Unnatural Selection in Chemical Systems†

LESLIE E. ORGEL

The Salk Institute for Biological Studies, P.O. Box 85800, San Diego, California 92186-5800

Received November 8, 1994

Unnatural Selection

The theory of evolution through natural selection was proposed by Darwin and Wallace to explain how the characteristics of populations of animals change with time. An examination of their assumptions shows that the theory has much broader application than they originally envisaged. We now know that in appropriate environments RNA molecules or computer viruses, for example, can evolve. The adventure with which we are concerned is the quest for chemical systems that undergo processes analogous to Darwinian selection in the test tube. The search is not restricted to systems that are closely related to nucleic acids, although most of the available experimental evidence concerns such systems.

The theory of natural selection is primarily concerned with the properties of populations of animals, not of individuals. It is implicitly assumed that the individuals are capable of reproduction. The possibility of evolution is dependent on variation, that is, on the existence of nonidentical individuals within the population. It also requires a mechanism of heredity, that, is a mechanism that ensures that an individual will be, on average, more similar to its parents than to a randomly chosen member of the population. If the above or analogous conditions are met, then successive generations of the population will adapt to a change in the environment in such a way as to maximize their ability to reproduce in the new environment.

A population of molecules satisfies all the requirements of the theory if there are different kinds of molecules in the population and if each individual molecule can direct the formation of copies of itself. Then a population of molecules will adapt to a varying environment by changing its composition so as to maintain as high as possible a rate of replication. Sol Spiegelman is the inventor of "unnatural selection". He showed clearly that populations of RNA molecules evolve when replicated repeatedly by $Q\beta$ RNA polymerase under a chosen set of adverse reaction conditions.[1] In the systems that he studied, the initial population was fairly homogeneous and much of the variation was created during the course of the experiment by mutation, that is, errors of replication.[2]

An example of "unnatural selection" analogous to examples of natural selection discussed by biologists is provided by the evolution of populations of RNA molecules replicated by $Q\beta$ RNA polymerase in the presence of ribonuclease A, an enzyme that cleaves single-stranded RNA selectively next to pyrimidine nucleotides.[3,4] In this case ribonuclease A can be considered as a predator. After 70 rounds of growth in the presence of the enzyme, unnatural selection leads to a population of replicating RNA molecules that is strongly resistant to cleavage. Sequence analysis shows that the two complementary RNA strands that make up the population are present in very different amounts. The abundant strand has eliminated almost all pyrimidine residues from single-stranded regions and so is resistant to degradation. The less abundant complementary strand is, of course, rich in pyrimidines and so would normally be highly vulnerable to cleavage. Apparently, it escapes at least in part by associating with the enzyme in such a way as to protect its vulnerable sites: the RNA hides on the back of the enzyme, out of reach of its "teeth"![4] In this, as in many other examples, unnatural selection has solved a problem in a way that could not be anticipated by the experimenter.

The term "unnatural selection" will be used loosely to describe evolution of nucleic acids or other replicatable polymers *in vitro*. The term "Natural Selection" will be reserved for the evolution of living organisms and their viruses. Natural Selection usually involves the coevolution of nucleic acids and proteins, while "unnatural selection", as practiced so far, allows replicating nucleic acids to evolve but holds constant the enzymes that catalyze replication.

It is widely believed that biology based on DNA, RNA, and proteins was preceded by the biology of an "RNA world" in which enzymes were composed of RNA alone.[5] The origin of RNA replication is thus the central puzzle of the origins of life. Consequently, RNA-catalyzed RNA replication is presently one of the main goals of experimental work on unnatural selection. However, there is also a more distant goal, namely, to achieve replication and selection in systems unrelated to RNA.

Unnatural Selection and the Search for "Masterpieces"

In a remarkably perceptive book *What is Life*? Schrödinger discussed the difference between periodic and aperiodic crystals.[6] I believe that he had in mind

Leslie Orgel was educated at Magdalen College, Oxford, where he obtained a D.Phil. degree in chemistry in 1951. Prior to his move to the U.S.A. he contributed to the development of ligand-field theory, a branch of inorganic chemistry. He now works on the chemistry of nucleic acids at The Salk Institute in La Jolla. For many years, much of his research has been concerned with chemistry that may be relevant to the origin of life. Dr. Orgel has published extensively on prebiotic chemistry. He is a fellow of the Royal Society and a member of the National Academy of Sciences.

† Part of the text of this article was originally published in *Nature*: Orgel, L. E. *Nature* **1992**, *358*, 203–209.

(1) Spiegelman, S.; Mills, D. R.; Kramer, F. R. In *Stability and origins of biological information*; Miller, I. R., Ed.; Halsted Press: New York, 1975.

(2) Kramer, F. R.; Mills, D. R.; Cole, P. E.; Nishihara, T.; Spiegelman, S. *J. Mol. Biol.* **1974**, *89*, 719–736.

(3) Strunk, G. Doctoral Thesis, University Braunschweig, Shaker, Aachen, FRG, 1993.

(4) Eigen, M. *Gene* **1993**, *135*, 37–47.

(5) *The RNA WORLD. The Nature of Modern RNA Suggests a Prebiotic RNA World*; Gesteland, R., Atkins, J. F., Ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1993.

0001-4842/95/0128-0109$09.00/0

repeating and random copolymers. He writes of periodic crystals:

> Yet compared with the aperiodic crystal, they are rather plain and dull. The difference of structure is of the same kind as that between an ordinary wallpaper in which the same pattern is repeated again and again in regular periodicity and a masterpiece of embroidery, ... (page 3)

This analogy is remarkably apt. A repeated copolymer, say the oligonucleotide sequence poly(A-U-C-G)$_{100}$, is interesting to physical organic chemists, but to others if is a rather dull molecule. The corresponding random copolymer poly(N)$_{400}$ (N = A, U, C, or G) is a random selection from an ensemble of 4^{400} sequences which includes everything in the RNA world and much more. Some molecules are clearly "masterpieces", for example, efficient and highly specific ribozyme catalysts. In the past, the only nonrepeating copolymers available for study as individual sequences were those that had been refined by natural selection, namely, DNA, RNA, and protein. This situation is changing rapidly.[7-9] Soon it should be possible to look for "masterpieces" among sequences present in such innocent-looking copolymers as poly(L-asp,L-glu) or even poly(L-asp,D-asp). It is not unlikely that ensembles as simple as these include molecules which, in the presence of appropriate metal ions, are highly specific catalysts in much the same way as ribozymes.

There are clearly severe obstacles to be overcome in the search for "masterpieces". Average properties of ensembles of different sequences are of little interest, while the properties of individual sequences are swamped out if measurements are made on unselected ensembles. It would be possible to synthesize defined sequences at random and study them one at a time, but this is a hopelessly inefficient way of finding "masterpieces" when their proportion is very small. Unnatural selection is the answer. One must invent protocols that pick out molecules with the desired character and replicate only those molecules.

The most unnatural approach to "unnatural selection" is to select the correct molecule in sufficient amounts to permit the determination of its sequence by traditional chemical methods. Then one can replicate the molecule by synthesizing much larger amounts of it. This approach is already being applied widely in the search for drugs.[7-9] A large library of short peptides, for example, is synthesized on beads of a polymer support in such a way that all the molecules on any bead have the same sequence. Beads carrying peptides with the desired property, for example, the ability to bind to a biological receptor, are readily selected and their sequences determined by microsequencing. Large amounts of the peptides are then obtained by solid-phase peptide synthesis. In principle this method could be applied to any linear copolymer for which methods of analysis and synthesis are available. Other ingenious protocols use replicatable and easily sequenced oligonucleotide tags to indicate the sequence of some other kind of copolymer which is co-localized with the oligonucleotide.[10,11] However, even in this approach replication of the selected copolymer molecule is achieved by an independent, large-scale chemical synthesis.

The full power of unnatural selection is revealed only in the context of a population of polymers that are able to replicate. There is then no need to carry out selection and a subsequent independent synthesis of the selected molecules. Instead, protocols are designed to select a subpopulation enriched in molecules having the required properties.[12,13] This selected subpopulation is then expanded to the size of the original population by exponential replication. The cycle can be repeated again and again, until only the most "adapted" molecules remain in the population. Provided monomeric substrates are available in adequate amounts, this whole process, once begun, is self-sustaining: it requires no intervention by the synthetic or analytical chemist. It is widely believed that this type of autonomous selection occurred on the primitive earth some 4 billion years ago and led, directly or indirectly, to the appearance of the RNA world.

Experimental Studies Relevant to Unnatural Selection

Studies Involving Protein Enzymes and Oligonucleotide Substrates. The only systems in which "unnatural selection" has been demonstrated *in vitro* involve protein enzymes as catalysts for replication. The earliest experiments were those of Sol Spiegelman and his colleagues, who used Qβ RNA polymerase to replicate short substrate RNAs.[1] The system has major drawbacks, since the substrates of RNA polymerase are a very restricted subset of all RNA sequences. Nonetheless, the system has yielded, and continues to yield, important insights into the power of unnatural selection.[4]

The full power of unnatural selection techniques only became apparent with the invention of methods of amplification of nucleotide sequences that are not restricted to special subclasses of RNA. The polymerase chain reaction (PCR) is one such technique. More important for unnatural selection are techniques that make use of reverse transcription to generate double-stranded DNA from RNA and then T7 RNA polymerase to produce multiple copies of the original RNA from the DNA.[12]

In one group of studies a ribozyme capable of cleaving RNA in the presence of Mg^{2+} is the starting point for selection. After a relatively small number of rounds of selection one can obtain ribozymes that will cleave an alternate substrate involving a phosphodiester bond to a deoxynucleotide in place of a ribonucleotide, or a ribozyme that will accept Ca^{2+} in place of Mg^{2+} as the activating metal ion.[14] Most impressively, a ribozyme that can cleave an unactivated amide bond can be selected in the same way.[15]

(6) Schrödinger, E. *What is Life?;* Cambridge University Press: Cambridge, 1944.
(7) Gallop, M. A.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gordon, E. M. *J. Med. Chem.* **1994**, *37*, 1233–1251.
(8) Gordon, E. M.; Barrett, R. W.; Dower, W. J.; Fodor, S. P. A.; Gallop, M. A. *J. Med. Chem.* **1994**, *37*, 1385–1401.
(9) Scott, J. K.; Craig, L. *Curr. Opin. Biotechnol.* **1994**, *5*, 40–48.
(10) Brenner, S.; Lerner, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5381–5383.
(11) Needels, M. C.; Jones, D. F.; Tate, E. H.; Heinkel, G. L.; Kochersperger, L. M.; Dower, W. J.; Barrett, R. W.; Gallop, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10700–10704.
(12) Joyce, G. F. *Curr. Opin. Struct. Biol.* **1994**, *4*, 331–336.
(13) Chapman, K. B.; Szostak, J. W. *Curr. Opin. Struct. Biol.* **1994**, *4*, 618–622.
(14) Lehman, N.; Joyce, G. F. *Nature* **1993**, *361*, 182–185.

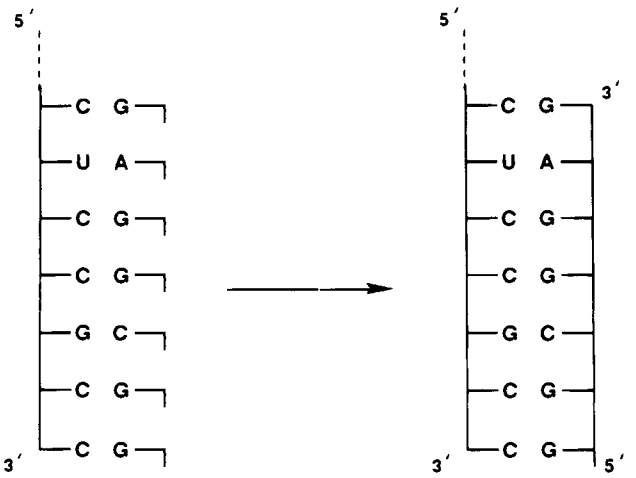


Figure 1. Template-directed synthesis. A preformed oligonucleotide chain forms a double-helical complex with complementary activated nucleoside 5′-phosphates. The geometry of the helix brings the 5′-phosphate of each monomer close to the 3′-hydroxyl of its neighbor, thus facilitating the formation of a phosphodiester bond. The outcome is the formation of an oligomer complementary to the template. This model is greatly oversimplified. The experimental results and their interpretation are reviewed in ref 19.

In another group of studies the starting point for selection is an ensemble of RNA molecules containing a randomized RNA sequence 50–200 bases long. It has been possible to select molecules that catalyze the template-directed formation of a phosphodiester bond from an oligonucleotide with a free 3′-terminus and a second oligonucleotide that carries a 5′-triphosphate group.[16] The rate of the catalyzed reaction exceeds that of the corresponding spontaneous reaction by a factor of 10^6–10^7. A similar technique has been used to "evolve" a kinase ribozyme that will transfer a phosphate group efficiently from ATP to the 5′-terminus of an RNA molecule.[17]

The catalytic potential of RNAs selected from random ensembles is not restricted to reactions involving oligonucleotides as substrates. Recently a ribozyme that catalyzes the isomerization of a bridged biphenyl has been described.[18]

Template-Directed Synthesis of Polynucleotides. Nonenzymatic replication of oligonucleotides using activated mononucleotide derivatives as substrates has not been achieved. However, substantial progress has been made in copying preformed oligonucleotide templates to form complementary oligonucleotide products (Figure 1). Work published before 1987 has been reviewed in detail.[19]

In describing how well a given template sequence is copied one must specify the efficiency, the fidelity, and the regiospecificity. The efficiency is measured by the proportion of template molecules that succeed in directing the synthesis of a complementary sequence; the fidelity is measured by the probability that a complementary base rather than a noncomplementary base will be inserted opposite a given base in the template; the regiospecificity is measured by the ratio of natural 3′–5′ linkages to unnatural 2′–5′ linkages. The first major conclusion is that the template-directed reactions of most activated nucleotides are inefficient and nonregiospecific. A considerable search was required to find a set of activated nucleotides, the 5′-phosphoro 2-methylimidazolides, that undergo efficient and highly regiospecific template-directed reactions.[20]

These reagents can be employed to copy faithfully many defined oligonucleotide sequences. These include sequences containing an excess of C residues and A, T, and G residues isolated from each other by at least three C residues. Runs of G residues are also copied into runs of C residues if the formation of self-structures by G residues can be avoided.[21–24]

The sequence 5′-CCCGCCCGCCCGCC-3′ is the longest oligomer that has been copied successfully at present. The yield of full-length products in this case never exceeded 2%, but the yield including truncated products can amount to 50% or more.[25]

It is unlikely that a pair of complementary sequences can be found each of which facilitates the synthesis of the other using nucleoside 5′-phosphoro 2-methylimidazolides as substrates. The major obstacles are the formation of intermolecular complexes involving a tetrahelix formed by sequences of consecutive G residues[26] and the formation of intramolecular complexes in which the molecules bend back on themselves and form Watson–Crick double-helical segments. Some of the obstacles may be attributable to the choice of reagents and reaction conditions, but others seem to be intrinsic to oligonucleotide replication.

Nonenzymatic Replication. The first successful demonstration of molecular replication, reported by G. von Kiedrowski,[27] was strictly modeled on nucleic acid replication (Figure 2). The two trinucleotide analogues **I** and **II** and the hexanucleotide analogue **III** were prepared by standard synthetic methods. The product formed by linking **I** and **II** via a phosphodiester bond is **III** (Figure 2). The sequences were chosen because they form the ternary complex **T** in which the 3′-phosphate of **I** is brought close to the 5′-hydroxyl of **II** (Figure 2). The addition of a water-soluble carbodiimide to the solution, therefore, efficiently joins together **I** and **II** to form **III** and converts the ternary complex **T** to the binary complex **B**. Thus a molecule of **III** catalyzes the production of another molecule of **III** from **I** and **II**.

This result and related results using different oligonucleotide-related substrates that replicate more efficiently[28–33] have demonstrated the possibility of a

(15) Dai, X.; De Mesmaeker, A.; Joyce, G. F. *Science*, in press.
(16) Bartel, D. P.; Szostak, J. W. *Science* **1993**, *261*, 1411–1418.
(17) Lorsch, J. R.; Szostak, J. W. *Nature* **1994**, *371*, 31–36.
(18) Prudent, J. R.; Uno, T.; Schultz, P. G. *Science* **1994**, *264*, 1924–1927.
(19) Joyce, G. F. In *Cold Spring Harbor Symposia on Quantitative Biology*; Cold Spring Harbor Press: Cold Spring Harbor, NY, 1987; Vol. LII, pp 41–51.
(20) Inoue, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1981**, *103*, 7666–7667.
(21) Wu, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1992**, *114*, 318–322.
(22) Wu, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1992**, *114*, 5496–5501.
(23) Wu, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1992**, *114*, 7963–7969.
(24) Wu, T.; Orgel, L. E. *J. Mol. Evol.* **1992**, *32*, 274–277.
(25) Acevedo, O. L.; Orgel, L. E. *J. Mol. Biol.* **1987**, *197*, 187–193.
(26) Kang, C.; Zhang, X.; Ratliff, R.; Moyzis, R.; Rich, A. *Nature* **1992**, *356*, 126–131.
(27) von Kiedrowski, G. *Angew. Chem., Int. Ed. Engl.* **1986**, *25*, 932–935.
(28) von Kiedrowski, G.; Wlotzka, B.; Helbing, J. *Angew. Chem., Int. Ed. Engl.* **1989**, *28*, 1235–1237.
(29) von Kiedrowski, G. In *40 Jahre Fonds der Chemischen Industrie, 1950–1990*; V. d. C. I. e. V., Ed.; VCI Publications: Frankfurt, Germany, 1990; pp 197–218.
(30) von Kiedrowski, G.; Wlotzka, B.; Helbing, J.; Matzen, M.; Jordan, S. *Angew. Chem., Int. Ed. Engl.* **1991**, *30*, 423–426.

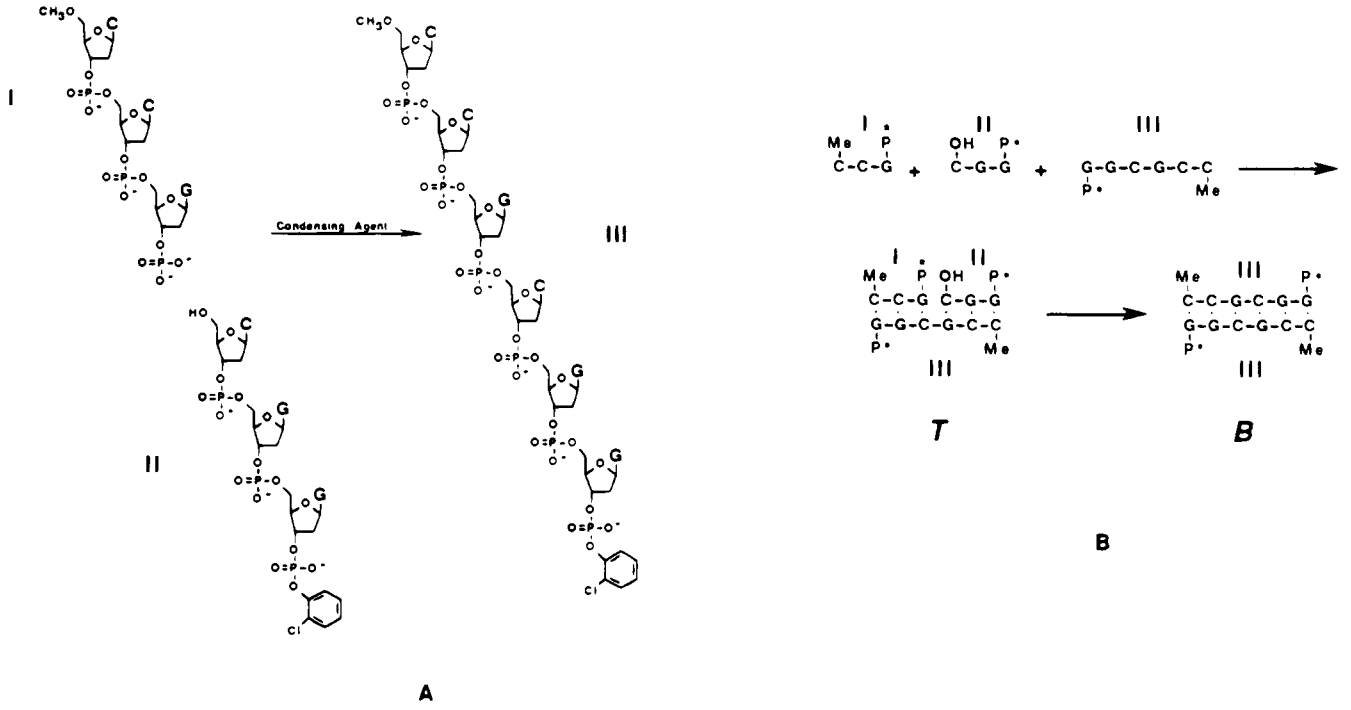


Figure 2. Replication of a hexanucleotide. (A) The fundamental reaction is the ligation of the two trinucleotide analogues **I** and **II** to form the hexanucleotide analogue **III**. (B) The catalytic mechanism: a Watson–Crick double-helical ternary complex **T** is formed from one molecule each of **I**, **II**, and **III**. Formation of a phosphodiester bond between adjacent 3′-phosphate and 5′-hydroxyl group converts **T** to the binary complex **B** made up from two hexamers. The binary complex **B** must dissociate into two molecules of **III** to initiate the next round of replication (see text). Adapted from ref 27.

limited degree of informational replication. They strongly imply that informational replication analogous to the enzymatic replication of nucleic acids is possible. First, it seems unlikely from the work of von Kiedrowski that the replication of **III** would be inhibited by the presence of oligonucleotides with sequences unrelated to **I**, **II**, and **III**. Second, a chain of nine or 12 residues would almost certainly form a double-helical complex analogous to **B** but with three or four trinucleotides. The trinucleotides could then be ligated with an appropriate condensing agent. Thus replication of such polymers is at least likely. Ribozyme-catalyzed copying of a polynucleotide using trinucleotides as substrates has already been achieved.[34]

The initial experiments using trinucleotides[27] revealed a major obstacle to replication. The increase in the rate of synthesis of **III** due to autocatalysis is proportional to the square root of the concentration of **III** present, rather than to the concentration of **III**. The reason for this is obvious. Most of the molecules of **III** present in the solution are tied up in unbroken double helices (Figure 2), but only isolated molecules can act as templates. The number of isolated molecules increases roughly as the square root of the total concentration. Oligomers containing more than six residues would form much more stable double helices than **B**, so autocatalytic synthesis would be expected to slow down before substantial concentrations of replication products had formed. To overcome this difficulty it will almost surely be necessary to cycle the temperature of the reaction mixture, or to cycle some other variable that controls the stability of double-helical complexes relative to single strands.

Julius Rebek and his co-workers have approached the problem of self-replication in a novel way.[35–40] Instead of working with molecules related to nucleotides in aqueous solution, they have synthesized relatively complicated organic substrates that replicate in nonaqueous solvents. The basis of their experiments[37] can be understood by reference to Figure 3. The subunits **1** and **2** contain an activated carboxyl group and a primary amine, respectively. When **1** and **2** react together, an amide bond is generated and the molecule **3** is formed. The triad of molecules **1**, **2**, and **3** have been designed with great ingenuity, for they form a hydrogen-bonded ternary complex **T** in which the carboxyl group of **1** and the amine function of **2** are held in close proximity. Thus **3**, the product of the reaction between **1** and **2**, facilitates the synthesis of further molecules of **3** from **1** and **2**. This is exactly what is required for self-replication.

The kinetics of the autocatalytic synthesis of **3** from **1** and **2** demonstrates a roughly square-root dependence on the concentration of **3**.[37] As in the case discussed by von Kiedrowski, this is the result of the

(31) von Kiedrowski, G.; Helbing, J.; Wlotzka, B.; Jordan, S.; Matzen, M.; Achilles, T.; Sievers, D.; Terfort, A.; Kahrs, B. C.; Ullrich, B. *Nachr. Chem., Tech. Lab.* **1992**, *40*, 578.
(32) Achilles, T.; von Kiedrowski, G. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 1198–1201.
(33) Sievers, D.; von Kiedrowski, G. *Nature* **1994**, *369*, 221–224.
(34) Doudna, J. A.; Couture, S.; Szostak, J. W. *Science* **1991**, *251*, 1605–1610.
(35) Tjivikua, T.; Ballester, P.; Rebek, J., Jr. *J. Am. Chem. Soc.* **1990**, *112*, 1249–1250.
(36) Rotello, V.; Hong, J.-I.; Rebek, J., Jr. *J. Am. Chem. Soc.* **1991**, *113*, 9422–9423.
(37) Nowick, J. S.; Feng, Q.; Tjivikua, T.; Ballester, P.; Rebek, J., Jr. *J. Am. Chem. Soc.* **1991**, *113*, 8831–8839.
(38) Rebek, J., Jr. *Chem. Ind. (London)* **1992**, March 2, 1992, 171–174.
(39) Hong, J.-I.; Feng, Q.; Rotello, V.; Rebek, J., Jr. *Science* **1992**, *255*, 848–850.
(40) Feng, Q.; Park, T. K.; Rebek, J., Jr. *Science* **1992**, *256*, 1174–1180.

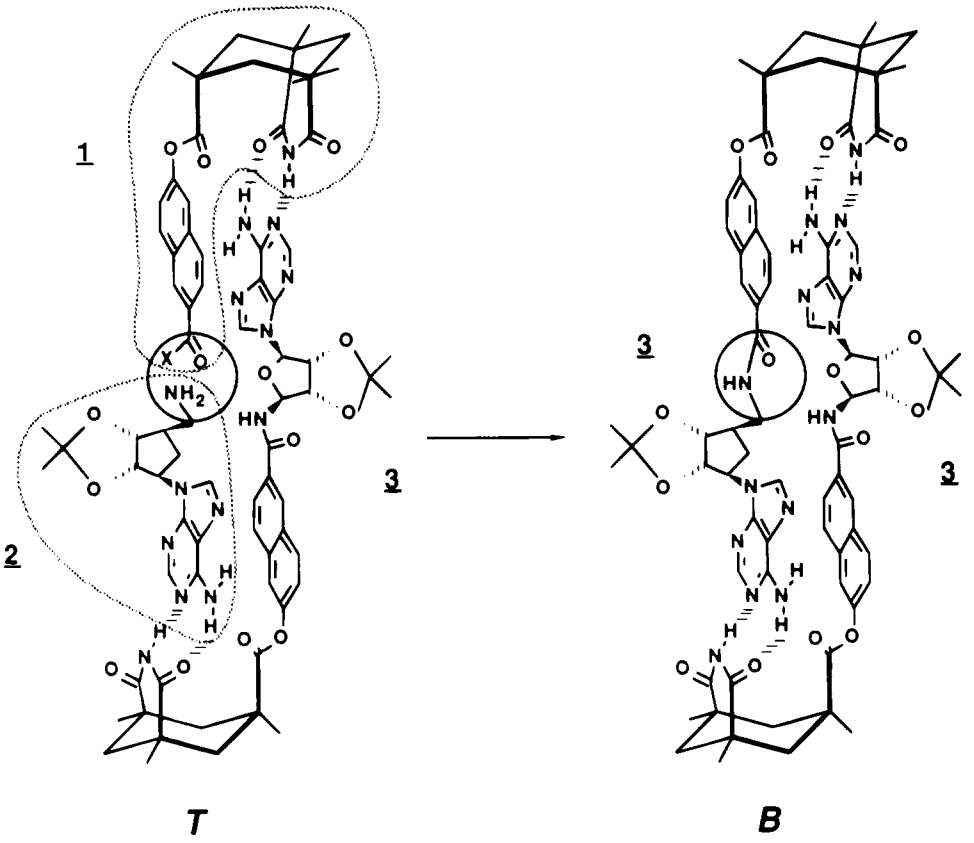


Figure 3. Replication of novel hydrogen-bonding substrates. The trio of molecules **1**, **2**, and **3** have been designed ingeniously to form the hydrogen-bonded ternary complex **T**, in which an activated carboxylic acid is brought close to an amine group. (The relevant groups are enclosed in a circle.) When an amide bond is formed, **T** is converted to the binary complex **B**, a dimer of **3**. Separation of **B** into two molecules of **3** makes possible another round of replication.

stability of dimer **B** formed by association of two molecules of **3**.

Rebek has extended his studies to other kinds of molecules that can form the basis for a replicating system and is beginning to study more complicated systems in which more than one substrate of replication can be formed in the reaction mixture.[38,39]

The experiments described by Rebek and his coworkers do not address the problem of informational replicating systems directly. However, like those of von Kiedrowski,[27] they could probably be generalized by arranging for several different substrate molecules to polymerize on a preformed linear template.

Terfort and von Kiedrowski have developed another self-replicating system which involves relatively simple organic molecules, derivatives of (2-formylphenoxy)-acetic acid **4** and 3-aminobenzamidine **5**, which condense autocatalytically in dimethyl sulfoxide to form anils **6** (Figure 4).[41] The initial experiments suggest that the rate of the autocatalytic reaction, under some conditions, depends linearly on the concentration of **6**. This is in sharp contrast to previously described replication reactions which display a square-root dependence on catalyst concentration. Presumably the ternary complex **T** is sometimes more stable than the binary complex **B**. This new system supports the idea that relatively simple molecules unrelated to nucleotides may provide the basis for an exponentially replicating system.

Li and Nicolaou have described a novel replication scheme that depends on the ability of palindromic oligopurine sequences to form stable triple helices with the complementary oligopyrimidine sequence at pH 6 but only the standard Watson–Crick double helices at pH 7.[42] The replication cycle (Figure 5) begins with a Watson–Crick double helix and two half-length pyrimidine strands at pH 6. Ligation using *N*-cyanoimidazole as condensing agent leads to the production of a stable triple helix. The pH is then adjusted to pH 7, causing the newly-formed triple-stranded product to dissociate to a double-stranded Watson–Crick helix and a free pyrimidine strand. The pyrimidine strand at pH 7 can then act as a template to generate a new polypurine strand from two half-length purine strands. Readjustment to pH 6 completes a cycle in which the original double helix has been replicated at the expense of two pyrimidine half-strands and two purine half-strands. At present this method is restricted to palindromic oligopurine and oligopyrimidine strands. It may be possible to overcome the first restriction, but the method is not obviously applicable if purines and pyrimidines are present in the same strand.

The Quest

We have seen, on the one hand, that template-directed copying of some but not all preformed informational oligonucleotides has been achieved in a very simple system. The only essential components are the oligonucleotide templates, the 5′-phosphoro 2-methylimidazolide substrates, and sufficient Na^+ and Mg^{2+} ions to stabilize the double-helical complexes. Unfor-

(41) Terfort, A.; von Kiedrowski, G. *Angew. Chem., Int. Ed. Engl.* **1992**, *31*, 654–656.

(42) Li, T.; Nicolaou, K. C. *Nature* **1994**, *369*, 218–221.

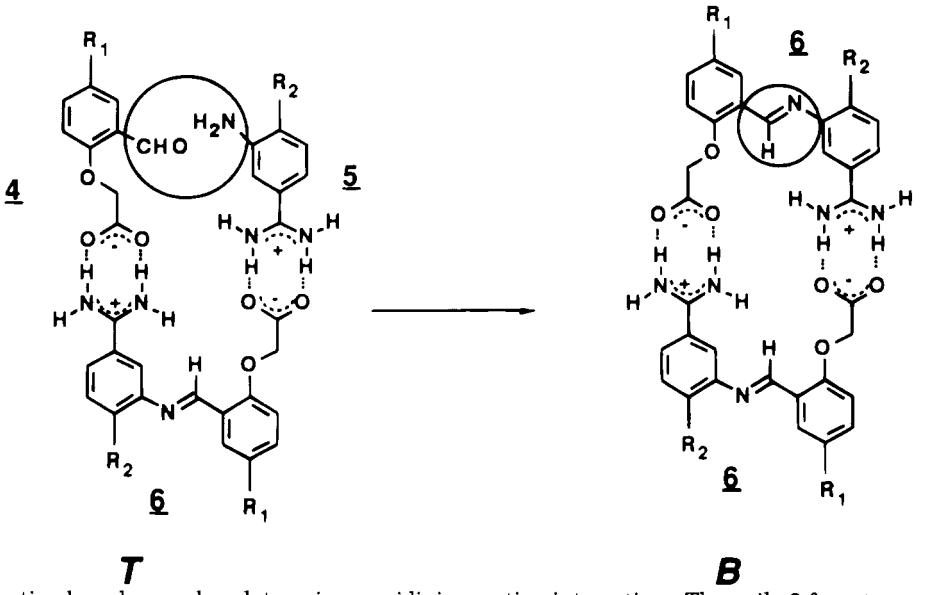


Figure 4. Self-replication based on carboxylate anion–amidinium cation interaction. The anils **6** form ternary complexes **T** with the carboxylate-containing ions **4** and the amidinium-containing ions **5** via hydrogen bonding and electrostatic interaction. Formation of a Schiff's base generates another molecule of **6** from **4** and **5** to give the binary complex **B**. The dissociation of **B** to two molecules of **6** begins a new round of replication.

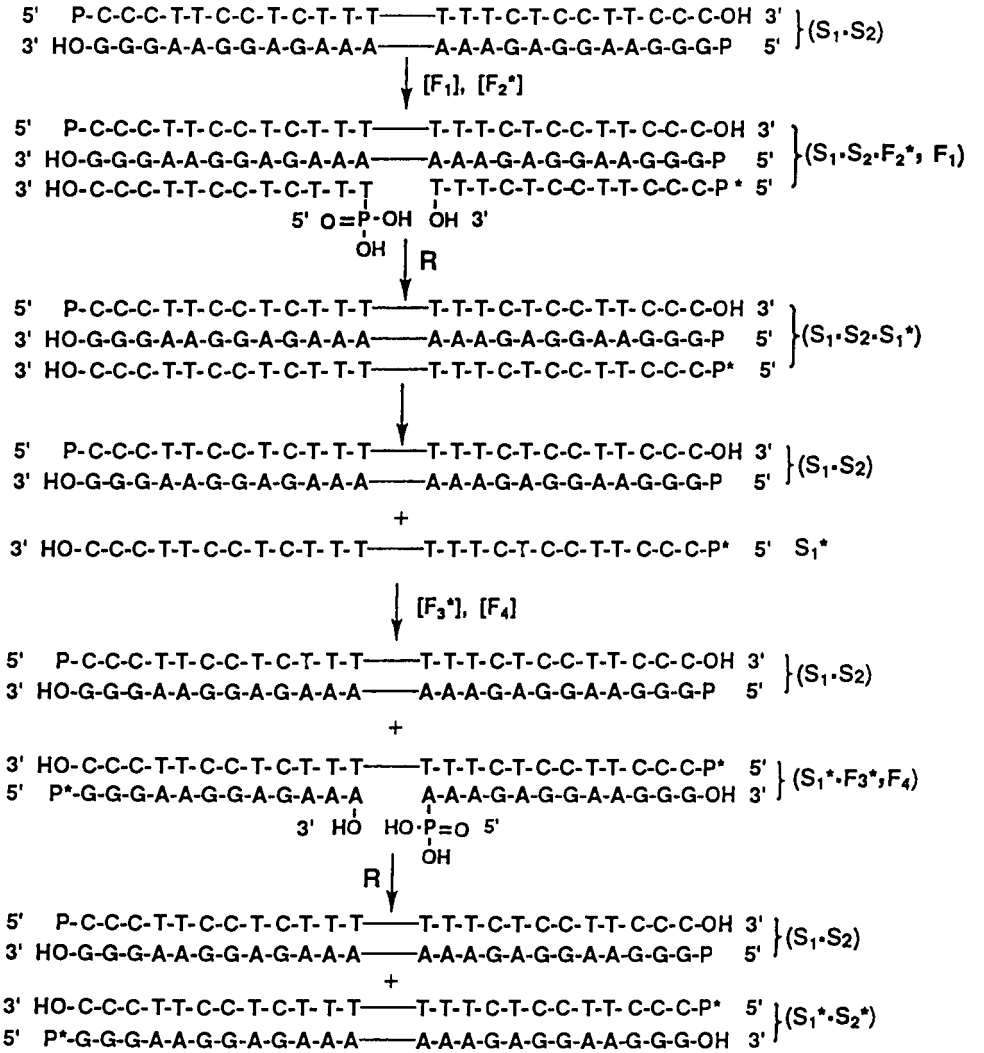


Figure 5. Replication of duplex DNA via sequential triple- and double-helix formation. Sequences of oligonucleotides used: F_1 = 5′ P-T-T-T-C-T-C-C-T-T-C-C-C-OH 3′; F_2 = 5′ P-C-C-C-T-T-C-C-T-C-T-T-T-OH$_3$ 3′,F_3 = 5′ P-G-G-G-A-A-G-G-A-G-A-A-A-OH 3′; F_4 = P-A-A-A-G-A-G-G-A-A-G-G-G-OH 3′. R = *N*-cyanoimidazole. For explanation, see text.

tunately, replication has not been achieved in this system; the complementary product formed on an effective template is never itself an effective template. Replication, on the other hand, has been achieved in

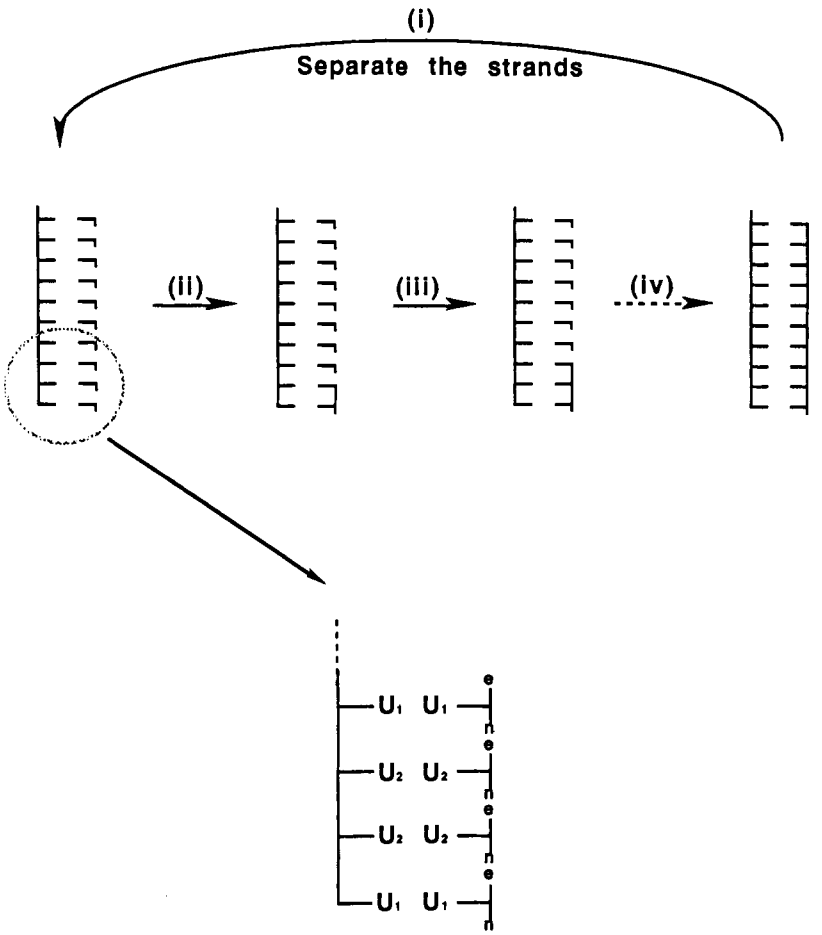


Figure 6. Idealized scheme for informational replication via a double-stranded intermediate. (a) (i) The double-stranded product formed in the previous round of replication dissociates into single strands, and monomers line up on them; (ii) a dimer is formed at the end of the template; (iii) this dimer acts as a primer for addition of another monomer; and (iv) the process continues until the template is filled. (b) The lower section of the figure represents a generalization of polynucleotide replication. The monomers are assumed to be bifunctional, involving nucleophilic (n) and electrophilic (e) groups that can react with each other to form a covalent bond. In the diagram U_1 pairs with itself and similarly for U_2.

a number of simpler chemical systems, but then information transfer is limited. The Holy Grail is the development of nonenzymatic reactions in which replication and information transfer occur together without the help of protein enzymes. Such systems would be capable of initiating *de novo* evolution by Darwinian selection and would open up a new field of chemistry, the nonequilibrium equivalent of the chemistry of self-organization.

The goal of a protein-free replicating system is closest to being achieved for RNA. The selection of a ribozyme that catalyzes the template-directed ligation of oligonucleotides using a 5′-triphosphate group as the source of energy is an important advance.[16] If the 5′-triphosphate-terminated oligonucleotide could be replaced by a mononucleoside 5′-triphosphate, copying of an arbitrary RNA strand using nucleoside triphosphates as substrates would be possible. This would be a major step toward an autonomous replicating system.

Progress toward template-directed synthesis and replication in systems related to nucleic acids, but with modified backbones, can also be expected in the near future. Nielsen and his co-workers have synthesized nucleic acid analogues (PNAs) in which the normal phosphodiester backbone is replaced by a backbone joined together by amide bonds. These oligomers bind very tightly to complementary RNA or DNA sequences to form complexes which are believed to have structures closely similar to those of double-stranded nucleic acids.[43] It is already known that pairs of PNA oligomers having complementary sequences form stable helical complexes,[44] so it would be surprising if much of the template-directed chemistry achieved with RNA could not be repeated with PNA substrates. We have recently shown that PNA template C_{10} will facilitate the extension of an oligonucleotide primer G_5 using 2-MeImpG as substrate.[45] It seems probable that further template-directed chemistry matching that of the oligonucleotides will be possible in this and other systems involving nucleotide analogues.

Studies of this kind are of particular interest in the context of the origins of life. Many chemists find it hard to believe that activated β-ribofuranoside 5′-phosphates could have formed on the primitive earth in sufficient quantity and purity to permit the beginning of RNA replication. However, the ease with which the purines and to a lesser extent the pyrimidines are formed prebiotically[46] suggests that they are likely to have been essential components of a very

(43) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. *Science* **1991**, *254*, 1497–1500.
(44) Wittung, P.; Nielsen, P. E.; Buchardt, O.; Egholm, M.; Nordén, B. *Nature* **1994**, *368*, 561–563.
(45) Böhler, C.; Orgel, L. E. Unpublished.
(46) Miller, S. L.; Orgel, L. E. *The Origins of Life on the Earth*; Prentice-Hall, Inc.: Englewood Cliffs, NJ, 1974.

early genetic system. These considerations have led to the suggestion that the very first genetic system involved base-pairing interaction, probably of the Watson–Crick type, in polymers with a backbone different from that of RNA. The challenge is to find purine- and pyrimidine-containing monomers that are much more easily synthesized under prebiotic conditions than ribonucleotides, but can still form a double-stranded structure stabilized by specific hydrogen bonds. The studies of Eschenmoser on oligonucleotides with alternative sugars[47] and Nielsen on polyamide nucleic acids[44] are intriguing, but in neither case has a simple prebiotic synthesis of a candidate monomer been demonstrated.

There is no obvious reason why accurate template-directed synthesis and consequent exponential replication should be restricted to systems involving nucleotide bases. The general requirements are illustrated in Figure 6, which is a straightforward generalization of polynucleotide replication.

A preformed chain made up by polymerizing the units U_1, U_2, etc. must be able to interact with these same monomeric subunits according to some well-defined pairing rule. The geometry of the complex formed by this interaction must facilitate the joining together of the subunits in a sequence that corresponds to the sequence of the preformed chain. It must then be possible to separate the two strands without breaking the backbone of either strand. It is important to note that the last requirement is kinetic rather than thermodynamic. The interstrand interactions need not be weaker than the intrastrand interactions holding the backbone together, but there must be some way of dissociating the interstrand bonds without breaking the intrastrand bonds.

Apart from the requirements described above, few of the features of RNA replication are essential for a general replication model. Clearly, there could be fewer or more than four different subunits, pairing might occur between identical or complementary subunits, each substrate molecule might contain several monomeric units (trinucleotides might be used in an RNA system, for example), chains could run in parallel or antiparallel directions, and replication might occur in a homogeneous aqueous or nonaqueous solution or on a surface. The interaction between chains could be mediated by hydrogen bonding, electrostatic interaction, hydrophobic forces, covalent bonds, or coordination of functional groups on the two chains to common metal ions. Chain growth could occur in either direction, or possibly in both directions from an internal initiation site.

The task for the organic chemist is to design two or more informational subunits that can be joined into indefinitely long chains such that a preformed chain acts as a template to facilitate the synthesis of its "complement" from monomers. There is no restriction on the nature of the subunits, the chemical functionalities involved in the polymerization, the nature of the interaction between chains, or the solvent system. Once such a system is available it should be possible to begin to design experiments on unnatural selection.

(47) Eschenmoser, A. *Origins Life* **1994**, *24*, 389–423.

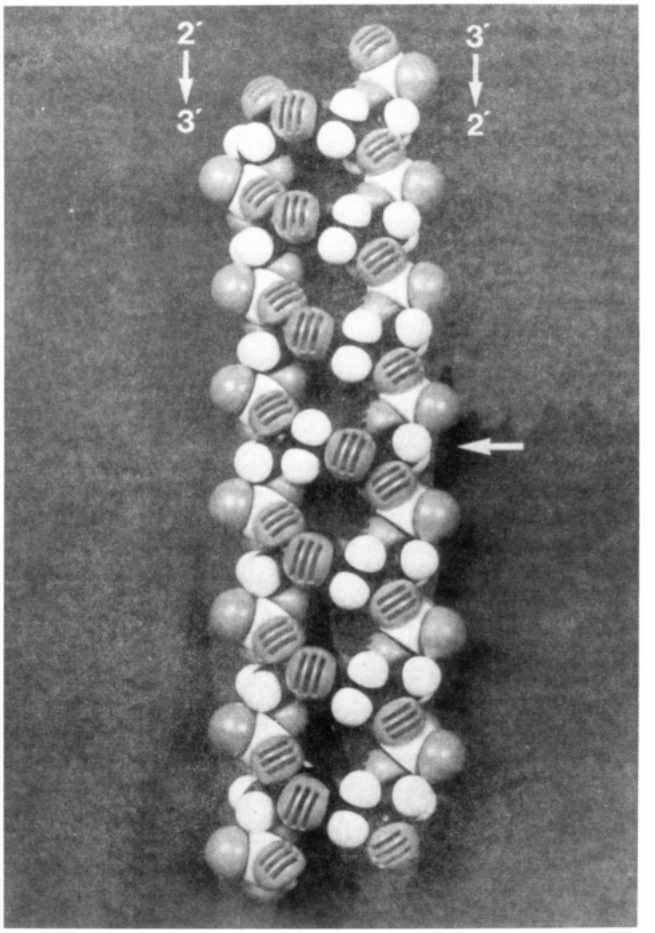


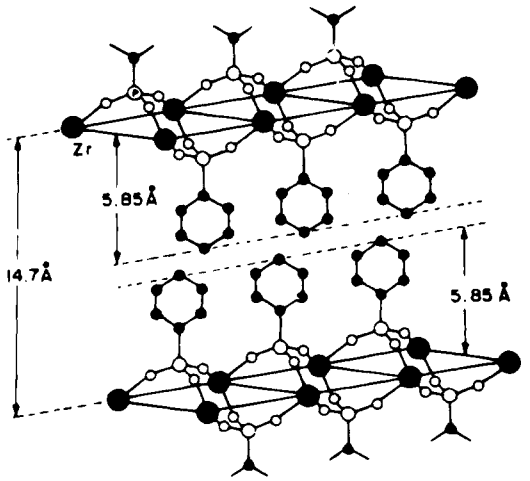
Figure 7. Antiparallel duplex of complementary copolymers of D-glycerate 3-phosphate and D-glycerol 3-phosphate having a phosphodiester backbone. Along the duplex, glycerate-3-P residues in the left strand are joined in an ester bond to glycerol-3-P residues in the right strand, except at the position marked by the arrow, where the position of the residues is reversed and a glycerate-3-P residue in the right is joined to a glycerol-3-P residue in the left strand.

Speculations Mostly Concerning Replication and the Origins of Life

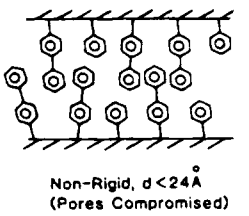
The progress and outcome of any respectable quest cannot be predicted in detail—quests are adventures—only the ultimate success of the enterprise is certain. The following speculations, some of my own and some published by other authors, are relevant mainly to studies of the origins of life.

It is clearly difficult to design polymers that associate specifically by hydrogen bonding in aqueous solution. It has therefore been suggested by Weber[48] that, in the earliest genetic systems, the two chains of a complementary structure were held together by covalent bonds. In particular he supposed that readily hydrolyzable ester bonds are used to hold together two complementary chains while more difficultly hydrolyzable phosphodiester bonds join the monomers to form the backbone. A copolymer of glycerol 3-phosphate and glyceric acid 3-phosphate[49] is a very simple example of a two-chain structure of this type (Figure 7). Weber also suggests an ingenious solution to many of the problems associated with optically active sub-

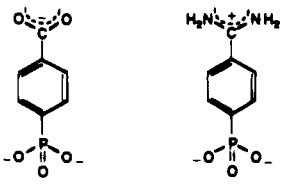
(48) Weber, A. L. *Origins Life* **1989**, *19*, 179–186.
(49) Weber, A. L. *Origins Life* **1987**, *17*, 107–119.



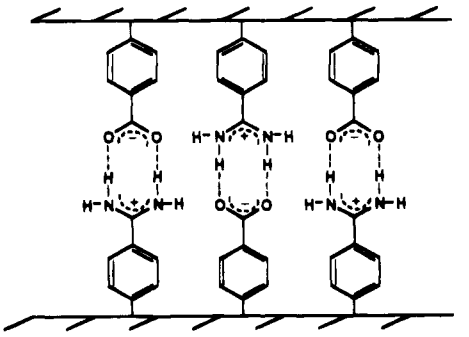
A



B



C



D

Figure 8. (A) The structure of the Zr^{4+} salt of phenyl phosphonate from ref 58. (B) The suggested structure of the Zr^{4+} salt formed with a mixture of phenyl phosphonate and biphenyl phosphonates from ref 60. (C) "Complementary" phosphonates that might form an informational two-dimensional system. (D) Hypothetical informational system based on the phosphonates in C above.

strates:[50] If you can't beat them, join them! More precisely, he suggests that the D- and L-enantiomers of a substrate should be regarded as two different informational subunits. Thus the sequence D-glyceric acid–D-glyceric acid–L-glyceric acid would be an informational trimer carrying 3 bits of information.

It has recently been demonstrated that the catalytic efficiency of some enzymes depends on the positioning of coordinated metal ions: no functional groups of the macromolecules are directly involved in the chemical transformation.[51] A similar function for the essential Mg^{2+} ions associated with ribozymes has been proposed.[52] Loosely speaking, the enzyme or ribozyme acts as a scaffolding, positioning ions such as Mg^{2+} or Zn^{2+} accurately in space. This suggests that metal ions might function similarly in replication. The Mg^{2+} ion, for example, could coordinate to the preformed template strand of a polyanionic linear copolymer and then stabilize its interaction with a growing "complementary" strand.[53] This mechanism is attractive because metal ion coordination to simple anions is much more labile than covalent interaction, but could provide a degree of thermodynamic stabilization in aqueous solution substantially greater than can be achieved easily by hydrogen bonding.

While polymerization reactions are usually carried out in homogeneous solution, the formation and replication of polymers on the primitive earth is more likely to have occurred on the surface of a mineral, or possibly at the interface between a micelle and water. The role of minerals has, therefore, attracted a great deal of attention, but the experimental results are somewhat sparse. Ferris has shown that certain clay minerals adsorb activated nucleotides and catalyze their regiospecific oligomerization to give short 3′–5′-

(50) Weber, A. L. *J. Mol. Evol.* **1987**, *25*, 191–196.
(51) Beese, L. S.; Steitz, T. A. *EMBO J.* **1991**, *10*, 25–33.
(52) Freemont, P. S.; Friedman, J. M.; Beese, L. S.; Sanderson, M. R.; Steitz, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 8924–8928.
(53) Joyce, G. F.; Orgel, L. E. In *The RNA World*; Gesteland, R. F., Atkins, J. F., Ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1993.

linked products.[54] We have found that oligonucleotides adsorbed on a few minerals such as hydroxyapatite and inorganic solids such as ferric hydroxide are fully competent as templates.[55] We have not observed catalysis of oligomerization by the minerals.

Finally, one should mention the speculations of Cairns-Smith,[56] who proposes that the most primitive genetic system was inorganic: a clay in which the two-dimensional clay layers are disordered, and the disorder is propagated from a preformed layer to a newly precipitated layer. While no convincing suggestions have been made concerning the nature of the replication mechanisms, two-dimensional replication systems clearly fall within the range of the quest.

A suggestion for such a two-dimensional system is indicated in Figure 8D.[57] Phosphonate complexes of zirconium and other tetravalent metal ions have layer structures of the type indicated in Figure 8A. Interaction between the phosphonate ligands determines the packing of the layers.[58,59] The structure suggested for the disordered crystal formed by Zr^{4+} ions with a mixture of phenyl phosphonate and diphenyl phosphonate is potentially informational (Figure 8B).[60] It might be possible to build accurately replicating structures (Figure 8D) based on phosphonates such as those in Figure 8C and a tetravalent ion such as Ti^{4+}, Zr^{4+}, or Hf^{4+}.

Unnatural Selection: A New Paradigm

One prototypical task of a traditional synthetic organic chemist is to obtain a target molecule in good yield and high purity. The acclaim with which work of this kind is received depends on the difficulty of the problem and on the skill with which novel procedures are invented and combined with established methods to achieve the final end. This attitude carries over into much polymer research, where the synthesis of stereoregular products is often the goal.

The prototypical task of a molecular biologist working in the relatively new field of *in vitro* selection is diametrically opposed to that of the organic chemist. The objective is now to obtain a polymeric molecule with a desired "phenotype", for example, an efficient and specific catalyst for a reaction that is slow and nonspecific in the absence of the catalyst. This is best achieved by selecting from as large and diverse a population of sequences as possible. The importance of the work is judged by the interest of the phenotype and the ingenuity of the procedures that are involved in selection. The structure of the selected molecules and the structural basis of their phenotypes are often of great interest, but selection proceeds without any *a priori* knowledge of them.

For the organic chemist, the Holy Grail of the field is an understanding of "unnatural selection" as a nonmysterious emergent property of populations of replicating macromolecules, and the concomitant ability to design novel chemical systems, unrelated to nucleic acids, that evolve. What sorts of polymers or solids can replicate and with what error rates? How rapidly can a population evolve? How does the nature of the monomers determine the range of properties, catalytic or otherwise, of the selected polymers? The quest for answers to these and related questions will require a combination of the techniques of organic chemistry, which will be essential for the design of replicating polymers and the synthesis of their monomers, and the methods of molecular biology, which will be equally essential for the selection of efficient catalysts etc.

Traditionally, the quest for the Holy Grail can only be accomplished successfully by a completely pure knight. Never mind; there will be plenty of spoils for those with interests in technology, once the quest is completed. Improved understanding of the origins of life on the primitive earth is likely to be another reward.

This work was supported by NSCORT/EXOBIOLOGY Grant No. NAGW-2881 from the National Aeronautics and Space Administration. Parts of the section entitled "Experimental Studies Relevant to Unnatural Selection" are based on a recent review in Nature, and Figures 1–6 are reprinted with permission from Nature (Figures 1–4 and 6: Orgel, L. E. ***1992****, 358, 203–209. Figure 5: Li, T.; Nicolaou, K. C.* ***1994****, 369, 218–221). Figure 7 is reprinted by permission of Kluwer Academic Publishers (Weber, A. L.* ***1989****, 19, 179–186). I thank Sylvia Bailey for manuscript preparation.*

AR940076Y

(54) Ferris, J. P.; Ertem, G. *Origins Life* **1992**, *22*, 369–381.
(55) Schwartz, A. W.; Orgel, L. E. *J. Mol. Evol.* **1985**, *21*, 299–300.
(56) Cairns-Smith, A. G. *Genetic takeover and the mineral origins of life*; Cambridge University Press: Cambridge, 1982.
(57) Orgel, L. E. In *Evolutionary Tinkering in Gene Expression*; M. Grunberg-Manago, M., Clark, B. F. C., Zachau, H. G., Ed.; Plenum Publishing Corporation: New York, 1989; pp 215–224.
(58) Alberti, G.; Constantino, U.; Allulli, S.; Tomassini, N. *J. Inorg. Nucl. Chem.* **1978**, *40*, 1113.
(59) Cao, G.; Hong, H.-G.; Mallouk, T. E. *Acc. Chem. Res.* **1992**, *25*, 420–427.
(60) Dines, M. B.; Griffith, P. C. *Inorg. Chem.* **1983**, *22*, 567.