

# The origin of life – a review of facts and speculations

Leslie E. Orgel

Three popular hypotheses attempt to explain the origin of prebiotic molecules: synthesis in a reducing atmosphere, input in meteorites and synthesis on metal sulfides in deep-sea vents. It is not possible to decide which is correct. It is also unclear whether the RNA world was the first biological world or whether some simpler world preceded it.

**THE PROBLEM OF** the origin of life on the earth has much in common with a well-constructed detective story. There is no shortage of clues pointing to the way in which the crime, the contamination of the pristine environment of the early earth, was committed. On the contrary, there are far too many clues and far too many suspects. It would be hard to find two investigators who agree on even the broad outline of the events that occurred so long ago and made possible the subsequent evolution of life in all its variety. Here, I outline two of the main questions and some of the conflicting evidence that has been used in attempts to answer them. First, however, I summarize the few areas where there is fairly general agreement.

The earth is slightly more than 4.5 billion years old. For the first half billion years or so after its formation, it was impacted by objects large enough to evaporate the oceans and sterilize the surface<sup>1,2</sup>. Well-preserved microfossils of organisms that have morphologies similar to those of modern blue-green algae, and date back about 3.5 billion years, have been found<sup>3</sup>, and indirect but persuasive evidence supports the proposal that life was present 3.8 billion years ago<sup>4</sup>. Life, therefore, originated on or was transported to the earth at some point within a window of a few hundred million years that opened about four billion years ago. The majority of workers in the field reject the hypothesis that life was transported to the earth from somewhere else in the galaxy and take it for granted that life began *de novo* on the early earth.

L. E. Orgel is at the Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA.  
Email: orgel@salk.edu

The uniformity of biochemistry in all living organisms argues strongly that all modern organisms descend from a last-common ancestor (LCA). Detailed analysis of protein sequences suggests that the LCA had a complexity comparable to that of a simple modern bacterium and lived 3.2–3.8 billion years ago<sup>5</sup>. If we knew the stages by which the LCA evolved from abiotic components present on the primitive earth, we would have a complete account of the origin of life. In practice, the most ambitious studies of the origins of life address much simpler questions. Here, I discuss two of them. What were the sources of the small organic molecules that made up the first self-replicating systems? How did biological organization evolve from an abiotic supply of small organic molecules?

## Abiotic synthesis of small organic molecules.

Miller, a graduate student who was working with Harold Urey, began the modern era in the study of the origin of life at a time when most people believed that the atmosphere of the early earth was strongly reducing. Miller<sup>6</sup> subjected a mixture of methane, ammonia and hydrogen to an electric discharge and led the products into liquid water. He showed that a substantial percentage of the carbon in the gas mixture was incorporated into a relatively small group of simple organic molecules and that several of the naturally occurring amino acids were prominent among these products. This was a surprising result; organic chemists would have expected a much less-tractable product mixture. The Urey–Miller experiments were widely accepted as a model of prebiotic synthesis of amino acids by the action of lightning.

Miller and his co-workers went on to study electric-discharge synthesis of amino acids in greater detail<sup>7,8</sup>. Using

more-powerful analytical techniques, they identified many more amino acids – some, but by no means all, of which occur in living organisms. They also showed that a major synthetic route to the amino acids is through the Strecker reaction – that is, from aldehydes, hydrogen cyanide and ammonia. Glycine, for example, is formed from formaldehyde, cyanide and ammonia – all of which can be detected among the products formed in the electric-discharge reaction.

In the years following the Urey–Miller experiments, the synthesis of biologically interesting molecules from products that could be obtained from a reducing gas mixture became the principle aim of prebiotic chemistry (Fig. 1). Remarkably, Oro and Kimble<sup>9</sup> were able to synthesize adenine from hydrogen cyanide and ammonia. Somewhat later, Sanchez, Ferris and I<sup>10</sup> showed that cyanoacetylene is a major product of the action of an electric discharge on a mixture of methane and nitrogen and that cyanoacetylene is a plausible source of the pyrimidine bases uracil and cytosine. This new information, together with previous studies that showed that sugars are formed readily from formaldehyde<sup>11,12</sup>, convinced many students of the origins of life that they understood the first stage in the appearance of life on the earth: the formation of a prebiotic soup of biomonomers.

As in any good detective story, however, the principle suspect, the reducing atmosphere, has an alibi. Recent studies have convinced most workers concerned with the atmosphere of the early earth that it could never have been strongly reducing<sup>13</sup>. If this is true, Miller's experiments, and most other early studies of prebiotic chemistry, are irrelevant. I believe that the dismissal of the reducing atmosphere is premature, because we do not completely understand the early history of the earth's atmosphere. It is hard to believe that the ease with which sugars, amino acids, purines and pyrimidines are formed under reducing-atmosphere conditions is either a coincidence or a false clue planted by a malicious creator.

Many of those who dismiss the possibility of a reducing atmosphere believe that the crime was an outside job. A substantial proportion of the meteorites that fall on the earth belong to a class known as carbonaceous chondrites<sup>14</sup>. These are particularly interesting because they contain a significant amount of organic carbon and because some of the standard amino acids and nucleic-acid

bases are present<sup>8</sup>. Could the prebiotic soup have originated in preformed organic material brought to the earth by meteorites and comets?

Supporters of the impact theory have argued convincingly that sufficient organic carbon must have been present in the meteorites and comets that reached the surface of the early earth to have stocked an abundant soup. However, would this material have survived the intense heating that accompanies the entry of large bodies into the atmosphere and their subsequent collisions with the surface of the earth? The results of theoretical calculations depend strongly on assumptions made about the composition and density of the atmosphere, the distribution of sizes of the impacting objects, etc.<sup>15</sup> The impact theory is probably the most popular at present, but nobody has proved that impacts were the most important sources of prebiotic organic compounds.

The newest suspects are the deep-sea vents, submarine cracks in the earth's surface where superheated water rich in transition-metal ions and hydrogen sulfide mixes abruptly with cold sea water. These vents are sites of abundant biological activity, much of it independent of solar energy. Wächtershäuser<sup>16,17</sup> has proposed a scenario for the origin of life that might fit such an environment. He hypothesizes that the reaction between iron(II) sulfide and hydrogen sulfide [a reaction that yields pyrites (FeS<sub>2</sub>) and hydrogen] could provide the free energy necessary for reduction of carbon dioxide to molecules capable of supporting the origin of life. He asserts that life originated on the surface of iron sulfides as a result of such chemistry. The assumptions that complex metabolic cycles self-organize on the surface and that the significant products never escape from the surface are essential parts of this theory; in Wächtershäuser's opinion, there never was a prebiotic soup!

Stetter and colleagues<sup>18</sup> have confirmed the novel suggestion that hydrogen sulfide, in the presence of iron(II) sulfide, acts as a reducing agent. They have reduced, for example, acetylene to ethane, and mercaptoacetic acid to acetic acid, but they have not reported reduction of CO<sub>2</sub>. However, in a new study, Wächtershäuser and co-workers<sup>19</sup> have shown that FeS spiked with NiS reduces carbon monoxide. Given that carbon monoxide might well have been present in large amounts in the gases escaping from the vents, Wächtershäuser's findings could well

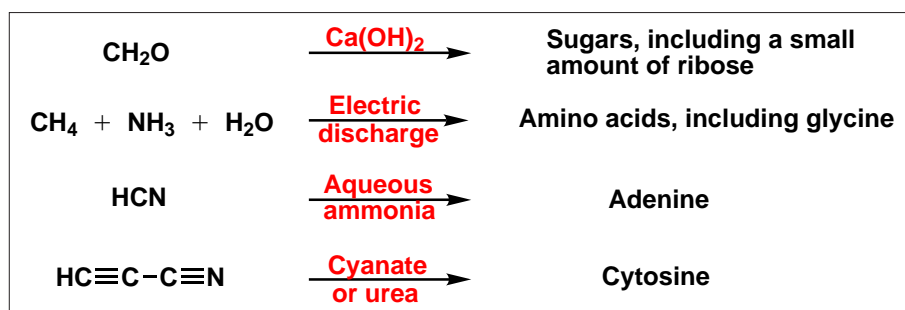


Figure 1

Early prebiotic syntheses of biomonomers.

prove important. If metal sulfides can be shown to catalyze the synthesis of a sufficient variety of organic molecules from carbon monoxide, the vent theory of the origins of biomonomers will become very attractive.

In summary, there are three main contending theories of the prebiotic origin of biomonomers (not to mention several other less-popular options). No theory is compelling, and none can be rejected out of hand. Perhaps it is time for a conspiracy theory; more than one of the sources of organic molecules discussed above may have collaborated to make possible the origin of life.

#### Self-organization

There is no general agreement about the source of prebiotic organic molecules on the early earth, but there are several plausible theories, each backed by some experimental data. The situation with regard to the evolution of a self-replicating system is less satisfactory; there are at least as many suspects, but there are virtually no experimental data.

The fairly general acceptance of the hypothesis that there was once an RNA world (i.e. a self-contained biological world in which RNA molecules functioned both as genetic materials and as enzyme-like catalysts) has changed the direction of research into the origins of life<sup>20</sup>. The central puzzle is now seen to be the origin of the RNA world. Two specific, but intertwined, questions are central to the debate. Was RNA the first genetic material or was it preceded by one or more simpler genetic materials? How much self-organization of reaction sequences is possible in the absence of a genetic material? I shall concentrate on the first question.

The assumption that a polymer that doubled as a genetic material and as a source of enzyme-like catalytic activity once existed profoundly changes the goals of prebiotic synthesis. The central issue becomes the synthesis of the first

genetic monomers: nucleotides or whatever preceded them. The synthesis of amino acids, coenzymes, etc. becomes a side issue, because there is no reason to believe that they were ever synthesized abiotically; some or all of them might have been introduced as direct or indirect consequences of the enzyme-like activities of RNA or its precursor(s). Supporters of the hypothesis that RNA was the first genetic material must explain where the nucleotides came from and how they self-organized. Those who believe in a simpler precursor have the difficult task of identifying such a precursor, but they hope that explaining monomer synthesis will then be simpler.

Returning to the idiom of the detective story, accumulating evidence suggests that RNA, a prime suspect, could have completed the difficult task of organizing itself into a self-contained replicating system. It has proved possible to isolate sequences that catalyze a wide variety of organic reactions from pools of random RNA<sup>21,22</sup>. As regards the origin of the RNA world, the most important reactions are those in which a preformed template-RNA strand catalyzes the synthesis of its complement from monomers or short oligomers. Eklund and co-workers<sup>23</sup> have isolated catalysts for the ligation of short oligonucleotides surprisingly easily, and the catalysts carry out ligation with adequate specificity. These molecules are the RNA equivalents of the RNA and DNA ligases. Considerable progress has also been made in selecting RNA equivalents of RNA polymerases<sup>24</sup>.

If the RNA world evolved *de novo*, it must have depended initially on an abiotic source of activated nucleotides. However, oxidation-reduction, methylation, oligosaccharide synthesis, etc., supported by nucleotide-containing coenzyme, probably became part of the chemistry of the RNA world before the invention of protein synthesis. Unfortunately, we cannot say just how complex the RNA world could have been until we know more about the

range of reactions that can be catalyzed by ribozymes. It seems likely that RNA could have catalyzed most of the steps involved in the synthesis of nucleotides<sup>25</sup>, and possibly the coupling of redox reactions to the synthesis of phosphodiester and peptides, but this remains to be demonstrated experimentally.

The experiments on the selection of ribozymes that catalyze nucleic acid replication (discussed above) use as inputs pools of RNA molecules synthesized by enzymes. Recently, Ferris and co-workers<sup>26,27</sup> have made considerable progress in the assembly of RNA oligomers from monomers, using an abundant clay mineral, montmorillonite, as a catalyst. The substrates that they use, nucleoside 5'-phosphorimidazolides, were probably not prebiotic molecules, but the experiments do indicate that the use of minerals as adsorbents and catalysts could allow the accumulation of long oligonucleotides once suitable activated monomers are available. We have shown that, using activated monomers, non-enzymatic copying of a wide range of oligonucleotide sequences is possible<sup>28</sup> and have obtained similar, but less extensive, results for ligation of short oligomers.

An optimist could propose the following scenario. First, activated

mononucleotides oligomerize on montmorillonite or an equivalent mineral. Next, copying of longer templates, using monomers or short oligomers as substrates, leads to the accumulation of a library of dsRNA molecules. Finally, an RNA double helix, one of whose strands has generalized RNA-polymerase activity, dissociates; the polymerase strand copies its complement to produce a second polymerase molecule, which copies the first to produce a second complement – and so on. The RNA world could therefore have arisen from a pool of activated nucleotides<sup>29</sup>. All that would have been needed is a pool of activated nucleotides!

Nucleotides are complicated molecules. The synthesis of sugars from formaldehyde gives a complex mixture, in which ribose is always a minor component. The formation of a nucleoside from a base and a sugar is not an easy reaction and, at least for pyrimidine nucleosides, has not been achieved under prebiotic conditions; the phosphorylation of nucleosides tends to give a complex mixture of products<sup>30</sup>. The inhibition of the template-directed reactions on D-templates by L-substrates is a further difficulty<sup>31</sup>. It is almost inconceivable that nucleic acid replication could have got started, unless there is a much

simpler mechanism for the prebiotic synthesis of nucleotides. Eschenmoser and his colleagues<sup>32</sup> have had considerable success in generating ribose 2,4-diphosphate in a potentially prebiotic reaction from glycolaldehyde monophosphate and formaldehyde. Direct prebiotic synthesis of nucleotides by novel chemistry is therefore not hopeless. Nonetheless, it is more likely that some organized form of chemistry preceded the RNA world. This leads us to a discussion of genetic takeover.

Cairns-Smith<sup>33</sup>, long before the argument became popular, emphasized how improbable it is that a molecule as high tech as RNA could have appeared *de novo* on the primitive earth. He proposed that the first form of life was a self-replicating clay. He suggested that the synthesis of organic molecules became part of the competitive strategy of the clay world and that the inorganic genome was taken over by one of its organic creations. Cairns-Smith's postulate of an inorganic life form has failed to gather any experimental support. The idea lives on in the limbo of uninvestigated hypotheses. However, Cairns-Smith also contemplated the possibility that RNA was preceded by one or more linear organic genomes<sup>34</sup>. This idea has

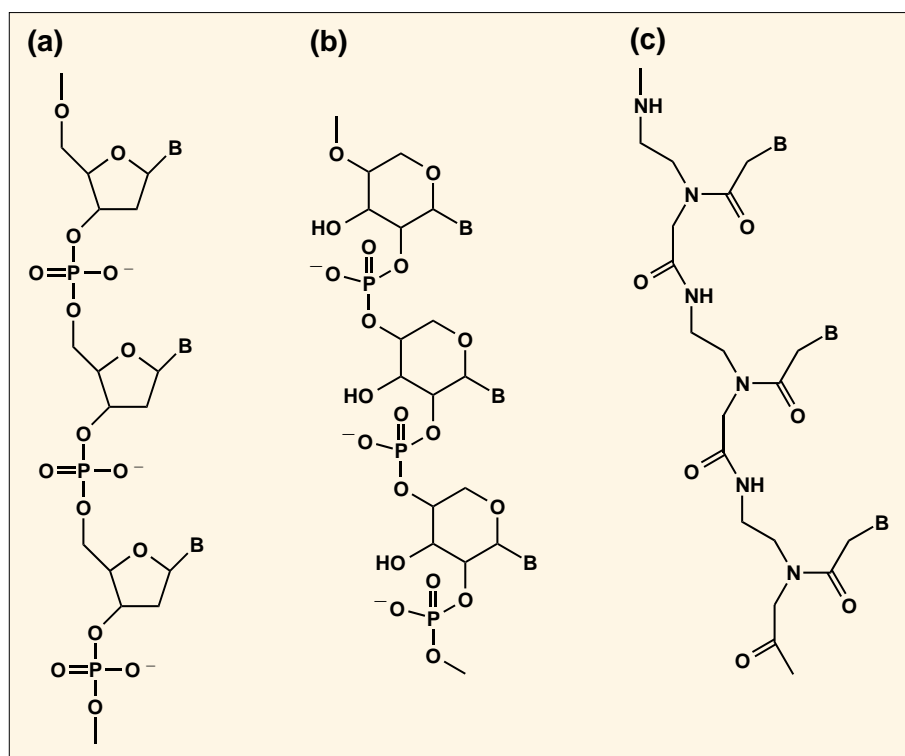


taken root, but its implications have not always been appreciated.

If RNA was not the first genetic material, biochemistry might provide no clues to the origins of life. Presumably, the biological world that immediately preceded the RNA world already had the capacity to synthesize nucleotides. This should help us to formulate hypotheses about its chemical characteristics. However, if there were two or more worlds before the RNA world, the original chemistry might have left no trace in contemporary biochemistry. In that case, the chemistry of the origins of life is unlikely to be discovered without investigating in detail all the chemistry that might have occurred on the primitive earth – whether or not that chemistry has any relation to biochemistry. This gloomy prospect has not prevented discussion of alternative genetic systems.

The only potentially informational systems, other than nucleic acids, that have been discovered are closely related to nucleic acids. Eschenmoser and his colleagues<sup>35</sup> have undertaken a systematic study of the properties of nucleic acid analogs in which ribose is replaced by another sugar or in which the furanose form of ribose is replaced by the pyranose form (Fig. 2b). Strikingly, polynucleotides based on the pyranosyl isomer of ribose (p-RNA) form Watson–Crick-paired double helices that are more stable than RNA, and p-RNAs are less likely than the corresponding RNAs to form multiple-strand competing structures<sup>35</sup>. Furthermore, the helices twist much more gradually than those in the standard nucleic acids, which should make it easier to separate strands during replication. Pyranosyl RNA seems to be an excellent choice as a genetic system; in some ways, it might be an improvement on the standard nucleic acids. However, prebiotic synthesis of pyranosyl nucleotides is not likely to prove much easier than synthesis of the standard isomers, although a route through ribose 2,4-diphosphate is being explored by Eschenmoser and his colleagues.

Peptide nucleic acid (PNA) is another nucleic acid analog that has been studied extensively (Fig. 2c). It was synthesized by Nielsen and colleagues<sup>36</sup> during work on antisense RNA. PNA is an uncharged, achiral analog of RNA or DNA; the ribose-phosphate backbone of the nucleic acid is replaced by a backbone held together by amide bonds. PNA forms very stable double helices with complementary RNA or DNA<sup>36,37</sup>. We have shown that information can be



**Figure 2**

DNA and potentially informational oligonucleotide analogs. (a) DNA. (b) Pyranosyl analog of RNA. (c) Peptide nucleic acid.

transferred from PNA to RNA, and vice versa, in template-directed reactions<sup>38,39</sup> and that PNA–DNA chimeras form readily on either DNA or PNA templates<sup>40</sup>. Thus, a transition from a PNA world to an RNA world is possible. Nonetheless, I think it unlikely that PNA was ever important on the early earth, because PNA monomers cyclize when they are activated; this would make oligomer formation very difficult under prebiotic conditions.

The studies described above suggest that there are many ways of linking together nucleotide bases into chains that can form Watson–Crick double helices. Perhaps a structure of this kind will be discovered that can be synthesized easily under prebiotic conditions. If so, it would be a strong candidate for the very first genetic material. However, another possibility remains to be explored: the first genetic material might not have involved nucleoside bases. Two or more very simple molecules could have the pairing properties needed to form a genetic polymer – a positively charged and a negatively charged amino acid, for example. However, it is not clear that stable structures of this kind exist. RNA is clearly adapted to double-helix formation: its constrained backbone permits simultaneous base pairing and stacking. It is unlikely that much simpler molecules could substitute for the

nucleotides. Perhaps some other interaction between the chains can stabilize a double helix in the absence of base stacking; binding to a mineral surface might supply the necessary constraints, but this remains to be demonstrated. In the absence of experimental evidence, little useful can be said.

The above discussion reveals a very large gap between the complexity of molecules that are readily synthesized in simulations of the chemistry of the early earth and the molecules that are known to form potentially replicating informational structures. Several authors have therefore proposed that metabolism came before genetics<sup>41–43</sup>. They have suggested that substantial organization of reaction sequences can occur in the absence of a genetic polymer and, hence, that the first genetic polymer probably appeared in an already-specialized biochemical environment. Because it is hard to envisage a chemical cycle that produces  $\beta$ -D-nucleotides, this theory would fit best if a simpler genetic system preceded RNA.

There is no agreement on the extent to which metabolism could develop independently of a genetic material. In my opinion, there is no basis in known chemistry for the belief that long sequences of reactions can organize spontaneously – and every reason to believe that they

cannot. The problem of achieving sufficient specificity, whether in aqueous solution or on the surface of a mineral, is so severe that the chance of closing a cycle of reactions as complex as the reverse citric acid cycle, for example, is negligible. The same, I believe, is true for simpler cycles involving small molecules that might be relevant to the origins of life and also for peptide-based cycles.

### Conclusion/outlook

In summary, there are several tenable theories about the origin of organic material on the primitive earth, but in no case is the supporting evidence compelling. Similarly, several alternative scenarios might account for the self-organization of a self-replicating entity from prebiotic organic material, but all of those that are well formulated are based on hypothetical chemical syntheses that are problematic. Returning to our detective story, we must conclude that we have identified some important suspects and, in each case, we have some ideas about the method they might have used. However, we are very far from knowing whodunit. The only certainty is that there will be a rational solution.

This review has necessarily been highly selective. I have neglected important aspects of prebiotic chemistry (e.g. the origin of chirality, the organic chemistry of solar bodies other than the earth, and the formation of membranes). The best source for such material is the journal *Origins of Life and Evolution of the Biosphere*, particularly those issues that contain the papers presented at meetings

of the International Society for the Study of the Origin of Life.

### Acknowledgements

This work was supported by NASA (grant number NAG5-4118) and NASA NSCORT/EXOBIOLGY (grant number NAG5-4546). I thank Aubrey R. Hill, Jr for technical assistance and Bernice Walker for manuscript preparation.

### References

- 1 Sleep, N. H., Zahnle, K. J., Kasting, J. F. and Morowitz, H. J. (1989) *Nature* 342, 139–142
- 2 Chyba, C. F. (1993) *Geochim. Cosmochim. Acta* 57, 3351–3358
- 3 Schopf, J. W. (1993) *The Earth's Earliest Biosphere: Its Origin and Evolution*, Princeton University Press
- 4 Mojzsis, S. J. et al. (1996) *Nature* 384, 55–59
- 5 Doolittle, R. F. (1997) *Proc. Natl. Acad. Sci. U. S. A.* 94, 13028–13033
- 6 Miller, S. L. (1953) *Science* 117, 528–529
- 7 Ring, D., Wolman, Y., Friedmann, N. and Miller, S. L. (1972) *Proc. Natl. Acad. Sci. U. S. A.* 69, 765–768
- 8 Wolman, Y., Haverland, H. and Miller, S. L. (1972) *Proc. Natl. Acad. Sci. U. S. A.* 69, 809–811
- 9 Oro, J. and Kimball, A. P. (1960) *Biochim. Biophys. Res. Commun.* 2, 407–412
- 10 Ferris, J. P., Sanchez, A. and Orgel, L. E. (1968) *J. Mol. Biol.* 33, 693–704
- 11 Butlerow, A. (1861) *Compt. Rend. Acad. Sci.* 53, 145–147
- 12 Butlerow, A. (1861) *Liebig's Ann. Chem.* 120, 295
- 13 Kasting, J. F. (1993) *Science* 259, 920–926
- 14 Cronin, J. R., et al. (1988) in *Meteorites and the Early Solar System* (Kerridge, J. F. and Mathew, M. S., eds), pp. 8191–1857 University of Arizona Press
- 15 Chyba, C. and Sagan, C. (1992) *Nature* 355, 125–132
- 16 Wächtershäuser, G. (1988) *Microbiol. Rev.* 52, 452–484
- 17 Wächtershäuser, G. (1992) *Prog. Biophys. Mol. Biol.* 58, 85–201
- 18 Blochl, E., Keller, M., Wächtershäuser, G. and Stetter, K. O. (1992) *Proc. Natl. Acad. Sci. U. S. A.* 89, 8117–8120

- 19 Huber, C. and Wächtershäuser, G. (1997) *Science* 276, 245–247
- 20 Gesteland, R. F. and Atkins, J. F. (1993) *The RNA World: The Nature of Modern RNA Suggests a Prebiotic World*, Cold Spring Harbor Laboratory Press
- 21 Pan, T. (1997) *Curr. Opin. Chem. Biol.* 1, 17–25
- 22 Breaker, R. R. (1997) *Curr. Opin. Chem. Biol.* 1, 26–31
- 23 Eklund, E. H., Szostak, J. W. and Bartel, D. P. (1995) *Science* 269, 364–370
- 24 Eklund, E. H. and Bartel, D. P. (1996) *Nature* 382, 373–376
- 25 Unrai, P. J. and Bartel, D. P. (1998) *Nature* 395, 260–263
- 26 Ferris, J. P. and Ertem, G. (1993) *J. Am. Chem. Soc.* 115, 12270–12275
- 27 Kawamura, K. and Ferris, J. P. (1994) *J. Am. Chem. Soc.* 116, 7564–7572
- 28 Hill, A. R., Jr, Wu, T. and Orgel, L. E. (1993) *Orig. Life Evol. Biosphere* 23, 285–290
- 29 Orgel, L. E. (1994) *Sci. Am.* 271, 52–61
- 30 Ferris, J. P. (1987) *Cold Spring Harbor Symp. Quant. Biol.* LII, 29–39
- 31 Joyce, G. F. et al. (1984) *Nature* 310, 602–604
- 32 Muller, D. et al. (1990) *Helv. Chim. Acta* 73, 1410–1468
- 33 Cairns-Smith, A. G. (1982) *Genetic Takeover and the Mineral Origins of Life*, Cambridge University Press
- 34 Cairns-Smith, A. G. and Davis, C. J. (1977) in *Encyclopaedia of Ignorance* (Duncan, R. and Weston-Smith, M., eds), pp. 397–403, Pergamon Press
- 35 Eschenmoser, A. (1997) *Orig. Life Evol. Biosphere* 27, 535–553
- 36 Egholm, M., Buchardt, O., Nielsen, P. E. and Berg, R. H. (1992) *J. Am. Chem. Soc.* 114, 1895–1897
- 37 Egholm, M. et al. (1993) *Nature* 365, 566–568
- 38 Schmidt, J. G., Christensen, L., Nielsen, P. E. and Orgel, L. E. (1997) *Nucleic Acids Res.* 25, 4792–4796
- 39 Schmidt, J. G., Nielsen, P. E. and Orgel, L. E. (1997) *Nucleic Acids Res.* 25, 4797–4802
- 40 Koppitz, M., Nielsen, P. E. and Orgel, L. E. (1998) *J. Am. Chem. Soc.* 120, 4563–4569
- 41 Kauffman, S. A. (1986) *J. Theor. Biol.* 119, 1–24
- 42 Wächtershäuser, G. (1988) *Microbiol. Rev.* 52, 452–484
- 43 De Duve, C. (1991) *Blueprint for a Cell: The Nature and Origin of Life*, Neil Patterson

## DOMO: a new database of aligned protein domains

Domains are autonomously folding units that combine to form modular proteins<sup>1</sup>. At a sequence level, accurately delineating the boundaries of homologous protein domains is essential for multiple sequence alignment. Tertiary-structural data that could guide visual determination of such domain boundaries are not available for most proteins.

Consequently, although many motif<sup>2</sup>, block<sup>3</sup> and full-sequence-alignment<sup>4</sup> databases exist, as yet there are only two domain-alignment databases (PRODOM and PFAM) that have been constructed by a fully automated process from sequence information alone<sup>5,6</sup>.

Here, we describe DOMO, a new database that contains 8877 multiple

sequence alignments, including 99 058 protein domains as well as repeating-sequence regions extracted from 83 054 non-redundant amino acid sequences from the SWISS-PROT<sup>7</sup> and PIR<sup>8</sup> databases. The domain boundaries and alignments were generated by a fully automated analysis process that involves the detection and clustering of amino acid sequence similarities and, subsequently, delineation of the domain boundaries and multiple sequence alignment of related protein segments<sup>9,10</sup>. The domain boundaries were not inferred from three-dimensional data. Instead, the relative positions of homologous segment pairs within the same protein (for repeats) or within