

Jack W. Szostak, David P. Bartel & P. Luigi Luisi

Figure 1. The study design. The study was a 2-year, 2-center, parallel, randomized, controlled trial. The study was conducted in two centers, the University of Michigan and the University of California, San Francisco. The study was conducted in two phases. In the first phase, the study was conducted in the University of Michigan. In the second phase, the study was conducted in the University of California, San Francisco. The study was conducted in two phases. In the first phase, the study was conducted in the University of Michigan. In the second phase, the study was conducted in the University of California, San Francisco.

How simple can a cell be and still be considered as living? The answer depends on what we consider to be the essential properties of life. Defining life is notoriously difficult<sup>4</sup>; its very diversity resists the confines of any compact definition. An operational approach focuses on identifying simple cellular systems that are both autonomously replicating and subject to darwinian evolution. Autonomous replication is understood as continued growth and division which is reliant on the input of small molecules and energy only, and does not depend on the products of pre-existing living systems such as protein enzymes. Darwinian evolution requires the essential biological aspects of genetic variation and its phenotypic expression as variation in survival and reproduction.

We can consider life as a property that emerges from the union of two fundamentally different kinds of replicating systems: the informational genome and the three-dimensional structure in which it resides. The simplest way to enable darwinian evolution is to begin with a nucleic acid genome.

**Figure 1** Role of vesicles in enabling darwinian evolution of a replicase. The vesicular compartment ensures that molecules related by descent are, on average, kept in physical proximity to each other, allowing a superior mutant replicase (red) to preferentially self-replicate, in comparison to the parental replicase (black). The evolutionary advantage of increased replication is amplified as vesicles with superior replicase molecules are more likely to give rise to vesicles with at least two replicase molecules (or a replicase and a template molecule). Vesicles with less than two replicase molecules (indicated by an X) and the progeny of these vesicles cannot continue RNA self-replication. In this way, vesicles with superior replicase molecules become an increasing fraction of the vesicles that maintain replicase activity.

But a replicase molecule by itself is not living, for two reasons. First, a single molecule could not actually replicate, as it cannot be both template and polymerase at the same time. Replication requires two RNA molecules — a replicase that acts as the polymerase, and another molecule, which could be either an unfolded replicase or an RNA complementary in sequence to the replicase, to act as a template. Some form of physical compartmentation is therefore required to keep replicase and template together, unless both are present at high concentration. Second and more subtly, a population of

All known cells use membranes composed of amphipathic lipids as their compartment-defining barriers, and therefore the easiest way to construct our simple protocell is to surround it with a lipid membrane. This also makes it easier to imagine how a simple cell could evolve into more complex cells, similar to present-day cells, without major architectural transitions. However, in the absence of any machinery for

cell growth and division, the membrane-bounded vesicle must itself be a spontaneously replicating entity.

So our simple protocell will consist of an RNA replicase replicating inside a replicating membrane vesicle. Both these components are self-assembling; the catalytically active structure of the replicase will form spontaneously as a consequence of its nucleotide sequence, while membrane vesicles assemble spontaneously as a result of interactions between the lipid molecules. As RNA molecules can become spontaneously encapsulated in vesicles as they form, the protocell as a whole could self-assemble. With compartmentation, the replicase component is not only capable of, but also inevitably subject to, variation, natural selection and thus darwinian evolution.

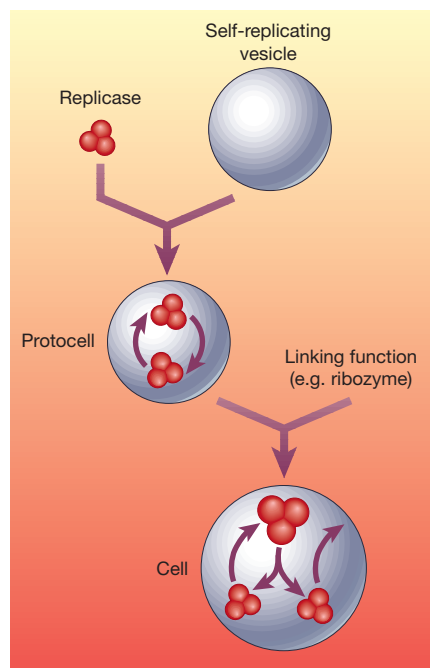
## From protocell to living cell

Such simple protocells would be nearly, but not quite, alive. When fed small-molecule precursors for membrane and RNA synthesis, they would grow and divide, and improved replicases would evolve. However, a vesicle carrying an improved replicase would itself not have improved capacity for survival or reproduction. For this to happen, an RNA-coded activity is needed that imparts an advantage in survival, growth or replication for the membrane component. A simple example would be a ribozyme that synthesizes amphipathic lipids and so enables the membrane to grow. The membrane and the genome would then be coupled, and the 'organism' as a whole could evolve (Fig. 2) as vesicles with improved ribozymes would have a growth and replication advantage. A simple cell with an interdependent genome and membrane would be a sustainable, autonomously replicating system, capable of darwinian evolution. It would be truly alive.

## The RNA replicase

The first experimental challenge is the evolution or design of an RNA replicase. Early attempts to derive an RNA replicase from the natural group I self-splicing introns produced ribozymes that could direct the assembly of oligonucleotide substrates on a template, and even the assembly of full-length RNA strands complementary to the ribozyme itself<sup>9,10</sup>. The low efficiency of the reaction, however, even when driven by vast substrate excess, suggested that it was probably essential to use activated nucleotides, such as nucleoside triphosphates, to provide an energetic driving force for polymerization. The use of oligonucleotide substrates would also make it difficult or impossible to maintain a high concentration of all the different substrates needed for the replicase to mutate and evolve.

No natural ribozymes are known that can catalyse the required chemistry and use nucleoside triphosphates as substrates. As this is a complex enzymatic function, attempts to



**Figure 2** Outline of proposed pathway for synthesis of a cell. The first major synthetic intermediates are an RNA replicase and a self-replicating vesicle. These are combined into a protocell, enabling rapid evolutionary optimization of the replicase. Addition of an RNA-coded linking function, such as a lipid-synthesizing ribozyme, completes the cellular structure.

evolve an RNA polymerase ribozyme experimentally have proceeded incrementally. First, *in vitro* selection was used to isolate, from a fully random set of starting sequences, a set of ribozymes that would carry out the correct chemistry—the attack of a 3'-hydroxyl on the  $\alpha$ -phosphate of a triphosphate, yielding a new phosphodiester bond—but in a more favourable context, the joining, or ligation, of two RNA sequences. This experiment yielded many ligases that carried out pyrophosphate-activated oligonucleotide-ligation reactions<sup>11</sup>. Of these, the class I ligase ribozyme synthesized the desired 3',5'-phosphodiester linkage in a template-directed reaction, and therefore carried out the same chemical transformation as protein-enzyme polymerases<sup>12</sup>. Mutations that increased catalytic activity<sup>13</sup> gave a highly active ligase that could join oligonucleotides with a rate of greater than one joining event per second<sup>12,14</sup>. Derivatives of this ribozyme were subsequently shown to act as primitive polymerases capable of template-directed extension of a 'primer' strand of RNA complexed to the RNA template, using nucleoside triphosphates as substrates<sup>15</sup>. In essence, the oligonucleotide providing the 5'-triphosphate could be replaced by a single nucleotide, which can still be ligated to the 3'-end of the growing primer. The cycle of primer extension can be repeated several times before steric constraints prevent

further chain growth. These advances bring the evolution of a true RNA replicase tantalizingly close.

What further improvements are required to obtain an RNA replicase suitable for incorporation into an artificial cell? In its current form, the polymerase ribozyme recognizes the primer-template complex through hybridization to a particular unpaired segment of the template. This pairing restricts ribozyme movement along the template and reduces severely the number of compatible template sequences. A ribozyme that can recognize the primer-template using non-sequence-specific contacts would enable more extensive and general RNA synthesis. The next hurdle will be to improve the fidelity and efficiency of polymerization. It is possible that only a 100-fold increase in the rate of polymerization and a 10-fold improvement in the Watson-Crick fidelity of this ribozyme would lead to an RNA polymerase able to faithfully copy templates of its own length<sup>7</sup>.

A more subtle problem is that a true replicase must function both as a polymerase and as a template. How could the same RNA sequence act both as a highly active ribozyme structure, presumably favoured by stable folding, and as a template available for copying, favoured by less stable folding? A potential solution comes from the finding that active ribozymes can be reconstituted by the spontaneous self-assembly of two or more oligonucleotides; the separate oligonucleotides can be more or less unstructured, while the assembled complex can be stable and enzymatically active<sup>16</sup>. This solution has the advantage that the average length of sequence that needs to be copied by the replicase can be fairly short (30–40 nucleotides), but the potential disadvantage that the relatively unstructured + and – strand fragments might rapidly reanneal to form double-stranded RNA.

This leads to the issue of strand separation during or after replication in order to reconstitute the replicase. In principle, thermal denaturation might separate the strands of a double-stranded replication product. The extreme stability of long RNA duplexes, especially in the presence of significant concentrations of divalent cations, makes this approach unattractive, however, as conditions that would lead to denaturation would probably also lead to chemical degradation of the RNA and disruption of the membrane vesicle. But the alternatives involve more complexity in the RNA replication machinery—either RNA helicase activities to carry out energy-dependent strand separation, or, as is the case with bacteriophage T7 RNA polymerase, a portion of the replicase that binds single-stranded RNA and peels the new RNA strand off the template as it is synthesized. The rapid formation of local secondary structure in the RNA strands would then prevent them re-forming a dead-end full-length duplex.

The RNA polymerase (a protein enzyme) of the Q $\beta$  virus is known to depend on the secondary structure of the viral + and – strands to kinetically block duplex formation<sup>17</sup>. A better understanding of the kinetics of RNA folding will be required to predict sequences that will be reasonably stable as unpaired + and – strands when packaged in the same vesicle.

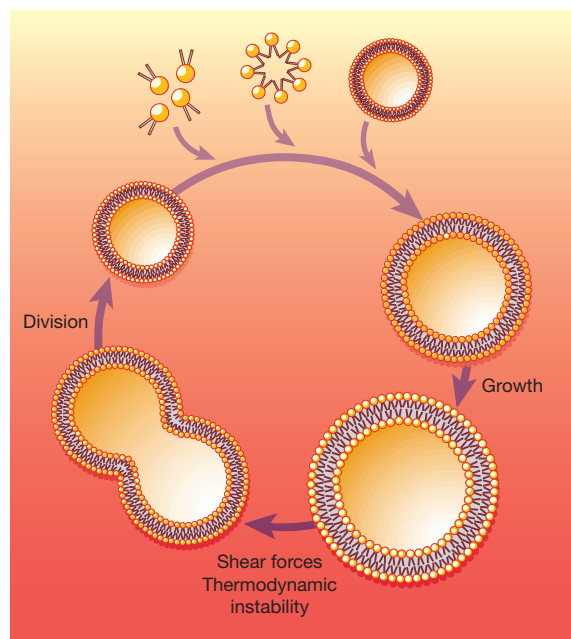
An attractive alternative strategy is replication by strand-displacement on a duplex template, which provides a mechanism for coupling the chemical energy released during polymerization to strand displacement. Replication would also be more likely to start at the beginning of the genome, because of transient ‘fraying’ of the termini of the duplex. However, specific oligonucleotide primers may be required to obtain significant initiation. These could be difficult to deliver to the vesicle interior and might not be considered ‘small-molecule’ substrates. Nevertheless, the use of primers would allow replication of the entire template, avoiding gradual shortening without having to enlist a telomerase-like activity or having to prime polymerization with a single nucleotide at the 3′-terminus of the template.

The class I ligase is an excellent starting point for attempts to evolve a replicase but does have one drawback. Its minimal catalytic domain is about 100 nucleotides long. When coupled to additional domains that may be required for proper template binding, fidelity and strand separation, the total length could approach 200–300 nucleotides. The longer the replicase, the more difficult the problem of replication, so shorter replicases should be looked for. One approach would be to use more highly activated nucleotide substrates, such as the phosphorimidazolides. As these substrates are more reactive, less rate enhancement would be required to achieve extension rates in the range of one nucleotide per minute or per second, and a simpler and shorter ribozyme might suffice. Clearly, there are many possible approaches to evolving an RNA replicase, all of which bear investigation.

### The membrane compartment

The vesicle component of a protocell or simple cell must possess a suite of rather unusual properties, including spontaneous growth, spontaneous division, permeability to nucleotide substrates, physical stability under conditions required for RNA replication and compatibility with ribozyme activity.

Spontaneous vesicle growth could in principle occur either gradually by the incorporation of single lipid molecules or micelles, or stepwise by fusion with other vesicles. Gradual growth, which is more biological, is possible if the rate of incorporation of lipid into pre-existing vesicles is greater than the rate of spontaneous assembly into new vesicles. Alternatively, the catalytic gen-



**Figure 3** Modes of vesicle growth and division. Self-replicating membrane vesicles can grow either gradually or in discrete steps, and may divide either spontaneously or under the influence of external environmental forces.

eration of new lipid molecules in a vesicle membrane could also lead to vesicle growth. Fatty acids such as oleate form micelles at high pH, an oil phase at low pH, and bilayer membrane vesicles at intermediate pH — these are thought to be stabilized by hydrogen bonding between the protonated and ionized carboxylates of the fatty acid. When the pH of a preparation of oleate micelles is lowered to the pK of oleate in a membrane (pH ~8) the micelles slowly rearrange to form vesicles. However, pre-existing oleate vesicles greatly accelerate the rate of formation of new vesicles<sup>18</sup>. Remarkably, the newly formed vesicles have a size distribution similar to that of the preformed vesicles, which can be quite different from the size distribution in the absence of preformed vesicles. The mechanism of this ‘matrix’ effect remains unknown, although some form of growth and division is an intriguing possibility. Subsequent work showed that oleate vesicles seem to grow in size in the presence of oleic anhydride, presumably as a result of vesicle-mediated catalysis of hydrolysis of the oleate precursor, followed by incorporation of the newly generated oleate into the vesicle membrane<sup>19</sup>.

These experiments could not distinguish between pre-existing vesicles that grew by incorporating new lipid, and newly formed vesicles. More recently, however, preformed oleate vesicles have been tagged by preparation in the presence of ferritin<sup>20</sup>. After exposure of these vesicles to additional oleate micelles, electron microscopic examination revealed that the tagged vesicles had grown to a larger average size, providing strong evidence for vesicle growth by spontaneous incorporation of new lipid.

Vesicle growth in discrete steps can occur by vesicle-vesicle fusion. The fusion of vesicles composed of acidic phospholipids is

mediated by low concentrations of Ca<sup>2+</sup> ions<sup>21–23</sup>, and short fusogenic peptides can catalyse the fusion of vesicles composed of neutral phospholipids<sup>24</sup>. The instability of small strained vesicles can provide a thermodynamic driving force for this process. However, these processes tend to operate most efficiently under conditions that are very close to those favouring a phase change for the component lipids, and thus catastrophic loss of vesicle integrity. The narrow range of conditions under which fusion is efficient but vesicle disruption is minimal indicates that it may be difficult to devise a robust cycle of growth and division using this approach. On the other hand, vesicle fusion could bring fresh supplies of nucleotide substrates to replicases encapsulated in a separate vesicle.

What about division? In the absence of the complex machinery that controls the division of modern cells, the division of growing vesicles must rely on the intrinsic properties of the vesicle and the physical properties of the environment<sup>25</sup>. Input of energy from the environment can generate a population of vesicles with a non-equilibrium size distribution. High environmental shear forces, for example, can cause vesicles to divide. Such a process, operating in conjunction with a spontaneous growth mechanism, could lead to a primitive cell cycle controlled entirely by the biophysical properties of the membrane and environmental forces. An intriguing possibility is that the process of division could be highly favoured, or even become spontaneous, with lipid compositions that yield vesicles of optimum size for thermodynamic stability (Fig. 3).

Addition of lipid to lipid micelles causes micellar reproduction because the growing micelles become thermodynamically unstable above a certain size<sup>26,27</sup>. For example, ethyl caprylate oil layered over alkaline water hydrolyses slowly at the interface. Once the



critical micelle concentration for caprylate is reached, substrate is solubilized in the micelles and the rate of hydrolysis greatly increases as a result of the increased area of lipid–water interface, resulting in a rapid increase in the number of micelles. Whether an analogous system for the spontaneous growth and division of vesicles can be achieved remains an open question.

The membrane of a synthetic cell should allow transport of small-molecule substrates such as nucleotides, while keeping macromolecules encapsulated. The demonstration of RNA synthesis catalysed by polynucleotide phosphorylase inside vesicles, using external nucleoside diphosphate substrates, indicates that this requirement may not be too stringent<sup>28,29</sup>. Short-chain lipids, lipid mixtures, and co-surfactants such as cholate can all make membranes more permeable to small molecules<sup>30,31</sup>. Amplification by polymerase chain reaction of DNA inside vesicles suggests that it may be possible to find vesicle compositions compatible with conditions for RNA replication<sup>32</sup>.

An alternative approach to feeding the replicase small-molecule substrates would be to encapsulate the substrates within vesicles, which could then be delivered to the replicase by vesicle fusion. With repeated cycles of fusion and fission, a replicase initially present in just a few vesicles would spread throughout the vesicle population. Although the vesicle component would not be growing and evolving, the replicase component would be. In some ways, the replicase of this system is analogous to bacteriophage, which clearly evolves as it propagates through a system of compartments (bacterial cells). Such a system could provide a powerful tool for the evolution of replicase activities.

### Replicase-vesicle coupling

Once robust self-replicating replicases and vesicles have been devised, they must be brought together in a compatible and interdependent union in which the vesicle–RNA system as a whole is subject to darwinian evolution. First, the replicase and vesicle must be compatible; the replicase must be able to replicate inside the vesicle, and the vesicle must be able to grow and divide unperturbed by its cargo of RNA. It is impossible to foresee all the problems, but some are already clear. Most obviously, both replication cycles must operate under a single consistent set of conditions. Many ribozymes have optimal activity in the presence of high concentrations of divalent metal ions. In such conditions, vesicles composed of acidic phospholipids would aggregate, possibly interfering with growth and division. The timescales of the RNA and vesicle replication cycles must also be approximately the same, so that the replicases can keep up with the vesicles but not reach such a high internal concentration that vesicle growth or division is disrupted.

Once a compatible membrane and replicase have been united to form a protocell, an additional RNA activity must emerge spontaneously or be added exogenously to couple the information coded by the genome to the fitness of the vesicle. The emergence of ribozymes that catalyse the synthesis of suitable amphipathic molecules would facilitate growth of the membrane compartment, and they could ultimately be used to alter and control the properties of the vesicle membrane. Structural RNAs with favourable interactions with the interior vesicle wall could stabilize the vesicle and lead to preferential survival. The evolution of RNA filaments analogous to actin filaments or microtubules could influence vesicle shape and dynamics, and begin to provide internal control of cell division. It is clear that genomic (that is, RNA) variation would influence the ability of the cell to grow, survive and reproduce, and thus would control the fitness of the cell as a whole.

With coupling accomplished, the living synthetic cell would be capable of evolving in ways that none of its components are, and we expect that the strong selective forces to which it will be subject will provide a powerful driving force for the emergence of biochemical complexity, which in turn will lead to increasingly tight coupling and interdependence of the RNA and membrane components. The emergence of better replicases would allow the replication of longer RNA genomes, and the incorporation (or internal generation) of new random sequences would provide a source of new RNA activities. The evolution of ribozymes that contribute to the synthesis of RNA precursors would enhance the efficiency of the RNA replication process, both by decreasing the need for externally supplied substrates, and perhaps also by decreasing the need to spontaneously transport complete nucleotides across the membrane. Particular RNA sequences might also selectively alter the membrane permeability<sup>33</sup>, increasing the intracellular availability of replicase substrates. Other ribozymes might lead to the formation of RNA–lipid conjugates, providing a mechanism for membrane anchoring of ribozymes and possibly more equal segregation of ribozymes into daughter cells. As the number of ribozymes and structural RNAs grows, the membrane compartment will be crucial in maintaining the spatial integrity of the assembly of cooperating RNA species.

Experimentally, the potential exists to jump-start the emergence of biochemical complexity by the *in vitro* selection and directed evolution of potentially adaptive ribozymes. Alternatively, by supplying a population of cells with random RNA sequences, one might observe the process of evolving complexity in real time, and thus determine experimentally what new ribozyme activities were most accessible and

advantageous for evolving simple cells. In the long run, it might even be possible to observe at least some aspects of the evolution of protein synthesis, possibly with different basis sets of amino acids. These experimental possibilities could provide fascinating insights into what is now a complete black box of early evolution.

Jack W. Szostak is at the Howard Hughes Medical Institute and Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. David P. Bartel is at the Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. P. Luigi Luisi is at ETH-Zentrum, Institut für Polymere, Universitätstrasse 6, CH-8092 Zürich, Switzerland.

- Hutchison, C. A. *et al.* *Science* **286**, 2165–2169 (1999).
- Gilbert, W. *Nature* **319**, 618 (1986).
- Joyce, G. F. & Orgel, L. E. in *The RNA World* (eds Gesteland, R. E., Cech, T. R. & Atkins, J. F.) 49–77 (Cold Spring Harbor Laboratory Press, New York, 1999).
- Luisi, P. L. *Origins Life Evol. Biosphere* **28**, 613–622 (1998).
- Joyce, G. F. *New Biol.* **3**, 399–407 (1991).
- Hager, A. J., Pollard, J. D. & Szostak, J. W. *Chem. Biol.* **3**, 717–725 (1996).
- Bartel, D. P. in *The RNA World* (eds Gesteland, R. E., Cech, T. R. & Atkins, J. F.) 143–162 (Cold Spring Harbor Laboratory Press, New York, 1999).
- Bartel, D. P. & Unrau, P. J. *Trends Cell Biol.* **9**, M9–M13 (1999).
- Doudna, J. A. & Szostak, J. W. *Nature* **339**, 519–522 (1989).
- Green, R. & Szostak, J. W. *Science* **258**, 1910–1915 (1992).
- Bartel, D. P. & Szostak, J. W. *Science* **261**, 1411–1418 (1993).
- Eklund, E. H., Szostak, J. W. & Bartel, D. P. *Science* **269**, 364–370 (1995).
- Eklund, E. H. & Bartel, D. P. *Nucleic Acids Res.* **23**, 3231–3238 (1995).
- Bergman, N. H., Johnston, W. K. & Bartel, D. P. *Biochemistry* **39**, 3115–3123 (2000).
- Eklund, E. H. & Bartel, D. P. *Nature* **382**, 373–376 (1996).
- Doudna, J. A., Couture, S. & Szostak, J. W. *Science* **251**, 1605–1608 (1991).
- Priano, C., Kramer, F. R. & Mills, D. R. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 321–330 (1987).
- Blochiger, E., Blocher, M., Walde, P. & Luisi, P. L. *J. Phys. Chem.* **102**, 10383–10390 (1998).
- Walde, P., Wick, R., Frezza, M., Mangone, A. & Luisi, P. L. *J. Am. Chem. Soc.* **116**, 11649–11654 (1994).
- Berclaz, N., Muller, M., Walde, P. & Luisi, P. L. *J. Phys. Chem. B* (in the press).
- Ingolia, T. D. & Koshland, D. E. Jr *J. Biol. Chem.* **253**, 3821–3829 (1978).
- Wilschut, J., Düzgünes, N., Freale, R. & Papahadjopoulos, D. *Biochemistry* **19**, 6011–6021 (1980).
- Ellens, H., Bentz, J. & Szoka, F. C. *Biochemistry* **24**, 3099–3106 (1985).
- Parente, R. A., Nir, S. & Szoka, F. C. Jr *J. Biol. Chem.* **263**, 4724–4730 (1988).
- Norris, V. & Raine, D. J. *Origins Life Evol. Biosphere* **28**, 523–537 (1998).
- Bachmann, P. A., Luisi, P. L. & Lang, J. *Nature* **357**, 57–59 (1992).
- Luisi, P. L. *Adv. Chem. Phys.* **XCII**, 425–438 (1996).
- Chakrabarti, A. C., Breaker, R. R., Joyce, G. F. & Deamer, D. W. *J. Mol. Evol.* **39**, 555–559 (1994).
- Walde, P., Goto, A., Monnard, P.-A., Wessicken, M. & Luisi, P. L. *J. Am. Chem. Soc.* **116**, 7541–7547 (1994).
- Ueno, M. *Biochemistry* **28**, 5631–5634 (1989).
- Albalak, A., Zeidel, M. L., Zucker, S. D., Jackson, A. A. & Donovan, J. M. *Biochemistry* **35**, 7936–7945 (1996).
- Oberholzer, T., Albrizio, M. & Luisi, P. L. *Chem. Biol.* **2**, 677–682 (1995).
- Khvorovova, A., Kwak, Y. G., Tamkun, M., Majerfeld, I. & Yarus, M. *Proc. Natl Acad. Sci. USA* **96**, 10649–10654 (1999).

**Acknowledgements.** We thank our colleagues and the members of our laboratories for helpful discussions and comments on the manuscript. J.W.S. is an Investigator of the Howard Hughes Medical Institute. This work was supported by grants from NASA and the NIH to J.W.S., grants from NASA and the NIH to D.P.B. and grants from COST Supramolecular Chemistry, Project D 11-2 to P.L.L.