Thoughts on the Synthesis of Life

Thomas S. Ray
School of Life & Health Sciences
University of Delaware, Newark, Delaware 19716
ray@vax1.acs.udel.edu
302-451-2753
February 2, 1990

ROUGH DRAFT

ABSTRACT

One of the most challenging problems in Artificial Life remains the creation of open ended evolution. A number of current efforts are directed at this problem. One characteristic that these efforts generally share is that they parallel the origin of life event, in that they attempt to create pre-biotic conditions from which Artificial Life may emerge spontaneously and evolve in an open ended fashion.

While the origin of life is generally recognized as an event of the first order, there is another event in the history of life that is less well known but of comparable significance. The origin of biological diversity and at the same time of macroscopic multi-cellular life, occurred abruptly in the Cambrian explosion 600 million years ago. This event involved a riotous diversification of life forms. Dozens of phyla appeared suddenly, many existing only fleetingly, as diverse and sometimes bizarre ways of life were explored in a relative ecological void.

I discuss an approach to AL that parallels this second major event in the history of life, the origin of diversity. Rather than attempting to create pre-biotic conditions from which life may emerge, this approach involves engineering over the first 3 billion years of life's history to design complex evolvable artificial organisms, and attempting to create the biological conditions that will set off a spontaneous evolutionary process of increasing diversity and complexity of organisms.

This is a very difficult undertaking, because in the midst of the Cambrian explosion, life had evolved to a level of complexity in which emergent properties existed at many hierarchical levels: molecular, cellular, organismal, populational and community. In order to define an approach to the synthesis of Artificial Life paralleling this historical stage of organic life, we must examine each of the fundamental hierarchical levels and abstract the principal biological properties from their physical representation, and determine how they can be represented in our chosen artificial media. I will discuss some of the outstanding biological properties of several of these levels, in the context of Artificial Life. This is an exercise in considering the extension of biological properties to life forms beyond the organic.

KEY WORDS

Artificial life, synthesis, evolution, ecology, diversity, complexity

INTRODUCTION

One pleasing characteristics of the field of Artificial Life is that it is defined by each act of creation of an instance of AL. Artificial life forms are as diverse as the imaginations and interests of their creators. Each time we create an example of Artificial Life, we create a parallel universe that reflects a conception of organic life; we extract some aspect of organic life that interests us, and render it in our artificial media.

One aspect of organic life that is lacking in Artificial Life is history. Organic life on earth has a history of nearly four billion years. While AL lacks its own history, each instance of AL is likely to parallel organic life at some specific stage of organic history, even though such parallels may not have been explicitly intended by its creators. Therefore instances of AL could be grouped by their relations to the various historical stages of organic life.

Some instances of Artificial Life reflect recognizable organisms (e.g., trees), and therefore can be considered to parallel the historical period in which these specific organisms were extant. Some instances of AL reflect prebiotic conditions, and therefore can be considered to parallel the period of greater than 3.6 billion years before present.

One of the most challenging problems in Artificial Life remains the creation of open ended evolution. A number of current efforts are directed at this problem. One characteristic that these efforts generally have in common is that they parallel the origin of life event, in that they attempt to create prebiotic conditions from which Artificial Life may emerge spontaneously and evolve in an open ended fashion.

While the origin of life is generally recognized as an event of the first order, on a par with the origin of the universe, there is another event in the history of life that is less well known but of comparable significance. The origin of biological diversity and at the same time of macroscopic multicellular life, ocurred abruptly in the Cambrian explosion 600 million years ago. This event involved a riotous diversification of life forms. Dozens of phyla appeared suddenly, many existing only fleetingly, as diverse and sometimes bizarre ways of life were explored in a relative ecological void (Gould, 1990). This period of intense experimentation is recorded in some of the rare fossil deposits of the period that were of a quality to preserve soft body parts (Morris, 1989).

I would like to discuss an approach to AL that parallels the second major event in the history of life, the origin of diversity. Rather than attempting to create prebiotic conditions from which life may emerge, this approach involves engineering over the first 3 billion years of life's history to design complex evolvable artificial organisms, and then attempting to create the ecological conditions that will set off a spontaneous evolutionary process of increasing diveristy and complexity of organisms.

This is a very difficult undertaking, because in the midst of the Cambrian explosion, life had evolved to a level of complexity in which emergent properties existed at many hierarchical levels: molecular, cellular, organismal, populational and community. In order to define an approach to the synthesis of Artificial Life paralleling this historical stage, we must examine each of the fundamental hierarchical levels and abstract the principal biological properties from their physical representation, and determine how they can be represented in our chosen artificial media, in this case the computer. The essay which follows discusses some of the outstanding biological properties of several of these levels, in the context of Artificial Life.

SOME BASIC PRINCIPLES

The Metaphor

Organic life is viewed as utilizing energy, mostly ultimately derived from the sun, to organize matter. By analogy, digital life could be viewed as using CPU time, to organize memory. Organic life evolves through natural selection as individuals compete for resources (light, food, space, mates) such that genotypes which leave the most descendants increase in frequency. Digital life might evolve through the same process, as replicating algorithms compete for resources including CPU time and memory space.

I will discuss Artificial Life in the context of this specific AL model, digital organism adapting to the CPU-memory-operating system environment. A primary problem with attempting to evolve von Neumann style code is its "brittleness". The ratio of viable to possible genotypes is very low. However my experiments with the Tierra Simulator demonstrate that this problem can be easily overcome.

Below I will discuss ideas on how to loosen up machine code in a "virtual bio-computer", in order to create a computational system based on a hybrid between biological and classical von Neumann processes. What is most critical in this essay are the relationships between AL and biological principles. These relationships should remain the same regardless of the underlying computational representations.

To the digital organism, the environment consists of the CPU, the memory, the simulator program, and the other digital organisms sharing the resources. Successful adaptation to this environment will involve strategies for obtaining and retaining access to CPU time and memory. Selection might result in the evolution of strategies to increase priority levels, or to examine other individuals and assess their potential as mates, threats, or to be exploited. Predatory or parasitic species might evolve adaptations to exploit other individuals by seizure of their processor threads or allocated memory. Prey or host species may evolve strategies to defend their memory or CPU resources. This is a sort of free form evolutionary core wars (Dewdney 1984, 1987), a core wars where the only rule is the survival of the fittest.

Below I will discuss biological principles and processes and how they relate to digital organisms. This is an exercise in consiering the extension of biological properties to life forms beyond the organic.

GENES — MOLECULES — CODES — INSTRUCTIONS

Mutation

Mutation is the ultimate source of novelty in evolution. Incorporation of subroutines into digital organisms that cause errors in the replication of code during reproduction could provide a source of mutations. However, the mutation subroutines would likely be selected against and should be rapidly eliminated from the population.

For this reason, mutation should be imposed on the system from without, by the simulator program. The system might run a process which randomly alters memory at some appropriate rate. Similarly, other routines might result in rearrangement of code within organisms, rather than point mutations. While most such alterations of code are likely to result in the death of the organism, some small percentage of the altered individuals would presumably survive and would provide the raw material for evolution.

Selfish Code — Transposons

The explosion of diversity in the Cambrian occurred in the lineage of the eukaryotes; the prokaryotes were not players in the explosion. One of the most striking genetic differences between eukaryotes and prokaryotes is that most of the genome of prokaryotes is translated into proteins, while most of the genome of eukaryotes is not. It has been estimated that typically 98% of the DNA in eukaryotes is neither translated into proteins nor involved in gene regulation, that it is simply "junk" DNA (Thomas, 1971). Orgel and Crick (1980) and Doolittle and Sapienza (1980) have suggested that this junk code is the result of the self-replication of pieces of DNA within rather than between cells.

Mobile genetic elements, transposons, have this intra-genome self-replicating property. It has been estimated that 80% of spontaneous mutations are caused by transposons (Green, 1988). Repeated sequences, resulting from the activity of mobile elements, range from dozens to millions in numbers of copies, and from hundreds to tens of thousands of base pairs in length. They vary widely in dispersion patterns from clumped to sparse (Jelinek and Schmid, 1982).

Transposons code for an enzyme, "transposase", which makes a copy of the transposon and inserts it somewhere else in the genome of the same cell, though not all mobile elements code for their own transposase. Larger transposons carry one or more genes in addition to those necessary for transposition. Transposons may grow to include more genes; one mechanism involves the placement of two transposons into close proximity so that they act as a single large transposon incorporating the intervening code. In many cases transposons carry a sequence that acts as a promotor, altering the regulation of genes at the site of insertion (Syvanen, 1984).

Transposons do produce gene products (e.g., transposase) and often are involved in gene regulation. However, they may have no effect on the external phenotype of the individual (Doolittle and Sapienza, 1980). Therefore they evolve through another paradigm of selection, one that does not involve an external phenotype. They are seen as a mechanism for the selfish spread of DNA which may become inactive junk after mutation (Orgel and Crick, 1980).

DNA of transposon origin can be recognized by their palindrome endings flanked by short non-reversed repeated sequences resulting from insertion after staggered cuts. In *Drosophila melanogaster* approximately 5 to 10 percent of its total DNA is composed of sequences bearing these signs. There are many families of such repeated elements, each family possessing a distinctive nucleotide sequence, and distributed in many sites throughout the genome. One well known repeated sequence occuring in humans is found to have as many as a half million copies in each haploid genome (Strickberger 1985). Elaborate mechanisms have evolved to edit out junk sequences inserted into critical regions. An indication of the magnitude of the task comes from the recent cloning of the gene for cystic fibrosis, where it was discovered that the gene consists of 250,000 base pairs, only 4,440 of which code for protein, the remainder are edited out of the messenger RNA before translation (Kerem et. al. 1989; Marx 1989; Riordan et. al. 1989; Rommens et. al. 1989).

It appears that repeated sequences in genomes originated as transposons favored by selection at the level of the gene, favoring genes which selfishly replicated themselves within the genome. However, some transposons coevolved with their host genome as a result of selection at the organsimal or populational level, favoring transposons which introduce useful variation through gene rearrangement. In this manner, "smart" genetic operators evolved, through the interaction of selection acting at two or more hierarchical levels (it appears that some transposons have followed another evolutionary route, developing mobility and becoming viruses). It is likely that transposons today represent the full continuum from purely parasitic "selfish DNA" to highly coevolved genetic operators and gene regulators. The posession of smart genetic operators must have contributed to the explosive diversification of eukaryotes by providing them with the capacity for natural genetic engineering.

In designing self replicating digital organisms, it would be worthwhile to introduce such genetic parasites, in order to facilitate the shuffling of the code that they bring about. The excess code generated by this mechanism provides a large store of relatively neutral code that can randomly explore new configurations through the genetic operations of mutation and recombination. When these new configurations confer functionality, they may become selected for.

Genotype — Phenotype — Ribotype

Modern evolutionary theory is firmly based on the duality of the genotype and the phenotype. However, Barbieri (1985) has described a new view, in which life is based on a trinity of genotype, phenotype and ribotype. At the molecular level, the genotype is the DNA, the phenotype is the proteins, and the ribotype is the collection of molecules and structures based on RNA, i.e., the mRNA, tRNA and the ribosomes. The latter group of molecules, referred to collectively as the ribosoids, perform the critical function of translating the genotype into the phenotype.

Barbieri provides a convincing and refreshingly gap-free speculation on the origin of life, which begins with "quasi-replication" of low molecular weight ribosomes which only polymerize amino acids. Apart from providing a plausible explanation for the origin of life, Barbieri's theory is valuable in focusing attention on the role of the ribosoids. It is the ribosoids that give meaning to the genetic code. In the computer metaphor, if the bit pattern in memory is the genotype, and the instruction that the pattern is executed as is the phenotype, then the ribotype is the decode unit. The paradigm of the trinity lends support to the notion that the mechanism that performs the decoding of genotype into phenotype should be incorporated into the software of the digital organism so that it can evolve. It may be stifiling to allow the translation to be performed by the hardware or by the software of the simulator, rather than by the software of the organism.

Skipping Instruction Pointers — Code Folding

This idea derives in part from the evolutionary truism that novel adaptations are the result of modification of existing structures that are already present for another function (for example, legs may be modified for flying as in the flying squirrel and the bats). The notion is applied at the molecular level.

A particular sequence of DNA will generally be translated into a chain of amino acids (i.e., a protein). In the case of enzymes, these long chains fold up into a complex three-dimensional structure of which only small portions must assume critical configurations to serve as active sites (i.e., for catalysis of chemical reactions or regulatory functions). Other regions of the chain serve to bring the critical regions into their proper positions. Mutations altering the critical amino acids are likely to result in a non-functional enzyme. However, mutations affecting amino acids in other parts of the same enzyme molecule are in many cases less critical (Reidhaar-Olson and Sauer, 1988).

The result is that the critical region of the DNA is not robust to genetic operations, however, the non-critical regions are able, within limits, to randomly explore other configurations through these operations. If a new configuration arises in the non-critical regions that confers some new catalytic activity, that new configuration can be selected for if it enhances the reproductive potential of the cell. Furthermore, if the gene for this protein becomes duplicated, the two copies can then specialize their structures to optimize their respective catalyses.

The flexibility of the non-critical regions of proteins is not found in traditional computer programs. In computers, the instruction pointer moves sequentially from one instruction to the next in the memory (consider instructions as analogous to amino acids). Every instruction is critical, so virtually any change results in a non-functional program, and there is no region of code that can be modified without devastating effects. In order to implement an analog to the biological properties described above, we may set up an instruction pointer that skips through memory.

The instruction pointer may skip along, executing for example every third instruction. This would force the code to include two parallel non-critical sequences, which serve the purpose of bringing the critical instructions into the proper alignment to be executed. Instruction pointers would pass down these non-critical sequences, interpreting their functions, but it does not matter if their function is non-sense because the critical functions are being taken care of by other sequences. Mutation and recombination will act on all instructions. They will generally destroy the function of the critical regions, however, they can cause the non-critical regions to explore new configurations.

The every third instruction example is simple and arbitrary, and we could experiment with a wide range of more sophisticated skipping patterns. For example, let us assign a spacer value ranging from 1 to 3 to each instruction in the instruction set of our virtual computer. The instruction pointer will skip a number of spaces equal to the product of the spacer values of the last two executed instructions. If the last two instructions had spacer values of 2 and 3, then the IP will skip 6 spaces. However, each instruction skipped counts as the number of spaces of its assigned spacer value, thus two instructions of value 3 would together provide the 6 spaces. This arrangement would cause sequences to be of irregular spacing, and would allow sequences to intertwine.

Address by Object

Machine codes must know the exact address of any object they wish to utilize, so in machine code we have direct, indirect, register-indirect etc. addressing. In biological systems, objects are often addressed by templates to a part of the object, and the system relies on diffusion and collision to bring the objects together. This could be implemented in a virtual bio-computer by allowing the system to perform the diffusion-collision service.

Suppose that a piece of code wants to address an object. The instruction can indicate that the addressing mode is by object, and provide as an operand, a template or other symbol for the object. The system then performs the service of searching out in both directions for the nearest instance of the object, and hands the address to the instruction. The search might be limited to a certain radius or to be within an allocated block of memory. The "address by object" mode has some of the flexible properties of classifier systems. However by incorporating it as an addressing option for an assembler type language, it would remain powerfully programmable.

Metabolism of Instructions — Informational Objects

The MOV instruction of conventional assembly languages is a thermodynamic anomaly in that it violates the laws of conservation. MOV doesn't only move information from one place to another, it duplicates it at the same time. Perhaps this behavior is appropriate for digital life, as it is clearly native to the computer. However it would be interesting to experiment with the properties of a MOV instruction that conserves information, by leaving all zeros behind. This system endows information with a real energetic value and opens new possibilites for ecological interactions between individuals through the transfer of informational objects.

Suppose that we let zero represent the low energy state. Let all informational objects be built through the manipulation of bits, through the use of just three instructions, as follows: ONE places a one in the low order (right-most) bit of the cx register. SHL shifts the bit pattern in the cx register one bit to the left, introducing a zero into the right-most bit. XOR performs an OR operation with the contents of the cx register on the address indexed by the ax register. The contents of the address will receive the bit pattern from the cx register, except where a one is present in particular bit of both patterns in which case the bit will be reversed to zero.

Through this process, informational objects can be created, but at considerable computational expense. It will generally be more economical to utilize existing objects rather than to synthesize new ones. In addition, this introduces the possibility of utilizing informational objects as energy (cpu time) storage objects. Each bit pattern can be considered to be equivalent to the minimun number of instructions that it takes to synthesize it from zero. The highest energy object would be that consisting of all ones. Informational objects could be metabolized by converting their bits to zero, thereby entitling the organism to the appropriate number of execution cycles by the system.

Polymerization of Informational Objects — Pointer Bonding

When polymerization of amino acids, nucleotides or sugars occurs in organic systems, the monomer diffuses to the site of polymerization, and through a condensation reaction is added to the end of a growing chain of physically contiguous bonded molecules. In the computational system we do not have diffusion, although we may allow addressing by object to simulate this process. However when we attempt to polymerize informational objects such as instructions, we run into another thorny problem. If we move a series of instructions into a linear sequence in memory in order to form a functional process, we must overwrite the information already in that space, or move it over to make room. A simulator with conservation as described above would not permit the writing over of non-zero information, therefore movement would be required. This would mean nightmarishly complex and computationally expensive shuffling of positions any time an instructional or genetic sequence were formed.

As an alternative, all polymerizable computational objects could include a pointer at each end, which would be analogous to the amine and carboxyl groups of amino acids. When two objects bond, the low pointer of the first object in the sequence would be made to point to the address of the next object, while the high pointer of the next object would be made to point to the address of the previous object. As a concession to the need for a diffusion system (as in addressing by object) the pointer addresses could be added to the objects for "free", that is they would not have to be synthesized through bit manipulations. The single instruction BOND could link the pointers.

INDIVIDUALS

Cellularity

Cellularity is one of the fundamental propeties of organic life, and can be recognized in the fossil record as far back as 3.6 billion years. The cell is the original individual, with the cell membrane defining the limits and preserving the chemical integrity. An analog to the cell membrane is need in digital organisms in order to preserve the integrity of the informational structure from being disrupted by the activity of other organisms.

The need for this can be seen in AL models such as cellular automata where virtual state machines pass through one another (Langton 1986), or in core wars type simulations where coherent structure that arise demolish one another when they come into contact (Rasmussen, Feldberg, Hindsholm and Knudsen, preprint). An analog to the cell membrane that can be used in the core wars type of simulation is memory allocation. An artificial "cell" could be defined by the limits of an allocated block of memory. Free access to the memory within the block could be limited to process within the block. Processes outside of the block would have limited access, according the the rules of "semi-permeability"; for example they might be allowed to read and execute but not write.

Multi-cellularity

The Cambrian explosion of diversity ocurred simultaneously with the appearance of macroscopic multi-cellular organisms. The transition from single celled to multi-celled life seems to have been an important component of the origin of biological diversity. It is important to consider then, what are the essential characteristic of multi-cellularity that could be incorporated into artificial life forms.

Buss (1987) provides a provocative discussion of the evolution of multi-cellularity, and explores the consequences of selection at the level of cell lines. From his discussion the following idea emerges (although he does not explicitly state this idea, in fact he proposes a sort of inverse of this idea, p. 65): the transition from single to multi-celled existance involves the extension of the control of gene regulation by the mother cell to successively more generations of daughter cells.

In organic cells, genes are regulated by proteins contained in the cytoplasm. During early embryonic development in animals, an initially very large fertilized egg cell undergoes cell division with no increase in the overall size of the embryo. The large cell is simply partitioned into many smaller cells, and all components of the cytoplasm are of maternal origin. By preventing several generations of daughter cells from producing any cytoplasmic regulatory components, the mother gains control of the course of differentiation, and thereby creates the developmental process. In single celled organisms by contrast, after each cell division, the daughter cell produces its own cytoplasmic regulatory products, and determines its own destiny independent of the mother cell.

Complex digital organisms will be self replicating algorithms, consisting of many distinct processes dedicated to specific tasks (e.g., locating free memory or mates, defense, replicating the code). The activity of these processes must be coordinated and regulated. After replication, the daughter organism will assume control of its own processes. However, if the mother organism can continue to regulate the processes of the daughter, so as to force the daughter to specialize in function and express only a portion of its potentiality, then the essence of multi-cellularity will be achieved.

Mortality — Immortality

There is no reason why a functional algorithm cannot live forever. Therefore, once the memory has filled with organisms, the process could grind to a halt, as there would be no free memory space to reproduce or grow into, and all existing individuals would persist forever. However, mutation processes imposed by the system would tend to limit the lifespan of individuals, thereby freeing memory space for new offspring as older individuals mutate and die.

Another possible source of mortality would be through predation and parasitism. If organisms are able to adopt strategies (through evolution or design) to co-opt the memory occupied by other organisms or their CPU access, they could in effect kill them and fill the newly freed memory space with their offspring.

SPECIES — POPULATIONS

The Species Concept — Mating Rituals

Biological species may be defined as a group of individuals capable of interbreeding freely under natural conditions. This biological species concept can best be applied to sexual species. In digital organisms, sexual reproduction of a chimeric offspring by two parent algorithms will be technically difficult to achieve, but will be essential to allow for rapid evolutionary change. Holland (1975) has demonstrated the efficiency of recombination in leading to adaptation in genetic algorithms. Digital sex may be implemented in a great variety of ways, so much experimentation will be needed to determine the most effective techniques.

Sexual organisms must search for and select mates. Digital organisms may place messages providing information about themselves in shared memory. These messages can be examined by other individuals and used to initiate communications. Once communication channels have been established, messages can be exchanged to determine the feasibility of producing viable chimeric offspring from a mating. These communications would presumably compare the code of the two individuals to determine the degree of similarity. However, the comparison of code may be replaced (either by the digital designer or by evolution if it occurs) by a more efficient symbolic exchange of signals, a mating ritual. If the algorithms determine according to some criteria that they are not sufficiently similar, they will not mate. In this case, they are in effect different species.

Geographic Isolation and Speciation

A common mechanism of speciation for organic life is geographic isolation. Similar isolation mechanisms will probably be needed in order to allow speciation in sexually reproducing digital organisms. This could be implemented by running digital communities in isolated memory compartments or on different machines, and then periodically allowing some individuals to migrate between these isolated systems.

COMMUNITIES — DIVERSITY

Major temporal and spatial patterns of organic diversity on earth remain largely unexplained, although there is no lack of theories. Diversity theories suggest fundamental ecological and evolutionary principles which may apply to digital or other forms of artificial life. In general these theories relate to artificial life in two ways: 1) They suggest factors which may be critical to the auto-catalytic increase of diversity and complexity in an evolving system. It may be necessary then to introduce these factors into an artificial system to generate increasing diversity and complexity. 2) Because it will be possible to manipulate the presence, absence, or state of these factors in an artificial system, the artificial system may provide an experimental framework for examining evolutionary and ecological processes.

Community Structure

The simulator program will determine the mechanisms of interprocess communication, memory allocation, and the allocation of CPU time among competing processes. In theory, algorithms will evolve so as to exploit these features to their advantage. More than being a mere aspect of the environment, the simulator program will determine the topology of possible interactions between individuals, such as the ability of pairs of individuals to exhibit predator-prey, parasite-host or mutualistic relationships.

The simulator program determines not only the physics and chemistry of the virtual universe that it creates, but the community ecology as well. A main thrust of this paper is that ecological interactions are critical driving forces in evolution, and Artificial Life models must incorporate these forces in order to generate spontaneously increasing diversity and complexity. Therefore I suggest that it will be necessary to experiment with the structure of the simulator program in order to facilitate the existence of the appropriate ecological interactions. For example, if the integrity of individuals is confered through the allocation of memory, then there needs to be a means whereby a predator or parasite can circumvent this integrity in order to take advantage of its prey or host.

The Red Queen

Consider that any organism that is not well adapted to its environment will quickly go extinct. Therefore it follows logically that all extant organisms are well adapted to their environment. This logic seems to present a paradox: if all extant organisms are well adapted, then what is evolution doing? The Red Queen hypothesis (Van Valen, 1973) suggests that in the face of a changing environment, organisms must evolve as fast as they can in order to simply maintain their current state of adaptation. "In order to get anywhere you must run twice as fast as that" (Carroll, 1865).

If we were only concerned with the abiotic environment, the race would not be so urgent. Species would only need to evolve as fast as the relatively gradual changes in the geology and climate. However, a critical component of the environment for any organism is the other living organsims with which it must interact. Given that the species that comprise the environment are themselves evolving, the race becomes rather hectic. The pace is set by the maximal rate that species may change through evolution, and it becomes very difficult to actually get ahead. A maximal rate of evolution is required just to keep from falling behind. This is the principal theme of this paper, that interactions with other evolving species provide the primary driving force in evolution. These interactions must be incorporated into Artificial Life models in order to move evolution.

The Enigma of Sex

The preponderance of sex remains an enigma to evolutionary theory (Bell 1982). Careful analysis has failed to show any benefits from sex that outweigh the high costs (e.g., passing on only half of the genome). The only obvious benefit of sex is that it provides diversity among the offspring, allowing the species to adapt more readily to a changing environment. However, quantitative analysis has shown that in order for sex to be favored by selection at the individual level, it is not enough for the environment to change unpredictably, the environment must actually change capriciously (Charlesworth 1976; Maynard Smith 1971). That is, whatever genotype has the highest fitness this generation, must have the lowest fitness the next generation, or at least a trend in this direction, a negative heritability of fitness.

One theory to explain the perpetuation of sex (based on the Red Queen hypothesis) states that the environment is in fact capricious, due to the importance of biotic factors in determining selective forces. That is, sex is favored because it is necessary to maintain adaptation in the face of evolving species in the environment (e.g., predators/parasites, prey/hosts, competitors) who themselves are sexual, and can undergo rapid evolutionary change. Predators and parasites will tend to evolve so as to favor attacking whatever genotype of their prey/host is the most common. The genotype that is most successful at present is targeted for future attack.

It may be possible to test this theory of sex by studying the evolution of reproductive strategies in digital organisms tested in environments with or without other sexual species. In the context of digital organisms, asexual reproduction means producing an exact copy of oneself in memory, sexual reproduction means mating with another individual to produce a chimeric offspring. A test species would be capable of reproducing both sexually and asexually. Natural selection could alter the proportionate reliance on the two methods, and this may evolve differently depending on whether there are other sexual species in the environment.

Disturbance and the Maintenance of Diversity

The Gaussian Principle of Competitive Exclusion states that no two species that occupy the same niche can coexist. The species which is the superior competitor will exclude the inferior competitor. The principle has been experimentally demonstrated and is considered theoretically sound. However, natural communities flaunt the principle. In tropical rain forests several hundred species of trees coexist without any dominant species in the community. All species of trees must spread their leaves to collect light and their roots to absorb water and nutrients. Evidently there are not several hundred niches for trees in the same habitat. Somehow the principle of competitive exclusion is circumvented.

There are many theories on how competitive exclusion may be circumvented. One leading theory is that periodic disturbance at the proper level sets back the process of competitive exclusion, allowing more species to coexist (Huston 1979). There is substantial evidence that moderate levels of disturbance can increase diversity. In a digital community, disturbance might take the form of freeing blocks of memory that had been filled with digital organisms. It would be very easy to experiment with differing frequencies and patch sizes of disturbance. This could provide a powerful means to examine the relationship between disturbance and diversity. This experiment would not require evolving organisms.

Heterotrophs and the Explosion of Diversity

Organic life has existed on earth for nearly 4 billion years. Until about 600 million years ago, life existed at a very low diversity, consisting only of single celled organisms. Then in the Cambrian, there was a sudden explosion of diversity. Macroscopic multicellular organisms appeared, both metazoans and metaphytes, and most of the major groups of modern organisms followed rapidly.

One theory to explain this explosive diversification (Stanley 1973) is that it was sparked by the appearance of the first organisms that ate other organisms. As long as all organisms were autotrophs, there was only room for a few species. In a community with only one trophic level, the most successful competitors would dominate. The process of competitive exclusion would keep diversity low.

However, when the first herbivore appeared it would have been selected to prefer the most common species of algae, thereby preventing any species of algae from dominating. This opens the way for more species of algae to coexist. Once the "heterotroph barrier" had been crossed, it would be simple for carnivores to arise, imposing a similar diversifying effect on herbivores. With more species of algae, herbivores may begin to specialize on different species of algae, enhancing diversification in herbivores. The process presumably was auto-catalytic, and set off an explosion of diversity.

In digital life, autotrophs will obtain CPU time and memory allocations directly from the simulator. Heterotrophs will attempt to seize control of the CPU from other individuals, such as by inserting a jump statement into the code of another individual. Heterotrophs may also attempt to seize memory resources from other individuals by freeing their memory or by directly altering memory allocation specifications for segments owned by other individuals. Heterotrophs may kill another process or seize its thread with a jump, and then grow into or produce an offspring in the memory space occupied by the dead individual.

Digital life may provide a powerful means of testing the theory that the origin of heterotrophs was responsible for the explosion of organic diversity. We may control the presence of heterotrophs in digital communities and determine if their presence generates a diversification that does not occur in exclusively autotrophic communities. If this experiment were positive, the newly diversified autotrophs could be introduced into an environment lacking in heterotrophs. This could determine if the presence of heterotrophs is necessary to maintain diversity as well as to generate it.

Species Area Effect

One of the most universal ecological laws is the species area relationship (MacArthur and Wilson 1967). It has been demonstrated that in a wide variety of contexts, the number of species occupying an "area" increases with the area. The number of species increases in proportion to the area raised to a power between 0.1 and 0.3. $S = KA^z$, where $0.1 < z < 0.3$. The effect is thought to result from equilibrium species number being determine by a balance between the arrival (by immigration or speciation) and local extinction of species. The likihood of extinction is greater in small areas because they support smaller populations, for which a fluctuation to a size of zero is more likely. If this effect holds for digital organisms it suggests that very large amounts of memory will be needed to generate significant diversification.

DISCUSSION

Through the process of designing digital organisms, or as a result of their adaptive evolution, it is likely that structures will arise that are clearly analogous to the physical forms of organic life. However, the designers of digital organisms should concentrate on creating algorithms that are well adapted to life in the computer environment, and should not be burdened at the outset by the baggage of arbitrary physical analogies. The critical step is to get the evolutionary process started, after that it should take care of itself, spontaneously generating the appropriate structures.

Ecological properties however, to some extent transcend the physical properties of organisms, and therefore may apply more broadly to life forms beyond the organic. Furthermore, ecological and evolutionary theory aimed at explaining patterns of organic diversity suggest factors which may be critical in the successful generation of artificial systems which spontaneously increase in diversity and complexity.

There exists a rich body of ecological and evolutionary theory that can be explored through artificial life. The few examples discussed here are not intended to be comprehensive, but simply represent some of my favorites. Some may argue that digital organisms would be inappropriate to test ecological and evolutionary theory based on organic life. This may turn out to be true. On the other hand, it is not clear that the analytical approach based on calculus and differential equations is any more appropriate. It has been suggested that the historical reliance of science on calculus may have been due not just to its merits, but also to the fact that before the computer, alternative languages of description were not practical (Toffoli 1984).

Properly constructed Artificial Life models can reflect the bottom up structure of hierarchically organized living systems, in which the behavior of the system at any level is an emergent property resulting from the interactions of a population of behavors of the next lower level (Langton 1989). Ecologists tend to treat organisms as black boxes, therefore Artificial Life models may turn out to be appropriate for the exploration of ecological processes as long as they preserve the proper relationships between individuals and species. AL models based on evolving digital organisms can be extended to the exploration of evolutionary theories as well.

The objective of this essay is to present an approach, not to elaborate the details. It is my hope that these ideas will stimulate exploration of the possible approaches to creating digital life and using it to examine evolutionary and ecological processes.

REFERENCES

Bell, Graham. 1982. The masterpiece of nature: the evolution and genetics of sexuality (University of California Press, Berkeley).

Barbieri, Marcello. 1985. The semantic theory of evolution. Harwood Academic Publishers, Chur, London, Paris, New York. Pp. 188.

Buss, Leo W. 1987. The evolution of individuality. Princeton University Press. Pp. 203.

Carroll, L. 1865. Through the Looking-Glass. London, MacMillan.

Charlesworth, B. 1976. "Recombination modification in a fluctuating environment", Genetics 83: 181–195.

Dewdney, A. K. 1984. "Computer recreations: In the game called Core War hostile programs engage in a battle of bits", Scientific American 250(5): 14–22.

Dewdney, A. K. 1987. "Computer recreations: A program called MICE nibbles its way to victory at the first core war tournament", Scientific American 256(1): 14–20.

Doolittle, W. Ford, and Carmen Sapienza. 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature 284: 601–603.

Gould, Steven J. 1990. New life.

Green, Melvin M. 1988. Mobile DNA elements and spontaneous gene mutation. *In* M. E. Lambert, J. F. McDonald, I. B. Weinstein [eds.]: Eukaryotic transposable elements as mutagenic agents. Pp. 41–50. Banbury Report 30, Cold Spring Harbor Laboratory.

Holland, John Henry. 1975. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence (Univ. of Michigan Press, Ann Arbor).

Huston, Michael. 1979. "A general hypothesis of species diversity", Am. Nat. 113: 81–101.

Jelinek, Warren R., and Carl W. Schmid. 1982. Repetitive sequences in eukaryotic DNA and their expression. Ann. Rev. Biochem. 51: 813–844.

Kerem, Bat-sheva, Johanna M. Rommens, Janet A. Buchanan, Danuta Markiewicz, Tara K. Cox, Aravinda Chakravarti, Manuel Buchwald, and Lap-Chee Tsui. 1989. Identification of the cystic fibrosis gene: genetic analysis. Science 245: 1073–1080.

Langton, C. G. 1986. Studying artificial life with cellular automata. Physica 22D: 120–149.

Langton, Christopher G. [ed.]. 1989. Artificial life: proceedings of an interdisciplinary workshop on the synthesis and simulation of living systems. Vol. 6 in the series: Santa Fe Institute studies in the sciences of complexity (Addison-Wesley).

Mac Arthur, Robert H., and Edward O. Wilson. 1967. The theory of island biogeography. Princeton University Press. Pp. 203.

Marx, Jean L. 1989. The cystic fibrosis gene is found. Science 245: 923–925.

Morris, S. Conway. 1989. Burgess shale faunas and the cambrian explosion. Science 246: 339–346.

Maynard Smith, J. 1971. "What use is sex?", J. Theoret. Biol. 30: 319–335.

Orgel, L. E., and F. H. C. Crick. 1980. Selfish DNA: the ultimate parasite. Nature 284: 604–607.

Rasmussen, Steen, Rasmus Feldberg, Morten Hindsholm, and Carsten Knudsen. Preprint. Core-evolution: emergence of cooperative structures in a computational chemistry.

Reidhaar-Olson, John F. and Robert T. Sauer. 1988. Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. Science 241: 53–57.

Riordan, John R., Johanna M. Rommens, Bat-sheva Kerem, Noa Alon, Richard Rozmahel, Zbyszko Grzelczak, Julian Zielenski, Si Lok, Natasa Plavsic, Jia-Ling Chou, Mitchell L. Drumm, Michael C. Lannuzzi, Francis S Collins, Lap-Chee Tsui. 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 245: 1066–1073.

Rommens, Johanna M., Michael C. Iannuzzi, Bat-sheva Kerem, Mitchell L. Drumm, Georg Melmer, Michhael Dean, Richard Rozmahel, Jeffery L. Cole, Dara Kennedy, Noriko Hidaka, Martha Zsiga, Manuel Buchwald, John R. Riordan, Lap-Chee Tsui, Francis S. Collins. 1989. Identification of the cystic fibrosis gene: chromosome walking and jumping. Science 245: 1059–1065.

Stanley, Steven M. 1973. "An ecological theory for the sudden origin of multicellular life in the late precambrian", Proc. Nat. Acad. Sci. 70: 1486–1489.

Strickberger, Monroe W. 1985. Genetics. Macmillan Publishing Co. New York.

Syvanen, Michael. 1984. The evolutionary implications of mobile genetic elements. Ann. Rev. Genet. 18: 271–293.

Thomas, C. A. 1971. The genetic organization of chromosomes. Ann. Rev. Genet. 5: 237–256.

Toffoli, Tommaso. 1984. "Cellular automata as an alternative to (rather than an approximation of) differential equations in modeling physics", Physica 10D: 117–127.

Van Valen, L. 1973. A new evolutionary law. Evolutionary Theory 1: 1–30.