs that diverged in the most-recent common ancestor of the species sharing the homologs. This strategy is useful, of course. But it is likely to be far less general than is widely thought. Two species living in the same space, almost by axiom, cannot have identical strategies for survival. This, in turn, implies that two orthologous proteins might not contribute to fitness in exactly the same way in two species.

Some examples are useful. Leptin example, is known from genetics to be related to the obesity phenotype in the

Glossary

Analogy: The perception of similarity of two objects based on a similarity between the two in a subset of their traits, leading to the inference that if the two share some features, they will probably share others.
Equivalency (or identity): Biological sequences that have the same function and can, therefore, be interchanged between two organisms without loss of fitness in either.
Functional behavior: The behavior that, if different, would lead to a loss of fitness.
Functional bioinformatics: The field that uses computational tools to infer biological function from genomic sequences without requiring additional experiments; a subset of functional genomics.
Functional divergence: A change in the way in which the protein contributes to the fitness of an organism.
Functional genomics: The field that uses computational tools to infer biological function from genomic sequences supplemented with experiments driven, in part, by hypotheses generated from genomic sequence analysis.
Molecular behavior: Characteristics of a molecule that can be measured; molecular phenotype.

http://tig.trends.com 0168–9525/01/$ – see front matter PII: S0168-9525(01)02320-4

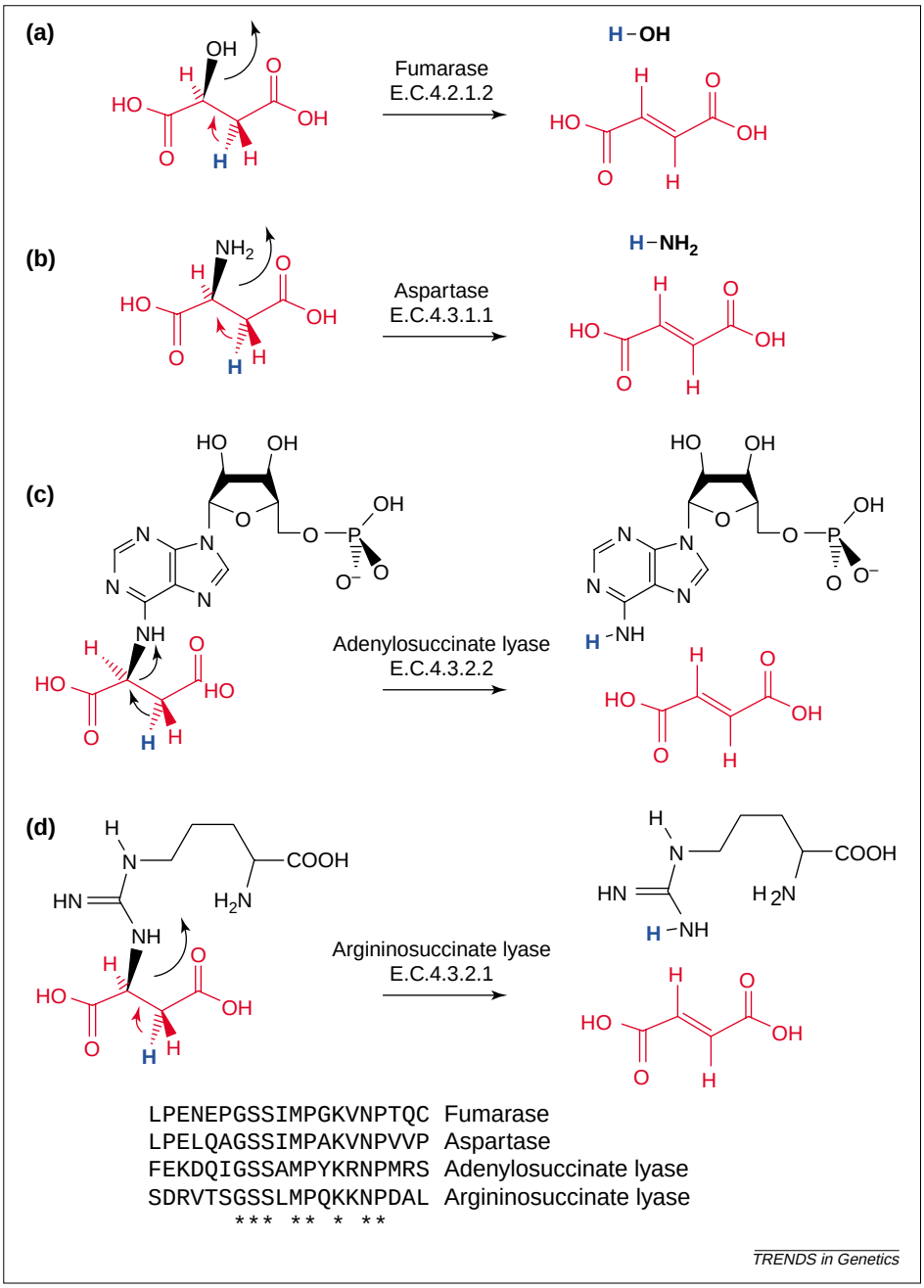


Fig. 1. Using analogy to determine function. Homologous enzymes catalyze four reactions: (a) in central metabolism (the citric acid cycle), (b) in amino acid degradation, (c) in nucleic acid biosynthesis and (d) in amino acid biosynthesis. The enzymes are indisputably homologous; even a simple sequence search identifies significant similarities. The colors show the analogy between the three catalyzed reactions from the perspective of organic chemistry. However, the functions of the proteins, from their roles in biological pathways, are quite different. An annotation strategy that assumes homologous proteins confer fitness in their host organisms in an analogous way would be defeated by this example.

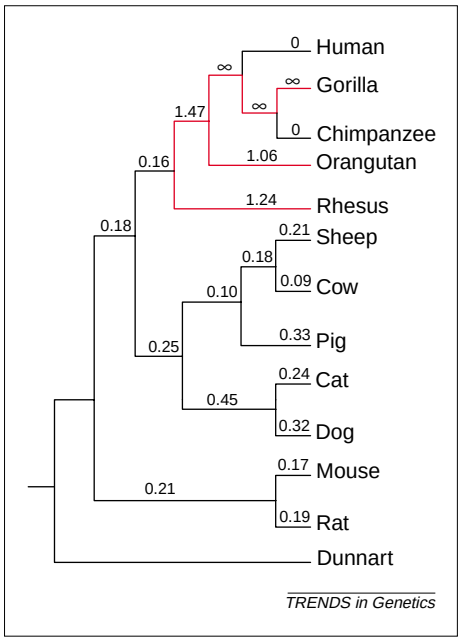


Fig. 2. Do all leptins have 'the same function'? The evolutionary tree for leptins extracted from the Master Catalog (Ref. 17). Numbers on the branches are K_a/K_s ratios, the ratio of nonsynonymous to synonymous changes, normalized for the number of nonsynonymous and synonymous sites in the gene. Undefined (∞) means that no silent substitutions occurred on the branch; calculating a K_a/K_s ratio would require division by zero. Reconstructed evolutionary sequences show rapid evolution of the leptin gene in primitive primates, consistent only with 'positive selection' (red), and implying a different 'function' in primate leptins than in the cenancestral leptins (and rodent leptins). The branch with the K_a/K_s ratio of 1.47 (leading to the apes) contains 7.63 and 2.57 nonsynonymous and synonymous substitutions, respectively (before normalization), meaning that the high ratio is quite significant. The branch with the K_a/K_s ratio of 1.24 (leading to the rhesus monkey) contains 7.68 and 3.61 nonsynonymous and synonymous substitutions, respectively. The branch with the K_a/K_s ratio of 0.21 (leading to the rat–mouse ancestor) contains 14.31 and 31.62 nonsynonymous and synonymous substitutions, respectively.

mouse. The human homolog, almost certainly the ortholog, is known, and is a target for drug development as an obesity gene in humans. Some details of the molecular history, however, suggested that it might not be such a gene. A reconstruction of the evolutionary history of the leptin family (Fig. 2) shows that as primates emerged from the cenancestor of mouse and human, the leptin gene underwent an episode of rapid sequence evolution involving many nonsynonymous substitutions in the leptin gene[7]. Indeed, the reconstructed evolutionary history[8] of the gene family shows that the number of nonsynonymous changes that accumulated in the gene during this episode, divided by the number of synonymous changes, normalized for the number of nonsynonymous and synonymous sites (the K_a/K_s ratio, sometimes referred to as ω, or d*N*/d*S*) is remarkably high. In fact, the K_a/K_s ratio in this episode is higher than that displayed by a pseudogene.

The only explanation consistent with darwinian theory for this episode is that leptin was under 'positive selection pressure'[9] as it entered the primate lineage 100 million years ago. Mutant forms of the primitive primate leptin evidently contributed more to the fitness of the primate descendants than unmutated forms of the protein. Four years ago[7], this suggested that human 'leptin' might not have a role in humans analogous to the one it has in mice. At the very least, a primate model is recommended for pharmacological analysis of compounds targeted towards this system. And now, articles are appearing with titles such as 'Whatever happened to leptin?'[10], noting that 'the

Box 1. Continuum of behavior in elongation factors

Recent studies demonstrate that the behaviors of various Elongation Factor Tu/1α proteins are different in different members of the family, and that these behavioral differences are functionally significant. Undoubtedly, 'participation in translation' is the language describing one behavior almost certainly important to function (fitness) in all of these. Specific features of the behavior have, however, changed (even to the point of being gained or lost entirely) in the evolutionary episodes separating Nodes 1–4 (Fig. I). Functional divergence in the GTP-, GDP-, tRNA- and actin-binding domains has been demonstrated for the highly conserved EF protein. Shifts in EF behavior are found throughout the phylogenetic tree; between bacteria, archaea, and eukaryotes[14] (Node 1), between ciliates and other eukaryotes[15] (Node 2), between plastid and nonplastid bacteria[16] (Node 3), and between photosynthetic and nonphotosynthetic bacteria[16] (Node 4).

At the core of these studies lies the notion that functional importance is highly correlated with conserved evolutionary patterns. For example, ciliate EFs display functional divergence in the domains proposed to interact with actin. This is consistent with actin being a quantitatively minor protein and having an accelerated mutation rate in ciliates (c.f. Ref. 15). A conventional homology-based search would have simply suggested simply that ciliate EFs have 'the same function' as other eukaryotic EFs owing to their high sequence identity, and a substantial level of analogy in other functional behaviors. But it is also clear that a full understanding of protein function requires an analysis of the differences between homologous proteins, and this understanding is best realized when sequences are placed within an historical, evolutionary comparative framework.

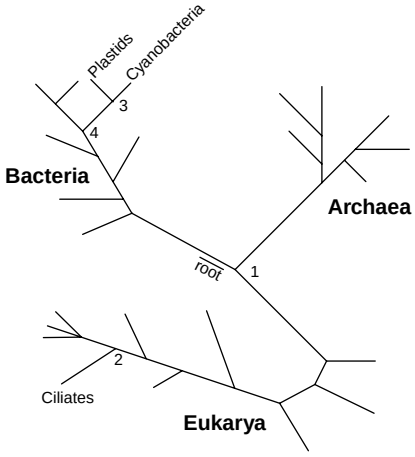


Fig. I

hormone's precise physical role seems to vary from species to species.'

Analogous statements can be made about other pairs of orthologs from mammalian species[11,12]. Just as we cannot accept confidently annotations made by homology, we cannot be confident that annotations based on orthology are correct either.

In fact, 'analogy', not 'equivalency' or 'non-equivalency', is the topic in these examples of annotation. Analogy involves selection of some features of a system as being more important than others and using these features to make a comparison. The Enzyme Commission views the structure of the substrate (fumarate) and the nature of the reaction being catalyzed (E2 addition, for example) as the features worth noting. The biologist (at least as represented above) considers the pathway as the noteworthy feature. The former is more likely to be predictable from the molecule formula derived from the genomic sequence. The latter is closer to the darwinian concept of fitness.

Again, neither view is 'right'. But the wittgenstinian view of functional genomics requires that we understand the process and language of 'analogy', recognize that it is not the same as 'equivalency', and appreciate that an analogy is frequently more informative about the culture of the individual drawing the analogy than it is about the systems between which the analogy is being constructed.

A behavioral–functional continuum

We can expect, almost from first principles, that the near continuum in molecular structure available to protein sequences is associated with a near continuum of molecular behavior[13]. This, in turn, should be associated with a near continuum in fitness. Within this continuum, the case can be frequently made that the differences are more interesting than the similarities, and need to be captured and understood to make a useful functional annotation.

Consider, for example, the family of elongation factors (EFs) represented by EF-Tu (in bacteria) and EF-1α (or eEF1A, in eukaryotes). All are annotated in the contemporary databases as having 'the same function'. After all, they all present a charged aminoacyl-tRNA to the ribosome. Closer inspection[14–16] shows, however, that the details of how this occurs and the specific behavior of the individual EFs, are different, in a manner that has an impact on any linguistic description of 'function', FUNCTIONAL DIVERGENCE (Box 1). For example, EF may function in eukaryotes by binding to uncharged tRNAs in the nucleus, being charged there and then being transported to the cytosol through binding to actin. Regardless of the ability of bacterial EFs to display these behaviors (this is underexamined), this language does not pertain to the function of EFs in bacteria, as bacteria do not have a nucleus.

With EFs, the first level of annotation will undoubtedly reflect the analog in the functions of different proteins from different species. At the next level, however, the annotation must capture the differences. With EFs, the signature of functional change can also be found in the sequences, when they are viewed with a sufficiently sophisticated evolutionary model[14].

A way forward

How might evolutionary analyses be used to generate linguistic statements concerning function for genomic sequences? The completeness of an organism's genomic sequence offers one advantage; it permits us to say what is not present. Furthermore, we can draw on classical descriptions of the history of life known from paleontology and geology to contrast with the molecular histories of protein families reconstructed from genomic sequence databases. Functional genomics must approach genomic sequences in a particular way to facilitate this process, outlined below.

Complete evolutionary models of a protein family

Reconstructed sequences of ancient proteins, intermediates in evolutionary history of a protein family, need to be added to evolutionary models that include a multiple sequence alignment and an evolutionary tree[17]. These ancestral

sequences increase the scope of functional inferences that can be made from reconstructed evolutionary biology.

Higher-order analyses of sequence evolution

Today, the evolution of protein sequences is modeled using simple stochastic mathematics that treats proteins as though they are formless, functionless strings of letters. These models are poor approximations for reality. Their use comes from their ability to provide a 'null hypothesis'. The difference in how real proteins evolve divergently and how the stochastic models expect them to evolve produces a signal, informative about form and function. Higher-order analyses[18] of sequence divergence capture this signal. These incorporate substitution rates that depend on the site (a gamma distribution)[19], interdependence of substitutions at different sites[20] and higher-order gap penalties[21]. Many examples are now available where these higher-order models support higher levels of sequence interpretation. Site-specific mutation rates are correlated to functionally important sites on a protein[14]. Shifts in functional constraints are evident when a specific site is rapidly evolving in one lineage but slowly evolving in another (covarion behavior[22]). Interdependence of sites is used for protein structure prediction[23]. We expect these types of analyses to be done routinely in the future[9,24–26].

Improve the dating of events in the reconstructed molecular history of the protein family

Genomics becomes especially powerful when events in the reconstructed molecular record are correlated with events in the geological and paleontological records. To make this correlation, however, requires a molecular clock. Amino acid sequences themselves are imprecise molecular clocks[27]. Metrics that use synonymous substitutions are frequently used to date molecular events. We expect to see new tools that reflect the complexities of the mutation process to make dating more reliable, especially within vertebrate evolution[28]. For example, the organic chemistry and selective mechanisms governing mutations within GC isochors can lead to spurious estimations when performing phylogenetic analyses. Incorporating more-complex evolutionary models that account for biases in synonymous substitution rates greatly enhances comparative analyses[29]. This, in turn, will open a new avenue for extracting information about function in an organismal and ecological context, FUNCTIONAL BIOINFORMATICS.

Interpret sequence evolution within the context of 3D structures

The three-dimensional structure of the protein connects sequence to reactivity. Permutations within primary sequences can be correlated to those sites that are responsible for protein–ligand interactions and, therefore, differences in behavior[14,25,26,30,31]. A three-dimensional structure, therefore, adds significantly to any story in molecular evolution and does so especially when complex phenomena are being analyzed.

Naturally structured protein sequence databases

After all of the genomes of all of the organisms on Earth are sequenced, all of the protein sequences will almost certainly be recognizable as being members of one of fewer than 10^5 protein families. Naturally structured databases reflect this fact, organizing sequences according to their natural history. This organizational principle is exploited by Hovergen (Ref. 32), COG (Ref. 33), DOMO (Ref. 34), and Pfam (Ref. 35), and the Master Catalog (Ref. 17).

Stories that combine part or all of these prescriptions are now emerging for many specific cases. These include: Myc and transferrins[36], elongation factors[14–16], ribonucleases[25], opsins[37], globins[26] and lysozymes[8]. The ultimate goal, however, will be to join these specific cases into a unified model that combines the molecular history of life on Earth with the record from natural history[38]. Such a large-scale analysis will incorporate dates in the past, places on the globe, and events in the molecular geological and paleontological records, in a way that connects genes and proteins, their host organisms, and their ecosystems in a planetary context.

References

1 Wittgenstein, L. (1993) Tractatus Logico-Philosophicus: German Text with English Translation, Routledge
2 Zhang, Y.Y. *et al*. (1994) Positional cloning of the mouse obese gene and its human homolog. *Nature* 372, 425–432
3 Eisenberg, D. *et al*. (2000) Protein function in the post-genomic era. *Nature* 405, 823–826
4 Benner, S.A. and Ellington, A.D. (1988) Interpreting the behavior of enzymes. Purpose or pedigree? *CRC Crit. Rev. Biochem*. 23, 369–426
5 Aimi, J. *et al*. (1990) Cloning of a cDNA encoding asenylosuccinate lyase by functional complementation in *Escherichia coli*. *J. Biol. Chem*. 265, 9011–9014
6 Gerlt, J.A. and Babbitt, P.C. (1998) Mechanistically diverse enzyme superfamilies: the importance of chemistry in the evolution of catalysis. *Curr. Opin. Chem. Biol*. 2, 607–612
7 Benner, S.A. *et al*. (1998) Post-genomic science. Converting primary structure into physiological function. *Adv. Enzyme Regul*. 38, 155–180
8 Messier, W. and Stewart, C.B. (1997) Episodic adaptive evolution of primate lysozymes. *Nature* 385, 151–154
9 Yang, Z.H. and Bielawski, J.P. (2000) Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol*. 15, 496–503
10 Chircurel, M. (2000) Whatever happened to leptin? *Nature* 404, 538–540
11 Chandrasekharan, U.M. *et al*. (1996) Angiotensin II forming activity in a reconstructed ancestral chymase. *Science* 271, 502–505
12 Liberles, D.S. *et al*. (2001) The Adaptive Evolution Database (TAED). *Genome Biol*. 2, 0003.1–0003.18
13 Hey, J. (1999) The neutralist, the fly and the selectionist. *Trends Ecol. Evol*. 14, 35–38
14 Gaucher, E.A. *et al*. (2001) Function–structure analysis of proteins using covarion-based evolutionary approaches: elongation factors. *Proc. Natl. Acad. Sci. U. S. A*. 98, 548–552
15 Moreira, D. *et al*. (1999) Unusually high evolutionary rate of the elongation factor 1a genes from the Ciliophora and its impact on the phylogeny of eukaryotes. *Mol. Biol. Evol*. 16, 234–245
16 Lockhart, P.J. *et al*. (1998) A covariotide model explains apparent phylogenetic structure of oxygenic photosynthetic lineages. *Mol. Biol. Evol*. 15, 1183–1188
17 Benner, S.A. *et al*. (2000) Functional inferences from reconstructed evolutionary biology involving rectified databases. An evolutionarily-grounded approach to functional genomics. *Res. Microbiol*. 151, 97–106
18 Thorne, J.L. *et al*. (1992) Inching toward reality. An improved likelihood model of sequence evolution. *J. Mol. Evol*. 34, 3–16
19 Yang, Z.H. (1996) Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol. Evol*. 11, 367–372
20 Olmea, O. *et al*. (1999) Effective use of sequence correlation and conservation in fold recognition. *J. Mol. Biol*. 293, 1221–1239
21 Benner, S.A. *et al*. (1993) Empirical and structural models for insertions and deletions in the divergent evolution of proteins. *J. Mol. Biol*. 229, 1065–1082
22 Miyamoto, M.M. and Fitch, W.M. (1995) Testing the covarion hypothesis of molecular evolution. *Mol. Biol. Evol*. 12, 503–513
23 Benner, S.A. *et al*. (1997) Bona fide predictions of protein secondary structure using transparent analyses of multiple sequence alignments. *Chem. Rev*. 97, 2725–2843
24 Golding, G.B. and Dean, A.M. (1998) The structural basis of molecular adaptation. *Mol. Biol. Evol*. 15, 355–369
25 Jermann, T.M. *et al*. (1995) Reconstructing the evolutionary history of the artiodactyl ribonuclease superfamily. *Nature* 374, 57–59

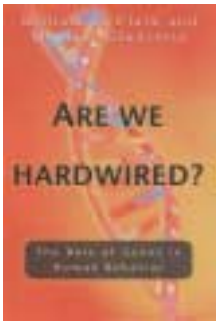
26 Naylor, G.J.P. and Gerstein, M. (2000) Measuring shifts in function and evolutionary opportunity using variability profiles: a case study of the globins. *J. Mol. Evol*. 51, 223–233
27 Ayala, F.J. (1999) Molecular clock mirages. *BioEssays* 21, 71–75
28 Peltier, M.R. *et al*. (2000) A comprehensive evolutionary analysis and structural prediction of the uterine serpin family. *J. Exp. Zool*. 288, 165–174
29 Mooers, A.O. and Holmes, E.C. (2000) The evolution of base composition and phylogenetic inference. *Trends Ecol. Evol*. 15, 365–369
30 Lichtarge, O. *et al*. (1996) An evolutionary trace method defines binding surfaces common to protein families. *J. Mol. Biol*. 257, 342-358
31 Landgraf, R. *et al*. (2001) Three-dimensional cluster analysis identifies interfaces and functional residue clusters in proteins. *J. Mol. Biol*. 307, 1487-1502
32 Duret, L. *et al*. (1994) HOVERGEN – a database of homologous vertebrate genes. *Nucleic Acids Res*. 22, 2360–2365
33 Tatusov, R.L. *et al*. (1997) A genomic perspective on protein families. *Science* 278, 631–637
34 Gracy, J. and Argos, P. (1998) DOMO: a new database of aligned protein domains. *Trends Biochem. Sci*. 23, 495–497
35 Bateman, A. *et al*. (2000) The Pfam protein families database. *Nucleic Acids Res*. 28, 263–266
36 Gu, X. (1999) Statistical methods for testing functional divergence after gene duplication. *Mol. Biol. Evol*. 16, 1664–1674
37 Chang, B.S.W. and Donoghue, M.J. (2000) Recreating ancestral proteins. *Trends Ecol. Evol.* 15, 109–114
38 Benner, S.A. (2001) Natural progression. *Nature* 409, 459

Steven A. Benner*
Eric A. Gaucher
Dept of Chemistry, NASA Astrobiology Institute, University of Florida, Gainesville, FL 32611-7200, USA.
*e-mail: benner@chem.ufl.edu

Book Review

Genes and environment: what does interaction really mean?

Are We Hardwired? The Role of Genes in Human Behaviour
by William R. Clark and Michael Grunstein
Oxford University Press, 2000.
£16.99 ($30.00) hbk (322 pages)
ISBN 0 19 513826 0



These are exciting times for genetics, but also times of serious misunderstandings. The complete sequencing of the human genome has been a crucial step, but it tells us nothing about what genes are or what they do. We know behavior is the result of the interaction between genes and environment, yet the complexities of interaction remain poorly understood. Even in the serious broadsheets and in nongenetic scientific literature, it is not unusual still to find mention of 'a gene for homosexuality' (or other traits, such as reading, intelligence, etc.). This is why *Are We Hardwired?* is welcome. It eloquently conveys the excitement of new discoveries and the subtleties of gene–behavior relationships, while treating the interpretations with great caution. The explanations are simple, but not simplistic, allowing the complexities of gene expression to become accessible to both the intelligent general public and academics.

The book explores behavior at its most basic level, from unicellular organisms to multicellular organisms, from roundworms, fruit flies and sea slugs, through to humans, stressing the remarkable evolutionary conservation of fundamental mechanisms of cell functioning and communication. When a single gene knockout disrupts a complex behavioral system, it is tempting to assume that the behavior is under the direct control of that gene. However, the book highlights the fact that the number of genes involved in a behavior is not the relevant dimension; seemingly direct gene–behavior mappings simply reflect the fact that biological systems underlying behaviors are tightly regulated, and breakdown of a single component can often shut down an entire pathway.

The authors proceed to evaluate the role of gene expression and neurotransmitter regulation in a variety of human behaviors: aggression, eating disorders, substance abuse, mental function and sexual preference. In doing so, they draw heavily on results from twin studies. Paradoxically, studies focused on locating specific genes often reveal significant information about the contribution of the environment. For example, even when the effect of environment seems minimal, results might actually reflect the fact that the identical genetic endowment of monozygotic twins reared apart predisposes them to select similar aspects of their different environments. In other words, 'genes' and 'environment' interact dynamically.

The authors argue adamantly against direct gene–behavior mappings in the chapter on human mental function. As they put it, "The likelihood that we will discover anything even remotely identifiable as 'intelligence genes' is just about nil". Human performance on IQ tests is likely to be governed by genes that are conserved across species and affect general cell functioning, such as the rate and direction of exchange between neuronal systems. However, what about the silencing of the single *FMR1* gene[1], which results in one of the most common forms of inherited mental retardation? Does this imply that *FMR1* is a 'gene for intelligence'? The authors suggest that this is not the case. *FMR1* is not involved in complex learning, but encodes a protein involved in the regulation of the expression of other genes affecting synaptic morphology – a crucial, basic cellular mechanism.

Although they give full reign to animal models and to human twin methodology, with the exception of the paragraph on Fragile X syndrome, the authors do not sufficiently discuss another important tool for exploring gene expression; that is, naturally occurring gene mutations in developmental disorders. For example, Williams syndrome (WS), which is caused by a hemizygotic microdeletion on chromosome 7q23.11, is characterized by an uneven cognitive profile. In the majority of individuals with WS, spatial cognition is seriously impaired, whereas language and face processing are surprisingly proficient[2]. Geneticists and psychologists have made strong claims about the role of a single gene (Limkinase1) in the spatial deficit. Their claims imply that the lack of Limkinase1 expression can impair one part of the brain, leaving other parts intact. But this ignores increasing evidence that the whole of the WS brain develops atypically. Different