From gene to protein

Cell-free translation systems are steadily changing the face of biochemistry.

BY BURKHARD ZIEBOLZ

Following genomics, protein research has gained tremendous importance over the past few years. In investigating the genetic causes of diseases, for example, it is not sufficient just to know the genes involved and their DNA sequence. If we are to understand and control these diseases, we need to investigate the several hundred thousand proteins in the body that are responsible for all biological processes.

Cellular conundrum

To study the function of a protein, it first needs to be synthesized in sufficient quantities. In the past, this step always represented one of the greatest obstacles to analysis. Proteins could not be synthesized within an acceptable time frame using traditional methods, and there was no universally applicable method for rapid and convenient protein expression. Processes using whole cells are time-consuming and are associated with a raft of problems, including the formation of insoluble inclusion bodies, toxic effects, and degradation by cellular enzymes.

The cellular systems used, including bacteria, yeast, insect cells infected with baculovirus, and mammalian cells, suffer from various limitations. Only limited success has been achieved in *E. coli* because expressed nonbacterial proteins must be compatible with the *E. coli* expression and processing systems. A common problem, for example, is that the expressed protein often aggregates.

To some extent, these limitations are addressed by using eukaryotic expression systems. However, whereas protein synthesis in *E. coli* is relatively simple and economical, the procedures used to produce proteins in mammalian cells require several complex steps and can take many weeks, if not months. Beyond the time requirements, eukaryotic expression systems are often much more expensive than *E. coli* systems.

Cell-free solutions

To avoid some of the complications involved in expressing proteins in living organisms, researchers developed various cell-free protein expression systems (Figure 1). The in vitro systems are usually based on extracts of *E. coli*, wheat germ, or rabbit reticulocytes that contain the components required for protein translation from an mRNA molecule, including amino acids, ribosomes, and energy molecules such as adenosine 5′-triphosphate. The amount of protein produced by these kits is often limited, and radioactive amino acids are frequently added as tracers to detect the product after synthesis. The main reason for the low productivity is a rapid exhaustion of the translation machinery and an accumulation of inhibitory substances that soon reach levels that prevent further protein production.

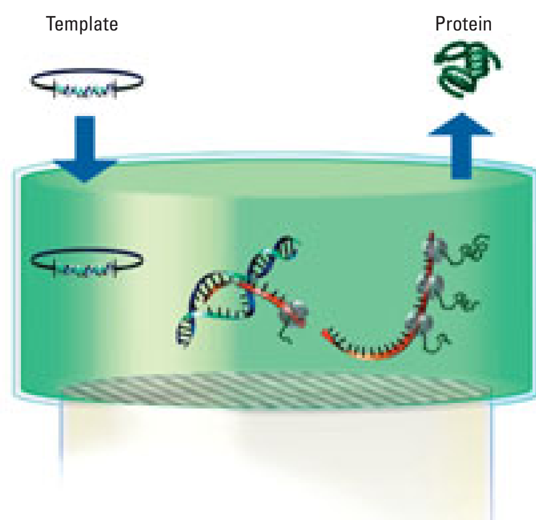

Figure 1. Stage to screen. Cell-free protein translation kits take short fragments of DNA (left), transcribe them into mRNA (middle), and synthesize proteins (right) in vitro for subsequent screening experiments.

The newer breeds of in vitro protein expression systems—such as the Roche Diagnostics Rapid Translation System (RTS)—are superior in several ways. First, they incorporate stronger energy systems that increase the biochemical yield. Second, some systems use a two-chamber principle, with a reaction chamber and a feeding chamber. The feeding chamber is 10 times as large as the reaction chamber and functions as a reservoir of protein building blocks and as an overflow chamber into which inhibitory substances can migrate. Therefore, the reaction can continue for longer time periods so that the final protein content is higher.

Promising proteins

The in vitro translation systems enable the user to proceed from a gene to the desired protein in as few as four steps. Starting from a DNA template, the system allows protein expression with yields ranging from micrograms (in 2 h) to tens of milligrams (in 24 h).

Step 1. Coding-sequence modification. Successful protein expression depends on a well-designed DNA template, but expression constructs can be tedious and difficult to optimize. ProteoExpert is an example of a Web-based service that can help overcome these challenges by modifying the normal or wild-type gene sequence with silent mutations that have the potential to improve protein expression (see box, "Silence is golden"). Each calculation generates a list of gene-specific oligonucleotide primer pairs. Using PCR, these primers generate mutated linear

KEY TERMS: cell biology, genomics, high throughput, informatics, proteomics, screening, technique

© 2003 AMERICAN CHEMICAL SOCIETY

DNA fragments containing the gene sequences that can be directly expressed with the in vitro system.

Step 2. Cloning-system choice. Translating an mRNA molecule into a protein requires the proper sequences for translation initiation and elongation. First, one needs a ribosomal binding site (RBS), from which the ribosome slides along the mRNA and starts protein translation. The DNA sequences at and around the RBS determine how efficiently the ribosomes bind to the DNA and initiate translation.

The RBS might be hidden by strong mRNA secondary structures such as hairpin formations, which prevent the ribosome from binding or efficiently translating the mRNA. For this reason, optimal protein yields can be achieved only with optimal sequences like those found in the upstream region of certain expression vectors. These sequences can be introduced using traditional cloning procedures, but it is often more efficient to introduce them by using overlap extension PCR.

Step 3. Screening and optimization. Small-scale, high-yield protein expression reactions are crucial for rapid screening and optimization studies. The RTS uses an advanced *E. coli* lysate supplemented to provide superior energy regeneration and adjusted for codon usage bias. Linear or circular DNA templates can be used without prior nucleic acid purification in a 50-µL coupled transcription/translation reaction to express up to 20 µg of protein in just 2–4 h.

Studies can be designed to analyze the expression levels of various constructs and their compatibility with the translation system. After maximizing the protein yield, researchers can consider the solubility and functional characteristics of the protein. The cell-free nature of these systems allows users to adjust reaction conditions easily and add components that might facilitate protein expression and stability.

Step 4. Scale up when needed. Because scale-up kits are typically based on the same lysate used in the smaller system, small-scale conditions are easily transferable to larger reactions. Using the two-chamber approach, the expression reaction can steadily produce proteins over periods up to 24 h, supplying the reaction with fresh components and removing inhibitory reaction byproducts.

The key to consistent and reproducible results is providing an easily controllable environment for the expression reactions and supporting the component-exchange reaction by gentle mixing. The incubation chamber of the RTS can be used with tubes or microwell plates, and it digitally controls the heating and cooling between 20 and 50 °C by distributing air throughout the reaction chamber. Similarly, shaking functions allow gentle mixing of reaction components and thereby support the exchange process. And some instruments include top heating, which minimizes artifacts that are caused by condensation of reaction components on the lid of the reaction tubes.

Silence is golden

The 20 amino acids are coded for by three-nucleotide codons or triplets, and of the 64 possible triplets, several code for the same amino acid. But the universal genetic code is not so universal, and different organisms exhibit preferential codon usage patterns (see the table), which leads to difficulties when expressing a mammalian protein in *E. coli*. Sometimes, however, this problem can be addressed by exchanging these codons for others that are more frequently used in *E. coli*.

For example, the sequence GCC codes for alanine, but so does GCG. Thus, a GCC to GCG change will not influence protein function. It might, however, be advantageous in certain experiments because there may be more tRNA molecules that recognize GCG instead of GCC in the protein expression environment used. When this occurs, the rate of protein expression will be much higher.

Codon bias in organisms

Organism	% AGG (arginine)	% AGA (arginine)	% CGA (arginine)
Bacterium (*E. coli*)	1.4[a]	2.1	3.1
Yeast (*S. cerevisiae*)	9.3	21.3	3.0
Worm (*C. elegans*)	3.8	15.6	11.5
Fly (*D. melanogaster*)	4.7	5.7	7.6
Human (*H. sapiens*)	11.0	11.3	6.1

[a]Percentages are based on codon frequency per 1000 arginine codons.
Adapted from Carstens, P.; Waesche, A. *Strategies* **2001**, *12* (2), 49–51; available at www.stratagene.com/vol12_2/p49-51.htm.

Applications

Many fields of research require the easy and flexible expression of proteins at high yields in a cell-free system. The cell-free systems also allow examination of protein–protein interactions by coexpressing proteins. Furthermore, mutant proteins (e.g., truncated proteins or point mutations) can be engineered and expressed rapidly, and even proteins that are toxic to cells can be produced.

Cell-free systems are also finding greater use in protein structure studies. Researchers at the Berkeley Structural Genomics Center (CA) recently used the RTS to determine the X-ray and NMR structures of proteins from *Mycoplasma pneumoniae* and *M. genitalium* (*1*). In a proof-of-principle experiment, the researchers produced milligram quantities of a phosphoserine phosphatase that was isotopically labeled with ^{15}N amino acids for NMR analysis, suggesting to them that the cell-free protein production system would become invaluable to high-throughput structural genomic studies.

Similarly, cell-free expression systems can be used to produce pharmaceutically important proteins. Masao Fukushima and Ryuji Kawaguchi at SRL, Inc. (Tokyo),

used the RTS to produce recombinant human interleukin-2 (IL-2), a cytokine that stimulates T-cell proliferation (*2*). When expressed in *E. coli*, IL-2 forms an insoluble aggregate that must be solubilized and refolded to be active, and even when conditions have been optimized, the *E. coli* system yields proteins in the nanogram to low-microgram range. By using the RTS, the Japanese researchers isolated near-milligram quantities of soluble and active IL-2.

And finally, even in vitro systems are hearing the clarion call of miniaturization. Recently, Mari Tabuchi and colleagues at the University of Tokushima and the Japan Science and Technology Corp. (also in Tokushima) modified the RTS method for use with a microfluidic poly(methyl methacrylate) microchip to produce a fatty acid binding protein (*3*). The researchers found that although the microfluidic chamber did not produce nearly as much protein as the full RTS, they could easily detect the protein that was synthesized by Western blot, suggesting that the miniaturization of protein synthesis was possible.

The key to consistent and reproducible results is providing an easily controllable environment for the expression reactions.

References

(1) Cho, H. S.; et al. In vitro protein production for structure determination with the Rapid Translation System (RTS); Application Note No. 4/2001; Roche Applied Science.
(2) Fukushima, M.; Kawaguchi, R. Cell-free expression of human interleukin-2 (IL-2) with the RTS 500 *E. coli* circular template kit; Application Note No. 3/2001; Roche Applied Science.
(3) Tabuchi, M.; et al. *Proteomics* **2002,** *2,* 430–435.

Burkhard Ziebolz is science communication manager with Roche Diagnostics (Mannheim, Germany). Send your comments or questions about this article to mdd@acs.org or to the Editorial Office address on page 3.