



## EVOLUTION ESCAPES RUGGED FITNESS LANDSCAPES BY GENE OR GENOME DOUBLING: THE BLESSING OF HIGHER DIMENSIONALITY\*

RICHARD GORDON

Departments of Botany, Radiology, Physics and Electrical & Computer Engineering,  
University of Manitoba, Winnipeg, Canada R3T 2N2

(Received 16 December 1993; in revised form 13 April 1994)

**Abstract**—Evolution on rugged landscapes ceases when a local maximum is attained. This poses the problem of how evolution could approach or attain a global maximum, especially for large genomes for which quasispecies are ineffective. I show how increasing the dimensionality of the landscape, which occurs every time there is a gene or higher order duplication (up to polyploidy), may solve this problem. Epistasis or complementarity between the duplicated genes provides an all uphill pathway towards the global maximum. The evolution of hemoglobin and other dimeric and tetrameric proteins provides a testable case, since fitness is readily defined.

### INTRODUCTION

The New Synthesis has paid only lip service to the role of increase in genome size as a mechanism of evolution. This is because the theory has concentrated on maximizing fitness in a given allele space (Wright, 1931, 1932; Kauffman, 1993; Macken, 1993; Stadler, 1993). Also, while we are adept at taking projections to get from higher to lower dimensional spaces (Penna & Patterson, 1986), our mathematics offers little of richness or elegance, other than perhaps visual hulls (Meisters & Ulam, 1967; Beyer & Ulam, 1968) and computed tomography (Gordon *et al.*, 1975), when we concatenate subspaces. However, as we learned with the invention of the airplane, and as all deep thinking flatlanders believe (Abbott, 1952), if they cannot perceive, there are incredible new opportunities to be gained by an increase in the dimension of the space one is prepared to explore. Since proteins (Ohno, 1970; Savageau, 1986; Lewis, 1986), genes (Ohta, 1987a, b, 1990), and genomes (Sparrow & Nauman, 1976) do appear to increase by doubling, it behooves us to tag along with them and discover the consequences.

Besides, the more we delve into the mathematics of adaptive landscapes, the more it becomes clear that achieving global optimization via natural selection is a tenuous proposition (*cf* Jacob, 1977). The problem,

as reviewed, albeit in a positive light by Kauffman (1993), is that adaptive or fitness landscapes are rugged, even fractal landscapes (*cf* Conrad & Ebeling, 1992; Weinberger & Stadler, 1993).

The quasispecies solution (Eigen, 1971; Demetrius, 1983), to send huge numbers of variant offspring out over the rugged landscapes, reaches a "limit at a sequence length of a hundred to a thousand repeating units" (Eigen & Winkler-Oswatitsch, 1992). It thus generally (*cf* Clarke *et al.*, 1993) works fine for viruses (Domingo *et al.*, 1985; Steinhauer & Holland, 1987; Epstein *et al.*, 1991; Eigen, 1993a, b), but not for us "higher" organisms with big genomes. Eigen & Winkler-Oswatitsch (1992) are driven to conclude "that the optimization of individual genes must have taken place before their integration into a giant molecule, the genome", perhaps unnecessarily.

In addition to the empirical evidence for genome doubling of Sparrow & Nauman (1976), we know that, at least for plants, change in ploidy is common in the creation of new species (Ray *et al.*, 1983). Lower vertebrates are also tolerant of ploidy changes [reviewed in Duellman & Trueb (1986)] and their speciation may sometimes involve ploidy change (Lewis, 1980).

Gene doubling is the basis of the evolutionary appearance of whole families of related genes and pseudogenes (Alberts *et al.*, 1994). Unfortunately, as each individual new gene may have a new function, it is hard to ascribe a fitness to each (Franklin & Lewontin, 1970) and estimate the overall improvement in the fitness of the organism, as much as we might believe that to occur.

\*Some topics included in this work have been discussed during the *Third International Workshop on Open Problems of Computational Molecular Biology*, Telluride, Colo., 11-25 July 1993.

### THE EVOLUTION OF HEMOGLOBIN

There is one category of genes, however, which permits us to examine improvement in fitness upon gene duplication. These are the dimeric and tetrameric proteins. Hemoglobin is an elegant example. In humans, it is a dimer of a dimer, consisting of alpha- and beta-globins. The two monomers are similar (Singer & Berg, 1991), which belies a common origin from a gene duplication.

Why should hemoglobin have evolved this way? We can phrase this question quantitatively, because there are functioning monomeric hemoglobins in lower organisms:

"Multiple genes and pseudogenes for the  $\alpha$ - and  $\beta$ -chains occur in all vertebrates, and these are part of a multigene superfamily that includes the coding sequences for invertebrate globins, vertebrate myoglobin, and plant leghemoglobin. All these proteins are functionally similar in that they contain heme and associate reversibly with oxygen. . . . A single 'globin' gene is assumed to have existed in a remote ancestor of contemporary species. An amplification of the gene is postulated more than 400 million years ago because, although the cyclostome fishes have a single globin chain, the jawed fishes, the gnathostomes, encode both  $\alpha$ - and  $\beta$ -globin chains" (Singer & Berg, 1991).

Quantification is available via the rate constants for the binding and release of oxygen, the presumed major functions of hemoglobin. Other cases, such as alpha- and beta-tubulin (Raff *et al.*, 1987; Joshi & Cleveland, 1990), do not yet permit such a clear statement of function and variation over phyla, let alone quantification.

Here then is our problem. "Primitive" hemoglobin represents a lower adaptive peak than "advanced" hemoglobin. How could evolution have gotten from the lower peak to the higher? It almost seems obvious, at least in retrospect, since we know this happened by gene duplication. But let us look at the problem from the point of view of rugged fitness landscapes.

First we must note that the primitive and advanced peaks are in different spaces. We could get around this inconvenience by considering the fact that the primitive hemoglobin's landscape is actually a subspace of the advanced hemoglobin's landscape. If we look at it this way, the improvement represented by advanced hemoglobin would seem to have little or nothing to do with the fact that it is a tetramer. We could conceive of the whole structure alternatively being built out of a single gene producing a long protein with four functional parts. For instance, we could imagine, to keep everything in one space, that the primitive hemoglobin gene happens to be adjacent to three times as much junk DNA (Ohno, 1972), which base by base is converted over evolutionary time to advanced hemoglobin. [This could even be construed to provide an "anticipatory purpose" or "preadaptation": Bock (1959); Eldredge

(1989) for junk DNA.] Of course, a few point mutations would also occur in the initial primitive segment as well.

We now see the absurdity of believing that hemoglobin could evolve this way. With a fixed dimensional landscape, we are forced to imagine that hemoglobin evolution started with a functional, if suboptimal molecule. Then step by step, every so many generations on average, the molecule accumulated additions which neither enhanced the functionality much, if any, and also did not hinder what functionality already existed. Moreover, it is hard to see why these incremental changes should be driven by selection towards providing improved functionality (better oxygen binding and release kinetics), especially since this property may depend on allosteric properties only ascribable to even multiples of the primitive protein. It is also hard to believe that there is any directionality along this pathway over the landscape. In fact, the valley between the two peaks is at best quite flat [a "golf course": Kauffman (1993, p. 112)], with no gradients pointing towards the advanced peak, or, even worse, deep crevices *en route*, if all that extra length of peptide got in the way of the (primitive) functional portion. It is indeed a rough landscape, but with little hope to offer the would be mountain climber, who is, so to speak, always in a fog, unable to see and direct herself towards the peak she would like to climb.

### INCREASING THE DIMENSIONALITY

I have had to use the conceptual artifact of adding junk DNA to the primitive gene, in order to visualize the hemoglobin evolutionary process in one space. While the C-value paradox is tantalizing (Cavalier-Smith, 1985), I would not really like to suggest that the function of junk DNA is to prepare for advanced evolution *cons hence* (*cf* Martin & Gordon, 1994). So now let's try to approach the problem from the point of view of actual increase in fitness landscape dimensionality, rather than as if all the dimensions were present "from the beginning", a concept that smacks too much of the homunculus (Gardner, 1972; Casti, 1990).

Let us consider the most elementary model, in which fitness varies along a single axis representing the alleles of a single genome (Fig. 1). Evolution has brought us to the point where the population is trapped on the lower of two peaks in the adaptive landscape. How do we get to the higher peak?

If we invoke an initial dimerization by gene duplication it would have the primary initial effect of producing more of the same protein. In general [though there are pathologies possible, as evidenced by Down's syndrome and other trisomies: Dyban & Baranov (1987)], we may presume that gene duplication represents only a minor genetic burden (if not an immediate advantage) to the individual in which the mutation first appeared. At this point, then, we

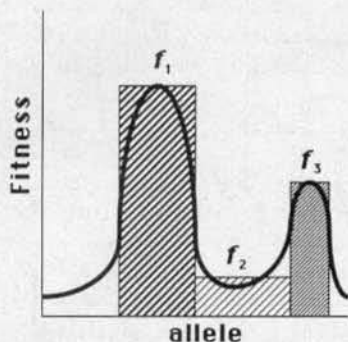


Fig. 1. Fitness is assumed to be a function of a more or less continuously variable allele. The fitness has a global maximum fitness of height  $f_1$  and a secondary peak in fitness of height  $f_3$  with a valley inbetween of fitness  $f_2$ . The peaks and valley are approximated by rectangles for further figures.

have two copies of the primitive gene with perhaps no particular advantage or disadvantage.

The subsequent divergence of the two monomers by point mutations, so that they both perhaps bind better to one another and provide greater overall functionality, may have occurred in a correlated manner. Such correlated point mutations would reduce the search space and perhaps aid in getting to the global maximum in the double dimension space. Thus, while epistasis is usually seen as providing a set of conflicting constraints that lead to a rough fitness landscape (Kauffman, 1993), in the case of gene doubling epistasis may actually smooth the way to the global maximum. This correlated evolution of two genes might be thought of as "intraspecific coevolution", for lack of a better name. A similar phenomenon may occur in a gene that is produced by the complementary strand to a protein, perhaps coding its chaperone (LéJohn *et al.*, 1994).

#### THE EPISTASIS LANDSCAPE

Let's try to generate a model, in the context of adaptive landscapes, of how epistasis between the primitive monomer and its duplicate might smooth the transition from a lower to a higher peak. The first thing to note is that our representative one-dimensional space in Fig. 1 is now a two-dimensional space (Fig. 2). Two simple models for combining fitnesses of two or more genes have been suggested, namely additive and multiplicative:

"The population can be thought of as a tight or loose cluster of individuals located at different points in the landscape. Mutations move an individual, or its offspring, to neighboring points in the space, representing neighboring genotypes. . . . I now introduce a simple formal model of rugged fitness landscapes, called the  $NK$  model. In this model,  $N$  refers to the number of parts of a system—genes in a genotype. . . . Each part makes a fitness contribution which depends upon that part and upon  $K$  other parts among the  $N$ . That is,  $K$  reflects how richly cross-coupled the system is. In the geneticist's term,  $K$  measures the richness of epistatic interactions among the components of the system. . . . The number of one-mutant neighbors, which I will symbolize by

' $D$ ' is the *dimensionality* of the genotype space and gives the number of directions in which each genotype can change to another neighboring genotype by a minimal alteration. . . . The most idealized genetic models assume that each locus contributes to the overall fitness of the genotype *independently* of all the other loci. Therefore, given the fitness contribution of the allele at each locus, the fitness of the genotype is just the sum of the  $N$  independent fitness contributions divided by  $N$ , hence the average of those contributions. . . . The theoretical problem is how to build useful models of epistatic interactions. One device commonly used has been to assume that genetic loci which interact can be represented by *multiplying* their fitness contributions (Franklin & Lewontin, 1970; Lewontin, 1974; Ewens, 1979)" (Kauffman, 1993, pp. 34–41).

These two options correspond to the simple additive and multiplicative backprojection algorithms of computed tomography when we go from one to two dimensions (Gordon & Herman, 1974; Gordon *et al.*, 1975). It should be noted that these algorithms have a tendency to reduce the contrast of the reconstructed image, in comparison to more accurate computed tomography algorithms, which treat pixels on an individual basis. This very contrast reduction could make the higher peak more accessible. This is another plausible mechanism for smoothing rugged landscapes.

If the two copies of the primitive gene now drift apart but also coevolve, we start taking a trajectory over the two dimensional adaptive landscape. But coevolution can mean epistatic interactions. Thus we cannot know *a priori* whether to add or multiply the fitnesses  $f_i$  and  $f_j$ . In general we can call the operator "&", so that the fitness of the organism is  $f_i$  &  $f_j$ . While we could take a linear combination of the addition and multiplication, for simplicity let us presume that we could have one or the other, and represent these cases by 0 (additive: "&" = "+") or 1 (multiplicative: "&" = "×"), respectively.

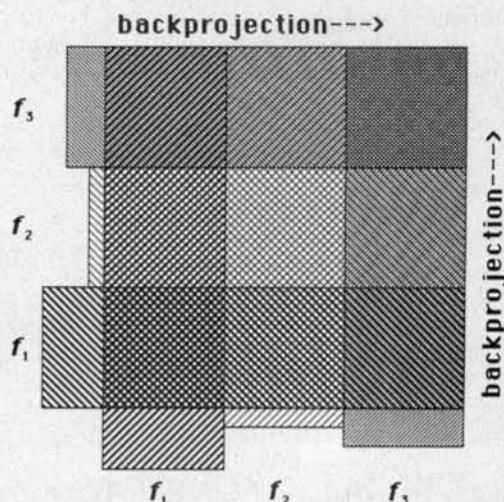


Fig. 2. Upon gene doubling, the adaptive landscape of Fig. 1 is converted from one to two dimensions. This is qualitatively represented by overlapping of the backprojected strips representing the two adaptive peaks and the valley inbetween.

0	1	1
1	0	0
0	1	0

Fig. 3. A representative epistasis landscape for Fig. 2.

It would be reasonable to assume that epistasis itself can evolve. Thus let us define an "epistasis landscape", whose scalar value is the degree of epistasis rather than the fitness scalar of the adaptive landscape. In our case, the epistasis landscape is simply a  $3 \times 3$  binary matrix, if we assume that epistasis either does or does not exist between two alleles of the two genes, and by assuming that we have only three alleles: the suboptimal peak, the optimal peak, and the valley inbetween (Fig. 3). Rather than create a theory of the evolution of epistasis, let us simply note that in our simplified case we have  $2^9$  epistasis landscapes possible. The  $2^9$  epistasis landscapes generate  $2^9$  adaptive landscapes, in which each square in Fig. 2 is either the sum or the product of the fitnesses, as shown in the example in Fig. 4.

As we can see in Fig. 5, there are six pathways involving one-mutant changes to get from the primitive (lower) to the advanced (higher) peak, without backtracking. Two of these pathways completely avoid the central valley. Thus for the double dimension case to have an advantage over the single dimension case of Fig. 1 (which has only one pathway, through the one dimensional valley), at least one of the six pathways should be easier to traverse than the pathway of the one dimensional case.

$f_3$	$f_1 + f_3$	$f_2 f_3$	$f_3^2$
$f_2$	$f_1 f_2$	$2f_2$	$f_2 + f_3$
$f_1$	$2f_1$	$f_1 f_2$	$f_1 + f_3$
	$f_1$	$f_2$	$f_3$

Fig. 4. The adaptive landscape corresponding to the epistasis landscape of Fig. 3.

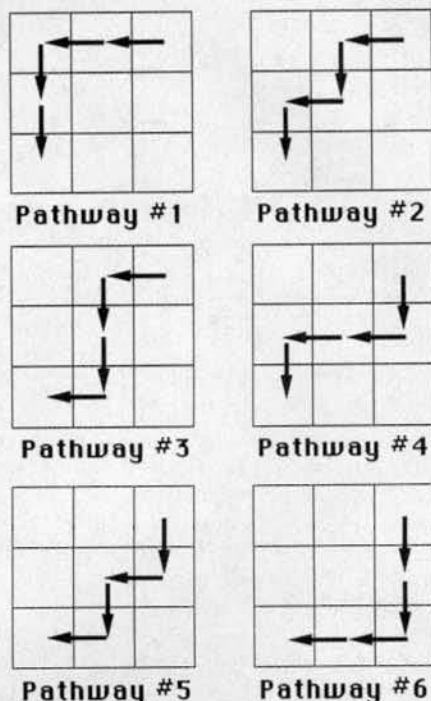


Fig. 5. The six pathways (without backtracking) for getting from the lower to the higher adaptive peak in our two dimensional adaptive landscape.

If we allowed more than three alleles, there would be many more pathways that avoid the central valley, so that I am actually handicapping the argument by restricting myself to three alleles. For the combinatorically inclined reader, I pose the (perhaps already solved) problem of determining the number of pathways for  $A$  alleles along each gene axis, and what fraction of those pathways avoid the central valley. This fraction would seem to approach 1 as  $A$  increases. Of course, the one-mutant changes relating the alleles of a single gene could no longer be presumed to lie along a single axis, which thus has to be expanded to a space of appropriate dimension. But this makes the number of pathways that avoid the central valley even larger.

Any pathway is going to involve some descent from the lower peak before the higher peak can be ascended. We thus need a model for the rate of proceeding along a pathway with ups and downs. Unfortunately, I have not been able to find this in Kauffman (1993), since all of his one-mutant algorithms can only go up:

"Simulations with this modified  $NK$  model were carried out under three 'move' procedures. In the first, adaptive steps occurred via the fittest one-mutant variant. This corresponds to a greedy gradient ascent and to the situation in which the fittest variant sweeps the population. In the second, one of the fitter one-mutant variants, if any existed, was chosen at random at each iteration. In the third, a random one-mutant variant was sampled and, if fitter, the adaptive process stepped to that variant" (Kauffman, 1993, p. 79).

While "Higher temperature is analogous to a higher mutation rate" (Kauffman, 1993, p. 111), it is not at all clear whether one should use the Boltzmann or some other distribution, or whether the kinetics should involve only the current state and the barrier or the current and final states (Gordon, 1968). We are presently lacking an acceptable model for the kinetics along a given pathway on an adaptive landscape, when that pathway involves both ups and downs. Such a model is needed before we can compare the relative rates along different pathways.

At this point I rest my argument along these lines, because it is time for constructing a kinetic model and then doing a computation. We could create a unit cube whose axes are  $f_1$ ,  $f_2$  and  $f_3$ , corresponding to the advanced peak, the valley, and the primitive peak, respectively. Over that region for which  $f_1 > f_3 > f_2$ , we could label the point  $(f_1, f_2, f_3)$  with the index of the pathway of choice. To determine this index we must evaluate the ease of getting from the lower to the higher peak over the 6 pathways (requiring a kinetics model) and the  $2^9$  epistasis landscapes, which gives us 3072 cases per point, a reasonable number to test by computer.

If we assume some genetic drift occurs, we might additionally model it by letting the two axes have different values for  $(f_1, f_2, f_3)$  and  $(f'_1, f'_2, f'_3)$ , in which case each term shown in the figures should be replaced by  $f_i$  &  $f'_i$ . It remains to try this by computation to see if such symmetry breaking is essential to establishing a two dimensional pathway that avoids the central valley.

#### THE MAX FUNCTION ON LANDSCAPES OF DUPLICATED GENES

So far I have accepted the approach of the New Synthesis, merely trying to generalize it in a minimal way to the case of gene duplication. But it is worth

$\max(f_i, f_j)$  with  $f_1 > f_3 > f_2$

$f_3$	$f_1$	$f_3$	$f_3$
$f_2$	$f_1$	$f_2$	$f_3$
$f_1$	$f_1$	$f_1$	$f_1$
	$f_1$	$f_2$	$f_3$

Fig. 6. The "max" operator applied to the duplication of the landscape of Fig. 1.

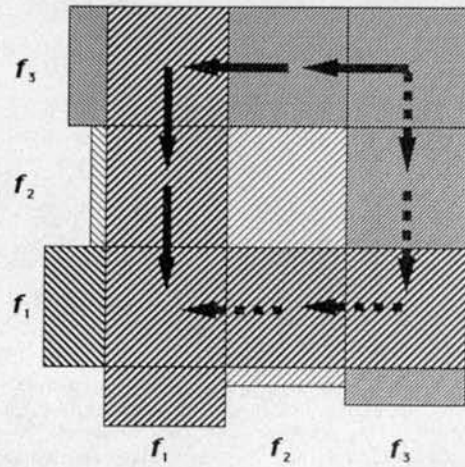


Fig. 7. A gray scale representation of Fig. 6 showing how the "max" operator skirts the valley between the peaks for a duplicated gene.

heeding Goldschmidt's warning:

"I do not want to create the impression that I underrate the mathematical study of selection problems, as found in the brilliant work of Fisher, Haldane, Volterra, Wright. . . . I am fully aware of the importance of this now-popular branch of evolutionary research. But it is necessary to remember an old remark of Johannsen in his criticism of Galtonian biometry; namely, that biology must be studied *with* mathematics but not *as* mathematics. . . . I am of the opinion that this criticism applies also to the mathematical study of evolution. This study takes it for granted that evolution proceeds by slow accumulation of micromutations through selection. . . . If, however, evolution does not proceed according to the neo-Darwinian scheme, its mathematical study turns out to be based on wrong premises" (Goldschmidt, 1940).

Taking this hint, let's reevaluate the fitness combination operator, "&". When a gene is duplicated, each of the pair initially makes the same gene product. As discussed above, while we can conceive

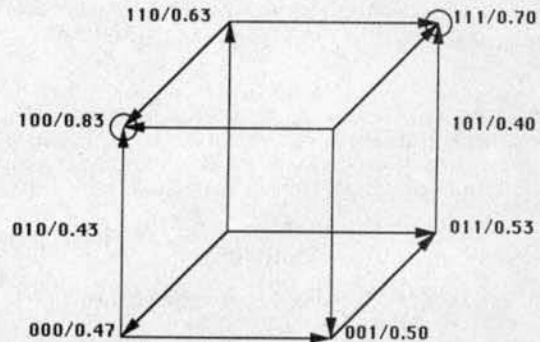


Fig. 8. An adaptive landscape "of three genes with random values for each of the eight combinations of [3] . . . alleles bearing on genes 1, 2 and 3. The fitness values then assign a fitness to each of the  $2^3 = 8$  possible genotypes as the mean value of the fitness contributions of the three genes. . . ." (Kauffman, 1993, p. 42, redrawn from Fig. 2.2c with permission from Oxford University Press, Inc.). The arrows indicate that trapping occurs at the local peak at (1, 1, 1), so that the global peak at (1, 0, 0) may not be accessible.

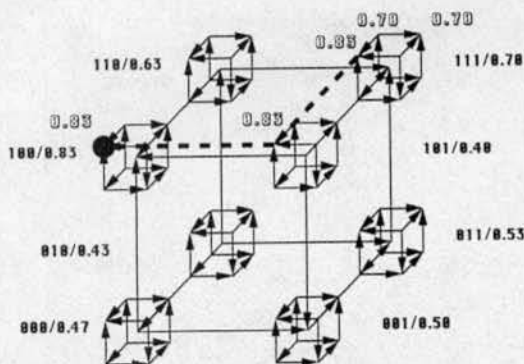


Fig. 9. The adaptive landscape of Fig. 8 is expanded in dimension via duplication of all three genes, resulting in a cube of cubes. The numbers in outline font are the "max" fitness values. The dashed line shows that there is now a path from the lower peak to the global peak with no valleys inbetween.

of pathological exceptions, in general the fitness of the organism initially bearing the mutation does not decrease nor increase. Therefore, both the additive model (giving an increased fitness) and the multiplicative model (giving a decreased fitness) are wrong for gene duplication. A simpler, and more appropriate operator for "&" would be "max", i.e. "&" = "max", so that the organism's fitness is  $f_i \max f_j$ , or, more conveniently,  $\max(f_i, f_j)$ . In other words, at any given moment, either gene will do, and whichever has the best gene product determines the fitness. With the maximum operator, there simply is no valley between the lower peak and the higher peak (Figs 6-9). Any nonbacktracking path will get to the higher peak, without loss of viability or fitness en route. Gene duplication thus smooths the rugged landscape, making higher maxima easily accessible. Instead of a curse (Marimont & Shapiro, 1979), at least in the case of gene doubling, and perhaps whole genome doubling, higher dimensionality turns out to be a blessing, perhaps the very one that makes increasing complexity and thus progressive evolution possible (Gordon, 1994).

**Acknowledgements**—I would like to thank Andrzej K. Konopka, Peter Shuster and Peter F. Stadler for taking the time in the midst of a fast paced and exciting conference to read the draft manuscript, and the referee for pointing out the need to relate these ideas to quasispecies.

## REFERENCES

- Abbott E. A. (1952) *Flatland, A Romance of Many Dimensions*, 6th rev. ed. Dover, New York.
- Alberts B., Bray D., Lewis J., Raff M., Roberts K. & Watson J. D. (1994) *Molecular Biology of the Cell*, 3rd edn. Garland Publishing, New York.
- Beyer W. A. & Ulam S. M. (1968) Note on the visual hull of a set. *J. Comb. Theory* **4**, 240.
- Bock W. J. (1959) Preadaptation and multiple evolutionary pathways. *Evolution* **13**, 194.
- Casti J. L. (1990) *Searching for Certainty, What Scientists Can Know About the Future*. William Morrow and Co., New York.
- Cavalier-Smith T. (Ed.) (1985) *The Evolution of Genome Size*. John Wiley & Sons, Chichester.
- Clarke D. K., Duarte E. A., Moya A., Elena S. F., Domingo E. & Holland J. (1993) Genetic bottlenecks and population passages cause profound fitness differences in RNA viruses. *J. Virol.* **67**, 222.
- Conrad M. & Ebeling W. (1992) M. V. Volkenstein, evolutionary thinking and the structure of fitness landscapes. *Biosystems* **27**(3), 125.
- Demetrius L. (1983) Selection and evolution in macromolecular systems. *J. Theor. Biol.* **103**, 619.
- Domingo E., Martinez-Salas E., Sobrino F., de la Torre J. C., Portela A., Ortin J., López-Galindez C., Pérez-Brena P., Villanueva N., Nájera R., VandePol S., Steinhauer D., DePolo N. & Holland J. (1985) The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance—a review. *Gene* **40**, 1.
- Duellman W. E. & Trueb L. (1986) *Biology of Amphibians*. McGraw-Hill Book Co., New York.
- Dyban A. P. & Baranov V. S. (1987) *Cytogenetics of Mammalian Embryonic Development*. Clarendon Press, Oxford.
- Eigen M. (1971) Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**, 465.
- Eigen M. (1993a) Viral quasispecies. *Sci. Am.* **269**(1), 42.
- Eigen M. (1993b) The origin of genetic information: viruses as models. *Gene* **135**, 37.
- Eigen M. & Winkler-Oswatitsch R. (1992) *Steps Towards Life, A Perspective on Evolution*. Oxford University Press, Oxford.
- Eldredge N. (1989) *Macroevolutionary Dynamics: Species, Niches, and Adaptive Peaks*. McGraw-Hill Publishing Co., New York.
- Epstein L. G., Kuilen C., Blumberg B. M., Hartman S., Sharer L. R., Clement M. & Goudsmit J. (1991) HIV-1 V3 domain variation in brain and spleen of children with AIDS: tissue-specific evolution within host-determined quasispecies. *Virology* **180**, 583.
- Ewens W. J. (1979) *Mathematical Population Genetics*. Springer-Verlag, New York.
- Franklin I. & Lewontin R. C. (1970) Is the gene the unit of selection? *Genetics* **65**, 707.
- Gardner E. J. (1972) *History of Biology*, 3rd edn. Burgess Publishing Co., Minneapolis.
- Goldschmidt R. B. (1940) *The Material Basis of Evolution*, 1960 reprint edn. 46 Pageant Books, Paterson, New Jersey.
- Gordon R. (1968) Steady-state properties of Ising lattice membranes. *J. Chem. Phys.* **49**, 570.
- Gordon R. (1994) *The Hierarchical Genome and Differentiation Waves: Novel Unification of Development, Genetics and Evolution*. World Scientific, Singapore, in preparation.
- Gordon R. & Herman G. T. (1974) Three dimensional reconstruction from projections: a review of algorithms. *Int. Rev. Cytol.* **38**, 111.
- Gordon R., Herman G. T. & Johnson S. A. (1975) Image reconstruction from projections. *Scient. Am.* **233**(4), 56.
- Jacob F. (1977) Evolution and tinkering. *Science* **196**, 1161.
- Joshi H. C. & Cleveland D. W. (1990) Diversity among tubulin subunits: toward what functional end? *Cell Motil. Cytoskeleton* **16**, 159.
- Kauffman S. A. (1993) *Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, New York.
- LéJohn H. B., Cameron L. E., Yang B., MacBeath G., Barker D. S. & Williams S. A. (1994) Cloning and analysis of a constitutive heat shock (cognate) protein 70 gene inducible by L-glutamine. *J. Biol. Chem.* **269**, 4513.
- Lewin R. (1986) Unexpected size pattern in bacterial proteins. *Science* **232**, 825.

- Lewis W. H. (Ed.) (1980) *Polyploidy: Biological Relevance*. Plenum Press, New York.
- Lewontin R. C. (1974) *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York.
- Macken C. (1993) Evolution on rugged landscapes. *Abstracts, Open Problems of Computational Molecular Biology Third International Workshop*, 11–25 July 1993, pp. 51–53. Telluride Summer Research Center, Telluride, CO.
- Marimont R. B. & Shapiro M. B. (1979) Nearest neighbour searches and the curse of dimensionality. *J. Inst. Math. Applics.* **24**, 59.
- Martin C. C. & Gordon R. (1994) Differentiation trees, a junk DNA molecular clock, and the evolution of neoteny in salamanders. Submitted.
- Meisters G. H. & Ulam S. M. (1967) On visual hulls of sets. *Proc. Natl. Acad. Sci. USA* **57**, 1172.
- Ohno S. (1970) *Evolution by Gene Duplication*. Springer-Verlag, New York.
- Ohno S. (1972) So much "junk" DNA in our genome. *Evolution of Genetic Systems* (Edited by Smith H. H.), pp. 366–370. Gordon and Breach, New York.
- Ohta T. (1987a) A model of evolution for accumulating genetic information. *J. Theor. Biol.* **124**, 199.
- Ohta T. (1987b) Simulating evolution by gene duplication. *Genetics* **115**, 207.
- Ohta T. (1990) How gene families evolve. *Theor. Population Biol.* **37**, 213.
- Penna M. A. & Patterson R. R. (1986) *Projective Geometry and Its Applications to Computer Graphics*. Prentice-Hall, Englewood Cliffs, N.J.
- Raff E. C., Diaz H. B., Hoyle H. D., Hutchens J. A., Kimble M., Raff R. A., Rudolph J. E. & Subler M. A. (1987) Origin of multiple gene families: are there both functional and regulatory constraints? *Development as an Evolutionary Process* (Edited by Raff R. A. & Raff E. C.), pp. 203–238. Alan R. Liss, New York.
- Ray P. M., Steeves T. A., Fultz S. A. & Jacobs M. (1983) *Botany*. Saunders College Publishing, Philadelphia.
- Savageau M. S. (1986) Proteins of *Escherichia coli* come in sizes that are multiples of 14 kDa: domain concepts and evolutionary implications. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1198.
- Singer M. & Berg P. (1991) *Genes & Genomes, A Changing Perspective*. University Science Books, Mill Valley, California.
- Sparrow A. H. & Nauman A. F. (1976) Evolution of genome size by DNA doublings. *Science* **192**, 524.
- Stadler P. F. (1983) Evolution on fitness landscapes. *Abstracts, Open Problems of Computational Molecular Biology Third International Workshop*, 11–25 July 1993, p. 64. Telluride Summer Research Center, Telluride, Colorado.
- Steinhauer D. A. & Holland J. J. (1987) Rapid evolution of RNA viruses. *A. Rev. Microbiol.* **41**, 409.
- Weinberger E. D. & Stadler P. F. (1993) Why some fitness landscapes are fractal. *J. Theor. Biol.* **163**, 255.
- Wright S. (1931) Evolution in Mendelian populations. *Genetics* **16**, 97.
- Wright S. (1932) The roles of mutation, inbreeding, cross-breeding, and selection in evolution. *Proceedings of the Sixth International Congress of Genetics*, Vol. I, pp. 356–366.