J. theor. Biol (1987) **128**, 11–45

Towards a General Theory of Adaptive Walks on Rugged Landscapes

STUART KAUFFMAN†

Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, *U.S.A.*

AND

SIMON LEVIN

Section of Ecology and Systematics, and Ecosystems Research Center, Cornell University, Ithaca, New York 14853, *U.S.A.*

(*Received* 10 *December* 1986)

Adaptive evolution, to a large extent, is a complex combinatorial optimization process. In this article we take beginning steps towards developing a general theory of adaptive "walks" via fitter variants in such optimization processes. We introduce the basic idea of a space of entities, each a 1-mutant neighbor of many other entities in the space, and the idea of a fitness ascribed to each entity. Adaptive walks proceed from an initial entity, via fitter neighbors, to locally or globally optimal entities that are fitter than their neighbors. We develop a general theory for the number of local optima, lengths of adaptive walks, and the number of alternative local optima accessible from any given initial entity, for the *baseline case* of an *uncorrelated* fitness landscape. Most fitness landscapes are correlated, however. Therefore we develop parts of a universal theory of adaptation on correlated landscapes by adaptive processes that have sufficient numbers of mutations per individual to "jump beyond" the correlation lengths in the underlying landscape. In addition, we explore the statistical character of adaptive walks in two independent complex combinatorial optimization problems, that of evolving a specific cell type in model genetic networks, and that of finding good solutions to the traveling salesman problem. Surprisingly, both show similar statistical features, encouraging the hope that a general theory for adaptive walks on correlated and uncorrelated landscapes can be found. In the final section we explore two limits to the efficacy of selection. The first is new, and surprising: for a wide class of systems, as the *complexity* of the entities under selection increases, the local optima that are attainable fall progressively *closer* to the mean properties of the underlying space of entities. This may imply that complex biological systems, such as genetic regulatory systems, are "close" to the mean properties of the ensemble of genomic regulatory systems explored by evolution. The second limit shows that with increasing complexity and a fixed mutation rate, selection often becomes unable to pull an adapting population to those local optima to which connected adaptive walks via fitter variants exist. These beginning steps in theory development are applied to maturation of the immune response, and to the problem of radiation and stasis. Despite the limitations of the adaptive landscape metaphor, we believe that further development along the lines begun here will prove useful.

† To whom correspondence should be addressed.

0022-5193/87/170011+35 $03.00/0

Introduction

Francois Jacob, in his 1977 essay, "Evolution and tinkering", eloquently explicated the ways in which adaptation is constrained by past history and by the nature of the evolutionary process itself: adaptation generally progresses through small changes involving a local search procedure in the space of possible genetic rearrangements. To understand the evolutionary process, we therefore must understand the constraints that such a local search procedure places on any optimization scheme. Moreover, by comparing local to more global search algorithms, we can gain some insight into how the nature of evolutionary change depends upon the genetic distances that can exist between parents and offspring. If, for example, the rate of mutation is increased, so that parents and offspring may differ at a larger number of loci, one result may be the appearance of qualitatively different patterns of evolution at the population level.

The simplest paradigm of evolutionary change is one of local "hill climbing". Despite this transparent and simplistic metaphor, such evolutionary hill climbing involves a complex combinatorial optimization process. In such optimization processes, typically, many parts and processes must become coordinated to achieve some measure of overall success, but conflicting requirements due to alternative simultaneous optimization goals, or conflicting constraints due to the natures of the different parts and processes to be coordinated, limit the end result achieved. Our aim in this article is to explore some initial steps in the formulation of a general theory of such constrained adaptive "walks", which must proceed along a path of "fitter variants" towards the attainable local or global optima. The theory we derive is motivated by organic evolution; but in fact, it reaches beyond evolutionary biology towards a general theory of optimization. It is essential to note that viewing evolution as an optimization process does not presuppose some master-plan, or require teleology. Rather, it recognizes the nature of the process, and observes that it has features in common with deliberate optimization schemes.

We note at the outset that the hill climbing analogy has many limitations (Levin, 1978). The adaptive landscape undulates in response to environmental variation. Adaptation is a response to past environments, rather than an anticipation of the future. Furthermore, the fitness of a genotype may depend upon its frequency in the population, or upon coevolving properties of other species or habitat characteristics. Thus "fitness" is not a property of a genotype alone, but depends upon its environmental context.

These caveats (and see also Provine, 1986) having been made, it remains the case that the powerful imagery created by Wright's adaptive landscape (Wright, 1932) and Fisher's Fundamental Theorem of Natural Selection are among the most important concepts in evolutionary theory. They apply strictly for a number of cases of interest, and are points of departure for others. When the genetic system is simple, involving a single locus or weak epistasis, and when fitnesses are constant in time, the mean fitness of individuals in the population will increase monotonically and asymptotically toward a maximum, or peak, in an adaptive landscape (see, for example, Levin, 1978; Ewens, 1979). In more realistic circumstances of strong epistasis in a complex genomic system, with intermediate levels of linkage and

recombination, the consequent genetical constraints, mirrored in the vigorous lability of phenotypic properties as the genome changes, guarantees that the fitness landscape will be very rugged, with many peaks, ridges and valleys. Yet we have essentially no theory of adaptation on such rugged landscapes. This paper, then, is only a beginning in developing a general theory of adaptive walks on rugged fitness landscapes.

In the first section of this paper we introduce a general framework, give some preliminary intuitive examples, and specify some of the questions that arise naturally. In the second section we derive universal properties for adaptive walks in a statistically *uncorrelated* "landscape". This important baseline case allows us to characterize the numbers of local optima, the lengths of adaptive walks prior to attaining a local optimum, the number of alternative optima accessible to a given initial entity, etc. However, most fitness "landscapes" are correlated, rather than uncorrelated, in the sense that neighboring entities, which differ by a single "mutation", tend to have similar fitness. Therefore, in the third section, we derive universal features of adaptive evolution in the "long jump" limit, where adaptation occurs via multiple simultaneous mutations that "jump beyond" the correlation lengths in the fitness landscape. These results are simple and very general; they should apply to adaptation on a vast number of statistically rugged landscapes. In the fourth section we consider adaptation on correlated landscapes. No universal theory is in sight, but our hope is that many correlated fitness landscapes exhibit similar general features. Two numerical examples that reflect apparently different optimization problems have similar underlying statistical features, lending preliminary support to this hope. In the fifth section, we consider the correlation between the *complexity* of the entities under selection, and the *difficulty* in maintaining entities that are very fit. More precisely, we identify two general limitations on adaptive evolution: as entities become more complex, the attainable local optima become less fit; and, in the face of constant mutations, selection becomes less able to maintain populations "at" or "near" such optima. We argue that these limitations have deep implications for evolutionary biology.

A General Framework

In evolution, the generation of variability is underlain by mutational processes. These include point mutations, which substitute, insert, or delete a single nucleotide in the DNA or RNA sequences comprising the genome, and chromosomal mutations such as deletions, inversions, duplications, translocations, transpositions, and conversions, by which entire chromosomal regions are altered, deleted, duplicated, or moved and integrated into novel chromosomal locations. In sexual organisms, recombination is another powerful process for generating diversity, shuffling the extant variants within the gene pool.

It is useful initially to restrict attention to simple point mutations that switch, insert, or delete single nucleotide bases, for such mutations are the fundamental form of mutational alteration in any contemporary population. All other forms can be thought of as mechanisms that achieve many such single mutations at once. Each

genotype is surrounded by a number of other genotypes, its 1-mutant neighbors, each accessible by a single mutational alteration. In turn, these new genotypes have some large number of further 1-mutant neighbors, which again have 1-mutant neighbors. More formally, there is a *space or ensemble* of genotypes, where each genotype is a point in that space, and has as its immediate neighbors those points representing 1-mutant neighboring genotypes.

Suppose, to be concrete, we restrict attention to haploid organisms with DNA genotypes of some fixed length in nucleotides, say 100 000. Each position in the DNA sequence can be occupied by 4 alternative bases, thus by 3 alternatives to the base present in a given DNA sequence. Then each genome has 300 000 1-mutant neighbors in the very high dimensional space of haploid genotypes. Thus, the basic notion is that of a space or ordered ensemble of entities, each "next to" all those alternative entities that are accessible by a single mutation step. Notice that a change in the elementary mutation mechanism could alter which genotypes are 1-mutant neighbors of one another. That is, the topology of the space of entities is determined by the mutational "move" generator that specifies which entities can mutate in one step to one another. Note also that in the simplest cases, the "moves" are symmetric: if genotype A is one step from B, then B is one mutational step from A. This restriction need not always hold.

The second basic notion to be made explicit is that each entity, here genotype, is associated with some set of attributes, or phenotypes, that may serve as the basis for selection. That is, a "fitness" with respect to those attributes can be ascribed to each entity. Whatever the mapping of attributes to fitness, we can ask about the distributions of the attributes across the space of entities, and derivatively, about the distribution of "fitness" across the space of entities. Any given attribute may be rare or common in the space, and may be scattered at random, or in a variety of correlated ways across the space of entities. Similarly, different fitness values may be rare or common, and scattered at random or in correlated ways across the space of entities. We shall speak of the distribution of fitness values across the space of entities as the *fitness landscape.* Note that this landscape is a discrete one, since values are assigned to points on a lattice rather than on a continuum. Given this notion of the distribution of fitness values across the space of entities, and the notion of 1-mutant accessible neighbors, it follows that each entity is surrounded by a large number of neighbors, all, some, or none of which may be of higher fitness value than that entity. Conceive of drawing an arrow from each entity to those of its 1-mutant neighbors with higher fitness. Then adaptive evolution in its simplest form, via 1-step *fitter* variants, can be thought of as a directed walk *uphill* from some initial entity through fitter variants until a local or global optimum is reached. Such an optimum at least must be fitter than its 1-mutant neighbors. That is, the fitness distribution across the space of entities induces a directed graph, given by the arrows between each entity and its fitter 1-mutant variants; the structure of the graph is fundamental for adaptive walks to or towards local or global optima. Similar ideas have been introduced by Agur & Kerszberg (1987).

For the situations described above, adaptive evolution by mutation and selection in a constant environment drives a population across a fitness landscape from some

initial distribution as a more or less dispersed cloud in the space to a distribution that, subject to the constraints imposed by the genetic system, in general, has a higher average fitness. Although individual genotypes are associated with points on a lattice, the population mean varies on a continuum formed by interpolation among the nodes on the lattice.

Adaptive hill climbing by mutation and selection, the most basic form of trial and error learning, is only the simplest of schemes used in dealing with complex optimization problems. It is of general interest to compare its capacities as an optimization process with other processes such as simulated annealing (Kirkpatrick *et al.*, 1983; Brady, 1985). We return in the final section to discuss some of the limitations of mutation and selection as a means to attain and maintain local or global fitness optima.

Preliminary Examples

By way of examples we mention four related optimization problems, two biological and two not, in each of which the process of improvement by local mutation and selection is important. We introduce these here to clarify the nature of the problem. We return later to apply our analyses to two of them.

1. PEPTIDE SPACE

Consider the space of all peptides of some specified length, e.g. 50 amino acids, each amino acid being one of 20 types, (Smith, 1970; Nino, 1979; Eigen, 1985). Then each peptide has 19×50 "neighboring" peptides that differ from the first peptide in a single amino acid. The set of all 20^{50} peptides can be arranged in the appropriate high-dimensional discrete space such that each point vertex in that space represents one peptide, and each vertex is adjacent to 19×50 neighboring peptides. Peptides differ in terms of capacity to catalyze a specific reaction, or to bind a specific ligand. Any such property induces a "fitness" function on the peptide space, and suggests an optimization process to find peptides that perform best.

2. THE TRAVELING SALESMAN PROBLEM

Consider N cities located on a plane, and connected by roads of defined length. The famous traveling salesman problem requires a salesman to start at any initial city, visit each city once, and return to the initial city, by the shortest possible total route. This simply stated problem, one of the most studied complex combinatorial optimization problems, is known to be NP complete (Johnson & Papadimitriou, 1985). That is, as N grows, exhaustive search for the shortest route grows faster than polynomially. Any potential solution to the traveling salesman problem is a cyclic permutation among the N cities, each of which occurs once. Consequently there are $N!$ possible solutions, if we relax the specification of the initial city. Each possible legitimate circuit can be deformed to $N(N-1)/2$ "neighboring" circuits

by elementary permutations exchanging the positions of two cities in the circuit. Thus, with this "1-mutant move generator", the possible solutions can be arranged in a high dimensional space, each next to its 1-mutant neighbors. Note that other "elementary" changes, such as exchanging 4 cities, or links between cities, would generate a different set of neighbors to each solution, hence alter the *topology* of the space of solutions, but not the set of possible solutions. Given any definition of "neighbors" specifying the space of solutions, an adaptive walk is one that starts at a solution and moves by some algorithm or procedure, including mutation and selection, toward better solutions. Techniques of this sort are commonly used as part of schemes to solve the traveling salesman problem (e.g. Lin & Kernighan, 1973).

3. SPIN GLASSES

Spin glasses derive from condensed matter physics and are models of disordered magnetic materials, where two adjacent spins might prefer to orient in the same direction, or might instead prefer to orient in the opposite direction (Sherrington & Kirkpatrick, 1975). A simple mathematical model of a spin glass consists of a lattice of spins, each only up or down, each adjacent to a fixed set of neighboring sites in the lattice. The disordered aspect of spin glasses is modeled by deciding, at random, for each pair of adjacent spins, if they prefer to orient in the same or opposite directions. Thus, the coupling between spins is symmetric. Further, the strength of the coupling is chosen at random from some fixed underlying distribution. This process creates a lattice that is a clear example of conflicting constraints.

Consider a single "square" loop of four adjacent neighbors in a cubic lattice. The random "preference" assignment might create a loop where all four spins prefer to be in the same orientation. This preference is satisfied easily for all four simultaneously. Alternatively all four might prefer the opposite orientation to their neighbors, a condition again easily satisfied by having the spins at two corners "up" and the spins at the remaining two corners of the square loop "down". But if three of the four pairs prefer to be in the same orientation, while one prefers to be in the opposite orientation, then no configuration of the four spins simultaneously satisfies all preferences. Such a loop is said to be *frustrated.* The widespread existence of frustrated loops in spin glasses means that, when an appropriate dynamics is imposed, the system as a whole has very many metastable states, that is, local energy minima. More precisely, a lattice of N spins, each up or down, has 2^N configurations, each with a defined energy. Each configuration has N neighboring configurations, attained by flipping one spin to the opposite orientation. Thus, at any finite temperature, the system evolves towards configurations with lower energy, via 1-flip neighbors. An enormous literature exists characterizing the number of local energy minima, which appears to rise exponentially in N; the energy spectrum of these minima with respect to the global minimum; the sizes of the potential wells for high and low energy local minima; the overlap between spin states in different minima, etc.

Consider now an adaptive game, whose aim is to obtain a spin glass with a defined configuration of its N spins as the global minimum for that spin glass, or more

generally, as a local minimum with some specified low energy. Then consider a space of spin glasses, where each vertex in the space is one entire spin glass, and has as its 1-mutant neighbors those spin glasses obtained by altering one preference rule for one pair of coupled spins. Since each spin glass assigns a specific energy to the desired configuration, the space of spin glasses, each a 1-mutant neighbor of the others, has assigned the energy of the desired configuration to each spin glass, and an adaptive walk from an initial spin glass towards those in which the desired configuration has lower energy is well specified. The abundance of frustrated loops leads to the intuition that such walks might not lead very far. We will return to this theme in a more general context.

4. GENETIC REGULATORY NETWORKS

Ontogeny is underlain by a complex "cybernetic" network by which genes and their products regulate one another's activities. The most famous example, and the first, is the lactose operon in *E. coli.* Another is the C1-tof mutually inhibitory control loop underlying lysogeny or lysis in bacteriophage lambda.

A higher eukaryote has on the order of 100 000 structural genes and perhaps as many regulatory (cis acting and trans acting) genes, coordinating the expression of the entire genome. Kauffman (1969, 1974, 1984) has modeled genetic regulatory networks by idealizing the activity of any gene to be "on" or "off", and any product to be "present" or "absent". The resulting Boolean models of genetic networks are spiritually similar to spin glasses, but predated that concept. A network with N binary genes has 2^N combinations of gene activities. A genetic regulatory network is a *dynamical system* that specifies for each such combination, or state of gene activities, what successive state of activities will occur. The analogue to the potential wells of spin glasses are *basins of attraction and attractors* in these network models. Any such system, started in an initial configuration or state of gene activities, settles into one of a small number of alternative recurrent patterns of gene activity, analogous to local potential minima. These alternative patterns can be considered as the alternative model "cell types" in the repertoire of one genomic system underlying cellular differentiation. A large body of work (Kauffman, 1969, 1974, 1984, 1985; Gelfand & Walker, 1984; Fogelman-Soulie *et al.*, 1982; Derrida & Weisbuch, 1986) has shown that, subject to the simple constraint that any gene is directly regulated by few other genes, such otherwise random networks have many features similar to real cells. Conceive again of an evolutionary game. New cell types can be supposed to evolve both by emergence of new structural genes, and by alteration in the cis and trans acting regulatory genes controlling existing genes. Then a simple picture of the evolution of genetic regulatory systems is portrayed as occurring in a space of genetic regulatory systems, each of which differs from its 1-mutant neighbors by a change in one regulatory connection, or one Boolean law prescribing the behavior of a regulated gene as a function of those regulating it. Thus, the lactose operator might mutate to be constitutively active. Then we can begin to explore obvious questions. For example, suppose a cell type were desirable that consisted in a specified steady state pattern of gene activities among the N

genes. Can mutations altering control connections and control rules start from some arbitrary genetic network and, via selection, attain a network with the desired pattern of gene activities as a dynamical steady state attractor of the system? Interestingly, the general answer appears to be "No", as will be described in more detail below.

Natural Questions

These preliminary examples from biology, physics and mathematics, suffice to indicate the host of questions that arise naturally. These fall into three classes: questions about the specific fitness landscape and its statistical structure, questions about what kinds of entities "live in" what kinds of landscapes and why, and questions about adaptive flow by mutation and selection (or other optimization algorithms) on any given landscape.

Concerning any specific fitness landscape, some of the obvious questions are these: (1) How many local (and global) optima are in the space with respect to 1-mutant moves, 2-mutant moves, etc? (2) If an adaptive process starts at one "entity" in the space and can move only via fitter neighbors, what is the expected number of improved variants passed on the way to a local optimum—that is, what is the expected length of the adaptive walk? (3) If a "greedy algorithm", always choosing the best of the neighboring variants, is used, how long are adaptive walks? (4) How many alternative fitter variants are 1-mutant neighbors to any entity and how does that vary as the "fitness" of an entity increases? Alternatively stated, if adaptation can only occur via fitter 1-mutant variants, how many ways can the process "branch" at each step uphill? (5) Correspondingly, how many alternative local optima are accessible to an arbitrary initial entity via adaptive walks through successively fitter variants? How does that vary with the fitness of the initial entity? (6) Is the global optimum attainable? More broadly, how "fit" are the attainable local optima with respect to either the global optimum, or to the mean fitness of entities in the space? (7) If "neutral mutations" are allowed, or if adaptation can pass through less fit variants, what are the effects on the character of adaptive walks in the space of entities?

With respect to the relations between the entities and the adaptive space in which they live, several basic questions arise. For example, we tend to suppose that similar peptides, differing by a single amino acid, will have strongly similar physicochemical properties. Thus we suppose that the fitness landscape in peptide space for each of many properties will be correlated. What is that correlation structure, and why? Similarly, a defined space of spin glasses, where each vertex is a neighboring spin glass due to alteration is a single spin coupling law, is expected to be correlated in many ways. We have no idea what that correlation structure is, nor how to characterize it. The same may be said of model or real genetic regulatory networks, where the structure of the landscape must be critical to the evolution of differentiation and ontogeny. That is, we have, at present, no theory saying what kinds of entities inhabit what kinds of adaptive landscapes and why, nor do we have a theory that indicates whether, among the possible correlated fitness landscapes, these fall into rather few equivalence classes with similar statistical structures and a number of

pathological but very rare cases, or whether there are vastly many different kinds of correlation structures. A related general question is this: For a given class of entities, peptides, genetic regulatory systems, etc., as those entities become more complex, what happens to the adaptive landscapes they inhabit? Do accessible local optima fall closer to or further from the mean fitness of the entities in the space? We shall see that for a large class of systems, increased complexity leads to accessible local optima that are closer to the mean of the entities in the space.

Given any specific adaptive landscape, the tools of population genetics provide a conceptual armamemtarium to ask many familiar questions (Ewens, 1979). Given specified fitness differences, can a haploid or diploid population, dispersed initially in some way across the space and driven by mutations occurring at a given frequency, be "pulled" to a given local or global optimum? Can the population be maintained by selection in the face of mutations at or near such an optimum? If near, how near? How long does it take for the population to flow to the optimum, or more generally to attain its stationary distribution, if any? As the complexity of the entities under selection increases, what happens to the capacity of selection to pull the population to local optima and hold it there? How does population evolution by mutation and selection compare to other complex combinatorial optimization processes? When is it more efficient and why? Again, to compare the efficiency of evolution with other optimization schemes does not suggest that there is any reason why evolution should have selected for an efficient scheme. Such an argument would require a higher order selection for which we see no evidence at present.

Obviously, these questions only scratch the surface. A new body of theory is needed. We turn to the first few steps.

Adaptation on an Uncorrelated Landscape

We consider in this section, the expected character of adaptive walks in uncorrelated spaces, i.e. those having the property that the fitness value of each entity is drawn at random from some fixed underlying distribution. This benchmark case can be made both simpler, and more universal, by ignoring the actual fitness values assigned to the entities in the space, and replacing those values by their *rank orders* counted from the least fit entity (rank 1) to the top rank entity, rank T. (For simplicity, we assume that there are no "tie values".) In so far as adaptation is constrained to occur along walks from an initial entity through *fitter* variants, passage to a rank order model preserves completely most statistical features of the resulting adaptive walks, including lengths of walks to maxima, numbers of alternative optima that are accessible by connected walks, etc. However, passage to rank orders throws away information about "distances" between different rank order fitness values in the underlying distribution from which fitness values were drawn. To be concrete, for the remainder of this section we make the explicit assumption that the underlying fitness values are uniformly distributed on the real line.

As partial biological justification for the concept of the population following an adaptive walk passing via the first fitter neighbor sampled by the mutational process, consider a fixed sized population of haploid entities that are initially identical;

hence "release" the population "at" a particular vertex or point in the space of possible entities. Suppose that mutations occur rarely. Then on a slow time scale the population will sample 1-mutant neighbors, and if a fitter variant is found, that variant either will die out due to random fluctuations; or once above a rough threshold frequency, will sweep through the population on a fast time scale. Gillespie has argued that in this plausible limit, one can think of the adaptive process as a continuous time, discrete state Markov process, in which the entire population is resident at one state—here entity—and then jumps with fixed probabilities to each of its 1-step mutant fitter variants (Gillespie, 1983, 1984). We return later to consider the effects of a larger mutation rate with respect to fitness differentials.

Consider as a concrete example, a space of length N peptides constrained to use two amino acids, say leucine and alanine. Then 1 and 0 can represent the two amino acids, and each peptide is a binary string, length N, of 1 and 0 values. Such strings are easily represented as vertices of an N-dimensional Boolean hypercube (Fig. 1). The number of such strings is just 2^N, and rank order fitness values can be assigned at random by assigning rank orders 1 through 2^N, without replacement, to each of the vertices.

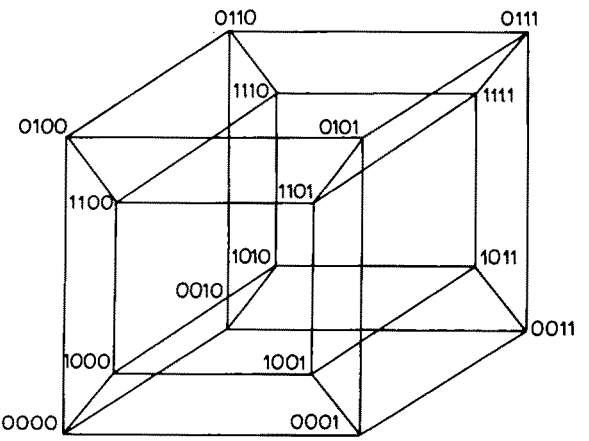

FIG. 1. A four-dimensional Boolean hypercube, in which each of the 16 vertices represents one of the possible strings of four 0 or 1 values. Here each such string is interpreted as a specific tetrapeptide with two possible types of amino acid at each position. Each peptide is a 1-mutant neighbor of the four peptides which can be obtained by altering a single amino acid at a single one of the four positions in the tetrapeptide.

The probability that any vertex, representing a specific peptide of length N, is a local optimum, is just the probability that that vertex is of higher rank order than any of the N 1-mutant neighboring vertices obtained by altering any 1 to a 0, or 0 to a 1. Since all rank orders are chosen at random, the probability that a vertex is a local maximum is just

$$P_m = 1/(N+1). \tag{1}$$

Since the total number of sequences, hence vertices, is 2^N, it follows that the expected number of local optima with respect to 1-mutant neighbors, M_1, is

$$M_1 = \frac{2^N}{N+1}. \tag{2}$$

This generalizes trivially to 1- and 2-step mutant neighbors, where the expected number of local optima, M_2, falls to

$$M_2 = \frac{2^{N+1}}{2+N(N+1)}. \tag{3}$$

The calculation is simple for any fixed number of k mutants, since the denominator in (2) simply is replaced by the cumulative binomial sum

$$\sum_{j=0}^{k} \binom{N}{j}$$

where k equals 1 in (2) and 2 in (3). Thus for any fixed k, the numerator rises exponentially in N, while the denominator rises much more slowly as a function of N. The first important result then is a *scaling law*; for any small fixed number k, of k-mutant neighbors, hence any fixed notion of "local", the number of local optima rises a bit less than exponentially as N increases. Thus, as N increases, there are a very large number of possible peptides. But for an uncorrelated fitness landscape, the number of local (1-step) optima is nearly as large. Uncorrelated landscapes are rife with local optima.

It is simple to generalize this result to peptides that use all 20 amino acids. If B represents the number of amino acids, any peptide of length N has $D=(B-1)N$ 1-mutant neighbors. Hence the number of local optima is just

$$M_1 = \frac{B^N}{D+1} = \frac{B^N}{N(B-1)+1}. \tag{4}$$

Again, the number of local optima increases almost exponentially in N.

The probability that a peptide sequence is a local optimum is related to its rank order, X, ranging from 1 to T, where T is the most fit. The fraction of the other vertices having lower rank order is $(X-1)/(T-1)$, or approximately X/T. For large T and X, the probability that a given vertex has a higher rank order than any of its D 1-mutant neighbors, hence the probability, P_m, that the vertex (sequence) is a local optimum approaches

$$\left(\frac{X}{T}\right)^D. \tag{5}$$

For small X or D, account must be kept of the lack of replacements in calculating P_m; thus, one obtains:

$$P_m = (X-1)!(T-D-1)!/(T-1)!(X-D-1)! \tag{5b}$$

These formulae imply that the probability that an entity in the space is a local optimum remains low until its rank order is high, then rises rapidly.

We estimate next the expected lengths of adaptive walks via fitter variants before a local optimum is attained. For concreteness, we assume that the process picks, with equal probability, any one of its fitter neighbors at each adaptive step, then iterates. Later we shall relax this assumption. In general, the starting position for an adaptive walk can be anywhere; but to get an upper bound, assume that we begin at the lowest rank entity, rank 1. All D neighbors are fitter, and their rank order positions are, on average, scattered uniformly between rank order 2 and rank order T. The adaptive process samples among the D neighbors, one at a time, until the first fitter variant is found, then moves to that variant. Thus, on average, the rank order of that first fitter variant is *halfway* between the present rank order, 1, and the top, T. Moreover, the frequency distribution for the position of the adaptive process is uniformly distributed about its mean. That symmetry will not last beyond the first step, after which the iterated distribution will become skewed to the right. Nonetheless, we can gain a bound on the adaptive process initially by assuming that at each adaptive step, the stochastic process is concentrated at its mean expected position. Thus, variance is ignored. On a second iteration, from an entity of relative rank order $X/T = 0{\cdot}5$, half the D neighbors are fitter, and the first sampled lies, on average, half way between the current entity and the top ranked entity, T. Thus, at each successive step, the process on average, moves half way from the current position toward the top rank, T; simultaneously, the expected number of the more fit neighbors is halved, on average, on each improvement step. Thus, the average position after "K" improvement steps is approximately $T[1-0{\cdot}5^K]$, and we shall use this below to compute the probability that the nearest (integer) rank order is a local maximum.

Before carrying out this calculation, however, what is the consequence of ignoring the variance in the increase in rank order at each adaptive step, on the calculated expected length of adaptive walks to local optima? Note that, at each step, the new conditional distribution is symmetric about its mean, and further deviations to lower rank order add much less to the conditional expected path length than do deviations to higher rank orders subtract from it. Therefore, the approach ignoring variance overestimates the expected lengths of adaptive walks to optima. We carry this calculation out next.

It is clear that, after each improvement step, the entity achieved at is fitter than at least one of its D neighbors: that from which it has just arrived. Therefore the probability that the current entity is a local optimum is then just its relative rank order raised to the $(D-1)$ power. The probability that the adaptive walk continues for one more step from a given entity is simply the probability that the current entity is *not* a local optimum. The upper bound discussed above for the probability that the adaptive walk continues for K steps without stopping, P_K, is

$$P_K = \prod_{R=0}^{K} \left[1 - \left(\frac{2^R - 1}{2^R} \right)^{(D-1)} \right]. \tag{6}$$

On average, walks continue until $P_K < 0{\cdot}5$. We can estimate the value of k where this occurs quite accurately. In particular, note that if $R = \log_2 (D-1)$, the corresponding term in P_R is $1-(1-1/(D-1))^{D-1}$, which is extremely well approximated

by $1-1/e=0{\cdot}63$. Moreover, if this represents the Rth term in P_K, then, working backwards, it is easily shown that the preceding terms are approximately $1-(1/e^2)=0{\cdot}86$, $1-(1/e^4)=0{\cdot}98$. Thus, there is very little probability of the adaptive walk stopping more than one or two steps before this R value, and to a very high degree of accuracy, the upper bound on average walk length is given by $R=\log_2(D-1)$.

Another thing that becomes clear from this calculation is that the distribution of stopping times is very tight: if an adaptive walk has achieved relative rank order $(D-2)/(D-1)$, it has probability approximately $1/e$ of being at a local maximum and stopping. If it proceeds one step farther then, on average its rank order is $(1-\frac{1}{2}(D-1))$, which has a probability of $1/e^{1/2}=0{\cdot}6$ of being a local maximum. Arguing similarly, we can see that the great majority of walks will stop within one or two steps of achieving the rank order of $(D-2)/(D-1)$. An upper estimate of the number of steps it takes the process to reach this rank is $R=\log_2(D-1)$.

As D increases, the number of neighbors to an entity increases linearly, but the lengths of walks increase only as the logarithm base 2 of the number of neighbors. Thus, in an uncorrelated space of peptides of length N, with $B=20$ amino acids, as N goes from 10 to 100, the numbers of peptides increases from 20^{10} to 20^{100}, a vast increase; but the bound on walk lengths increases from $\log_2(190)$ to $\log_2(1900)$, or from about 7 or 8 to 11 steps. This, of course, reflects the fact that uncorrelated spaces are very richly dotted with local optima.

It is of interest to compare these walks, which progress via the first fitter 1-step mutant neighbor found, to "greedy" walks, which proceed from any entity via the fittest of its D neighbors. In general, such walks will be shorter than those just considered, since they rise faster, and in fact reach their apogee rapidly. Begin at the lowest rank entity. On average, the fittest of its D neighbors is located at a relative rank order of $(D-1)/D$. The chance that this entity itself is fitter than its remaining $D-1$ untried neighbors is just

$$\left(\frac{D-1}{D}\right)^{D-1}$$

which tends to $1/e$ for large D. Thus, in a single step, the expected position of a greedy process is already in the region where most adaptive walks end, and in fact has probability 0·37 of stopping. In 2 steps, this conditional probability rises to $(1/e^{1/2})=0{\cdot}6$, and in 3 steps to $(1/e^{1/4})=0{\cdot}8$. The average walk length associated with this process is less than 2, and this still represents an upper bound. It is a surprising result. Greedy walks are very short in an uncorrelated space. Gillespie has independently derived this result recently (Gillespie, personal communication).

Branching Probabilities

Because the predecessor to any entity in an adaptive walk necessarily was less fit, the expected number of fitter neighbors for any entity is derived simply from its rank order, X: approximately it is given by $(T-X)(D-1)/T$. Thus, the number of alternative pathways towards increased fitness values decreases linearly with rank order. From this, it is easy to calculate an upper bound on the expected number of

local optima accessible from the lowest ranked entity. D of its neighbors are fitter. On average, after a single improvement step, $(D-1)/2$, or almost $D/2$ of the neighbors of that first step variant are still fitter. After successive steps, on average $D/4$, $D/8$ neighbors are fitter. On average, as described before, walks continue for no more than $\log_2 D$ steps, in which time the process could have taken on the order of $B = D \times D/2 \times D/4 \ldots \times D/D$ alternative branching adaptive steps. Here we assume, without loss of generality, that $\log_2 D$ is an integer. Since many of these branching adaptive walks might rejoin one another, B is an overestimate of the true number of alternative local optima accessible from the lowest ranked entity. This upper bound is

$$B = \frac{D^{(\log_2 D)}}{2^{(\log_2 D)((\log_2 D)+1)/2}}$$

$$= \frac{D^{\log_2 D}}{D^{(\log_2 D+1)/2}}$$

$$= D^{(\log_2 D-1)/2}. \tag{7}$$

While the bound is an upper estimate, it serves our purpose in demonstrating what a small fraction of the true local optima are attainable from any single starting entity. Thus, for a peptide space, using 20 amino acids, and $N = 64$ as the length, this gross upper bound on the number of reachable local maxima from the lowest ranked peptide is about 10^{18}; by contrast, the total number of local maxima in such an uncorrelated space is $1{\cdot}5 \times 10^{80}$. Thus, a critical conclusion is that only a tiny fraction of all local optima are accessible from any entity on adaptive walks via 1-mutant fitter variants in uncorrelated spaces.

Some Implications

We pause at this stage to consider some implications of this baseline case of adaption on uncorrelated fitness landscapes.

1. MATURATION OF THE IMMUNE RESPONSE

Adaptive walks to local optima in uncorrelated fitness landscapes are short, of the order of $\log_2 D$, where D is the number of mutant neighbors. We will see below that walks on correlated landscapes often tend to be somewhat longer, depending upon the particular correlation structure. Uncorrelated landscapes are, as noted, the limiting baseline case to consider. Physico-chemical properties in peptide space may well be highly correlated, but it is possible that, for short peptides, the correlation is relatively weak: each amino acid alteration in a pentapeptide should make a larger difference in properties than in a protein of length 500. This intuition suggests that adaptation for such properties as ligand binding, or epitope features, might be rather uncorrelated for small peptides.

While no data as yet bear directly on this issue, data are available for a rather similar problem arising in maturation of the immune response. Faced with a specific exogenous antigen, the immune system undergoes clonal selection and amplification for a number of antibody-secreting B cells, each of which secretes an antibody molecule that has reasonably high affinity for the antigen (Tonegawa, 1983). Early in the immune response, each clonally selected B cell secretes an antibody molecule derived by combinatorial association of several disjoint genes, which when ligated together, form the gene for the variable (V) region of the heavy or light chain of the antibody molecule. These combinatorially prefabricated variants can be thought of as "roughed in" antibody binding sites, more or less complementary in shape to the antigenic determinants. Subsequent *maturation* of the immune response consists of somatic mutations to the heavy and light chain coding regions of these B cells. Those somatic mutations, which result in an alteration of the protein sequence of the V region, alter the binding affinity of the antibody molecule for the antigenic determinant. Subsequently, those mutated B cells whose antibodies bind the antigen with higher affinity than the initial roughed in V region proliferate more rapidly and come to dominate the immune response by clonal selection. Over a succession of somatic mutations to the V region of the initial roughed in B cell, the mean affinity of the antibodies increases to some maximum. Typical changes in affinity over the course of maturation are increases from 5×10^4 to 5×10^7.

The V region, in the heavy or light chain, consists of about 100 amino acids embedded in the primary structure of the entire antibody heavy or light chain. Thus, the DNA coding sequence of the V region for either chain is on the order of 300–350 base pairs. Each V region therefore has 3×300 or $3 \times 350 \approx 1000$ 1-mutant neighboring DNA sequences. If the affinities of 1-mutant V regions for a given antigenic determinant are nearly uncorrelated with one another, the theory derived above is applicable. The expected number of improvement steps to a local optimum is about $\log_2 D$; hence the expected number of improvement steps to a local optimum is about $\log_2 (1000)$ or about 10. In fact, the available data suggest that the typical range in the number of nucleotide changes is on the order of 6–8 during maturation of the immune response (Crews *et al.*, 1981; Bothwell *et al.*, 1982; Tonegawa, 1983; Heinrich *et al.*, 1984; Berek *et al.*, 1985; Clark *et al.*, 1985). Since the immune response begins with antibodies of intermediate fitness, walk lengths should be shorter than the expected upper bound, $\log_2 D$, from the lowest rank order.

This correspondence between theory and data obviously is good, but must be considered cautiously. First, it rests on the interpretation that a B cell found secreting a high affinity antibody at the mature stage of the immune response actually is derived by a sequence of mutations from a progenitor with a roughed in V region early in the immune response. The intermediate lineages are rarely if ever in hand. Second, some workers have observed up to 30 or so mutants, although these authors favor the view that this large number of mutations reflects, not 30 point mutations, but a small number of recombination events (Wysocki *et al.*, 1986). It is less important at this stage to try to assert that the theory of adaptation on uncorrelated landscapes correctly predicts the behavior of the immune system, than to note that the correlation structure of 1-mutant variants with respect to binding a single antigenic determinant

can be discovered, and should yield a better theory predicting the lengths of adaptive walks in the immune system.

2. RADIATION AND STASIS

It is a fundamental feature of adaptation via 1-step fitter variants on an uncorrelated landscape, that the number of fitter neighbors of an entity decreases (linearly) as its relative rank order fitness increases. That is, an initial low fit entity has many neighbors that are fitter, a highly fit entity has few, and a locally optimal entity has no fitter neighbors. This implies that any adaptive process constrained to begin at some entity, and proceed via fitter 1-mutant variants can branch to many alternative entities early in the process, and to successively fewer as adaptation achieves higher fitness. Thus, without being careful yet with respect to the ecological conditions of isolation etc. that might be needed to sustain biological *branching lineages*, we conclude that adaptation on uncorrelated landscapes inherently favors branching radiation, slowing to ultimate stasis. Further, it is characteristic that the radiation from an initial entity of low fitness is "bushy" at the base, less bushy at higher levels, then becomes confined to single lineages that wend upward to local optima. Standard accounts of radiation and stasis in biological populations ignore the structure of the underlying fitness landscapes (Valentine, 1980). We believe, and shall discuss further below, that this generic feature of adaptation on uncorrelated landscapes extends to many correlated landscapes for complex combinatorial optimization problems and must hold relevance for branching phylogenies in biological and technological evolution.

Changing Environments and Molecular Clock Data

Gillespie (1983, 1984) has used a variant of the model presented here to consider evolutionary data bearing on the molecular clock hypothesis. Gillespie notes that several sources support the claim that the *variance* in the rate of nucleotide or amino acid substitutions, compared to the *mean* rate of substitutions, is too high to be accounted for by the neutral theory. Rather, it appears as if substitutions occur in small bursts. Gillespie has presented an elegant way of thinking about this problem. He supposes that in a given environment, DNA sequences or proteins may be more or less optimal for some given tasks. In a manner similar to ours, he considers a space of DNA sequences or protein sequences, and rank orders the sequences with respect to the task. He then suggests that when the environment changes, it does so in a correlated way, such that the rank order of any peptide changes a little, but not a lot. Specifically, a peptide which previously was a local optimum with respect to its 1-mutant neighbors might now find itself third or fourth best among the D neighbors. The global consequence of this, Gillespie shows, is that the number of adaptive steps from the old locally optimal sequence to the new locally optimal sequence after the environment changes, is a small number. This leads to a model in which adaptation occurs in small bursts, running from a former to a new local optimum. Gillespie shows that the results are robust for a number of different

underlying fitness distributions, and suggests, in conclusion, that burst-like evolution fits with a selectionist theory better than with a neutral theory.

Whether or not Gillespie's interpretation is correct, it is attractive to adopt his idea that alterations in the environment can be modeled as more or less correlated alterations in the rank orders assigned to entities in the (correlated or uncorrelated) fitness landscape.

Universal Features of "Long Jump" Adaption on Statistically Correlated Landscapes

Most fitness landscapes are locally correlated in some way: 1-mutant neighbors will tend to have similar properties, hence similar rank orders. We return below to attempts to characterize adaptation via 1-mutant neighbors on correlated landscapes. First, however, it is of considerable interest to investigate a universal theory for adaptation on correlated landscapes in the limit of "long jumps", where the number of mutants occurring simultaneously in any one entity, J, is sufficiently large to "jump beyond" the correlation lengths in the fitness landscape. (We show below that such processes are not rare.) The underlying intuition is simple. Consider a point in a very mountainous region, such as the Alps. If one moves horizontally 1 metre, the altitude of the point at which one lands is, even in the Alps, highly correlated with the altitude of the initial point. If one moves horizontally for 50 kilometres, the altitude is essentially uncorrelated. That is, if an adaptive process jumps far enough from the current entity in a correlated landscape, then the adaptive process actually experiences an *uncorrelated* landscape. Therefore, it should be possible to construct a universal theory for adaptation via sufficiently long jumps on any correlated, but rugged landscape. It will differ from the previous theory for uncorrelated landscapes in that, in the long jump limit, the notion of a local optimum disappears: all points are accessible. As the number of mutants per individual, J, increases towards that limit, and wider and wider neighborhoods are explored, the correlation structure becomes weaker and weaker while the number of local optima diminishes.

Consider peptide space, restricted to 20 amino acids and peptides of length 100. The number of 1-mutant variants of any peptide is 1900. But the number of J-mutant variants, $J = 10$, or 50, is enormous. The implication of the enormous number of J-mutant variants is that is is not possible, in a reasonable time, for the adaptive process to sample *all* of the J-mutant variants. Therefore, it is not possible to determine, in that reasonable time, whether a given variant is actually a local optimum with respect to its J-mutant variants. In turn, this implies that the theory for adaptation via 1-mutant fitter variants on an uncorrelated landscape, must be recast for long jump adaptation on a correlated landscape. We must replace the idea of a local optimum with the concept of the metastability of a given entity, expressed as the expected waiting time to find a fitter J-mutant variant. To develop this theory we shall simplify a real population and consider a fixed population of size N^*, and assume that at each generation, or trial $(N^* - 1)$ J-mutant variants of the current entity are formed, while one entity remains unaltered. Below we will

generalize to consider a population with a spectrum of number of mutants per individual from 0 to many, rather than exactly J. For N^* large, $N^* \approx N^* - 1$. Further, we shall assume that if more than one fitter J-variant is found at a given generation, then the *fittest* among these is chosen in one generation, and the process iterates at the next generation with all $N^* - 1$ exploring J-mutant variants of this new fitter variant. The assumption that the population "moves" to the fittest variant encountered is related to the biological fact that the rate at which a fitter variant sweeps through a population is proportional to the fitness difference between the wild type and the new fitter variant. The fittest of several fitter alternatives will tend to dominate by sweeping the population fastest.

Consider an adaptive process begun with the entire population of N^* individuals located at a particular, randomly chosen entity. The relative rank order of that entity, X/T, is, on average, 0·5. Thus 0·5 N^* of the J-mutant variants sampled in the first generation are fitter than the initial entity. The best among these has an expected relative rank order of $(N^*-2)/N^*$. At the second generation, (N^*-1) J-mutant variants are sampled, and the expected number of fitter variants is 2. The best of these, on average, is 2/3 of the rank order distance from the current entity, to the top ranking entity. At the third generation, the expected number of fitter variants uncovered by the (N^*-1) J-mutants searched is 0·66. Consequently, the expected waiting time until a third fitter variant is found is the reciprocal, $1/0{\cdot}66 = 1{\cdot}5$ generations. On average, a single fitter variant is found, and on average it lies halfway between the last fittest variant and the top rank, T. Therefore after the third improvement step, the expected waiting time to find the next, fourth, improved variant, doubles to 3 generations. Again, on average the fourth fitter variant lies halfway between the third variant and the top rank. Thus, after each successively fitter variant is found, the waiting time to find the next fitter variant doubles. The cumulative number of improved variants S uncovered in G generations is therefore approximately

$$S = \log_2 G + 1. \tag{8a}$$

Remarkably this simple expression is almost independent of the population size, N^*. The above was derived for large N^*. At the opposite extreme, in the limit of $N^* = 1$, the result is not much different. The waiting time to find a first fitter variant is 2 generations, then becomes 4, 8, ... generations. The expression simplifies to

$$S = \log_2 G. \tag{8b}$$

Equations (8a) and (8b) can be thought of as very simple, universal aspects of long jump adaptation on, presumably, any (sufficiently) rugged, multipeaked fitness landscape. The fact that, for N^* large, we considered a process in which the fittest variant found is chosen makes a difference only initially, since after the first few steps it is unlikely that more than a single fitter variant is found per generation. Thus, if the process selects randomly one from among all fitter variants found, it will take slightly longer to reach rank $(N^*-2)/N^*$. An upper bound for the average number of improvements is $\log_2 N^* - 2$. Thereafter, the rate of finding new improved variants will yield the same slope as that found in (8a) or (8b). The difference is

not great: for $N^* = 16$, for example, the greedy algorithm saves less than one step on average.

The simplicity of the derivation does not vitiate the conclusion, which appears to be related to a similar simple and general result in the theory of records (Feller, 1971). Among the obvious implications of (8a) and (8b), is that during long jump adaptation, the rate of finding fitter variants is rapid at first, then slows. Since the waiting time doubles after each improvement step, after a modest number of improvement steps, apparent stasis sets in. Note that the mean expected number of fitter variants found after G generations is rather insensitive to the rank order position of the initial entity, as long as that rank order is 0·5 or less fit. However, if the initial relative rank order is above 0·5, the waiting time for the first few improvements is longer than given.

Branching to more than one alternative is more common initially for long jump adaptation and progressively harder later. If we require, for branching, that more than a single fitter variant be uncovered in one generation, or in a fixed number of adjacent generations, it is easy to calculate how many branches to fitter variants could occur as a function of time and population size, N^*. Thus, long jump adaptation on rugged but correlated landscapes, will tend to yield branching radiation quieting to stasis. We return to this theme again.

Numerical Tests of Long Jump Adaptation

In order to test this general "law", we have carried out numerical trials, attempting to evolve model genetic regulatory networks, via mutations to regulatory connections or the Boolean laws governing specific genes within the network, to attain a network having a model "cell type" with a defined steady state pattern of gene expression. Our simulations confirm the above theory of long jump adaptation. More precisely, we considered Boolean model genetic networks with N binary genes, either active or inactive, each receiving $K = 2$ regulatory inputs from 2 genes chosen at random among the N. We assigned at random to each gene one of the 16 possible Boolean functions specifying the activity of that gene as a function of the activities of its two input genes the moment before. For example, a gene might be active if either or both of its inputs were active before, hence the OR function, or active only if both were active, hence the AND function, etc. Such a network is a deterministic dynamical system, sampled at random from the ensemble of NK Boolean networks. With N genes, a network admits 2^N possible combinations of gene activities, each a *state* of the network. At each clocked instant, each gene assesses the activities of its inputs, and, according to its own Boolean function, assumes the proper next activity value; hence the network passes from a state to a unique successor state. There are a finite number of states, and hence eventually the network arrives at a state visited previously. Thereafter, the network cycles through this re-entrant sequence, called a *state cycle.* Many different states may flow onto the same state cycle, which is thus a *dynamical attractor* for that collection of states, called its basin of attraction. One network may have a number of different state cycle attractors. As noted earlier, these attractors, the asymptotic behaviors in the repertoire of the

model genomic system, are interpreted naturally as *cell types.* Previous work (Kauffman, 1969, 1974, 1984) shows that $K = 2$ networks, sampled at random from the ensemble of N, $K = 2$ networks, typically exhibit biological like order, even in the absence of selection. For example, the number of attractors predicts the number of cell types as a function of genomic complexity; the similarity of gene activity patterns between cell types mimics real cells; the numbers of cell types into which one cell type can differentiate by transient reversal of the activity of any one gene is small, mirroring the fact that any cell type can differentiate directly into few other cell types in ontogeny, thus that branching pathways of differentiation occur in all metazoan and metaphyten ontogenies. These and other features suggest that *NK* Boolean models are interesting, hopefully plausible models for real genomic regulatory systems (Kauffman, 1974, 1984).

Evolution of novel cell types, utilizing the *same* structural genes, has required alterations in the cybernetic network regulating genetic activity. Thus, it has been attractive to examine whether mutation of regulatory connections between genes, or Boolean rules determining each gene's activity as a function of its inputs, coupled with selection, can begin with an arbitrary genetic network and evolve towards networks whose cell types exhibit predefined "target" patterns of gene activity.

The natural space of *NK* Boolean models, with N and K held at fixed values, such as 100 and 2, used two ideas of an elementary mutation. The first alters one "end" of a regulatory connection between two genes, changing the regulating gene. This models chromosomal mutations altering cis acting regulatory elements. The second elementary mutation alters the Boolean control rule, by changing one "bit" in the four bit table specifying the response of the regulated gene to each of the four activity patterns of its two inputs. In order to test the effects of *long jump* adaptation in Boolean genetic networks, at each generation we worked with Robert Smith at the University of Pennsylvania, and mutated either half the connections, or one quarter of the bits in the Boolean functions. We fixed a population size of networks. We measured the closest state of the closest state cycle to a predetermined "target" pattern of gene activities among the N genes, and took the overlap with the target pattern as the fitness of that network. At each generation, all networks were initiated as identical to the best network of the previous generation, all but one network were mutated as described, each was assessed, and the population as a whole moved to the best network found at that generation. Clearly this implements the adaptation procedure used to derive the simple, but presumably widely applicable law in eqns (8a) and (8b).

Figures 2(a) and 2(b) show the results for networks with $N = 50$ and $N = 100$ genes, mutating half the connections at each generation. Figures 2(c) and 2(d) show the results of mutating $\frac{1}{4}$ of the bits in the Boolean functions at each generation. As predicted by eqn (8a), the cumulative number of improved variants, on average, is equal to 1 plus the base 2 logarithm of the number of generations. The experimental and theoretical curves are closely coincident. In fact, the experimental curve is slightly displaced to the left, reflecting the fact that, if the initial network happened to be well above 0·5 in fitness, the waiting time until the first improved variant was encountered was increased. The fact that the observed curves fit the theoretical

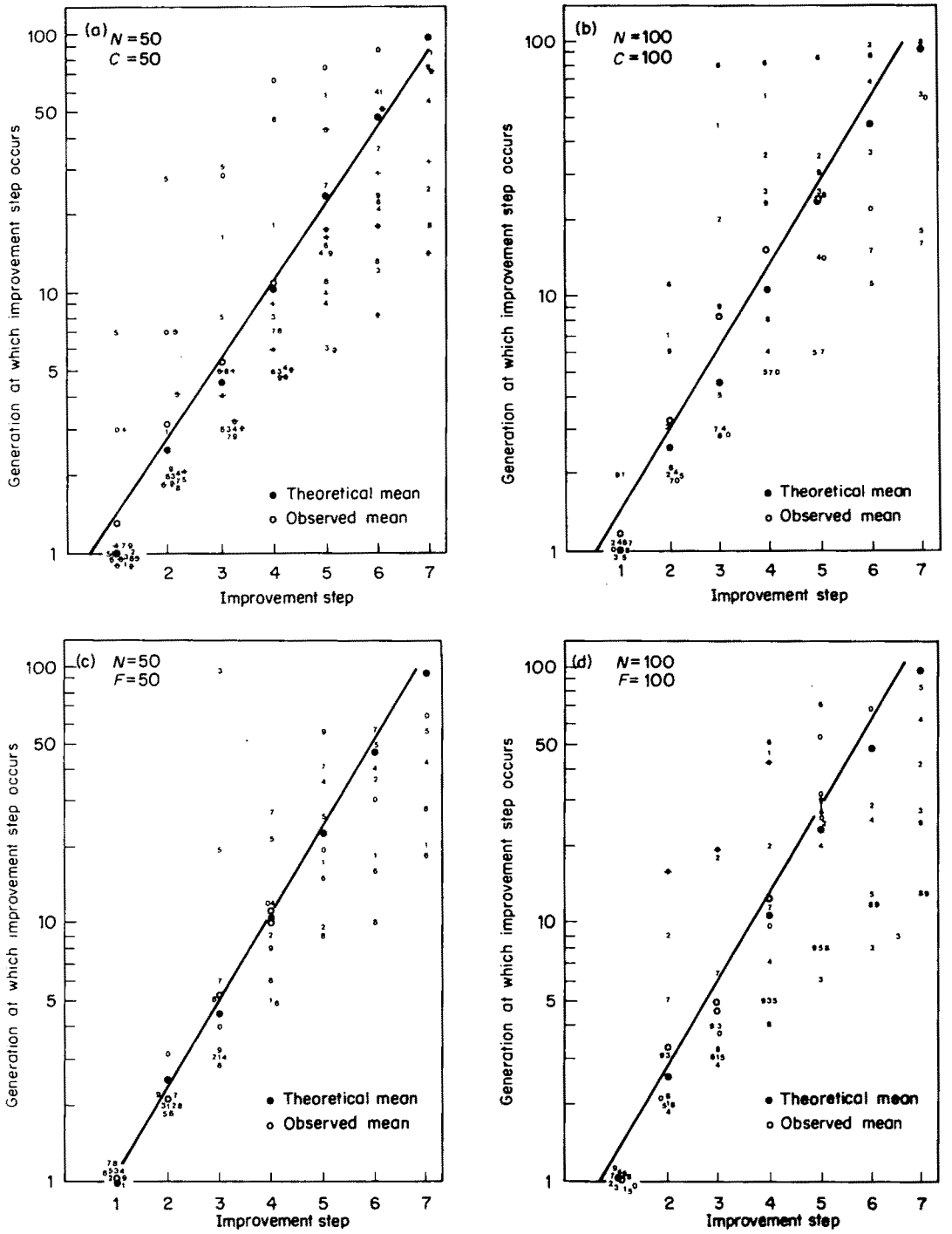

FIG. 2. (a)-(d) Long-jump adaptation in genetic regulatory networks with $N = 50$ or 100 genes, mutating half the regulatory connections per network, $C = 50$ or 100, or one quarter of the $4N$ "bits" in the N Boolean functions in each network, $F = 50$ or 100. Small numbers and symbols each refer to a particular selection experiment with a fixed size population of 50 networks, and show the generation at which each improvement step occurred.

curve, both for mutations to connections, and to bits in Boolean functions, supports the belief that the "universal law" should be very widespread indeed. Presumably, the topology of the fitness landscape under the two forms of mutation differs, yet for sufficiently long jump evolution, an uncorrelated landscape is encountered, and the general theorem applies.

While the bounds on the applicability of eqn (8) need to be found, it presumably applies to vastly many complex combinatorial optimization problems where the landscape is sufficiently rugged and multipeaked, even if the landscape is correlated, so long as an adaptive process of sufficiently long jumps is considered. Thus, it should apply to the traveling salesman problem, and to the intervals between finding configurations of lower energy in spin glasses in the long jump limit of flipping many spins at once, etc. In fact, for the traveling salesman problem, the agreement is extremely good. Our procedure, working with Dennis Swaney at Cornell, was to simulate the long jump case by considering random rearrangements of the current tour, and accepting, as replacements "mutations" that represent improvements. The results, summarized in Figs 3 and 4 indicate that the average cumulative number of improvement steps rises in a linear relationship proportional to $\log_2 G$, with a slope of 0·8. As noted above, the approach we adopted should slightly overestimate the number of improvement steps to be expected, and that is what is observed. The slope of the curves one standard deviation above and below the mean, 1·2 and 0·6, bracket the theoretical slope nicely.

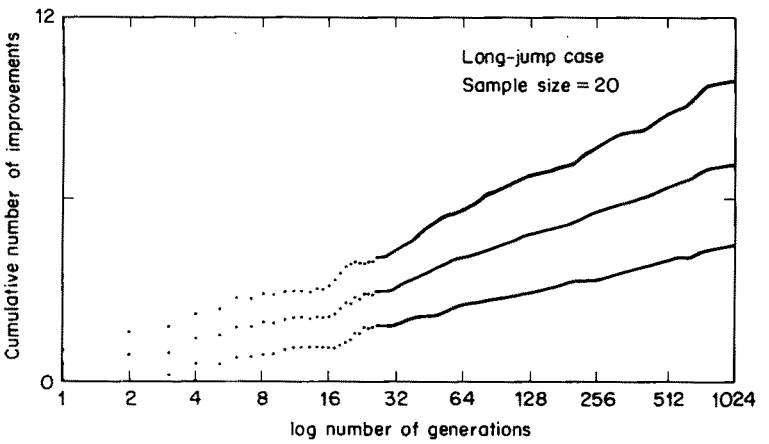

FIG. 3. Long-jump adaptation in the traveling salesman problem. At each generation, a population of 19 of 20 trial solutions was generated at random, while the current best solution was left unmodified. Slopes show mean and the +1 and −1 standard deviation curves for the cumulative number of improvements found plotted against the logarithm of the number of generations. Number of cities in tour is 100.

Figures 5(a)-(d) show another revealing feature of adaptation on correlated landscapes as the mutations per individual, J, is increased. For the 1-mutant case, improvement is rapid initially, but after 60 steps or so settles down to stasis. For 2 and 3 mutants per individual, the process continues to find improvements after the early flurry, mirroring the 1-mutant case, and only slows slightly thereafter. As the

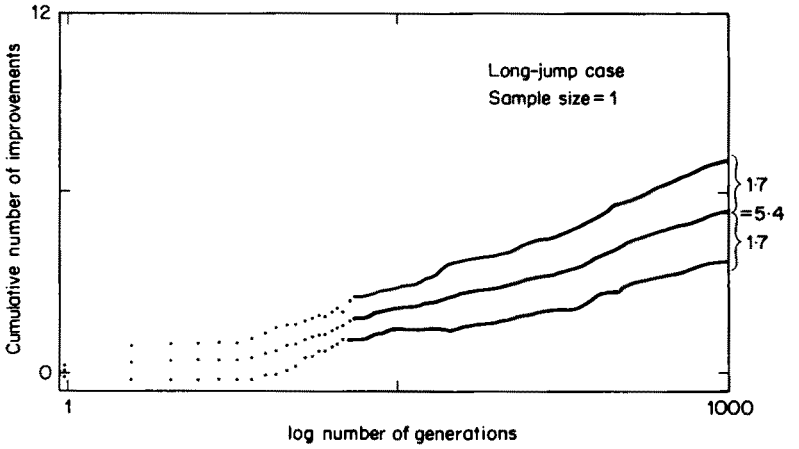

FIG. 4. Long-jump adaptation as in Fig. 3, but the population size was a single trial solution. At each generation a single random solution was tried and adopted if better than the current solution.

number of mutants per individual is increased further, however, the asymptotic rate of finding improvements slows. Thus the rate is slower for the 4-mutant case than the 3-mutant case, and slower still for the long jump limit. The explanation, presumably, is that, having found a "good hill" to climb, the 4-mutant or long jump process makes the mistake of ignoring that information. An efficient adaptive procedure should ignore the correlation structure initially, but capitalize on it later. Rather strikingly as we shall see next, just such a process appears to occur almost inevitably in a population with a *spectrum* of mutations per individual.

Long-jump evolution in biology is not merely a conceptual contrivance. A number of mechanisms assure that a single mutational event of a specific kind can achieve at once, a long jump. For example, a single nucleotide deletion causes a frame shift mutation, hence alters almost all of the amino acids incorporated into a protein downstream of the deletion. Duplication of structural genes, followed by recombination, is widely suspected of breaking and joining different protein domains into new combinations. Such rearrangements are long-jump mutations from the point of view of nucleotide substitutions. The entire panoply of chromosomal mutations, shuffling cis acting genes into proximity to new sets of structural genes are long jump mutations, for a single event can drastically rearrange regulatory cascades. Most broadly, sex and recombination, which shuffle sets of dominant or recessive regulatory and structural genes present in a diploid organism, constitute long-jump mutation across the space of genomes, although these may be difficult to maintain.

Generalizations Toward Correlated Landscapes

Most statistically rugged landscapes are correlated. Presumably, there are many ways in which such correlations can occur. A theory of long jump adaptation on rugged landscapes can be "universal" by jumping far enough to escape the correlation structure. But adaptation via 1-mutant neighbors cannot avoid such statistical features. The first step we take in considering this difficult problem is entirely

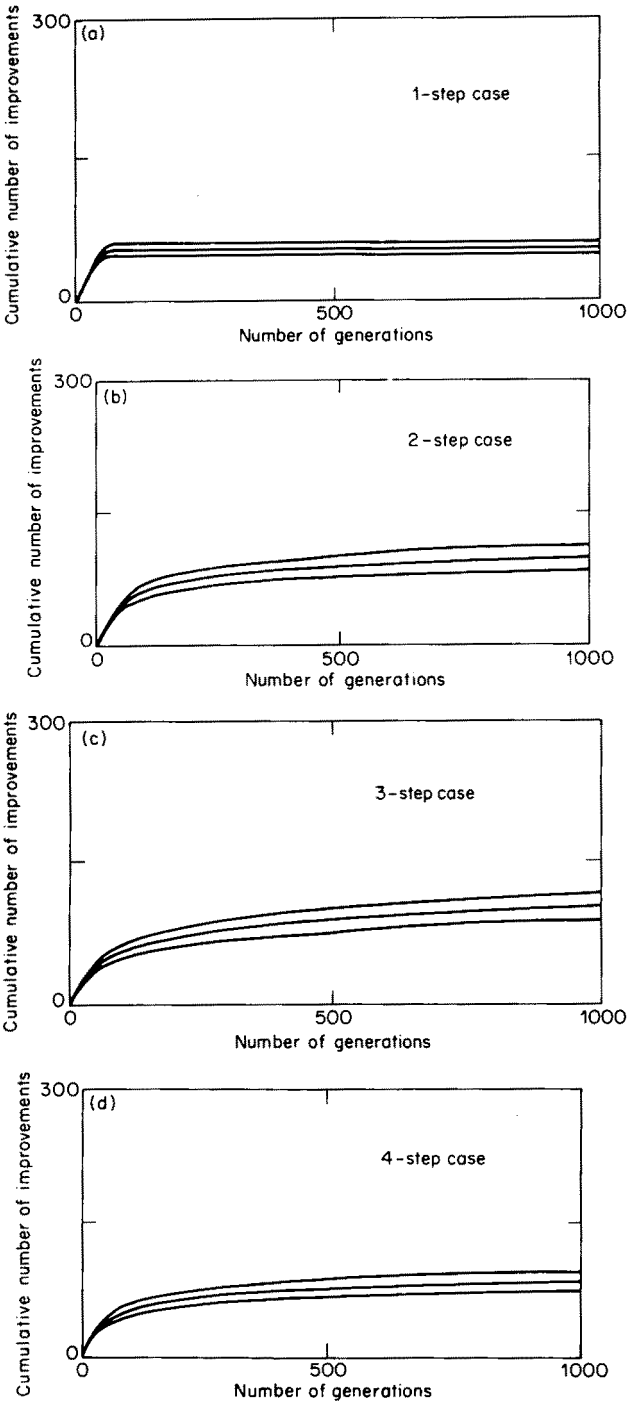

FIG. 5. (a)-(d) Cumulative number of improvements in the traveling salesman problem, subjected to 1, 2, 3, or 4-mutant moves, in 19 of 20 trial solutions at each generation, plotted against the number of generations. Note the data are not plotted here against the logarithm of the number of generations.

qualitative, but we hope, interesting. In a real biological population faced with some fixed mutation rate, the number of mutations per individual is given by some distribution from 0 to perhaps many. What, crudely, will be the evolutionary behavior of a population in a statistically correlated landscape, if "released" at a specified entity, then subjected to a spectrum of mutations? We recall the general result that the rate of invasion of the population by a fitter variant is proportional to the fitness differential. Thus, fitter variants sweep the population fast if the fitness differential is large.

Begin with a randomly chosen entity. Its fitness, on average, is average. Early in the adaptive process, the population samples both near and far from that entity in the space of entities. Half of all mutant entities sampled will be fitter than the first entity. However, the nearby entities will typically be only slightly better than the initial entity, due to the correlations in the fitness landscape, which assures that nearby entities tend to have similar fitness values. By contrast, the more distant mutant entities are less constrained by the correlation structure, hence include variants of higher fitness than the nearby variants. Consequently, *early* in the adaptive process, distant long jump mutants will dominate the evolutionary process. Further, because early branching to many fitter mutants is characteristic of long jump adaptation, radiation will tend to occur. But as more improved long jump variants are encountered, the waiting time to find still fitter long jump variants increases by doubling each time. Thus, eventually the rate of finding distant fitter variants is substantially less than the rate of finding nearby but only slightly fitter variants. These nearby variants therefore will have the opportunity to sweep through the population. Consequently, in the *midterm* of the process, adaptation occurs by familiar local hill climbing to or toward local optima. As the local hill is climbed, or the local optimum is reached, the rate of finding nearby fitter variants dwindles; thus in the *long term* the process must wait until a long jump mutation lands on the side of some distant hill, whereafter local hill climbing recommences.

Local hill climbing is a familiar idea in evolutionary biology. An intriguing feature of this new picture is that it enables a marriage of local and global search, coupled naturally to time scales in the adaptive process. A theory of long jump evolution gives us a picture of how long it takes the process to leave a local peak or its vicinity, and how that interval increases after each long jump. Further, since after each long jump typically a local hill is climbed, the interval before the next long jump fitter variant is found must typically more than double.

It is also worth marrying this image to Gillespie's (1983, 1984) concept of modeling environmental change by more or less correlated alterations in the fitness landscape. In so far as such shuffling of rank orders is extensive, the process is repeatedly cast back to more average or slightly above average fitness, where long jump adaptation dominates the dynamics. This combined qualitative picture has a number of possible implications; we focus on one.

Radiation and (relative) stasis are inherent in this picture of adaptation. For example, long jump adaptation and radiation should dominate early in the process. It is interesting that in the Cambrian–Ordovician radiation, it is well known that the major *phyla* were well established before their subsidiary *classes*, which in turn

were well established before their subsidiary *orders*, and so on (Raup, 1983). This is what our simple qualitative picture would suggest.

It is of interest to contrast our picture with classical accounts of radiation. Valentine (1980), for example, pictures an initially empty niche space. Organisms can occupy niches, mutate to distant or nearby niches, and survive in niches that are free, but cannot occupy a niche if already filled. To this he appends the hypothesis that most viable mutants are similar to the parental type, hence most jumps occur to nearby niches. This adaptive process supports radiation by jumps to distant niches; but as the niche space fills, the rate of branching radiation slows or ceases. Thus Valentine sees the rate of radiation as dominated by filling of niche space, hence tied to competition both as the motor of selection, and as the means of vacant niche elimination. By contrast, we locate the rate of radiation in the very structure of the adaptive landscape: it is the rate at which fitter distant variants are found and sweep all or part of the population. Mechanisms of isolation of subpopulations, which may be further required to stabilize alternative branches, are important, but are secondary to the fundamental mechanism.

Our ideas are very general, and are not confined to biological evolution. They should apply to all sorts of complex combinatorial optimization problems, including technological evolution. Thus, shortly after the invention of the first crude bicycle, it was easy to conceive of many alternative ways to improve the design. Branching radiation begins early. As successively refined versions are built, it becomes harder to keep making substantive radical improvements to the design, and eventually hard to make even minor improvements to each of the well wrought machines. Thus, it is of interest to examine parallels between biological and technological evolution.

Adaptation via 1-Mutant Neighbors on Correlated Landscapes

We stressed earlier that it is possible to derive a universal theory of adaptation on uncorrelated landscapes, and perhaps a universal theory of long jump adaptation on correlated landscapes, but that the properties of adaptation via 1-mutant neighbors on correlated landscapes promises to be more difficult. There may be indefinitely many families of different correlation structure, each with its characteristic statistical features and implications for adaptive walks. Alternatively, there may turn out to be rather few dominant kinds of correlation structures. In the latter case, we might eventually hope to classify those structures, discover what kinds of entities live in what kinds of landscapes, and build a generally useful body of theory. In this section we describe briefly an analytic approach to this problem, which is simple but apparently inappropriate for the concrete cases we consider below numerically. Then we explore two numerical examples which, despite apparent differences, appear to reveal a rather similar underlying statistical structure.

The simplest picture we can imagine of a correlated landscape is based on a rank ordering of the entities in a space, subject to the constraint that an entity and all its 1-mutant neighbors have fitness values drawn at random from some defined "W Ball" or range of rank orders W. That is, we assume that neighboring points do not differ by more than $W/2$. It is easy to verify that if the W ball range is large enough

with respect to the number of neighbors of an entity, D, it is possible to assign values consistent with the constraint.

The implications of this model are derived simply. Begin at an arbitrary entity of average fitness. Half of its neighbors are expected to be fitter. Pick the first fitter variant encountered. On average, it lies half way from the current entity to the top of the local W ball, hence moves upward a distance of $W/4$. This new variant is fitter than that from which it just arrived, which is in the W ball of the new variant. Thus on average, the new variant is slightly above the middle of its own W ball. Therefore, at each step, almost half the neighbors of an entity are fitter. At each step, a local W ball roughly centered on the entity found is carried along. If the total range contains T entities, the adaptive process continues until the percentile rank order is within the top $W/2T$. Thereafter, continued progression cannot carry the local W ball along centered on each fitter variant, and the process reverts to adaptation on an uncorrelated landscape, due to the hypothesis that fitness values are assigned at random within the W ball. This simple model shows that this type of correlation structure can greatly increase the lengths of adaptive walks before local optima are found, compared to that in an uncorrelated landscape. Further, in the present case, the rate of finding fitter variants is constant as fitness increases, until the top W ball is attained. For the same reason, branching radiation is constant as fitness increases until the top W ball is attained. As we will see shortly, this does not fit the cases we shall describe, where the rate of finding fitter variants slows as fitness increases, as in the uncorrelated landscape and long jump adaptation. This means that the correlation structures of these problems are more complicated than are those we have just discussed.

Lacking a good theory, we have resorted to numerical simulations to discover the features of landscapes for two apparently dissimilar problems: (1) adaptation of desired cell types via 1 to 5 connection or 1 to 5 Boolean function mutants in $K = 2$ input genetic networks; (2) heuristic approaches to the traveling salesman problem by 1-mutant to 4-mutant moves, where a one-mutant move involves exchanging the positions of two cities in the circuit through the N cities. Both complex combinatorial optimization problems nevertheless exhibit similar features.

Figure 6 shows typical results for genetic networks in which we mutated 1, 2 or 5 connections in genetic networks with $N = 50$ genes and $K = 2$ inputs per gene. We mutated connections in all but one member of a fixed size population, and selected the best in each generation with respect to matching a predefined cell type pattern of gene expression to seed the next generation. The first important result is that fitness increases faster initially, then slows, and reaches a local optimum well below a perfect match. In the many simulations we have done, we have not yet encountered a case where the adaptive process actually created a network with the desired cell type matched perfectly (Kauffman & Smith, 1986). The rapid initial rate of improvement reflects the fact that initially, many neighbors of low fit entities are fitter. Thus had we followed branching radiation, it would here again be bushy at the base, then quiet as adaptation proceeded upwards. We ignore many features of this process, and focus on only one. Figures 7(a) and 7(b) show the effect of network size, N, and numbers of mutations per individual, on the rate of finding

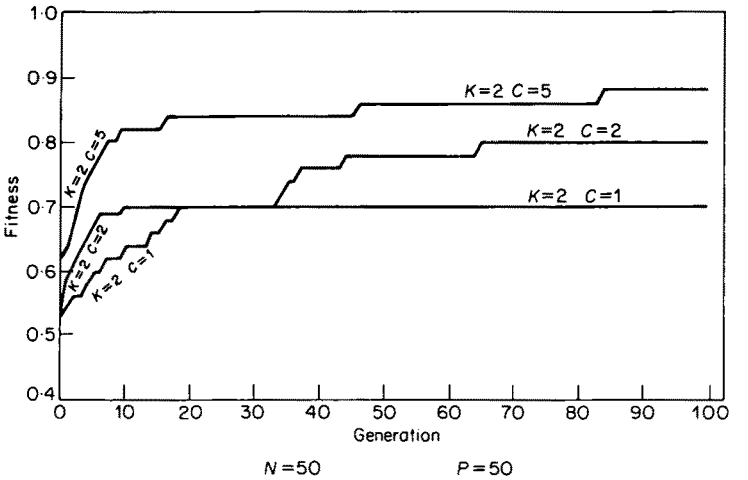

FIG. 6. Fitness, measured as fractional overlap of the closest attractor state cycle to a predetermined "target" pattern of gene activities, in genetic networks with $N = 50$ genes, and $K = 2$ inputs per gene. C, number of regulatory connections mutated in all but one copy of the current best network in each generation. P, population size in each selection run.

fitter variants, plotted against the logarithm of the number of generations; they exhibit the cumulative numbers of improvement steps. The interesting feature is that, in the midterm of the process, the data are roughly log linear, suggesting that the number of improvement steps is a function of the logarithm of the number of generations. This parallels the case for long jump adaptation where the slope is log 2; however, in this case, the slope is far shallower. Many more steps, each a smaller increment in fitness, appear to be taken, but a general log linear relation is again evident.

Figures 8(a)-(d) show similar results, replotting Figs 5(a)-(d) on a logarithmic scale of generations, for the traveling salesman problem. Adaptation is rapid at first, slows and reaches a local optimum for the 1-mutant case. For the 2, 3 and 4-mutant cases, the plots reveal a tendency for a rough log-linear relation in the midterm of the adaptive process. These results seem to indicate that these two correlated landscapes, and the cases of long jump adaptation and adaptation on uncorrelated landscapes, share very general statistical features. Although we cannot explain why these features are common, we can hope that they are sufficiently universal to build a useful general theory.

We conclude this section with a simple statement of the view that discovering the kinds of correlation structures that are common to different real complex combinatorial optimization problems is a first step towards developing a useful theory of adaptive walks via 1-mutant neighbors on such spaces.

Two Limitations on Adaptation in Rugged Landscapes in the Face of Complexity

In the final section of this article, we show two general tendencies: (1) As the complexity of entities under selection increases, the fitness of the *attainable* local

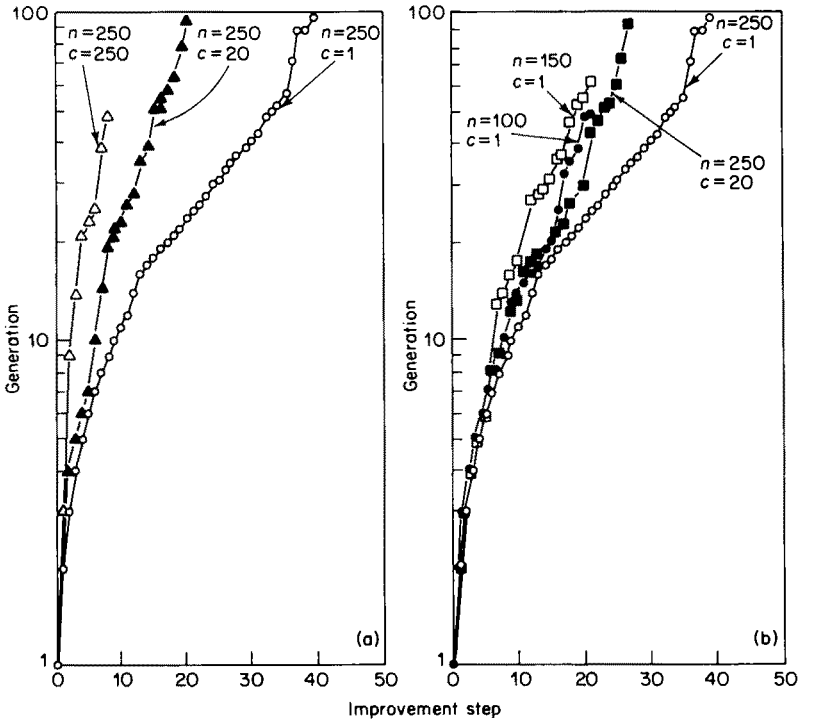

FIG. 7. (a) Cumulative number of improvements in genetic networks with $N = 250$ genes and $K = 2$ inputs per gene as $C = 1$, 20 or 250 connections per network were mutated in each generation in all but one copy of the current best genetic network. (b) Cumulative number of improvement steps in genetic networks with $n = 100$, 150, 200 and 250 genes per network, as $c = 1$ regulatory connection per network was mutated in each generation in all but one copy of the current best genetic network.

optima to which pathways exist falls closer to the unselected mean of the space. That is, as complexity increases, the local optima become more typical of the space of entities. (2) As complexity increases, selection becomes a weaker force relative to mutation, and becomes unable to *hold* an adapting population at those local optima to which pathways exist.

Adaptation on the Random Boolean Hypercube

To be concrete, consider an N dimensional Boolean hypercube. In the normal hypercube, each vertex is labeled with a vector of N entries each a 0 or 1 value, in the natural order, with (0000...00) at one vertex, (1111...111) at the antipode, etc. We here relabel the 2^N vertices by assigning to each at random, and without replacement, one of the 2^N vectors of N 0 and 1 values. We consider the fitness of each vertex to be the fraction of 1 values in the vector assigned to it. Two features characterize the distribution of fitness values. First it is uncorrelated on the space. Second, the fitness values are drawn from an underlying binomial distribution. In order to break ties, let the 0 and 1 values at each position in the vectors be slightly above or below 0 or 1 by small values chosen at random from a distribution. We

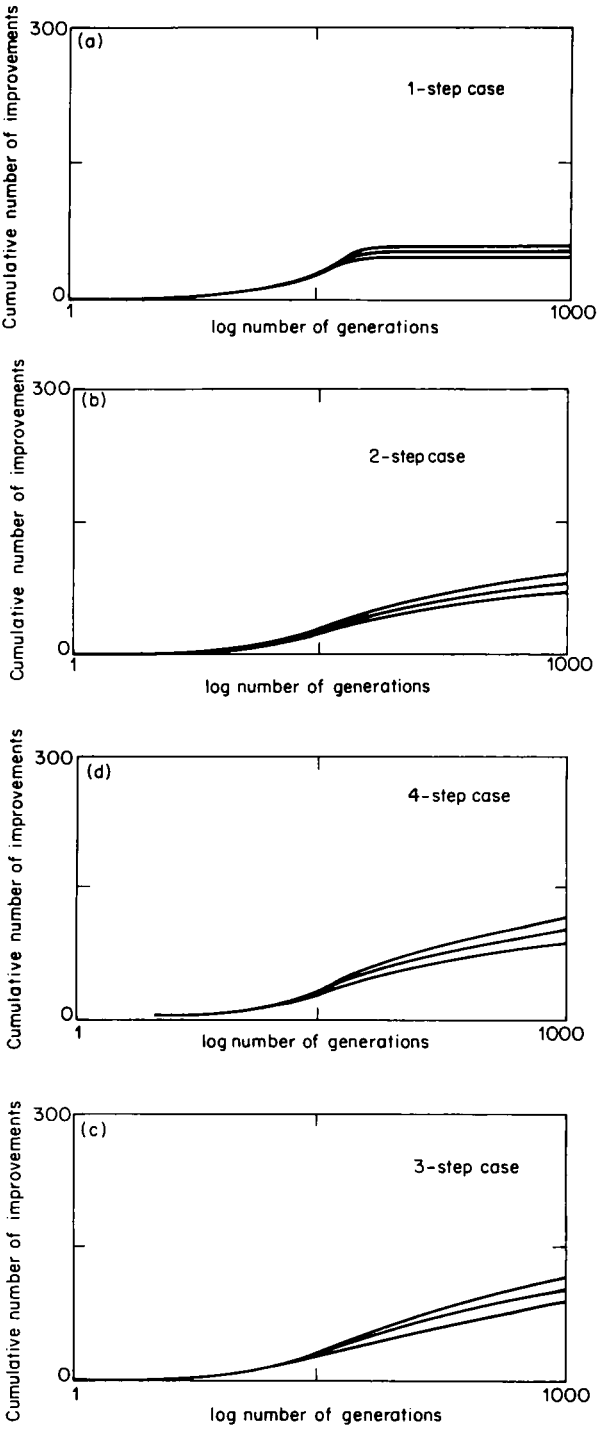

FIG. 8. (a)-(d) Replots of Figs 5(a)-(d), showing cumulative number of improvement steps plotted against the logarithm of the number of generations elapsed.

explicitly note that we no longer are considering a landscape in which fitness is only rank ordered, but one in which the landscape, though uncorrelated, consists of fitness values drawn at random from a defined range, approximately 0·0 to 1·0.

We already understand the character of adaptive walks on such an uncorrelated landscape. Begin with an arbitrary, hence average, vertex. Its fitness is about 0·5. Half of its neighbors are fitter. Adaptation will walk upward for about $\log_2 N$ steps, where $N = D$ is the number of neighbors of each vertex. The next question to consider is this: What happens to adaptive walks as N increases? N can be thought of as the complexity of the entities (here vectors) under selection. As N increases, the number of neighbors to each entity increases and the lengths of adaptive walks will increase. But what happens to the *actual fitnesses* of the local optima attained as N increases?

A simple way to think about the lengths of adaptive walks is that such walks will continue until the expected number of fitter neighbors to an entity drops below 1. Each entity has N neighbors. So the stopping criterion is that fitness at which the expected number of fitter neighbors is $1/N$. As fitness increases, the expected fraction of neighbors having a fitness greater than the fitness of that entity is given by the upper tail of the cumulative binomial distribution. For example, if N is 100, and a given entity has 60 1-values, then the cumulative binomial distribution showing the fraction of vectors having 61, 62, ... 100 1-values among the total 2^N vectors gives the probability that each neighbor will have a fitness higher than 0·6. As fitness increases, say to 70 1-values, this cumulative fraction decreases. Thus, walks will continue to higher fitness values until this probability falls to $1/N$. But, it is a simple consequence of the Central Limit Theorem that, as N increases, the cumulative fraction of vectors with 0·6 N 1-values or more, decreases *faster* than linearly. The critical implication is that, as N increases, the fitness values of attainable local optima *fall* toward 0·5, the unselected average fitness of entities in the space.

This is a very simple, but very general result. It rests only on two features of the space: First fitness is normalized in some sense, here as the *fraction* of 1-values. Second the fitness values are assigned from an underlying distribution such as Gaussian, to which the Central Limit Theorem applies. More precisely, the distribution has an upper tail for above average fitness values having the property that as the complexity of the entities in the space increase, the cumulative distribution of entities above any fixed above average fitness value *decreases* faster than the number of neighbors to the entity increases. For any such space, as complexity increases, the actual local optima that are attainable from initially average entities, fall ever closer to the fitness of the average entities in the space.

This result is not universal; landscapes that escape it can be constructed, in particular by having appropriated correlation features. Nevertheless, the general result seems to us to be of great significance. First, it should apply to long-jump adaptation on many correlated landscapes. In fact, in long-jump adaptation with model genetic nets, we have confirmed that as N, the number of genes in a network increases, the fitness attained after any fixed number of generations falls closer to the unselected mean fitness of an average network, 0·5 (Table 1). If this property holds true for adaptation on model genetic nets, it is reasonable to expect it to hold

TABLE 1

Mean fitness achieved in ten independent selection experiments with genetic networks with N = 20, 50, *or* 100 *genes, subjected to long-jump adaptation by mutating half of the regulatory connections in all but one of the current best network in each generation. The population size in each selection experiment was* 20 *networks. Mean fitness at each generation decreases as the complexity of the genetic network,* N, *increases, as predicted*

	N		
Generation	20	50	100
10	0.82	0.695	0.612
20	0.831	0.724	0.625
50	0.865	0.448	0.653
100	0.88	0.766	0.669

true quite widely. We have therefore identified a simple but probably powerful limitation of adaptation in complex systems. We are only beginning to understand the structures of fitness landscapes. However, for uncorrelated landscapes, and for long-jump adaptation on correlated landscapes where the underlying fitness values are drawn from something like a Gaussian or other appropriate distribution, it is to be expected that increased complexity is associated generically with a *decreased capacity* of selection to escape the average properties of the underlying space of entities. The more complex the entities, the more the best attainable by adaptive walks are typical of the entities in the space. Among the possible implications is this: If, as complexity increases, selection typically is unable to escape the average properties of the kinds of entities upon which it operates, and *if* those entities exhibit spontaneous order even in the absence of selection, then that spontaneous order may be present in the entities not because selection has achieved it, but because selection is unable to *avoid* it. Such ordered properties, in short, might be "ahistorical universals", reflecting, not selection, but the inherent properties of the class of systems upon which selection operates. Genetic networks in which each gene is regulated by rather few other genes are only one among many examples of complex systems which exhibit surprising order in the absence of selection. Many typical features of such networks are surprisingly close parallels to real genetic regulatory networks. It is not a trivial question whether such parallels reflect the self organized properties of a class of genetic regulatory systems in the *absence* of selection, which persist in contemporary organisms because selection is unable to avoid those properties.

We close by drawing attention to a classical result in population genetics which shows that as the complexity of a system increases, selection becomes progressively less able to hold an adapting population at local optima in the face of persistent mutation (Ewens, 1979). Consider a system of N loci, each with two alleles, one of higher fitness than the other, each contributing additively to overall fitness, and with equal forward and reverse mutation rates between the two alleles. As the number of loci, N, increases, the proportional contribution to overall fitness of the

more favored allele decreases inversely to N. Thus the selective force acting on any allele decreases, while the mutational force transforming it into the less favored allele remains constant. The consequence is that, as N increases, the expected number of less favored alleles that accumulates among the N increases as a square root function of N. For small N, selection can hold the population at the optimally fit entity with all N favored alleles. As N increases, a value is reached beyond which the population diverges away from that optimum.

This result has been rediscovered by Eigen & Schuster (1979) who call it the error catastrophy, and by (Kauffman, 1985) in investigations into the precision with which selection can mold genetic regulatory networks. The general implication is that as complexity increases, selection often becomes too weak a force to hold a population "at" those local or global optima to which connected adaptive walks exist. Thus, for two rather different reasons, there must be a strong general tendency, as the complexity of entities increases, for the best attainable entities to become more typical of the underlying space of entities: (1) Local optima for many fitness landscapes fall closer to the mean properties of the space as complexity increases. (2) Selection becomes too weak to hold populations at rare improbable local or global optima, and the population falls toward the mean properties of the underlying space.

Conclusions

This article has begun to develop a theory of adaptation on statistically rugged landscapes. The simplest benchmark case is an uncorrelated landscape of rank ordered fitness values. Simple mean properties of adaptive walks via 1-mutant fitter neighbors have been worked out; the variances and higher moments of those features have not. We have not included the concept of "neutral mutations", which might be thought of as equivalence classes of neighbors of "nearly" the same fitness. We have been able to give simple general results for long-jump adaptation on correlated rugged landscapes. The requirements on the underlying landscapes such that the theorems hold are not worked out. We have examined two cases of adaptation via fitter 1-mutant neighbors on two cases of correlated landscapes and found that they exhibit statistical features similar to one another and to the case of adaptation on an uncorrelated landscape. This encourages the hope that quite general features may characterize a wide variety of fitness landscapes for diverse complex combinatorial optimization problems. We have identified an unexpected limitation on selection: for many landscapes, as complexity increases, local optima fall toward the mean of the underlying space of entities.

Obviously, the work reported is the crudest beginning of an area that deserves careful development. Evolutionary theory has grown without a body of theory concerning the structure of adaptive landscapes for the kinds of complex combinatorial optimization that must occur in molding the "design" of organisms. We need to develop such a body of theory, exhibiting the "structures" of these landscapes, then examine whether and when selection is sufficiently powerful to pull evolving populations across such spaces to or toward fitness optima.

It is a pleasure to note useful conversations with John Gillespie, and help in simulations by Robert Smith and Dennis Swaney. This work was partially supported under D.N.R. Contract N00014-85-0258 to SK and NSF grant DMS-840647 to SL.

REFERENCES

AGUR, Z. & KERSZBERG, M. (1987). The emergence of phenotypic novelties through progressive genetic change. *Am. Nat.* (in press.)

BEREK, C., GRIFFITHS, G. M. & MILSTEIN, C. (1985). Molecular events during maturation of the immune response to oxazolone. *Nature* **316,** 412.

BOTHWELL, A. L. M., PASKIND, M., RETH, M., IMANISHI-KARI, T., RAJEWSKY, K. & BALTIMORE, D. (1982). Somatic variants of murine immunoglobulin lambda light chains. *Nature* **298,** 380.

BRADY, R. M. (1985). Optimization strategies gleaned from biological evolution. *Nature* **314,** 804.

CLARK, S. H., HUPPI, K., RUEZINSKY, D., STAUDT, L., GERHARD, W. & WEIGERT, M. (1985). Inter- and intraclonal diversity in the antibody response to influenza hemagglutinin. *J. Exp. Med.* **161,** 687.

CREWS, S., GRIFFIN, J., HUANG, H., CALAME, K. & HOOD, L. (1981). A single V gene segment encodes the immune response to phosphorylcholine: somatic mutation is correlated with the class of the antibody. *Cell* **25,** 59.

DERRIDA, B. & WEISBUCH, G. (1986). Evolution of overlaps between configurations in random Boolean networks. *J. Physique* **47,** 1297.

EIGEN, M. (1985). Macromolecular evolution: dynamical ordering in sequence space. In: *Emerging Synthesis in Science.* (Pines, D. ed.). Proceedings of the Founding Workshops of the Santa Fe Institute, pp. 25-69. Santa Fe, New Mexico.

EIGEN, M. & SCHUSTER, P. (1979). *The Hypercycle.* New York: Springer-Verlag.

EWENS, W. J. (1979). *Mathematical Population Genetics.* New York: Springer-Verlag.

FELLER, W. (1971). *Introduction to Probability Theory and Its Applications, Vol. 2,* 2nd edn. New York: Wiley.

FOGELMAN-SOULIE, F., GOLES-CHACC, E. & WEISBUCH, G. (1982). Specific roles of the different Boolean mappings in random networks. *Bull. Math. Biol.* **44,** 715.

GELFAND, A. E. & WALKER, C. C. (1984). *Ensemble Modeling.* New York: Dekker.

GILLESPIE, J. H. (1983). A simple stochastic gene substitution model. *Theor. Pop. Biol.* **23,** 202.

GILLESPIE, J. H. (1984). Molecular evolution over the mutational landscape. *Evolution* **38,** 1116.

HEINRICH, G., TRAUNECKER, A. & TONEGAWA, S. (1984). Somatic mutation creates diversity in the major group of mouse immunoglobulin *k* light chains. *J. Exp. Med.* **159,** 417.

JACOB, F. (1982). *The Possible and the Actual.* New York: Pantheon Books.

JOHNSON, D. S. & PAPADIMITRIOU, C. H. (1985). Computational complexity. In: *The Traveling Salesman Problem.* (Lawler, E. L., Lenstra, J. K., Rinnooy Kan, A. H. G. & Shmoys, D. B. eds.). pp. 37-85. Chichester: Wiley.

KAUFFMAN, S. A. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *J. theor. Biol.* **22,** 437.

KAUFFMAN, S. A. (1974). The large scale structure and dynamics of gene control circuits: an ensemble approach. *J. theor. Biol.* **44,** 167.

KAUFFMAN, S. A. (1984). Emergent properties in random complex automata. *Physica* **10D,** 145.

KAUFFMAN, S. A. (1985). Self organization, selective adaptation, and its limits: a new pattern of inference in evolution and development. In: *Evolution at a Crossroads.* (Depew, D. J. & Weber, B. H. eds.). pp. 169-207. Cambridge: MIT Press.

KAUFFMAN, S. A. & SMITH, R. (1986). Adaptive automata based on Darwinian selection. *Physica* 22 *D,* 68.

KIRKPATRICK, S., GELATT, C. D. JR & VECCHI, M. P. (1983). Optimization by simulated annealing. *Science* **220,** 671.

LEVIN, S. A. (1978). On the evolution of ecological parameters. In: *Ecological Genetics: The Interface.* (Brussard, P. F. ed.). pp. 3-26. New York: Springer-Verlag.

LIN, S. & KERNIGHAN, B. W. (1973). An effective heuristic algorithm for the traveling salesman problem. *Oper. Res.* **21,** 498.

NINO, J. (1979). *Approaches Moleculaires de l'Evolution.* Paris: Masson.

PROVINE, W. B. (1986). *Sewall Wright and Evolutionary Biology.* Chicago: The University of Chicago Press.

RAUP, D. M. (1983). On the early origins of major biologic groups. *Paleobiology,* **9,** 107.

SHERRINGTON, D. & KIRKPATRICK, S. (1975). *Phys. Rev. Lett.* **35,** 1792.

SMITH, J. M. (1970). Natural selection and the concept of a protein space. *Nature* **225,** 563.

TONEGAWA, S. (1983). Somatic generation of antibody diversity. *Nature* **302,** 575.

VALENTINE, J. W. (1980). Determinants of diversity in higher taxonomic categories. *Paleobiology* **6,** 444.

WRIGHT, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. In: *Proceedings 6th International Congress on Genetics* **1,** 356.

WYSOCKI, L., MANSER, T. & GEFTER, M. L. (1986). Somatic evolution of variable region structures during an immune response. *Proc. natn. Acad. Sci. U.S.A.* **83,** 1847.