

## REVIEW

# Towards synthesis of a minimal cell

Anthony C Forster<sup>1,\*</sup> and George M Church<sup>2,\*</sup>

<sup>1</sup> Department of Pharmacology and Vanderbilt Institute of Chemical Biology, Vanderbilt University Medical Center, Nashville, TN, USA and

<sup>2</sup> Department of Genetics, Harvard Medical School, Boston, MA, USA

\* Corresponding authors. AC Forster, Department of Pharmacology, Vanderbilt University Medical Center, 23rd Ave.S. at Pierce, Nashville, TN 37232, USA. Tel.: +1 615 936 3112; Fax: +1 615 936 5555; E-mail: a.forster@vanderbilt.edu or GM Church, Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, NBR 238, Boston, MA 02115, USA. Tel.: +1 617 432 1278; Fax: +1 617 432 6513

Received 7.5.06; accepted 26.7.06

**Construction of a chemical system capable of replication and evolution, fed only by small molecule nutrients, is now conceivable. This could be achieved by stepwise integration of decades of work on the reconstitution of DNA, RNA and protein syntheses from pure components. Such a minimal cell project would initially define the components sufficient for each subsystem, allow detailed kinetic analyses and lead to improved *in vitro* methods for synthesis of biopolymers, therapeutics and biosensors. Completion would yield a functionally and structurally understood self-replicating biosystem. Safety concerns for synthetic life will be alleviated by extreme dependence on elaborate laboratory reagents and conditions for viability. Our proposed minimal genome is 113 kbp long and contains 151 genes. We detail building blocks already in place and major hurdles to overcome for completion.**

*Molecular Systems Biology* 22 August 2006;  
doi:10.1038/msb4100090

*Subject Categories:* synthetic biology

*Keywords:* cell; genome; minimal; RNA; self-replication; translation

## Overview

### ‘How far can we push chemical self-assembly?’

This question was posed recently as one of the big 25 questions in science for the next 25 years (Service, 2005). Nowadays, big questions often are addressed by big experimental efforts. But before embarking on a big project, it is helpful to get specific. What push in chemical self-assembly might be most worthwhile and practical? Self-assembly *in vitro* of viruses and the ribosome, achieved decades ago, taught us some of the principles assumed to be used in general by cells (Lewin, 2004). For example, self-assembly occurs in a definite sequence and is generally energetically favored, obviating the need for enzymes and an energy source. Assembling some type of cell (i.e. a self-replicating, membrane-encapsulated collection of biomolecules) would seem to be the next major

step, yet detailed plans have not been published. Here, we attempt to outline the synthesis of a minimal cell containing the core cellular replication machinery, review the pertinent literature and highlight gaps in knowledge that need filling.

## Utility

Synthesizing a minimal cell will advance knowledge of biological replication. Many hypotheses in replication and its subsystems can only be tested in such a synthetic biology project. The meaning of ‘synthetic’ (from Greek *sunthesis*, to put together) discussed here bypasses the current reliance of synthetic biology on cells or macromolecular cell products: the aim is to put together an organism from small molecules alone. The simplest approach for creating an artificial cell may be by evolving an RNA polymerase made exclusively of RNA (Szostak *et al*, 2001) to replace all protein components of *in vitro* replicating and evolving systems (e.g. to replace Q $\beta$  replicase; Mills *et al*, 1967). But in comparison with a purified protein-based system, it is neither guaranteed to arrive sooner nor tell us more. A protein-based system will connect with, and reveal more about, existing biological systems. Life, like a machine, cannot be understood simply by studying it and its parts; it must also be put together from its parts. Along the way to synthesizing a cell, we might discover new biochemical functions essential for replication, unsuspected macromolecular modifications or previously unrecognized patterns of coordinated expression.

How good a model would an artificial, protein-based, minimal cell be for natural cells? The only cellular alternative is a perturbed natural cell, an incredibly complex system even for the simplest of cells. A much simpler purified system based on a real cell would thus be easier to model and understand. It could certainly answer questions that cannot be answered *in vivo* or in crude extracts, such as which macromolecules and macromolecular modifications are sufficient for subsystem function. However, even the simplest minimal cell would still be highly complex; so its construction and study would be facilitated by substituting some of the necessary subsystems with simpler analogs. Should the simpler *in vitro* model turn out to be a poor model for the more complex *in vivo* system, one could always construct a more complex *in vitro* system that may better reflect *in vivo*.

Synthesizing a cell will also lead to new applications. Purified biochemical systems already offer major advantages, such as the polymerase chain reaction (PCR) and *in vitro* transcription. A better understanding and manipulation of all cellular replication subsystems (molecular biology’s tool kit) should spin off new technologies. For example, *in vitro* genome replication may be useful for replicating very large segments of DNA with high fidelity. Combined *in vitro* transcription, RNA processing and RNA modification would allow preparation of rRNAs and tRNAs with defined modifications to test the roles of the modifications, and modified tRNAs

to aid incorporation of unnatural amino acids into proteins. Purified translation systems have enabled reassignment of mRNA codons to encode unnatural amino acids by omission of competing natural amino acids (Forster *et al*, 2003); further improvements of the purified translation system could enable the genetic selection of protease-resistant, peptide-like ligands for drug discovery by pure translation display (Forster *et al*, 2004). The purified translation system may also facilitate expression of proteins difficult to express by standard approaches. Better control of lipid vesicle synthesis could advance liposome-based drug delivery. Since bacterial translation is the main target of antibiotics, greater understanding may assist development of new drugs to fight mounting antibiotic resistance. Ultimate success in cell synthesis could generate useful microorganisms, for example, for renewable production of biodegradable plastics (Pohorille and Deamer, 2002).

## Approach

The ideal approach for synthesizing a cell would allow all of the machine parts to be understood and tested. Like any engineering project, this requires detailed blueprints, raw synthetic capabilities and an overall diagnostic and debugging strategy. The use of entire genomes as the blueprints, some of which are small enough to synthesize *de novo*, is inconsistent with this approach. Self-replication of an unadulterated genome, however impressive, would not define the unnecessary genes, and the functions of about a third of the genes would remain unknown (Fraser *et al*, 1995; Jaffe *et al*, 2004). Building a machine from mysterious parts can only create a mysterious machine. What is needed is some way of defining a near-minimal genome and then a strategy that will lead inexorably to an understanding of all of its parts.

Theoretical and experimental studies have attempted to establish a minimal set of genes needed for a self-replicating system in a cushy constant environment of unlimited, small molecule nutrients. Three basic approaches present themselves.

## Comparative genomics

Comparative genomics searches for genes that have homologs in the genomes of groups of organisms. The approach estimates from 50 to 380 genes in a minimal genome (Mushegian and Koonin, 1996; Tomita *et al*, 1999; Koonin, 2000; Jaffe *et al*, 2004). It has the caveat that, among closely related genomes, some genes appear 'required' for those species (e.g. many of the genes retained in the synthetic reduced genome *Escherichia coli* (Posfai *et al*, 2006)) although they are not required for basic life. If one goes to longer evolutionary distances, many gene functions are replaced by non-homologous genes, hence making some essential genes look dispensable (e.g. some of the tRNA modification enzymes used by *Mycoplasma* are either different from *E. coli* or unidentified by sequence identity, but that does not mean the different ones are dispensable). An additional challenge is that about a third of the essential genes have unknown functions. It is thus expected that a minimal genome based on this

approach alone would be inviable, and it would not be possible to identify the missing essential genes.

## Genetics

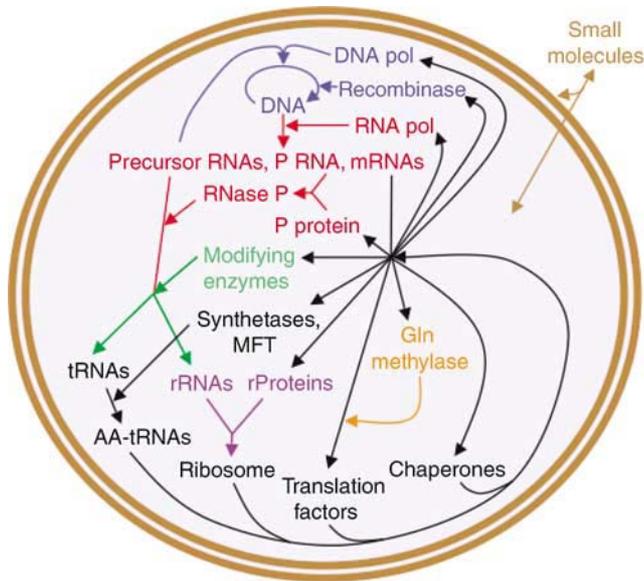
Genetics searches for essential genes by mutating one gene at a time. This approach estimates 430 genes in a minimal genome (out of *Mycoplasma genitalium*'s total of 528; Supplementary Table S3; Hutchison *et al*, 1999; Glass *et al*, 2006). About a fifth of these essential genes have unknown functions. It is limited by false 'essentials' due to the fraction of genes that were never mutated in the screen, due to creation of toxic partial complexes or pathways, and due to inadvertent effects on adjacent genes. The latter effects are prevalent in bacteria because a primary RNA transcript typically encodes multiple gene products. At the other extreme, false 'dispensables' are disastrous when trying to assemble a viable minimal genome that lacks all of the individual 'dispensables'. For example, most RNA modification enzymes are individually dispensable, but simultaneous deletion of tens of them would be expected to be unsustainable due to cumulative reductions in efficiency or fidelity (a useful working definition of essentials for a minimal genome should encompass such lethal 'dispensables'). Again, in using this approach alone, it would not be possible to identify the missing essential genes.

## Biochemistry

Biochemistry identifies from cell fractions those gene products essential for the reconstitution of biochemical reactions. It does not suffer from the above problems (except creation of toxic partial complexes), gives access to details of kinetic steps and allows debugging of isolated subsystems. However, the cellular subsystems must be integrated and thoroughly tested for accuracy on long templates before they can be considered physiological. Nevertheless, the biochemical approach has been successful at identifying macromolecules sufficient for reconstituting DNA, RNA and protein syntheses and, based on individual subtraction experiments, the components have either been shown to be necessary or could be so tested. Mindful of the remaining self-replication functions that need to be discovered (see below), it seems likely that a largely biochemical approach, now further empowered by mass spectrometry analyses and genetic and comparative genomic information, will be the most practical route to define a near-minimal, well-understood genome. We now review the relevance of current knowledge and technology to this new minimal cell project (MCP; Luisi, 2002).

## A minimal genome

An MCP may be realized by reconstituting the macromolecular catalysts that synthesize DNA, RNA and protein. However, this overlooks the formation of the membrane compartment and the poorly understood process in which it is divided by membrane proteins (Gitai, 2005), both of which are required for life. But lipids alone have been shown to be sufficient for formation of rudimentary membranous compartments capable of both transmembrane transport of small molecules and



**Figure 1** A minimal cell containing biological macromolecules and pathways proposed to be necessary and sufficient for replication from small molecule nutrients. The macromolecules are all nucleic acid and protein polymers and are encapsulated within a bilayer lipid vesicle. The small molecules (brown) diffuse across the bilayer. The macromolecules are ordered according to the pathways in which they are synthesized and act. They are colored by biochemical subsystem as follows: blue=DNA synthesis, red=RNA synthesis and cleavage, green=RNA modification, purple=ribosome assembly, orange=post-translational modification and black=protein synthesis. MFT=methionyl-tRNA<sup>Met</sup> formyltransferase. The system could be bootstrapped with DNA, RNA polymerase, ribosome, translation factors, tRNAs, MTF, synthetases, chaperones and small molecules.

fission autocatalytically (Szostak *et al*, 2001), so membrane proteins may be dispensable. Polysaccharides should also be dispensable. If the simplest and best-characterized examples of DNA, RNA and protein synthesis are selected, if translation of all codons is enabled for generalizability and if efficiency and accuracy are not compromised, then this leads to the macromolecules and pathways of Figure 1.

A detailed list of the gene products in the hypothetical synthetic minimal cell of Figure 1 is shown in Table I (left column). This list overlaps with a computational model of minimal cell genes largely derived from a minimal organism, *M. genitalium* (Tomita *et al*, 1999; Supplementary Table S4), but differs by omitting enzymes for synthesizing small molecules (e.g. lipids and glycolysis substrates) and by including DNA replication, RNA processing, RNA modification, extra tRNAs to decode the whole genetic code, some additional essential translation components and chaperones. It should be emphasized that Table I is a working model only and that strict adherence will likely hamper progress. Examples of omitted, potentially stimulatory genes are given below and in Supplementary Table S1. Conversely, examples of included, potentially dispensable genes may be gleaned by comparison with the streamlined *Mycoplasma* genome (Fraser *et al*, 1995; Table I, middle column; Supplementary Table S2).

Several conclusions can be drawn from the provisional list of genes selected for a minimal cell, most of which are attractive when contemplating an MCP. In genomic terms, the

list is very short, containing only 151 genes and 113 kbp. All of the genes are derived from *E. coli* and its bacteriophages (except for the hammerhead RNA from a plant virus; Forster and Symons, 1987), implying that the individual subsystems will be compatible. In contrast to lists derived by comparative genomics or genetic approaches, the biochemically based list does not contain any genes of unknown function or challenging membrane proteins; so it is close to a fully understood, accurately replicating 'platform' for life. The few known gaps constitute only about seven genes, all of which are predicted to be for RNA modification (Table I, bold in the left column). From the viewpoint of structural biology, courtesy of recent breakthroughs in ribosome structure determination (Diaconu *et al*, 2005; Ogle and Ramakrishnan, 2005), significant three-dimensional information is lacking for only 3% of the products: a few RNA modification proteins and aminoacyl-tRNA synthetases (Table I, right column). While some of the states and complexes remain to be solved at high resolution, a draft three-dimensional structure for any replicating system is a major milestone in the history of biology.

## Tools

Genes for an MCP could be synthesized using either natural or unnatural gene sequences as starting points. Using natural gene sequences, genes can be readily synthesized by PCR, and large cloned operons of essential genes can be fused using synthetic linkers and homologous recombination. However, gene synthesis by cloning and PCR will soon be more expensive than raw synthesis from synthetic oligodeoxyribonucleotides (oligos). The latter also allows unnatural sequences, such as versions with altered codon bias to adjust mRNA secondary structures (Tian *et al*, 2004). Scalability and cost limitations of established methods for gene synthesis from synthetic oligos are now being overcome by oligo synthesis on chips followed by PCR amplification and error correction (Carr *et al*, 2004; Richmond *et al*, 2004; Tian *et al*, 2004; Zhou *et al*, 2004).

## Biochemical subsystems

Several biochemical subsystems are required to synthesize a minimal cell, and they are reviewed here. For each subsystem, possible examples from natural systems will be compared, gaps in knowledge will be identified and diagnostic and debugging strategies to fill the gaps will be suggested. Mindful of the goal of integration of the subsystems, emphasis is placed on subsystems that are homologous and that operate under standard physiological conditions.

## Genome replication

In principle, the genetic material for an MCP could be either DNA or RNA. Although an RNA genome has the advantage of obviating genes for DNA replication, the challenges of preventing inhibitory double-stranded RNA structures and replicative mutations in artificial RNA genomes (Mills *et al*, 1967) are unsolved. So the genetic material for an MCP should be DNA.

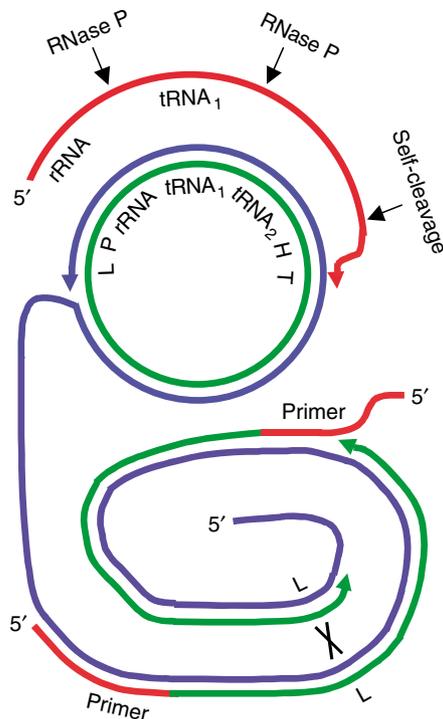
**Table 1** Biochemically derived list of genes that may encode a useful, near-minimal, self-replicating system dependent only on small molecule nutrients

<i>Escherichia coli</i>	Mycoplasma	3D structure
Coliphage f29 DNA polymerase	+	+
Coliphage P1 Cre recombinase	–	+
>Coliphage Lox/Cre recombinase site	–	+
Coliphage T7 RNA polymerase	Analog	+
>Coliphage T7 RNA polymerase initiation site	Analog	+
>Coliphage T7 RNA polymerase class II termination site	Analog	+
Lucerne viral hammerhead RNA	–	+
RNase P RNA	+	+
RNase P protein	+	+
>RNase P site/RNA primer for DNA polymerase	+	+
Small subunit 16S ribosomal RNA	+	+
All 21 small subunit ribosomal proteins (1–21)	+ except 1, 21	+
Large subunit 5S ribosomal RNA	+	+
Large subunit 23S ribosomal RNA	+	+
<b>Large subunit 23S rRNA G2445 &gt; m2G methylase: unidentified</b>	<b>Unknown</b>	–
<b>Large subunit 23S rRNA U2449 &gt; dihydroU synthetase: unidentified</b>	<b>Unknown</b>	–
Large subunit 23S rRNA U2457 > pseudoU synthetase	<b>Unknown</b>	–
<b>Large subunit 23S rRNA C2498 &gt; Cm methylase: unidentified</b>	<b>Unknown</b>	–
<b>Large subunit 23S rRNA A2503 &gt; m2A methylase: unidentified</b>	<b>Unknown</b>	–
Large subunit 23S rRNA U2504 > pseudoU synthetase	<b>Unknown</b>	–
All 33 large subunit ribosomal proteins (1–7, 9–11, 13–25, 27–36)	+ except 25, 30	+
Translational initiation factor 1	+	+
Translational initiation factor 2	+	+
Translational initiation factor 3	+	+
Translational elongation factor Tu	+	+
Translational elongation factor Ts	+	+
Translational elongation factor G	+	+
Translational release factor 1	+	+
Translational release factor 2	–	+
Translational release factor Gln methylase	+	+
Translational release factor 3	–	+
Ribosome recycling factor	+	+
<b>33/45 tRNAs (see Figure 3)</b>	Set of 29	+
tRNA C34 > lysidine synthetase	<b>Unidentified</b>	+
tRNA A34 > I deaminase	<b>Unidentified</b>	+
<b>tRNA U34 &gt; cmo5U (=V) synthetases: unidentified</b>	–	–
tRNA U34 > 2sU Cys desulfurase	–	+
tRNA U34 > 2sU synthetase	<b>Unidentified</b>	+
tRNA U34 > cmnm5U GTPase	<b>Unidentified</b>	+
tRNA U34 > cmnm5U synthetase	<b>Unidentified</b>	+
tRNA cmnm5U34 > nm5U > mnm5U synthetase	<b>Unidentified</b>	–
tRNA G37 N1-methylase	+	+
<b>tRNA A37 &gt; t6A N6-threonylcarbamoyl-A synthetase: unidentified</b>	<b>Unidentified</b>	–
tRNA A37 > i6A synthetase	–	+
tRNA i6A37 > s2i6A > ms2i6A synthetase	–	+
All 22 aminoacyl-tRNA synthetase subunits (20 enzymes)	+ except Gly sub., Gln	+ except Gly sub., Ala
Met-tRNA formyltransferase	+	+
Chaperonin GroEL	+	+
Chaperonin GroES	+	+
151 genes=38 RNAs + 113 proteins		

Gaps in knowledge are in bold. Left column: chosen gene products and DNA sites. Middle column: relationship to the minimal genome of *M. genitalium*; clear sequence homolog='+'; known enzyme product without an evident sequence homolog='unidentified'; no functional homolog='–'. Right column: high-resolution, three-dimensional, structural information; >25% of the structure solved='+', <25%='–'. The small molecules known to be required are four dNTPs, four NTPs, 20 amino acids, N-5,10-methenyltetrahydrofolate, S-adenosylmethionine and isopentenyl pyrophosphate). Note: a full version listing the nomenclature, database link, length and sequence of each individual product is available in Supplementary Tables S1 and S2.

A simple possible scheme for DNA replication that could be completely integrated with biological systems is shown in Figure 2. It shows rolling-circle DNA strand displacement (Zhong *et al*, 2001) initiated with RNA transcript primers synthesized *in situ* by an RNA polymerase. Processing of the

resulting double-stranded DNA concatemers into monomeric DNA circles occurs by homologous recombination at Lox sites catalyzed by Cre recombinase (Sauer, 2002). This approach has advantages over existing rolling-circle (Dahl *et al*, 2004) or PCR (Mitra and Church, 1999) replication methods, as it



**Figure 2** A generalizable, physiologically compatible, theoretical scheme for accurate DNA replication and RNA synthesis *in vitro*. Polymerase movements are illustrated by colored arrowheads. **DNA synthesis:** a nicked double-stranded DNA circle (middle) undergoes rolling-circle DNA synthesis by coliphage  $\phi$ 29 DNA polymerase (Dahl *et al*, 2004) to give an oligomeric single-stranded DNA (bottom, blue). RNA primers (red) then hybridize at two sites to prime lagging strand DNA synthesis (bottom, green). When two Lox sites (bottom, L) are completed, recombination occurs between them catalyzed by coliphage P1 Cre recombinase (black cross) to form a duplicate of the original circular template. **RNA synthesis:** the circular genetic operon (middle) contains a promoter for T7 RNA polymerase (P), a ribosomal RNA (rRNA) gene, two transfer RNA (tRNA) sequences, a self-cleaving hammerhead sequence (H) and a T7 terminator (T). RNA synthesis from P generates a precursor RNA (top, red) containing three cleavage sites (thin black arrows). The second tRNA sequence merely serves as a recognition site for RNase P cleavage. Cleavages yield the mature rRNA and tRNA<sub>1</sub>. Any cleavage product containing a 3' hydroxyl group or primary RNA transcript can serve as a primer for DNA synthesis (bottom, red).

requires neither solid-phase oligo synthesis nor changes in temperature, and is far simpler than natural DNA replication systems (Khan, 1997).

Rolling-circle DNA strand displacement could be engineered in a stepwise manner. First, a simpler version could be constructed in which the T7 RNA polymerase and RNA processing are substituted by addition of short RNA primers to test the effect of multiple initiation sites. The efficiency of synthesis of monomeric DNA circles would be followed by gel electrophoresis (Dahl *et al*, 2004), and replication fidelity at the base pair and whole genome levels should be tested with different polymerases. The biggest challenge anticipated is boosting the efficiency of monomeric circular template generation over by-products, such as linear DNAs or oligomeric circles. Such defective by-products would also be replicated and compete for nutrients (like PCR deletion products or defective interfering viruses). Defective by-

products potentially could be weeded out with appropriate selection schemes. For example, encapsulation of individual genomes within membranous cells would result in non-viability of cells containing deleted genomes. However, encapsulation would raise new challenges, especially for large genomes. This might be aided by compacting the DNA through addition of DNA gyrase.

## Transcription

A single RNA polymerase should suffice for an MCP. *E. coli*'s multi-subunit enzyme (Lewin, 2004) or the single polypeptide enzyme encoded by coliphage T7 (Studier *et al*, 1990) seems to be the best, with the choice influenced by several considerations that also determine possible modes of regulation. In considering the whole transcription cycle for a minimal replicating system, the simpler, more predictable T7 RNA polymerase is arguably a better starting point than the *E. coli* RNA polymerase (a detailed comparison is provided in Supplementary information).

## RNA processing

A host of RNases cleave precursor RNAs *in vivo* (Li and Deutscher, 1996) with a complexity that could be reproduced in an MCP. However, inclusion of these RNases comes with the risks of cryptic cleavages, and a simpler approach may be easier to engineer (Figure 2, top). This approach generates all required unadulterated termini: tRNA 5' and 3' ends (Forster and Altman, 1990) and, if necessary, the 3' end of an rRNA. The self-cleaving sequence (Forster and Symons, 1987) is included because precursor tRNAs with substantial 3' extensions can be poor substrates for RNase P (Li and Deutscher, 1996) and RNA polymerase terminators are inefficient. The efficiency of RNA processing, monitored by gel electrophoresis, could be improved by trying several different precursor-specific sequences.

## A minimal translome

The most complex universal biological machinery is clearly translation. Translation-associated genes (the 'translatome') account for a large fraction of cellular genes, 96% of the genes in Table I, and all of the currently predicted gaps in knowledge of an MCP. The eukaryotic version is less attractive for engineering than the bacterial version because it contains some 30 initiation factor proteins and because eukaryotic ribosome assembly *in vitro* awaits the coordination of more than a hundred non-ribosomal macromolecules (Fromont-Racine *et al*, 2003). Of the bacterial systems, *Mycoplasma* has advantages over *E. coli* owing to its eight-fold-smaller minimal genome and its simple set of 29 tRNAs that is the only completely characterized set (Andachi *et al*, 1989). Unfortunately, other important biochemical information for *Mycoplasma* is essentially unknown in areas where it is well studied in *E. coli* (e.g. reconstitution of ribosomes and translation, characterization and functional assays of rRNA modifications, characterization of RNA modification enzymes). Presently, this seems to favor the *E. coli* translome for an MCP.

## Purified translation

Efficient synthesis of proteins has been reconstituted from purified natural components (Kung *et al*, 1978) or recombinant His-tagged translation factors (Shimizu *et al*, 2005) from *E. coli*, but not yet from eukaryotes. The next steps with the *E. coli* system will be verifying accuracy by mass spectrometry and extending the short lifetime of the batch mode by continuous dialysis (Spirin *et al*, 1988). The versatility of the system will become apparent as more mRNAs are translated. If stronger mRNA secondary structures prove inhibitory despite the helicase activity of the ribosome (Takyar *et al*, 2005), introduction of an RNA helicase may be helpful. Given that aminoacyl-tRNA synthetases, translation factors and ribosomal proteins are among the most abundant proteins in the cell, it will be important to verify that the purified system can produce high concentrations of all of these proteins.

## An *in vitro* ribosome

The ribosome of choice is from *E. coli* because, in contrast with its eukaryotic cousins, it has been self-assembled from its purified components (Traub and Nomura, 1968; Nomura and Erdmann, 1970; Nierhaus and Dohme, 1974) and is homologous with the other components of the gene list (Table I). Reconstituted ribosomes have only been assayed by synthesis of phenylalanine polymers from polyU templates (Lietzke and Nierhaus, 1988); so future assays need to test initiation and elongation at non-UUU codons, and also termination. Furthermore, the self-assembly protocol is finicky and non-physiological. *In vitro* assembly of the 30S subunit under physiological temperatures has been attained recently by adding the DnaK/DnaJ/GrpE chaperone system (Maki and Culver, 2005), although this system is dispensable *in vivo* (El Hage *et al*, 2001). Perhaps addition of natural polyamines might overcome the requirement for an unphysiologically high concentration of magnesium ions. All 54 of the ribosomal proteins have been cloned (Culver and Noller, 1999; Semrad *et al*, 2004); the hypothesis that they (and other proteins in Table I) can be synthesized in a purified translation system in active forms warrants testing.

rRNA production in a purified system is complicated by post-transcriptional nucleoside modifications. Since 5S rRNA lacks nucleoside modifications and is short, it is not surprising that it is active when transcribed *in vitro* (Zvereva *et al*, 1998). But the other two rRNAs are modified by about 20 enzymes in *E. coli*, half of which are unidentified. All 11 modifications of the *E. coli* small subunit 16S rRNA are dispensable for subunit assembly and aminoacyl-tRNA binding (Krzyszosiak *et al*, 1987). However, *E. coli* 23S rRNA lacking its 23 modifications is 30-fold less active than the natural version in N-Ac-Met-puromycin synthesis (Semrad and Green, 2002) due to one to six modifications in a relatively small RNA domain (Green and Noller, 1996). The enzymes that catalyze these six modifications are therefore included in Table I, although the two known ones are individually dispensable (Del Campo *et al*, 2001). Other bacteria should also be entertained for an MCP, as these six *E. coli* modifications are not conserved and the unmodified 23S RNAs from two other eubacteria are quite active (Green and Noller, 1999; Khaitovich *et al*, 1999).

## *In vitro* tRNAs

Which of the myriad tRNA genes and tRNA modification enzymes are likely to be sufficient to decode all 61 sense codons in an MCP? There are some 85 tRNA genes in *E. coli* coding for some 45 different tRNAs each bearing post-transcriptional modifications on about 10% of their nucleosides (Supplementary Table S5), and a fifth of the tRNAs still remain to be characterized at the modification level. At least 27 different types of nucleoside modifications are present in *E. coli* (Bjork, 1995). There are an estimated 40–50 tRNA modification enzymes in *E. coli*, about half of which remain to be identified. To make matters worse (or more interesting) for an MCP, the roles of the tRNA modifications are controversial.

Arguments for choosing essential tRNA modification activities are highly speculative (detailed in Supplementary information). As few as 33 *E. coli* tRNAs may be sufficient to translate the entire genetic code accurately (Table I, left, and Figure 3). *E. coli* tRNAs could be substituted with the completely characterized set from *Mycoplasma capricolum* (Supplementary Table S7), which contains only 14 types of nucleoside modifications (Andachi *et al*, 1989), some of which differ from *E. coli* (Supplementary Table S1). However, the predicted savings in the number of essential tRNAs and modification enzymes are minor (Table I, middle column), and full compatibility with the heterologous *E. coli* translation apparatus seems unlikely (e.g. the codon UGA in *Mycoplasma* encodes Trp, not stop).

Each *in vitro*-synthesized nascent tRNA transcript should be modified with different combinations of modification enzymes and tested for efficiency and accuracy of codon recognition in translation, initially in a simplified purified translation system (Forster *et al*, 2001). Identification of the unknown modification enzymes is being hastened by bioinformatic and genomic approaches (Soma *et al*, 2003). It is also conceivable, although unlikely, that unknown small molecules would need to be identified biochemically for RNA modification (or other reactions). The remaining *E. coli* tRNA modification enzymes not listed in Table I might be predicted to be dispensable based on available data (Bjork, 1995; Giege *et al*, 1998). But given the uncertainties, it may be faster to get to a working near-minimal cell by using every known *E. coli* modification enzyme. Such a system would be ideally suited for freeing up codons to encode unnatural amino acids: this would be carried out by omission of one or more codons from all mRNAs and omission of their cognate tRNAs.

## Post-translation

An MCP must promote correct protein folding and any necessary post-translational amino-acid modifications. Early versions of a purified replicating system will contain cell-derived macromolecules, so establishing that such systems can be completely weaned from cells will require enough rounds of replication for 'infinite' dilution of the starting macromolecules. This will test for dependence on folding by chaperones and on post-translational modifications. It is unclear which, if any, chaperones will be necessary, but GroEL/ES (El Hage *et al*, 2001; Kerner *et al*, 2005) are likely

		2nd base					
		U	C	A	G		
5' base	U	Phe GAA <sup>ms<sup>2</sup>i<sup>6</sup>A</sup>	Ser GGAA		Tyr GUA <sup>ms<sup>2</sup>i<sup>6</sup>A</sup>	Cys GCA <sup>ms<sup>2</sup>i<sup>6</sup>A</sup>	CUA UAG
		Leu cmnm <sup>5</sup> UAA <sup>ms<sup>2</sup>i<sup>6</sup>A</sup>		Ser ms <sup>2</sup> i <sup>6</sup> A VGA	RF1	RF2	
	C	Leu GAG <sup>m<sup>1</sup>G</sup>	Pro GGG <sup>m<sup>1</sup>G</sup>		His GUGA	Arg ICGA	CUA UAG
		Leu UAGG		Pro VGG <sup>m<sup>1</sup>G</sup>	Gln cmnm <sup>5</sup> s <sup>2</sup> UUG <sup>A</sup>	Arg CCG <sup>m<sup>1</sup>G</sup>	
	A	Ile GAU <sup>t6A</sup>	Thr GGU <sup>t6A</sup>		Asn GUU <sup>t6A</sup>	Ser GCU <sup>t6A</sup>	CUA UAG
		Ile k2CAU <sup>t6A</sup>		Thr VGU <sup>t6A</sup>	Lys SUU <sup>t6A</sup>	Arg mnm <sup>5</sup> UCU <sup>t6A</sup>	
		fMet CAU <sup>A</sup>	Met CAU <sup>t6A</sup>				
	G	Val GACA		Ala GGCA	Asp GUCA	Gly GCCA	CUA UAG
			Val VACA	Ala VGCA	Glu SUCA	Gly U <sup>*</sup> CCA	

**Figure 3** All nucleoside modifications of all 33 synthetic tRNAs that may be sufficient for accurate translation. Outside (shaded): mRNA codons of the genetic code are illustrated in the standard format, except that the 3' U and C are switched to simplify depiction of decoding. Inside: tRNA nucleotides 34–37 (from 5' to 3') and their cognate amino acids. Nucleotides 34–36 are the anticodons, and the 37th nucleotides are represented by black superscripts. Codon and anticodon positions that base pair with each other are colored similarly. Stop codon specificities of release factor (RF) proteins are included. The portions of the tRNA sequences not shown in the figure are unmodified. Expected modifications of *in vitro* transcripts by the enzymes in Table I, and expected amino-acid and codon specificities are given. \* = unspecified modification, \_ = unknown modification status, ms<sup>2</sup>i<sup>6</sup>A = 2-methylthio-*N*<sup>6</sup>-isopentenyladenosine, m<sup>1</sup>G = 1-methylguanosine, t6A = *N*<sup>6</sup>-threonylcarbamoyl-adenosine, cmnm<sup>5</sup>U = 5-carboxymethylaminomethyluridine, V = cmo<sup>5</sup>U = uridine 5-oxoacetic acid, I = inosine, cmnm<sup>5</sup>s<sup>2</sup>U = 5-carboxymethylaminomethyl-2-thiouridine, k2C = lysidine, S = mnm<sup>5</sup>s<sup>2</sup>U = 5-methylaminomethyl-2-thiouridine, mnm<sup>5</sup>U = 5-methylaminomethyluridine.

candidates (Table I). The only known examples of required post-translational modifications for the proteins in Table I are the recently discovered methylations of translation release factors 1 and 2 catalyzed by release factor Gln methylase (Table I) (Heurgue-Hamard *et al*, 2002; Nakahigashi *et al*, 2002). Other possibilities include ribosomal protein acetylations. Mass spectral comparisons between proteins made in the purified system and those made *in vivo* will expose modifications and also assess fidelity, while the inactivity of a protein of expected mass would suggest a protein-folding deficit and the need for an additional chaperone. Any necessary missing components could be identified biochemically by mixing with fractionated crude extracts or through genetics.

### Compartments and division

Membranes would allow evolution without serial transfers and purifications, extension of the system to new environments and better modeling of cells. On the other hand, membranous boundaries are unnecessary for directed evolution (Mills *et al*, 1967) or, in theory, self-replication. Membranes also restrict applications (e.g. delivery of unnatural amino acyl-tRNAs, selection schemes based on binding and

spatial arraying for nanofabrication). Addition to self-replicating macromolecules of lipids alone may be sufficient for encapsulation of the macromolecules within bilayer membrane vesicles, synthetic cell division and transmembranous small molecule transport (Szostak *et al*, 2001). The choice of lipids is wide open, but one should not underestimate the challenges involved in working with them (Luisi, 2002) nor the advantages in regulation to be gained by adding membrane-modeling proteins (e.g. pores, transporters and the yet-to-be-discovered complement of cell division proteins; Gitai, 2005).

### Integrating the subsystems

How might all of the biochemical subsystems in Figure 1 be combined to generate a self-sustaining system? This is clearly a new level of complexity in comparison with prior self-assembly projects. None of the subsystems described above are completed, yet their selection is based on a reasonable plan for their ultimate integration. The approach again would be stepwise, and there are many possible pathways that could be integrated in parallel (Figure 1). For example, transcription by T7 RNA polymerase couples well with a purified *E. coli* translation system (Shimizu *et al*, 2005). Theoretical integration of DNA synthesis, RNA synthesis and RNA processing was

discussed above (Figure 2). These four different subsystems could then be combined to synthesize part of a fifth system (the ribosome) by synthesis of an antibiotic-resistant 16S rRNA and His-tagged versions of all 21 small subunit ribosomal proteins (Tian *et al.*, 2004). The products of these integrated subsystems could then be assayed for correct *in vitro* reconstitution of small ribosomal subunits by (i) selecting for resistance of protein synthesis to the antibiotic, and (ii) detecting the presence of tagged proteins in purified small ribosomal subunits by Western blot with anti-His antibodies. As another example, rudimentary vesicles encapsulating replicating systems (e.g. Q $\beta$  replicase) were shown to be capable of multiplication (Luisi, 2002).

Numerous fine-tuning strategies can be envisioned. Relative strengths of DNA promoters and mRNA ribosome-binding sites for different genes could be modeled on the *in vivo* strengths, with necessary adjustments of synthetic rates (and thus concentrations of products) achieved by mutations in the binding sites (see Supplementary information on transcription). Additional modules might be useful, such as catabolism (nucleases and proteases), active conversion or removal of waste products (e.g. by energy regenerating enzymes (Supplementary Table S1) or membrane transporters) and regulatory feedback (e.g. excess transcription  $\rightarrow$  excess T7 lysozyme mRNA  $\rightarrow$  excess lysozyme  $\rightarrow$  lysozyme binding to and inhibition of T7 RNA polymerase). Control of macromolecular concentrations will be aided by *in silico* modeling and design (Tomita *et al.*, 1999). Given that the subsystems discussed above were selected with integration in mind by choosing physiological reaction conditions and homologous components, and given that additional subsystems could always be borrowed from living cells as needed (e.g. *E. coli* RNA polymerase (Supplementary Table S1) and regulatory modules such as riboswitches (Isaacs *et al.*, 2004)), it seems likely that this approach will eventually produce synthetic self-replication and ultimately a self-sustaining minimal cell.

It is important to note that a minimal cell would be intentionally fragile. For example, the vesicle would be easily lysed and the small molecule feeding mix would be highly specialized indeed (including unstable cofactors such as N-5,10-methenyltetrahydrofolate and S-adenosylmethionine). These built-in safety features will prevent a minimal cell from replicating outside the laboratory. However, some or all of the synthetic genes for an MCP would be intentionally passaged through living cells for construction of recombinant DNA clones and for amplification. Constantly upgraded ethical and safety regulations in place for existing biohazards would also encompass this research (Cho *et al.*, 1999; [http://arep.med.harvard.edu/SBP/Church\\_Biohazard04c.htm](http://arep.med.harvard.edu/SBP/Church_Biohazard04c.htm)).

## Completion

In conclusion, a stepwise biochemical approach lends itself to the eventual identification of any remaining functions essential for the synthesis of a minimal cell sustained solely by small molecules. Five states of completion present themselves as tractable goals of an MCP. Namely, the identification of

(1) the genes listed as missing in Table I,

- (2) any additional genes and organization necessary experimentally for minimal cell synthesis,
- (3) any dispensable genes,
- (4) biochemical parameters and computational models sufficiently detailed to predict the effects of alterations and
- (5) the missing three-dimensional structures of the gene products and their relevant complexes.

It is difficult to predict how long it will take to debug each of the individual biochemical subsystems or to put them all together; so it is important to bear in mind that there are short-term goals (see the Utility section). Intermediate assembly steps could also be pursued while the gaps in RNA modification knowledge (Table I) are being filled. For example, the project to assemble a ribosome under physiological conditions could be carried out without the missing 23S rRNA modification enzymes (Table I) by substituting in natural 23S rRNA. Similarly, assembly of self-replication in the absence of functional *in vitro*-synthesized tRNA substrates could be carried out using cellular total tRNA to enable self-replication from substrates (rather than just small molecules) as a major step towards understanding biological self-replication. This would also allow directed evolution of all of the components except the tRNAs in a more flexible manner than is possible *in vivo* (e.g. for selecting ribosome mutants that incorporate unnatural amino acids more efficiently).

The biochemical subsystems necessary for an MCP are central, old fields that have lost impetus. Completion within a decade will only be possible through a coordinated filling of the key gaps in knowledge by the cutting-edge laboratories scattered around the world in these fields. It will also require stimulation of rate-limiting fields. For example, although rRNAs and tRNAs can constitute more than 70% of the dry weight of a cell, half of the estimated 60–70 RNA modification enzymes of *E. coli* and one-fifth of the tRNAs remain to be characterized (Supplementary Tables S5 and S6), despite the recent completion of about 300 bacterial whole genome sequences. The momentum of genomics and consequent deluge of computed hypotheses cries out for comparable breakthroughs in experimental tests. Synthetic systems biology projects such as an MCP promise such tests with the added bonus of new applications.

## Supplementary information

Supplementary information is available at *Molecular Systems Biology* website ([www.nature.com/msb](http://www.nature.com/msb)) and at <http://arep.med.harvard.edu/SBP>

## Acknowledgements

We acknowledge Dr Glenn Björk for help with the compilation and review of the tRNA modification data, the late Dr James Ofengand and Dr William Studier for advice and many colleagues for comments on the manuscript. This work was supported by an NIH K08 grant (to ACF) and a DOE GTL Center grant (to GMC). ACF's work was largely performed in the Department of Pathology, Brigham and Women's Hospital, Harvard Medical School under the exceptional mentorship of Dr Stephen Blacklow.

## References

- Andachi Y, Yamao F, Muto A, Osawa S (1989) Codon recognition patterns as deduced from sequences of the complete set of transfer RNA species in *Mycoplasma capricolum*. Resemblance to mitochondria. *J Mol Biol* **209**: 37–54
- Bjork GR (1995) Biosynthesis and function of modified nucleosides. In Dieter Söll, Uttam RajBhandary (eds), *tRNA: Structure, Biosynthesis, and Function*, pp 165–205. Washington, DC: ASM Press
- Carr PA, Park JS, Lee YJ, Yu T, Zhang S, Jacobson JM (2004) Protein-mediated error correction for *de novo* DNA synthesis. *Nucleic Acids Res* **32**: e162
- Cho MK, Magnus D, Caplan AL, McGee D (1999) Policy forum: genetics. Ethical considerations in synthesizing a minimal genome. *Science* **286**: 2087–2090
- Culver GM, Noller HF (1999) Efficient reconstitution of functional *Escherichia coli* 30S ribosomal subunits from a complete set of recombinant small subunit ribosomal proteins. *RNA* **5**: 832–843
- Dahl F, Baner J, Gullberg M, Mendel-Hartvig M, Landegren U, Nilsson M (2004) Circle-to-circle amplification for precise and sensitive DNA analysis. *Proc Natl Acad Sci USA* **101**: 4548–4553
- Del Campo M, Kaya Y, Ofengand J (2001) Identification and site of action of the remaining four putative pseudouridine synthases in *Escherichia coli*. *RNA* **7**: 1603–1615
- Diaconu M, Kothe U, Schlunzen F, Fischer N, Harms JM, Tonevitsky AG, Stark H, Rodnina MV, Wahl MC (2005) Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. *Cell* **121**: 991–1004
- El Hage A, Sbai M, Alix JH (2001) The chaperonin GroEL and other heat-shock proteins, besides DnaK, participate in ribosome biogenesis in *Escherichia coli*. *Mol Gen Genet* **264**: 796–808
- Forster AC, Altman S (1990) External guide sequences for an RNA enzyme. *Science* **249**: 783–786
- Forster AC, Cornish VW, Blacklow SC (2004) Pure translation display. *Anal Biochem* **333**: 358–364
- Forster AC, Symons RH (1987) Self-cleavage of virusoid RNA is performed by the proposed 55-nucleotide active site. *Cell* **50**: 9–16
- Forster AC, Tan Z, Nalam MNL, Lin H, Qu H, Cornish VW, Blacklow SC (2003) Programming peptidomimetic syntheses by translating genetic codes designed *de novo*. *Proc Natl Acad Sci USA* **100**: 6353–6357
- Forster AC, Weissbach H, Blacklow SC (2001) A simplified reconstitution of mRNA-directed peptide synthesis: activity of the epsilon enhancer and an unnatural amino acid. *Anal Biochem* **297**: 60–70
- Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, Fritchman RD, Weidman JF, Small KV, Sandusky M, Fuhrmann J, Nguyen D, Utterback TR, Saudek DM, Phillips CA, Merrick JM, Tomb JF, Dougherty BA, Bott KF, Hu PC, Lucier TS, Peterson SN, Smith HO, Hutchison III CA, Venter JC (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**: 397–403
- Fromont-Racine M, Senger B, Saveanu C, Fasiolo F (2003) Ribosome assembly in eukaryotes. *Gene* **313**: 17–42
- Giege R, Sissler M, Florentz C (1998) Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res* **26**: 5017–5035
- Gitai Z (2005) The new bacterial cell biology: moving parts and subcellular architecture. *Cell* **120**: 577–586
- Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, Hutchison CA, Smith HO, Venter JC (2006) Essential genes of a minimal bacterium. *Proc Natl Acad Sci USA* **103**: 425–430
- Green R, Noller HF (1996) *In vitro* complementation analysis localizes 23S rRNA posttranscriptional modifications that are required for *Escherichia coli* 50S ribosomal subunit assembly and function. *RNA* **2**: 1011–1021
- Green R, Noller HF (1999) Reconstitution of functional 50S ribosomes from *in vitro* transcripts of *Bacillus stearothermophilus* 23S rRNA. *Biochemistry* **38**: 1772–1779
- Heurgue-Hamard V, Champ S, Engstrom A, Ehrenberg M, Buckingham RH (2002) The hemK gene in *Escherichia coli* encodes the N(5)-glutamine methyltransferase that modifies peptide release factors. *EMBO J* **21**: 769–778
- Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC (1999) Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* **286**: 2165–2169
- Isaacs FJ, Dwyer DJ, Ding C, Pervouchine DD, Cantor CR, Collins JJ (2004) Engineered riboregulators enable post-transcriptional control of gene expression. *Nat Biotechnol* **22**: 841–847
- Jaffe JD, Berg HC, Church GM (2004) Proteogenomic mapping as a complementary method to perform genome annotation. *Proteomics* **4**: 59–77
- Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang HC, Stines AP, Georgopoulos C, Frishman D, Hayer-Hartl M, Mann M, Hartl FU (2005) Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* **122**: 209–220
- Khaitovich P, Tenson T, Kloss P, Mankin AS (1999) Reconstitution of functionally active *Thermus aquaticus* large ribosomal subunits with *in vitro*-transcribed rRNA. *Biochemistry* **38**: 1780–1788
- Khan SA (1997) Rolling-circle replication of bacterial plasmids. *Microbiol Mol Biol Rev* **61**: 442–455
- Koonin EV (2000) How many genes can make a cell: the minimal-gene-set concept. *Annu Rev Genomics Hum Genet* **1**: 99–116
- Krzyzosiak W, Denman R, Nurse K, Hellmann W, Boublik M, Gehrke CW, Agris PF, Ofengand J (1987) *In vitro* synthesis of 16S ribosomal RNA containing single base changes and assembly into a functional 30S ribosome. *Biochemistry* **26**: 2353–2364
- Kung H-F, Chu F, Caldwell P, Spears C, Treadwell BV, Eskin B, Brot N, Weissbach H (1978) The mRNA-directed synthesis of the alpha-peptide of beta-galactosidase, ribosomal proteins L12 and L10, and elongation factor Tu, using purified translational factors. *Arch Biochem Biophys* **187**: 457–463
- Lewin B (2004) *Genes VIII*, 8th edn. Upper Saddle River, NJ: Pearson Prentice Hall
- Li Z, Deutscher MP (1996) Maturation pathways for *E. coli* tRNA precursors: a random multienzyme process *in vivo*. *Cell* **86**: 503–512
- Lietzke R, Nierhaus KH (1988) Total reconstitution of 70S ribosomes from *Escherichia coli*. *Methods Enzymol* **164**: 278–283
- Luisi PL (2002) Toward the engineering of minimal living cells. *Anat Rec* **268**: 208–214
- Maki JA, Culver GM (2005) Recent developments in factor-facilitated ribosome assembly. *Methods* **36**: 313–320
- Mills DR, Peterson RL, Spiegelman S (1967) An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc Natl Acad Sci USA* **58**: 217–224
- Mitra RD, Church GM (1999) *In situ* localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Res* **27**: e34
- Mushegian AR, Koonin EV (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc Natl Acad Sci USA* **93**: 10268–10273
- Nakahigashi K, Kubo N, Narita S, Shimaoka T, Goto S, Oshima T, Mori H, Maeda M, Wada C, Inokuchi H (2002) HemK, a class of protein methyl transferase with similarity to DNA methyl transferases, methylates polypeptide chain release factors, and hemK knockout induces defects in translational termination. *Proc Natl Acad Sci USA* **99**: 1473–1478
- Nierhaus KH, Dohme F (1974) Total reconstitution of functionally active 50S ribosomal subunits from *Escherichia coli*. *Proc Natl Acad Sci USA* **71**: 4713–4717
- Nomura M, Erdmann VA (1970) Reconstitution of 50S ribosomal subunits from dissociated molecular components. *Nature* **228**: 744–748
- Ogle JM, Ramakrishnan V (2005) Structural insights into translational fidelity. *Annu Rev Biochem* **74**: 129–177
- Pohorille A, Deamer D (2002) Artificial cells: prospects for biotechnology. *Trends Biotechnol* **20**: 123–128

- Posfai G, Plunkett G, Feher T, Frisch D, Keil GM, Umenhoffer K, Kolisnychenko V, Stahl B, Sharma SS, de Arruda M, Burland V, Harcum SW, Blattner FR (2006) Emergent properties of reduced-genome *Escherichia coli*. *Science* **312**: 1044–1046
- Richmond KE, Li MH, Rodesch MJ, Patel M, Lowe AM, Kim C, Chu LL, Venkataramaiah N, Flickinger SF, Kaysen J, Belshaw PJ, Sussman MR, Cerrina F (2004) Amplification and assembly of chip-eluted DNA (AACED): a method for high-throughput gene synthesis. *Nucleic Acids Res* **32**: 5011
- Sauer B (2002) Cre/lox: one more step in the taming of the genome. *Endocrine* **19**: 221–228
- Semrad K, Green R (2002) Osmolytes stimulate the reconstitution of functional 50S ribosomes from *in vitro* transcripts of *Escherichia coli* 23S rRNA. *RNA* **8**: 401–411
- Semrad K, Green R, Schroeder R (2004) RNA chaperone activity of large ribosomal subunit proteins from *Escherichia coli*. *RNA* **10**: 1855–1860
- Service RF (2005) How far can we push chemical self-assembly? *Science* **309**: 95
- Shimizu Y, Kanamori T, Ueda T (2005) Protein synthesis by pure translation systems. *Methods* **36**: 299–304
- Soma A, Ikeuchi Y, Kanemasa S, Kobayashi K, Ogasawara N, Ote T, Kato J, Watanabe K, Sekine Y, Suzuki T (2003) An RNA-modifying enzyme that governs both the codon and amino acid specificities of isoleucine tRNA. *Mol Cell* **12**: 689–698
- Spirin AS, Baranov VI, Ryabova LA, Ovodov SY, Alakhov YB (1988) A continuous cell-free translation system capable of producing polypeptides in high yield. *Science* **242**: 1162–1164
- Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* **185**: 60–89
- Szostak JW, Bartel DP, Luisi PL (2001) Synthesizing life. *Nature* **409**: 387–390
- Takyar S, Hickerson RP, Noller HF (2005) mRNA helicase activity of the ribosome. *Cell* **120**: 49–58
- Tian J, Gong H, Sheng N, Zhou X, Gulari E, Gao X, Church G (2004) Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* **432**: 1050–1054
- Tomita M, Hashimoto K, Takahashi K, Shimizu TS, Matsuzaki Y, Miyoshi F, Saito K, Tanida S, Yugi K, Venter JC, Hutchison III CA (1999) E-CELL: software environment for whole-cell simulation. *Bioinformatics* **15**: 72–84
- Traub P, Nomura M (1968) Structure and function of *E. coli* ribosomes. V. Reconstitution of functionally active 30S ribosomal particles from RNA and proteins. *Proc Natl Acad Sci USA* **59**: 777–784
- Zhong XB, Lizardi PM, Huang XH, Bray-Ward PL, Ward DC (2001) Visualization of oligonucleotide probes and point mutations in interphase nuclei and DNA fibers using rolling circle DNA amplification. *Proc Natl Acad Sci USA* **98**: 3940–3945
- Zhou X, Cai S, Hong A, You Q, Yu P, Sheng N, Srivannavit O, Muranjan S, Rouillard JM, Xia Y, Zhang X, Xiang Q, Ganesh R, Zhu Q, Matejko A, Gulari E, Gao X (2004) Microfluidic PicoArray synthesis of oligodeoxynucleotides and simultaneous assembling of multiple DNA sequences. *Nucleic Acids Res* **32**: 5409–5417
- Zvereva MI, Shpanchenko OV, Dontsova OA, Nierhaus KH, Bogdanov AA (1998) Effect of point mutations at position 89 of the *E. coli* 5S rRNA on the assembly and activity of the large ribosomal subunit. *FEBS Lett* **421**: 249–251