Selection for Modularity in the Genome: Reading Frame Evidence for Exon Shuffling*

Lee Altenberg**

and

Douglas Brutlag***

Department of Biochemistry
Stanford University
Stanford, CA 94305 U.S.A.

Key Words: exon shuffling, modularity, reading frame, evolution

Abstract. During the addition of new, adaptive genes to the genome, if certain sequences are better donors in creating new genes than others, then a kind of selection process results that can improve the ability of the genome to generate new, useful genes. When applied to the evolution of exons, this idea predicts that there should be a predominance, among translated exons, of exons and pairs of exons that are multiples of 3 bases in length. It also predicts that more introns should fall between codons rather than splitting them. Empirical verification of these predictions is described.

Introduction

In the "Modern Synthesis" of evolutionary theory, evolution is usually taken to proceed through substitutions in the alleles of a gene. Yet some fraction of the time, adaptations must occur through the creation of new genes. The creation of new genes is subject to the same forces of variation and selection as allelic evolution, but it has one important additional effect: new genes add new dimensions of variation to the genome. In this way, selection acting in the creation of new genes may play a role in determining the very nature of genetic variation. When we apply this idea to the evolution of exons, it offers a new resolution to the controversy of "evolutionary foresight" (1) generated by Gilbert's exon shuffling hypothesis (2).

 This is a LaTeX transcription of the unpublished 1986 manuscript. Figures are not yet converted from native formats. This paper was cited in: Doolittle [1987]. These results were first predicted in an unpublished letter to *Nature*, 1983, and first presented at the 1985 Genetics Society of America meeting, Boston (Altenberg [1985]). Subsequent confirmation of these results have been reported by Patthy [1987], Smith [1988] and Gelfand [1992], and followed up by Tomita, Shimizu, and Brutlag [1996]

** Currently at the University of Hawaii at Manoa. Web site: http://dynamics.org/Altenberg/, e-mail: altenber@santafe.edu

*** Web site: http://cmgm.stanford.edu/~brutlag, e-mail: brutlag@stanford.edu

From an “engineering point of view”, a genome composed of modular components, which have a good chance of producing functional genes when rearranged, is very attractive. Gilbert (2) proposed that split genes exist in order to allow such a modular configuration of the genome, and hence speed the evolution of new proteins. Doolittle (3) and others (1) have pointed out, though, that such “engineering optimality” for evolutionary potential cannot be used as a Darwinian explanation for split genes, because this optimality confers no immediate selective advantage to the organism, being rather a property affecting the future evolution of the lineage.

Fortunately, this problem with Gilbert’s modular “exon shuffling” hypothesis did not deter workers from looking for modularity in the exon/intron patterns of eukaryotic genes, because in many cases they have found it. Evidence of modularity includes:

1. spatial compactness of the exon product in the folded peptide, producing what Go calls a “module” (4-6);
2. introns falling between functional or structural peptide units (7-12,16- 19,23,25-35);
3. hydrogen bonding or disulfide bonding found mainly within, and not between, exon products (19-21);
4. duplicated exons within a protein corresponding to repeated structural or functional units (7-10,36);
5. homologous exons in otherwise non-homologous genes corresponding to similar peptide structures or functions (11-18).
6. experimental manipulation of gene or peptide structure showing functional autonomy of the product of the exon or set of exons (22- 24).

Blake (37) suggested an explanation of these findings in light of Doolittle’s objection to Gilbert’s explanation by proposing that the correspondence found between exons and peptide structural units (38) is a historical *remnant*: today’s exons are the descendants of the earliest coding sequences, which coded for minimal units of stable peptide structure, and were subsequently rearranged into larger, multi-exon proteins. Blake points out that the structural correlations would not be rigid, but instead, when an exon is shuffled to a new peptide, its product might assume some other shape. This theory has direct implications: a continuation of this process would be expected to blur the original correspondence between exon and peptide structure. It would seem, therefore, that as exon shuffling continued, exons would become more and more random with regard to peptide structure and less well suited for exon shuffling.

A critical factor in the decay of this correspondence would be the degree to which the correspondences were inherent in the exons, maintained between exon shuffles. It is often implicitly assumed that any correspondence between exon and peptide structures or functions would be independent from the surrounding peptide. But the importance of tertiary interactions in folding and function would caution against such an *a priori* assumption. The term “module” has been employed to describe exons or groups of exons corresponding to independent structure or function (5,11,33). The independence of an exon’s (or set of exons’) properties from the rest of the gene can be considered as a feature distinct from the particular structural or functional properties themselves. We will use the term “modularity” to refer specifically to this degree of independence. Modularity is a *prerequisite* for the advantage of exon shuffling in Gilbert’s hypothesis.

Here we propose that modularity could evolve from exon shuffling itself, and does not require that exons originated as separate, functional "protogenes". Modularity would result from a process we call "constructional selection", which is a simple consequence of the accumulated construction of useful genes over the course of evolution (39).

Consider two exons, one which has a good chance of producing a functional protein when rearranged with other exons, and another which, although functional in its current gene, is less likely to produce a functional protein when rearranged with other exons. They differ, therefore, only in their degree of modularity. These two exons can be viewed as being in a competition for which is most likely to be a donor sequence to any new genes being added to the genome. The result of this continued competition over the course of evolution would be a differential proliferation of more modular exons (Fig. 1). This is reminiscent of the ideas about "selfish DNA" spreading through the genome (40,41), except that in this case, modularity is merely hitchhiking, both through the population (1) and through the genome, with the new genes it helps create, and such hitchhiking is an adjunct to organismal adaptation, in no way opposed to it. Even if the earliest exons in the genome were random with respect to modularity, the differential proliferation of modular exons could account for their abundance in genomes today. Neither "evolutionary foresight" nor a "protogene" origin for exons are therefore necessary in Gilbert's exon shuffling hypothesis. "Constructional selection" for modular exons is a Darwinian process that would improve the ability of the genome to generate new, useful genes.

Fig. 1. [CAPTION ONLY] Competition between exons during the creation of new genes. Starting with a genome whose exons were random with respect to protein structure, the exons which have greater *modularity* (open boxes) should proliferate in evolution.

We have sought to test the hypothesis of constructional selection by looking for properties of exons that have nothing to do with their functioning, but are features associated only with their modularity. We have considered two properties of exons on the DNA level rather than the peptide level. First, property independence involves not only an exon's own properties, but its effect on the properties of the rest of the gene. An exon whose length was a multiple of 3 nucleotides would thereby have greater modularity than exons of length $3n+1$ or $3n+2$, because it could be tandemly duplicated, or inserted into other genes, without shifting the reading frame downstream. Second, modularity would be enhanced by the ends of the exon falling between rather than within codons, because only in this case is a terminal amino acid of the exon independent from its adjoining exons. These features are not visible during the expression of the gene, but come into play only during a rearrangement of exons. They each would give the exon greater modularity, and such exons should have been more prolific donors to the construction of new genes through exon shuffling.

One would predict, therefore, that if constructional selection had left its mark on the exons in the genome, there should be an excess of exons that are multiples of 3 in length, and an excess of introns that fall between codons. We have tested these predictions by going through the large DNA sequence databases now available and counting exon lengths. Both of these predictions are found to hold.

Results

Table 1. G-Test of fit for frequencies of exon lengths, mod-3. Database is NIH GenBank. Tabulated are all the exons in the database excluding viral DNA, tDNA, rDNA, pseudogenes and putative exon duplications. H_0: $P_0 = 1/3$, $P_1 + P_2 = 2/3$.

	5' nontranslated exons			N-terminal exons			Middle coding exons			C-terminal exons			3' nontranslated exons		
Mod-3:	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2
Obs:	9	9	9	30	25	34	65	39	36	33	34	25	2	2	7
Exp:	9	9	9	29.6	29.6	29.6	46.6	46.6	46.6	30.6	30.6	30.6	3.6	3.6	3.6
G:	G = 0			G = 0.006			G = 10.3			G = 0.26			G = 1.26		
Prob.:	P > 0.9			P > 0.9			P < 0.005 **			0.9 > P > 0.5			0.5 > P > 0.1		

Table 1 shows a tally of all the exons in the database, excluding pseudogenes, tDNA, rDNA, and viral DNA. The exons can be divided into 5 categories: exons outside the coding region (both 5' and 3'), exons with the amino terminus within them, exons with the carboxy terminus within them, and exons that only have translated sequence. This latter category, the "middle coding exons", would include exons that had been inserted between two pre-existing exons. It is only in this category that there would be a "constructional" advantage to exons a multiple of 3 nucleotides in length. In Table 1 the number of exons that are 0, 1 or 2 nucleotides longer than a multiple of 3 (i.e., of length 0 mod-3, 1 mod-3 or 2 mod-3, respectively) are tallied within each of the 5 categories of exon. We find that in the middle coding exons there is a significant excess of exons that are multiples of 3 long (P< 0.005), and in the other categories of exons, no significant deviation from a random expectation, when a G-test is performed for goodness of fit. In the data collection, we have not counted exons that are suspected duplications of other exons, as in the collagen gene, so this figure would actually be on the conservative side.

Table 2 shows a tally of all the introns in the database. They are tallied by the phase within the codon that they fall, whether after the first (I), second (II), or third (III) nucleotide of the codon. A G-test on these frequencies based on a random expectation shows a significant excess of introns falling between rather than within codons.

There is a significant deficit of introns between the second and third nucleotides of codons. We do not see how constructional selection would account for this bias, and therefore consider an alternative hypothesis: that introns originated through insertion into intact exons, and either selection or sequence preference has biased the locations of introns within the gene. Craik et al.(42) and Rogers (43,44) have conjectured that such biased intron insertion could account for several observations on the locations of introns in split genes. Another possibility is that introns have been non-randomly deleted from genes according to their codon phase. Suppose that introns had a higher chance of being inserted or deleted in some codon phases over others. Such biased

Table 2. G-test of fit for frequencies of intron reading-frame phases. Introns in reading-frame phase III are between codons. Introns in phase I occur after the first nucleotide of the codon. Introns in phase II occur after the second nucleotide. For the set of exons in Table 4, G-tests for two null hypotheses (H_0) are given. Both are significant.

	Intron reading-frame phase		
	III	I	II
Obs:	99	74	38
Exp:	70.3	70.3	70.3
H_0:	$P_{\text{III}} = \frac{1}{3}$ $G = 8.3$ $*P < 0.005$		$P_{\text{II}} = \frac{1}{3}$ $G = 12.4$ $*P < 0.001$

intron insertion or deletion would confound the test on exon lengths by also producing a statistical excess of exons a multiple of 3 long. Table 3 shows the goodness of fit between the actual frequencies of exons lengths and those predicted from independent, biased intron insertion or deletion. The observed number of exons a multiple of 3 long is in excess of that predicted by this alternative hypothesis, but the departure is not statistically significant.

Table 3. G-test of fit for frequencies of exon lengths, mod-3,expected from intron reading-frame phase frequencies. The sample is the exons in Table 4. H_0 derives from the assumption that there is no correlation between intron reading-frame phases of adjacent introns.

	Exon lengths, mod-3		
	0	1	2
Obs:	74	47	49
Exp:	63.8	53.1	53.1
H_0:	$P_0 = P_{\text{III}}^2 + P_{\text{I}}^2 + P_{\text{II}}^2$ $G = 1.3$ $0.5 > P > 0.1$		

To distinguish between these two hypotheses– that of constructional selection and that of biased intron insertion or deletion– we need to examine another statistic, the number of pairs of exons whose combined length is a multiple of 3 nucleotides. Constructional selection would produce an excess of exon pairs of total length 0 mod-3, which can be distinguished from any effects of biased intron insertion. Such exon pairs would include single exons a multiple of 3 in length that had been inserted in a gene and later been split by an intron insertion, and exon pairs of combined length 0 mod-3 that had been inserted together into the gene. In both cases, the pair of exons would have experienced a constructional advantage, and would be expected to be in excess.

To test for such an excess, we have examined all the genes in the database with two or more middle coding exons. These are shown schematically in Table 4. The numbers under the gene names represent the lengths in nucleotides (mod-3) of the middle coding

Table 4. Schematic of genes with two or more middle exons. Some names of genes are abbreviated. Numbers indicate middle exon nucleotide lengths, mod-3. Pairs of exons with total length 0 mod-3 are in parentheses. Suspected duplicated exons are excluded.

lysozm	Igg	fos	lactalb	soy act	humil2	humgg
0 0	0 0	0 0	0 1	(1 2)	0 0	0 0

rattubal	slmras	soylbgI	chkx	humifng	humtbbm40	H-ras
(1 2)	1 0	1 0	2 0	0 0	1 0	(2 1)

prolact	actin	prokallik	humpla	bovops	musafp	ovalbumin
0 0 0	0 (1 2)	(1 2) 2	2 0 0	1 1 0	1 (1 2)	0 0 (1 2) 0

ratgh	H2	actin	MtCytB	humfol	humsisa	thr
2 0 0	0 0 0 0	1 0 0 2	1 0 1 0	(2 1)(1 2)	1 0 0 1	0 0 1 0 0

pyrkinase	cytochr oxidase	elastase	humfixg
1 1 1 0 2 2	0 2 0 (1 2) 0	2 2 0 2 2 0	(2 1) 0 0 (2 1)

mushprt	pepsinogen	ACH receptor
(2 1) 0 0 2 2 2	1 (1 2) 2 0 1 0	2 0 (2 1) (1 2) 0

rat cytochrome p451	hamster vimentin	celmyumc
1 0 2 0 (1 2) 1	1 0 0 0 2 2 2	(2 1)(1 2) 0 0 0

complement factor B	LDL receptor
0 0 (2 1)(1 2)(2 1) 2	0 0 1 0 2 (2 1) 0 0 2

exon. In most of the cases where an exon is not a multiple of 3, we observe that it is part of a pair of exons whose total length is a multiple of 3. These exons are in parentheses.

Fig. 2. [CAPTION ONLY] Distribution of the number of unpaired middle exons in a stochastic simulation. Intron reading-frame phases from exons in Table 4 were randomly permuted to generate a new set of exons. After grouping pairs of exons whose total length in nucleotides was 0 mod-3, unpaired exons of length 1 mod-3 or 2 mod-3 were counted. This was replicated 5,000 times. H_0: the intron reading frame phases are randomly distributed with respect to each other along the gene.

To test whether this excess of paired exons is statistically significant, a stochastic simulation was developed in which the intron phases from the actual set of genes were randomly permuted, creating a new set of exons. Figure 2 shows the distribution of this statistic for 5,000 replicates of this intron scrambling. The observed number of unpaired exons of lengths 1 mod-3 and 2 mod-3 is compared to the number of unpaired exons one obtains from these randomized sets of exons. The observed value falls in the lower 3 percentile of the simulation distribution, making it significantly less than expected. The only way a model of biased intron insertion or deletion could explain this result is by proposing that intron insertion (or deletion) is negatively (positively) correlated with the codon phases of introns that are one exon away, or positively (negatively) correlated with the codon phases of introns that are two exons away, or has other higher order interactions with existing introns. Current knowledge about intron splicing and intron origins does not suggest such dependencies.

Thus, the hypothesis that the composition of the exons in the genome has been impacted by the process of constructional selection remains the more plausible among those under consideration.

Discussion

The results here show that, based on the composition of exons among eukaryotes, their genomes stand a better than random chance that exon shuffling will produce a new gene with the correct reading frame, and with non-mutated amino acids on the exon termini. Previously, examples of non-modular exons have been cited as evidence against Gilbert's hypothesis (45). But we are recommending a population genetic approach be taken to modularity in the genome; just as there is a distribution of fitnesses in a population that changes under organismal selection, there is a distribution of modularity in the genome that would change under constructional selection. The excesses found here of exons and pairs of exons that are multiples of three nucleotides in length, and of introns between codons, are evidence for this effect on the distribution of modularity. Our theory does not rule out in any way Blake's plausible "protogene" hypothesis for the origin of exons, but it does not require it to account for modularity. Even if the genome began with contiguous genes which were later split up randomly into exons, it would still be possible for features such as exons being a multiple of 3 in length, or coding for peptide folding units, to become prevalent in the genome, because those exons that accidentally had these features would be the ones that later proliferated when new, useful genes were added to the genome.

It is more difficult to show that the protogene hypothesis is actually insufficient to account for our data, because one cannot know whether a feature that confers modularity might not at one time also have been necessary for function. Even though today exons needn't be multiples of 3 in length, one cannot exclude the possibility that this was necessary at some point in the evolution of translation. Assuming, however, that exons are remnants of the original protogenes, the features we see in exons today would, without constructional selection, be slowly decaying properties of the original, autonomous exons, whereas under constructional selection, the features we see today would be an amplification of the modularity fortuitously present in the earliest exons. It should be possible to develop statistics that could distinguish between these scenarios.

The underlying reason that modularity should increase the constructional advantage of an exon is a principle of "low pleiotropy" (47): the more features of a protein that exon shuffling affects, the less likely it is that the resulting product is useful. Greater modularity means lower pleiotropy on the molecular level, and hence, a higher expected chance of producing useful products.

Once a new gene or exon has become a part of the genome, constructional selection no longer acts on it, so an exon can retain its modularity only through fortuitous results of standard organismal selection. The high level of conservation of secondary and tertiary structure in proteins, even in the presence of large primary sequence divergence (46), is a prerequisite for the evolution of modularity on this level. The high level of conservation of exon lengths, where usually (27,48,49) but not always (27,44) documented changes in length have been by multiples of 3 , is a prerequisite for the accumulation of exons that are multiples of 3 in length.

Constructional selection as we have defined it is simply the differential spread of certain sequences due to their better than average chance of producing something useful for the organism when duplicated.

We have focused on modularity as a basic feature under constructional selection, but we should mention two others. One is seen in gene amplification, where copies of a gene are maintained simply to produce more transcript. The second is seen in the immunoglobulins, where the adaptive potential for new copies of the variable domains, due to selection for immunological diversity, is the likely cause of their proliferation in the genome. We have focused on exons as the sites for the selection for modularity, because this yields some immediate and testable predictions. But the idea would apply to other kinds of genetic units, and even to the organization of the organism's phenotype itself. The fact that regulatory and structural sequences of genes are almost always in separate segments, which can be interchanged with other such segments, is a good example of modularity in the genome. In promoters, for example, there is no biophysical requirement for modularity, as is shown by the existence of internal promoters in tRNA and 5s rRNA genes. But the fact that internal promoters have been found in no other genes is consistent with them having a constructional disadvantage due to their non-modularity. Within this view, external promoters have been the more prolific donors of promoters to the generation of new genes.

The addition of new functions to an organism would also have to pass the filter of constructional selection. New functions will not evolve if this would entail pleiotropic disruption of other functions that the organism still requires. Whether such disruption was due to pleiotropy in the underlying genetics, or due simply to functional linkages in the physiology or morphology of the organism, evolution will be prevented from following that path. When new functions are acquired by organisms over evolution, we expect therefore that they should exhibit modularity, an ability to vary without altering many other functions of the organism. Riedl (50) developed the basic idea of constructional selection in arguing that much of morphological evolution requires the creation of new developmental genes exhibiting high modularity. Empirically, modularity in the general phenotype of the organism is a much more difficult issue to deal with than the modularity of exons; nevertheless, Cheverud (51) and Wagner (52) examine correlations in quantitative genetic variation which may be interpretable as support for modularity being present in the genetic basis of morphology. In these cases, traits under common genetic control, as determined by principal component analysis, tend to be involved with the same function. The evolution of segmental differentiation in Drosophila and its control by homeotic genes, and of cell differentiation in Caenorhabditis (53), though yet to be fully elucidated, may stand as examples of modular genetic control of development.

Constructional selection is not restricted to any particular epoch in genome evolution. It would act to the degree that new genes are being stably added to the genome. Some evidence suggests that modular exons existed prior to the divergence of prokaryotes and eukaryotes (9), which is consistent with substantial genome expansion having occurred between the origin of exons and that time. The evolution of the serine proteases, which appears to have involved the permutation and combination of a small set of exons

(11,21), suggests that exon shuffling continues to be an important source of variation in the generation of new proteins, and that the modularity of these exons has continued to allow them to proliferate as donors to new, useful genes.

Many phenomena in evolution have been approached by considering how different levels of selection may be involved. Altruism has been considered as an effect of kin or group selection (54); phylogenetic change has been considered as an effect of species selection (55); and various repeated units in genomes have been considered as examples of "selfish DNA" (40,41). Here we are proposing that the property of modularity in the set of functioning genes in the genome be viewed as an effect of selection in the gene construction process itself. Constructional selection is different in that it applies only to the evolutionary expansion of the genome, and rather than being in any way opposed to selection on the individual organism, constructional selection has organismal selection as a sub-process. The target of its action is the relation between genetic changes and functional changes in the phenotype. Its predicted consequence is that the production of genetic variation should be biased toward ways that would be more likely to produce adaptive changes (56). As investigations into the architecture of genes and proteins and of development continue, it should be possible to discern the degree to which this process may have played a part in evolution.

Bibliography

1. F. Crick, Science 204, 264-271 (1979).

2. W. Gilbert, Nature 271, 501 (1978).

3. W.F. Doolittle, Nature 272, 581-582 (1978).

4. M. Go, Nature 291, 90-92 (1981).

5. M. Go, Proc. Natl. Acad. Sci. U.S.A. 80, 1964-1968 (1983).

6. W. Gilbert, M. , G. McKnight, Cell 46, 151-154 (1986).

7. J.P. Stein, J.F. Catterall, P. Kristo, A.R. Means, B.W. O'Malley, Cell 21, 681-687 (1980).

8. F. Alexander, F.R. Young, S.M. Tilghman, Mol. Biol. 173, 159-176 (1984).

9. M. Marchionni, W. Gilbert, Cell 46, 133-141 (1986).

10. W.K. Jones, L. Yu-Lee, S.M. Clift, T.L. Brown, J.M. Rosen, J. Biol. Chem. 260, 7042-7048 (1985).

11. L. Patthy, Cell 41, 657-663 (1985).

12. K. Takio, et al., Biochemistry 23, 4207-4208 (1984).

13. P.J.Artymiuk, C.C.F. Blake, A.E. Sippel, Nature 290, 287-288 (1981).

14. R. van den Heuvel, W. Hendriks, W. Quax, H. Bloemendal, J. Mol. Biol. 185, 273-284 (1985).

15. G. Wistow, FEBS 181, 1-6 (1985).

16. J. Plutzky, J.A. Hoskins, G.L. Long, G.R. Crabtree, Proc. Natl. Acad. Sci. U.S.A. 83, 546-550 (1986).

17. A.M. Michelson, C.C.F. Blake, S.T. Evans, S.H. Orkin, Proc. Natl. Acad. Sci. U.S.A. 82, 6965-6969 (1985).

18. T.C. Südhof, J.L. Goldstein, M.S. Brown, D.W. Russell, Science 228, 815- 822 (1985).

19. E.M. Stone, K.N. Rothblum, R.J. Schwartz, Nature 3l3, 498-500 (1985).

20. C.C.F. Blake, Intl. Rev. Cytology 93, 149-185 (1985).

21. R.F. Doolittle, Trends Biochem. Sci. 10, 233-237 (1985).

22. C.S. Craik, S.R. Buchman, S. Beychok, Proc. Natl. Acad. Sci. U.S.A. 77, 1384-1388 (1980).

23. G.D. Sanctis, G. Falcioni, B. Giardina, F. Ascoli, M. Brunori, J. Mol. Biol. l88, 73-76 (1986).

24. A.-J. Van Zonneveld, H. Veerman, H. Pannekoek, Proc. Natl. Acad. Sci. U.S.A. 83, 4670-4674 (1986).

25. C.-I. Bränden, H. Eklund, C. Cambillau, A.J. Pryor, EMBO J. 3, 1307-1310 (1984).

26. J. Douglass, O. Civelli, E. Herbert, Ann. Rev. Biochem. 53, 665-715 (1984).

27. M. Rieger, J.L. Jorcano, W.W. Franke, EMBO J. 4, 2261-2267 (1985).

28. S.S.-l. Li, et al., Eur. J. Biochem. l49, 215-225 (1985).

29. D. Straus, W. Gilbert, Mol. Cell. Biol. 5, 3497-3505 (1985).

30. F. de Ferra, et al., Cell 43, 721-727 (1985).

31. J.T. Den Dunnen, R.J.M. Moormann, N.H. Lubsen, J.G.G. Schoenmakers, Proc. Natl. Acad. Sci. U.S.A. 83, 2855-2859 (1986).

32. P. Van Den Elsen, K. Georgopoulos, B.-A. Shepley, S. Orkin, C. Terhorst, Proc. Natl. Acad. Sci. U.S.A. 83, 2944-2948 (1986).

33. N. Lonberg, W. Gilbert, Cell 40, 8l-90 (l985).

34. M. Malissen, B. Malissen, B.R. Jordan, Proc. Natl. Acad. Sci. U.S.A. 79, 893-897 (1982).

35. J. Nathans, D.S. Hogness, Cell 34, 807-814 (1983).

36. Y. Yamada, et al., Cell 22, 887-892 (1980).

37. C.C.F. Blake, Nature 306, 535-537 (1983).

38. C.C.F. Blake, Nature 273, 267 (1978).

39. Altenberg, L, Genetics 110, s41 (1985).

40. W.F. Doolittle and C. Sapienza, Nature 284, 601-603 (1980).

41. L.E. Orgel, and F.H.C. Crick, Nature 284, 604-607 (1980).

42. C.S. Craik et al., in Gene Regulation (ed O'Malley, B.W.) pp. 35-55 (1982).

43. J. Rogers, Nature 315, 485-486 (1985).

44. J. Rogers, Trends in Genetics 2, 223 (1986).

45. T. Cavalier-Smith, Nature 3l5, 283-284 (1985).

46. D. Marchuk, S. McCrohon, E. Fuchs, Cell 39, 491-498 (1984).

47. J.T. Bonner, On Development p. 60 (Harvard Press, 1974).

48. J.A. Lewis, Mol. Cell. Biol. 6, 1998-2010 (1986).

49. C.S. Craik, W.J. Rutter, R. Fletterick, Science 220, 1125-1129 (1983).

50. R. Riedl, Q. Rev. Biol 52, 351-370 (1977).

51. J.M. Cheverud, J. Theor. Biol. 110, 155-171 (1984).

52. G.P. Wagner, J. Math. Biol. 21, 77-95 (1984).

53. C.J. Kenyon, Trends Biochem. Sci. 8, 349-351 (1983).

54. W.D. Hamilton, J. Theor. Biol. 7, 1-52 (1964).

55. S.M. Stanley, Proc. Natl. Acad. Sci. U.S.A. 72, 646-650 (1975).

56. Conrad has taken what is essentially a modifier approach that produces a similar outcome, in what he calls the "self-facilitation" of evolution, in M. Conrad, Biosystems 15, 83-85 (1982).

57. Support for this research was provided by NIH Grant GM21498 and the Dean's Postdoctoral Fellowship, C.F. Aaron Endowment.

References to Annotations

[1985] Altenberg, L. 1985. Knowledge representation in the genome: new genes, exons, and pleiotropy. *Genetics* 110, supplement: s41. Abstract of paper presented at the 1985 Meeting of the Genetics Society of America.

[1987] Doolittle, W. F. 1987. The origin and function of intervening sequences in DNA: A review. *American Naturalist* 130: 915–928.

[1987] Patthy, L. 1987. Intron-dependent evolution: preferred types of exons and introns. *Febs (Federation Of European Biochemical Societies) Letters* 214(1): 1–7.

[1988] Smith, M. W. 1988. Structure of vertebrate genes: a statistical analysis implicating selection. *Journal of Molecular Evolution* 27: 45–55.

[1992] Gelfand, M. S. 1992. Statistical analysis and prediction of the exonic structure of human genes. *Journal of Molecular Evolution* 35: 239–252.

[1996] Tomita, M. and Shimizu, N. and Brutlag, D. L. 1996. Introns and Reading Frames: Correlations Between Splice Sites and Their Codon Positions. *Molecular Biology and Evolution* 13: 1219–1223.