Effects of selective neutrality on the evolution of molecular species

M. E. J. Newman[1] and Robin Engelhardt[2]

[1]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA*
[2]*Centre for Chaos And Turbulence Studies, Department of Chemistry, University of Copenhagen, Universitetsparken 5, Copenhagen Ø, Denmark*

We introduce a model of evolution on a fitness landscape possessing a tunable degree of neutrality. The model allows us to study the general properties of molecular species undergoing neutral evolution. We find that a number of phenomena seen in RNA sequence-structure maps are present also in our general model. Examples are the occurrence of 'common' structures that occupy a fraction of the genotype space which tends to unity as the length of the genotype increases, and the formation of percolating neutral networks that cover the genotype space in such a way that a member of such a network can be found within a small radius of any point in the space. We also describe a number of new phenomena that appear to be general properties of systems possessing selective neutrality. In particular, we show that the maximum fitness attained during the adaptive walk of a population evolving on such a fitness landscape increases with increasing degree of neutrality, and is directly related to the fitness of the most fit percolating network.

Keywords: fitness landscapes; RNA structure; neutral evolution; molecular evolution

1. INTRODUCTION

Biological molecules such as proteins and RNAs undergo evolution just as organisms do, selected for their ability to perform certain functions by the reproductive success which that ability imparts to their hosts. It is believed that many mutations of a molecule are evolutionarily neutral in the sense that they do not change the fitness of the molecule to perform the function for which it has been selected. We have many examples of proteins that appear to possess approximately the same conformation and to perform the same function in different species, but which have different sequences. Such proteins may differ only by a single amino acid or may have whole regions that have been substituted or inserted, or they may even be so different as to appear completely unrelated. A mutation is said to be neutral if it changes a molecule into one of these functional equivalents, leaving the viability of its host unchanged. This idea was first explored in detail by Kimura (1955, 1983).

In fact, as Ohta (1972) has pointed out, it is not necessary that the fitness of a molecule remains precisely the same under a given mutation for that mutation to be considered neutral. In populations of small size, genetic drift becomes the more important factor and small differences in fitness can be neglected by comparison. In effect, drift places a limit on the resolution with which selection can detect changes of fitness, so that small fitness changes are effectively, if not precisely, neutral.

It is possible that the concept of selective neutrality can also be applied to the evolution of entire organisms. Certainly there are changes possible in an organism's genome which have no immediate effect on its reproductive success, or which produce an effect sufficiently small that selection cannot detect it in a given population. In this paper we will primarily use the language of molecular evolution, but the reader should bear in mind that the ideas described may have wider applicability.

Despite the long history of the idea, many aspects of neutral evolution are still not well understood. In particular, we have very little idea of the general behaviours that can be expected of systems (molecules or organisms) with a significant degree of evolutionary neutrality. The primary reason for this gap in our understanding is that, despite many decades of hard work, we still have a rather poor idea of the way in which genomic sequences map onto molecular structures and hence onto a fitness measure. In the case of entire organisms the equivalent problem is that of calculating the genotype–phenotype mapping, which is even less well understood. One simple case in which neutral evolution has been investigated in some detail is that of RNA structure (Schuster *et al.* 1994; Grüner *et al.* 1996a,b; Huynen *et al.* 1996; Reidys *et al.* 1997), although calculations so far are limited to secondary structures, and even these cannot be calculated with any reliability, so that these studies should be taken more as a qualitative guide to the behaviour of systems undergoing neutral evolution than an accurate representation of the real world. The trouble with this approach, however, is that RNAs are not a sufficiently general model that the results gained from their study can be applied to other systems, such as protein evolution or the evolution of whole organisms.

At the other extreme, studies have been performed of extremely simple mathematical models of neutral

Proc. R. Soc. Lond. B (1998) **265**, 1333–1338
Received 5 January 1998   *Accepted* 11 March 1998

1333

© 1998 The Royal Society

evolution in the context of genetic algorithms (Prügel-Bennett & Shapiro 1994; Mitchell 1996). An example is the 'Royal Road' genetic algorithm studied by van Nimwegen *et al.* (1997, 1998), in which individuals' fitness is a sum of contributions from a number of genes, but each gene only makes a contribution if it has one particular 'correct' value. (All other values make zero contribution.) Such models possess a high degree of evolutionary neutrality, while at the same time being simple enough to yield to analytic methods. Like RNA secondary structures, these models have given us some insight into the type of effects we may expect neutral evolution to produce, but, like RNAs, they are not sufficiently general to be sure that these insights apply to other systems as well.

In this paper, therefore, we propose a new mathematical model of neutral evolution. This model is an abstract model of a genotype-to-fitness map in the spirit of the Royal Road model. This approach allows us to sidestep the problems of incorporating the chemistry of real molecules in our calculations and to investigate the properties of the system more quickly and in greater detail than is possible with, for example, RNA structure calculations. In addition, the model is more general than either the Royal Road fitness function or the RNA sequence-structure maps of Huynen *et al.* (1996) and Grüner *et al.* (1996*a*). In fact, it possesses regimes in which it mimics the behaviour of both of these systems, as well as protein- and organism-like regimes. Because the behaviours of our model cover such a wide range of possibilities, it seems reasonable to conjecture that generic features of the model which span all of these regimes may be common to most systems possessing selective neutrality. This is the power of our model, and these general results are the results that we will concentrate on in this paper; we believe that the generic behaviours of our model should be visible in the evolution of real systems such as proteins, which are, as yet, beyond our ability to study directly.

In §2 we introduce our landscape model of neutral evolution. In §3 we discuss its properties and compare these with previous results for other systems undergoing neutral evolution. In §4 we discuss the implications of our results for evolving molecular species. In §5 we give our conclusions.

2. THE MODEL

Selective neutrality arises as a result of the many-to-one nature of the sequence-structure or genotype-phenotype maps found in biological systems. Many protein sequences, for example, map onto the same tertiary structure, and as the fitness is primarily a function of structure, such sequences possess (at least approximately) the same fitness (Tacker *et al.* 1996). We wish to construct a model of this phenomenon without resorting to actual calculations of the structure of any particular class of molecules. Our approach is to employ a 'fitness landscape' model of the type first proposed by Wright (1967, 1982), which maps sequence (or genotype) directly to fitness. Structures (or phenotypes) appear in our model as contiguous sets or 'neutral networks' of sequences possessing the same fitness.

Our model is a generalization of the NK model introduced by Kauffman & Johnsen (1991). Consider a sequence of N loci, which correspond to the nucleotides in an RNA or to amino acids in the case of a protein. At each locus i we have a value x_i drawn from an appropriate alphabet, such as $\{A,C,G,U\}$ for RNAs, or the set of 20 amino acids in the case of proteins. We denote the size of the alphabet by A. Each locus interacts with a number K of other 'neighbour' loci, which may be chosen at random or in any other way we wish. (Kauffman refers to these interactions as epistatic interactions, although this nomenclature is strictly only appropriate to the case where we are modelling the fitness of whole organisms.) In the case of RNAs, bases most often interact with one other base to form either a Watson–Crick or a G-U pair. Some bases have both pairing and tertiary interactions. Some, in the single-stranded regions, have very little interaction with any others. Thus a value of $K = 1$ might be approximately correct for RNA. For proteins, which have more complex types of interactions, a higher value of K may be appropriate. Each locus i makes a contribution w_i to the fitness of the sequence, whose magnitude depends on the value x_i at that locus and also on the values at each of the K neighbouring loci. There are A^{K+1} possible sets of values for the $K + 1$ loci in this neighbourhood, and hence A^{K+1} possible values of w_i. Following Kauffman & Johnsen (1991) we choose this set of values at random. However, Kauffman & Johnsen chose the values to be random real numbers in the interval $0 \leqslant w_i < 1$. We by contrast choose them to be integers in the range $0 \leqslant w_i < F$. Thus if $F = 2$, for example, each contribution w_i is either zero or one. Now we define the fitness W of the entire sequence to be proportional to the sum of the contributions at each locus:

$$W = \frac{1}{N(F-1)} \sum_i w_i. \tag{1}$$

The fitness of all sequences thus falls in the range from zero to one, and there are $NF - N + 1$ possible fitness values in this range.

In the limit in which $F \to \infty$, the probability that two sequences will possess the same fitness becomes vanishingly small, and our model therefore possesses no neutrality and is in fact exactly equivalent to the NK model. However, when F is finite, the probability of two sequences possessing the same fitness is finite, so that the model possesses neutrality to a degree that increases as F decreases. Neutrality is greatest when F takes the smallest possible value of 2. Two sequences with the same fitness may be equivalent either to molecules that fold into the same conformation and perform the same function, or to molecules with different conformations but approximately the same contribution to the reproductive success of the host organism. The ruggedness of the landscape is controlled by the parameter K, and is largest when K takes the maximal value of $N - 1$ (Kauffman & Johnsen 1991; Weinberger 1991). In the next section, we investigate the properties of the landscapes generated by our model, and show that with the right choice of parameters they can be used to mimic real biological systems, such as RNAs.
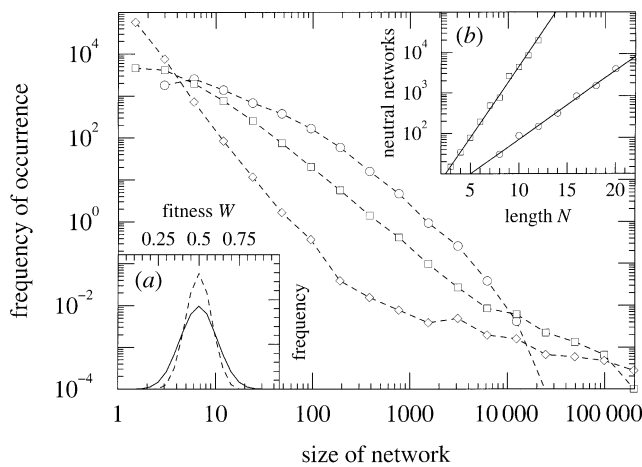
Figure 1. The frequency of occurrence of neutral networks as a function of size for a landscape with $\mathcal{N}=20$, $A=2$, $F=2$, and $K=1$ (circles), $K=2$ (squares) and $K=5$ (diamonds). Inset (*a*): the frequency of occurrence of sequences as a function of fitness for $\mathcal{N}=20$, $A=2$, $F=2$, and $K=19$ (solid line), $K=0$ (dotted line). Inset (*b*): the number of neutral networks as a function of $\mathcal{N}$ for $K=1$, $F=2$, and $A=2$ (circles), $A=4$ (squares).

3. EVOLUTION ON NEUTRAL LANDSCAPES

The topology of a fitness landscape depends on the types of mutation allowed to molecules evolving on it. In biological evolution, point mutations—mutations of the value at a single locus—are the most common. In this case, a neutral network is defined to be a set of sequences that all possess the same fitness and that are connected together via such point mutations. In the molecular case, we assume that closely similar sequences have the same fitness because they fold into the same conformation, so that these neutral networks correspond to (tertiary) structures. In the organismal case, they correspond to phenotypes.

The model described in the last section possesses neutral networks of exactly this type. The total fitness W in the model ranges from zero to one, but the greatest number of sequences have fitness close to $W=0.5$. (In the extreme case where $K=\mathcal{N}-1$ the distribution of W is binomial. When $K<\mathcal{N}-1$ it is approximately but not exactly so. Examples of these distributions are shown in figure 1*a*.) We would therefore expect the largest neutral networks to be those with fitness close to $W=0.5$, and this is indeed what we find in practice.

Typically, there are a large number of small neutral networks and a small number of large ones. In figure 1 we show histograms of the sizes of the neutral networks for $\mathcal{N}=20$ and various values of K. For the RNA-like case $K=1$, the histogram appears to be convex, indicating a distribution that falls off faster than a power law. The same behaviour has been seen in RNA studies by Grüner *et al.* (1996*a*). As K increases the distribution flattens, and by the time we reach $K=5$ it is markedly concave. Thus the behaviour seen in RNAs is not in this case generic. For some intermediate value of K close to $K=2$, the distribution appears to be power law in form, perhaps indicating the divergence of some scale parameter governing the distribution, in a manner familiar from the study of critical phenomena (Binney *et al.* 1992).
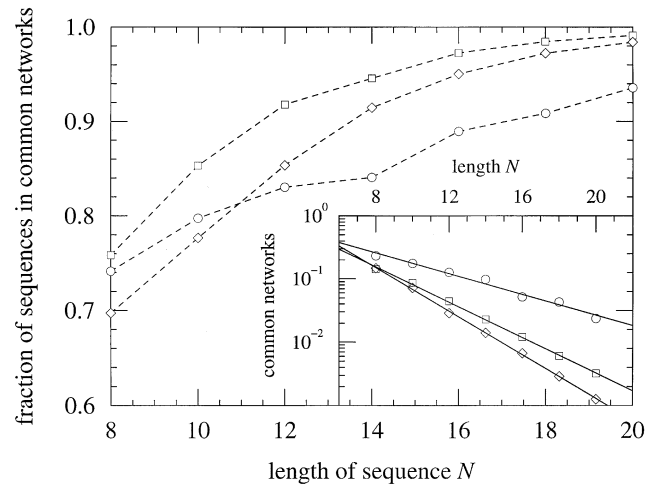

Figure 2. The fraction of sequences that fall in common networks as a function of $\mathcal{N}$ for a landscape with $A=2$, $F=2$, and $K=2$ (circles), $K=4$ (squares) and $K=6$ (diamonds). Inset: the number of common networks as a fraction of the total number of networks for the same landscapes.

We find that the total number of neutral networks $S_{\mathcal{N}}$ grows exponentially as $a^{\mathcal{N}}$ with sequence length. In figure 1*b* we show the number of networks in our model for $K=1$, both for two-letter {G,C} alphabets, and for a four-letter {A,C,G,U} alphabet. We find that $a \approx 1.5$ for the $A=2$ case, and $a \approx 2.3$ for the $A=4$ case. Interestingly, Stadler and co-workers (Grüner *et al.* 1996*a,b*; Stadler & Haslinger 1998) have studied the same distributions for RNA sequences using the full secondary-structure calculation and also find an exponential increase in the number of structures with sequence length, with values of $a=1.6$ and $a=2.35$ for the two- and four-letter cases, respectively. This suggests that this behaviour is more general than the specific secondary-structure map employed in the Stadler calculations.

The largest neutral networks on our landscapes percolate, which is to say, they fill the sequence space roughly uniformly, in such a way that no sequence is more than a certain distance away from a member of the percolating network. Determining which networks are percolating is not an easy task, however, so Grüner *et al.* (1996*a*) introduced instead the idea of a 'common' network, which is one that contains greater than the average number of sequences. In our model we find that the common networks form a small fraction of the total number of networks, that fraction decreasing exponentially as $\mathcal{N}$ increases, as shown in the inset to figure 2. The same result is found in RNAs (Grüner *et al.* 1996*a*).

In the main frame of figure 2 we show numerical results for the fraction of sequences that fall in the common networks as a function of $\mathcal{N}$. As the figure shows, this fraction increases with sequence length, tending to one in the limit of large $\mathcal{N}$. Even though the common networks form a smaller and smaller fraction of all networks as $\mathcal{N}$ becomes large, they nonetheless cover more and more of the sequence space. These results have interesting evolutionary implications: they imply that as sequences become longer, a larger and larger majority of structures (the small networks) are vanishingly unlikely to occur through natural selection. Evolution can only find the smaller and smaller fraction of 'common'
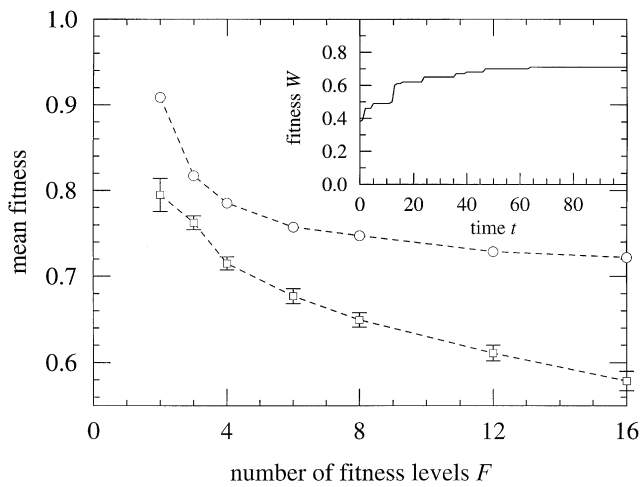
Figure 3. The maximum fitness attained by a random hill-climber averaged over 100 simulations with $\mathcal{N}=20$, $K=4$ and $A=2$, as a function of the neutrality parameter F (circles). The lower curve (squares) is the fitness of the most fit percolating neutral network averaged over the same 100 runs. Inset: the fitness of one of the hill-climbers in the simulation as a function of time.

structures. Similar behaviour has been found in RNAs and also in simple models of protein folding (Li *et al.* 1996). The results presented here suggest that the same conclusions may hold for other systems undergoing neutral evolution.

Next we have examined the dynamics of populations evolving on our landscapes. In their studies with $\mathcal{N}K$ landscapes, Kauffman & Johnsen (1991) made a useful approximation in representing evolving populations by a single dominant sequence. This approximation is only valid in the case in which the time-scale for mutation is much longer than the time-scale on which selection acts. For the moment we will assume this to be the case. A 'random hill-climber' is a population of this type, represented by a single dominant strain, which tries mutations—point mutations in the present case—until it finds one with higher fitness than the current strain. In this way the hill-climber performs an adaptive walk through sequences of ever-increasing fitness until it reaches a local fitness optimum. To study neutral landscapes we modify this strategy so that the hill-climber samples adjacent sequences at random until it finds one of fitness greater than or the same as itself. Such a climber will move at random on a neutral network until it finds a mutation that takes it to a network of higher fitness. In the upper curve of figure 3 we show the average fitness attained by a hill-climber over 100 simulations on our landscapes as a function of the neutrality parameter F. Recall that neutrality increases with decreasing F. As the figure shows, the climber, on average, finds higher fitness maxima for higher degrees of neutrality. In other words, neutrality helps the population to attain a greater fitness. This is certainly an idea that has been entertained before in the literature, but it is lent a new conviction when we see it emerge in the behaviour of a general model such as this.

The lower curve on figure 3 shows the fitness of the most fit percolating network averaged over the same 100 landscapes. This curve follows quite closely the form of the fitness of the local maxima found by the hill-climber.

Our explanation of this result is as follows. The climber moves diffusively on a neutral network until it finds a one-mutant neighbour that belongs to a network of greater fitness, at which point it shifts to that network. This process continues until it reaches a non-percolating network, at which point it is confined to the region occupied by the network and can only get as high as the local maximum within that region. Thus the highest fitness attainable on a landscape with neutrality depends directly on the highest fitness at which there are percolating networks. The landscapes with the greatest degree of neutrality also have the fittest percolating networks, which explains why higher fitnesses are attained on landscapes with lower values of F.

The inset to figure 3 shows the fitness of one of our hill-climbers as a function of time, and we can clearly see the jumps in this function where the climber finds its way onto a network of higher fitness. Similar jumps have been seen, for example, in laboratory experiments on the evolution of bacteria (Lenski & Travisano 1994). In the periods between jumps the climber diffuses around its network, testing new mutations to find one of higher fitness. Van Nimwegen *et al.* (1997*a*) have dubbed these periods 'epochs'. (They also bear some similarity to the palaeontological 'punctuated equilibria' described by Eldredge & Gould (1972), although there are many other possible explanations for the intermittency seen in fossil evolution.) The length of the epochs increases, on average, with increasing fitness. This behaviour was also seen in the Royal Road model, and occurs because as the fitness increases the number of structures with higher fitness still dwindles. The length of the epochs also depends on the rate of diffusion across the neutral network and on the density of 'points of contact' between the network and other networks of higher fitness (Kimura 1983).

Another interesting feature of the epochs seen in the Royal Road is that the average fitness of a population does not correspond exactly to the fitness of any of the networks. Typically, the average fitness is a little lower than the fitness of the dominant structure in the population because deleterious mutants are constantly arising. Even though these mutants are selected against, there are at any time enough of them in the population to make a noticeable difference to the average fitness. We would expect to see similar behaviour for populations evolving on our landscapes. Because the number of possible mutants with lower fitness than the dominant sequence increases with increasing fitness, it is also possible to get error threshold effects with increasing fitness (Eigen & Schuster 1979; Swetina & Schuster 1982). As the fitness increases, there may come a point where the rate at which deleterious mutants arise in the population exceeds the rate at which they are suppressed by selection, and at this point further improvement in fitness becomes impossible. (This effect has been studied in some detail for some simple landscape models (Woodcock & Higgs 1996).) Thus there may be a dynamical limit on the fitness of populations, independent of the limit imposed by the structure of the landscape discussed above. (This is true of landscapes without neutral evolution too, although the effect is much more prominent in the neutral case.)

Simulations similar in spirit to ours have been performed for populations of tRNAs by Fontana and co-workers (Fontana & Schuster 1987; Huynen *et al.* 1996). In these simulations the authors chose a 'target structure' which was artificially selected for, and they also observed epochs in the evolution as the population passed through a succession of increasingly fit structures on its way to the target.

4. DISCUSSION

The aim of this work is to study a model of neutral evolution which is general enough to encompass behaviours typical of other, more specific models that have been employed in the past. In this way we can reproduce, in a general context, the results that have been observed as special cases, and hence investigate the extent to which these results are general properties of fitness landscapes possessing neutrality, or particular to the systems in which they were first observed. In this spirit, we put forward the following conjectures about the fitness landscapes on which biological molecules evolve, based on the results of the investigations outlined in this paper.

1. The total number of possible structures increases exponentially with sequence length. The exponential constant of this increase appears to be approximately numerically equal in the general model and the only specific case in which it has been studied, that of RNA secondary structure.
2. There are a large number of structures that correspond to a small number of sequences, and a small number of structures that correspond to a large number of sequences. The exact form of the histogram of structure frequency, shown in figure 1, varies depending on the parameters of our model. However, for certain values of the parameters it has a form similar to that seen in RNA studies, whereas for others it appears to follow a power law.
3. The 'common' structures—ones that correspond to a large number of sequences—constitute an exponentially decreasing fraction of the total number as sequence length increases. Conversely, however, they cover a fraction of the sequence space which tends to unity for long sequences.
4. At least on short time-scales, evolution is dominated by the presence of neutrality. Neutrality helps populations to find structures of high fitness without having to cross fitness barriers. The highest fitness that can be found in this way is closely related to the fitness of the fittest percolating neutral network, which itself depends on the amount of neutrality. Therefore, for landscapes with a higher degree of neutrality, the population typically reaches a higher fitness.
5. The fitness may be limited by error threshold effects, which are particularly severe for landscapes of this type, because the size of the neutral networks (and hence the ratio of numbers of beneficial and harmful mutants) falls exponentially with increasing fitness.
6. Evolution proceeds in jumps separated by 'epochs' in which the fitness changes very little. An evolving population uses these epochs to diffuse across the current neutral network, allowing it to search a larger portion of sequence space for beneficial mutations.

5. CONCLUSIONS

To conclude, we believe that by studying a simple and general model of a neutral landscape, we should be able to distinguish properties of specific systems undergoing neutral evolution from properties common to all such systems. We have found a number of potential candidates for inclusion in a list of such common properties. There are many interesting lines of investigation which we have not been able to pursue in this short work, including details of the structure and size of the neutral networks such as percolation measures and covering radii, details of population dynamics on these networks including entropy and other statistical measures of the structure of such populations, calculations of the length of epochs, of the maximum fitness obtainable on these landscapes, of the effects of the error threshold on maximum fitness, and many effects of the variation of the parameters of the model, particularly the variation of the level of epistasis K and the neutrality parameter F. Some of these questions will be addressed in a forthcoming work.

After this work was completed, the authors became aware of Barnet's (1997) thesis, which suggests a different generalization of the NK model to landscapes with neutrality. We are grateful to Inman Harvey for bringing this to our attention.

REFERENCES

Barnet, L. 1997 *Tangled webs: evolutionary dynamics on fitness landscapes with neutrality.* MSc thesis, University of Sussex.

Binney, J. J., Dowrick, N. J., Fisher, A. J. & Newman, M. E. J. 1992 *The theory of critical phenomena.* Oxford University Press.

Eigen, M. & Schuster, P. 1979 *The hypercycle: a principle of natural self-organization.* New York: Springer.

Eldredge, N. & Gould, S. J. 1972 Punctuated equilibria: an alternative to phyletic gradualism. In *Models in paleobiology* (ed. T. J. M. Schopf), pp. 82–115. San Francisco: Freeman.

Fontana, W. & Schuster, P. 1987 A computer model of evolutionary optimization. *Biophys. Chem.* **26**, 123–147.

Grüner, W., Giegerich, R., Strothmann, D., Reidys, C., Weber, J., Hofacker, I. L., Stadler, P. F. & Schuster, P. 1996*a* Analysis of RNA sequence structure maps by exhaustive enumeration: neutral networks. *Mh. Chem.* **127**, 355–374.

Grüner, W., Giegerich, R., Strothmann, D., Reidys, C., Weber, J., Hofacker, I. L., Stadler, P. F. & Schuster, P. 1996*b* Analysis of RNA sequence structure maps by exhaustive enumeration: structure of neutral networks and shape space covering. *Mh. Chem.* **127**, 375–389.

Huynen, M., Stadler, P. F. & Fontana, W. 1996 Smoothness within ruggedness: the role of neutrality in adaptation. *Proc. Natn. Acad. Sci. USA* **93**, 397–401.

Kauffman, S. A. & Johnsen, S. 1991 Coevolution to the edge of chaos: coupled fitness landscapes, poised states, and co-evolutionary avalanches. *J. Theor. Biol.* **149**, 467–505.

Kimura, M. 1955 Solution of a process of random genetic drift with a continuous model. *Proc. Natn. Acad. Sci. USA* **41**, 144–150.

Kimura, M. 1983 *The neutral theory of molecular evolution*. Cambridge University Press.

Lenski, R. E. & Travisano, M. 1994 Dynamics of adaptation and diversification: A 10 000-generation experiment with bacterial populations. *Proc. Natn. Acad. Sci. USA* **91**, 6808–6814.

Li, H., Helling, R., Tang, C. & Wingreen, N. 1996 Emergence of preferred structures in a simple model of protein folding. *Science* **273**, 666–669.

Mitchell, M. 1996 *An introduction to genetic algorithms*. Cambridge, MA: MIT Press.

Ohta, T. 1972 Population size and rate of evolution. *J. Molec. Evol.* **1**, 305–314.

Prügel-Bennett, A. & Shapiro, J. 1994 An analysis of genetic algorithms using statistical mechanics. *Phys. Rev. Lett.* **72**, 1305–1309.

Reidys, C. M., Stadler, P. F. & Schuster, P. 1997 Generic properties of combinatory maps and neutral networks of RNA secondary structures. *Bull. Math. Biol.* **59**, 339–397.

Schuster, P., Fontana, W., Stadler, P. F. & Hofacker, I. L. 1994 From sequences to shapes and back: a case study in RNA secondary structures. *Proc. R. Soc. Lond.* B **255**, 279–284.

Stadler, P. F. & Haslinger C. 1998 RNA structures with pseudo-knots. *Bull. Math. Biol.* (Submitted.)

Swetina, J. & Schuster, P. 1982 Self-replication with error: a model for polynucleotide replication. *Biophys. Chem.* **16**, 329–345.

Tacker, M., Stadler, P. F., Bornberg-Bauer, E. G., Hofacker, I. & Schuster, P. 1996 Algorithm-independent properties of RNA secondary structure prediction. *Eur. Biophys. J.* **25**, 115–130.

van Nimwegen, E., Crutchfield, J. P. & Mitchell, M. 1997 Finite populations induce metastability in evolutionary search. *Phys. Lett.* A **229**, 144–150.

van Nimwegen, E., Crutchfield, J. P. & Mitchell, M. 1998 Statistical dynamics of the Royal Road genetic algorithm. *Theor. Comp. Sci.* (In the press.)

Weinberger, E. D. 1991 Local properties of the *NK* model, a tunably rugged energy landscape. *Phys. Rev.* A **44**, 6399–6413.

Woodcock, G. & Higgs, P. G. 1996 Population evolution on a multiplicative single-peak fitness landscape. *J. Theor. Biol.* **179**, 61–73.

Wright, S. 1967 Surfaces of selective value. *Proc. Natn. Acad. Sci. USA* **58**, 165–172.

Wright, S. 1982 Character change, speciation and the higher taxa. *Evolution* **36**, 427–443.