

Chapter 5

Cell-Free Protein Production for Structural Biology

Takaho Terada, Seisuke Kusano, Takayoshi Matsuda, Mikako Shirouzu,
and Shigeyuki Yokoyama

Abstract

Cell-free protein synthesis using *E. coli* cell extracts has successfully been applied to protein sample preparation for structure determination by X-ray crystallization and NMR spectroscopy. The standard reaction solution for *E. coli* cell-free protein synthesis by coupled transcription-translation contains the S30 extract of *E. coli* cells, T7 RNA polymerase, and the DNA template (either plasmid or PCR-amplified linear DNA). Milligram quantities of proteins can be synthesized by the dialysis mode of the cell-free reaction in several hours. The *E. coli* cell-free protein synthesis method is suitable for the production of mammalian proteins, heteromultimeric protein complexes, and integral membrane proteins and features numerous advantages over the recombinant protein expression methods with bacterial and eukaryotic host cells. We present examples of structure determinations of mammalian and bacterial heteromultimeric protein complexes prepared by the cell-free production method.

Keywords Cell-free protein synthesis, *Escherichia coli*, Protein complexes, Structure determination

1 Introduction

Proteins are synthesized in cells by translation of their messenger RNAs (mRNAs), which are transcribed from the encoding genes. Large-scale protein synthesis can be performed not only by the recombinant DNA methods using host cells, such as *Escherichia coli*, yeast, insect, and mammalian cells, but also by the cell-free or in vitro protein synthesis methods. Cell-free protein synthesis can be accomplished with cell extracts prepared from a variety of organisms, including *E. coli* [1–10], wheat germ [11–13], insects [14–16], and humans [17, 18]. The cell extracts contain the ribosomes, transfer RNAs (tRNAs), various translation factors, and downstream factors, such as molecular chaperones. As for mRNA, cell-free translation may be performed by either using separately prepared mRNA or coupling translation with transcription from the template DNA (“coupled transcription-translation”) by T7 or SP6 RNA polymerase [1, 2]. The reaction solution for

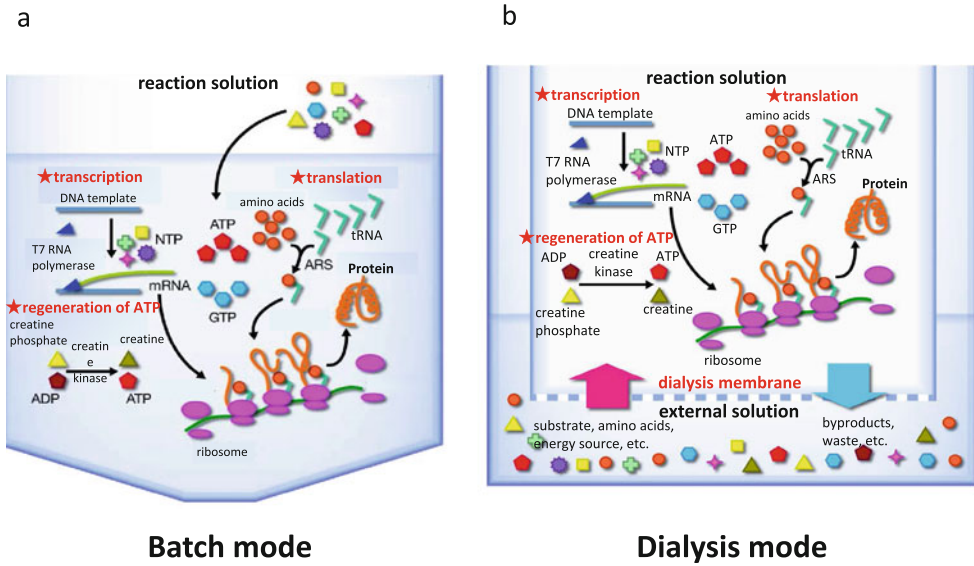


Fig. 1 Schematic illustration of the cell-free protein synthesis reaction modes. **(a)** The batch mode and **(b)** the dialysis mode

cell-free protein synthesis contains the cell extract, the template DNA for coupled transcription-translation or the pre-prepared mRNA, the low-molecular-mass substrates such as amino acids, the ATP regeneration system, and other components (Fig. 1). The cell-free synthesis reaction in a tube (the batch mode) continues for about one hour (Fig. 1a). To produce larger amounts of proteins, the reaction solution is dialyzed against the external solution containing the low-molecular-mass substrates (the dialysis mode) (Fig. 1b) [3, 4, 8–10, 19]. In this dialysis mode, the synthesis reaction continues for several hours, as the reaction solution is replenished with the low-molecular-mass substrates through the dialysis membrane, while the low-molecular-mass by-products are removed by dialysis (Fig. 1b) [3, 4, 8–10, 19].

The cell-free protein synthesis method has a number of advantages over conventional recombinant expression methods with host cells. For example, physiologically toxic proteins can be synthesized well by the cell-free method. The cell-free protein synthesis method actually has a much longer history than that of the host-vector recombinant protein expression, mainly for small-scale synthesis. However, drastic improvements of the cell-free protein synthesis method over the past decade have expanded its use for large-scale protein preparation [8–10, 20–24]. In fact, target proteins are frequently produced at levels of about 1 mg per ml cell-free reaction solution [25, 26]. In the case of the *E. coli* cell-free method, 1 ml of reaction mixture corresponds roughly to 50 ml of *E. coli* cell culture. This high yield of the cell-free protein synthesis method

makes it cost-effective, and cell-free protein synthesis systems for large-scale protein production are now commercially available. The DNA template for mg quantity protein production by coupled transcription-translation with an *E. coli* cell extract can be either a pre-prepared plasmid or a PCR-amplified linear DNA template, encoding the protein [27]. This “cloning-free” nature enhances the efficiency of the cell-free method. For example, it only takes a few hours to perform the steps from PCR to cell-free protein synthesis [22–24, 27, 28]. Thus, the cell-free protein synthesis method has become one of the standard methods for protein sample preparation.

For structural biology, the cell-free synthesis method used to be regarded as the “salvage” method, which was only tried when other methods were unsuccessful. In contrast, the cell-free method is now considered as the “first-line” method for structural biology, which should be tried before other methods because of its various advantages over recombinant DNA methods using live host cells. First of all, large amounts of highly purified, homogeneous proteins are characteristically needed for structural biology analyses by X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. In this regard, the cell-free protein synthesis method using the *E. coli* cell extract is much more suitable for mammalian protein production, in terms of both quality and quantity, than the host cell-based recombinant expression methods and cell-free synthesis methods using eukaryotic cell extracts. For example, *E. coli* cells may be engineered at the genome level, to tag a nuclease for removal from the cell extract [29]. Therefore, the *E. coli* cell-free method is by far the most frequently chosen for structural biology.

Naturally, the cell-free protein synthesis system is “open” with respect to the addition or subtraction of components. Many parameters, such as the reaction temperature, the incubation time, and the substrate and template concentrations, can be optimized easily. Intra- and intermolecular disulfide bonds may be formed by controlling the redox status of the reaction solution. Molecular chaperones may be added to the reaction solution, in order to facilitate proper folding. The cell-free protein synthesis method is suitable for the production of protein complexes consisting of two or more different components or subunits [30]. First, the components can be co-expressed simply by including their templates in stoichiometric amounts in the reaction solution, which is more strictly controllable than the cell-based recombinant methods. Moreover, a larger number of components may be co-expressed by the cell-free methods than by the recombinant cell-based methods. Otherwise, protein complexes may be reconstituted in a step-wise manner; for example, one or more components may be synthesized in the presence of a subcomplex consisting of the others [30]. Proteins can also be synthesized in complex with ligand(s), such as a low-molecular-mass cofactor, zinc ion [31],

substrate, inhibitor, peptide fragment of the binding partner protein, and nucleic acids [30]. The formation of such complexes frequently improves the qualities of the products with respect to proper folding, as compared with the synthesis of the proteins by themselves. Furthermore, the cell-free synthesis of membrane proteins is particularly more advantageous than the recombinant cell-based methods, as described in Chap. 7.

For structural biology, the flexibility of the cell-free method in terms of nonstandard amino acids is very useful. For the multiwavelength anomalous diffraction (MAD) method in protein crystallography, the methionine residues in the protein may be almost completely replaced with selenomethionine, simply by using the same cell extract and selenomethionine in place of methionine in the reaction and external solutions [20, 23, 24, 32], while the recombinant expression method uses a methionine auxotrophic mutant strain of *E. coli*. Stable isotope (SI) labeling of proteins with nitrogen-15 (^{15}N), carbon-13 (^{13}C), and/or deuterium (^2H) for NMR measurements can easily be performed by cell-free protein synthesis [5, 7–9, 21–24, 28, 31]. Uniform SI labeling of proteins may be accomplished by using a mixture of uniformly labeled amino acids [8]. In addition, a variety of selective labeling techniques have been developed, using the advantages of the cell-free synthesis method [5, 7, 8, 33, 34]. For instance, unnatural amino acids may be incorporated site specifically into proteins by the cell-free method, using an engineered pair of a tRNA, specific to a special codon such as the UAG “stop” codon, and an aminoacyl-tRNA synthetase, specific to the unnatural amino acid [35–38]. The engineered pair of tRNA and pyrrolysyl-tRNA synthetase from *Methanosarcina mazei* was used, along with an extract of the *E. coli* RFzero strain [39], which lacks the gene-encoding release factor 1 recognizing the UAG and UAA stop codons, to introduce an epigenetic modification, acetyl-lysine, at four sites in the human histone H4 N-terminal tail [38].

Table 1 summarizes the structures of proteins produced by our group, using the cell-free method with the *E. coli* cell extract, deposited in the Protein Data Bank (PDB) as of June 22, 2015. The organisms range from human and mouse to viruses and bacteria. The number of NMR structures is much larger than that of the X-ray crystallographic structures, because most of the NMR structures were determined for human and mouse functional domains in the framework of the Japanese structural genomics project, “The Protein 3000 Project,” from 2002 to 2007 [40–42]. We have deposited about 100 crystallographic structures of human and mouse proteins in the PDB, and eight of them are heteromultimeric protein complexes. For the human and mouse proteins with crystal structures determined with cell-free-produced samples, the average molecular masses are about 40 kDa. Therefore, the cell-free synthesis method is applicable for much larger proteins than the

Table 1

The numbers of PDB-deposited structures of proteins produced by *E. coli* cell-free protein synthesis method in our group (Deposited from Apr. 2001 to Dec. 2014)

Source organism	X-ray	NMR
Vertebrate		
<i>Homo sapiens</i>	74	1029
<i>Mus musculus</i>	30	261
<i>Rattus rattus</i>		1
Invertebrate	1	4
Yeast		2
Plant	1	33
Bacteria	17	3
Virus	5	
Total	128	1333

Table 2

The numbers of PDB-deposited structures of proteins produced by the wheat germ cell-free synthesis method (Available from <http://www.uwstructuralgenomics.org/structures.htm>, accessed on Jun. 22, 2015)

Source organism	X-ray	NMR
Vertebrate		
<i>Homo sapiens</i>	1	5
<i>Mus musculus</i>		1
<i>Danio rerio</i>		2
Plant	3	9
Bacteria	1	
Total	5	17

functional domains analyzed by NMR spectroscopy (12 kDa). Table 2 summarizes the structures of proteins produced by the cell-free method using wheat germ extract, from The Center for Eukaryotic Structural Genomics (USA), deposited in the PDB as of January 13, 2015. We expect that the *E. coli* cell-free protein synthesis method will be used more extensively in the future, particularly for difficult proteins, such as mammalian proteins, protein complexes, and membrane proteins.

2 Cell-Free Protein Production Methods for Structural Biology

2.1 Workflow of Cell-Free Protein Production

The overall workflow of cell-free protein production is shown in Fig. 2. The preliminary experiment is performed through small-scale reactions to optimize various conditions. Using these optimized conditions, the reaction scale can be increased to the large-scale protein production. Selenomethionine and stable isotope-labeled amino acids may be used to label the product for X-ray crystallography and NMR spectroscopy, respectively.

2.2 Template DNA for Cell-Free Coupled Transcription-Translation

The template DNA for cell-free coupled transcription-translation in the *E. coli* extract contains the coding region of the target protein and the flanking sequences for transcription, translation, and purification. A typical template DNA is shown in Fig. 3. The flanking sequences may be provided by the plasmid vector or PCR primer (s). The two-step PCR method [27] is useful to efficiently construct the designed template DNA and particularly for the preparation of a large number of constructs for comparison. The tag is selected by considering not only the ease of purification but also the folding and/or solubility, from a variety of tags, such as 6×histidine (6His),

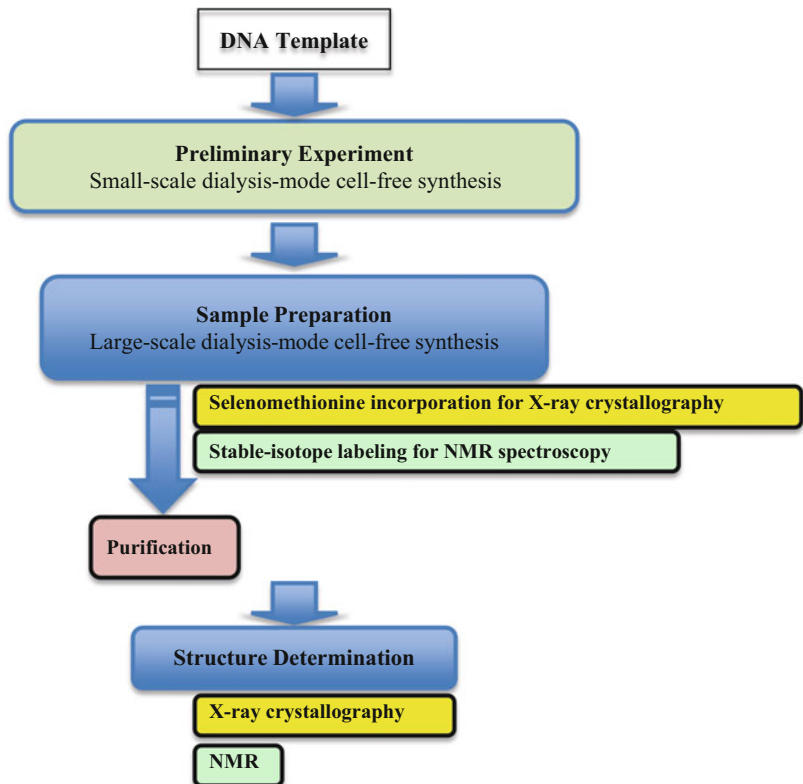


Fig. 2 Workflow of the cell-free protein production

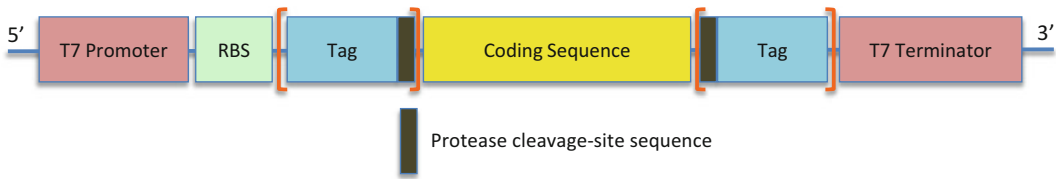


Fig. 3 Typical design of the template DNA for coupled transcription-translation in *E. coli* cell-free protein synthesis. The sequence encoding the target protein (coding sequence) and the preceding RBS (ribosome-binding site) for translation are flanked on the 5' and 3' sides by the T7 Promoter and T7 Terminator sequences, respectively. The N- and/or C-terminal tag sequence (usually including a protease cleavage site sequence) may be introduced not only for detection and purification but also for increasing folding and/or solubility

streptavidin-binding peptide (SBP), glutathione-S-transferase (GST), maltose-binding protein (MBP), and small ubiquitin-related modifier (SUMO). The template DNA, in the form of either plasmid DNA or PCR-amplified linear DNA, may be used in the cell-free reaction. The use of a linear template DNA significantly shortens the total duration of the experiment, and a higher protein yield is generally obtained with the use of plasmid DNA. In the latter case, the plasmid DNA must be well purified with a commercially available kit (Qiagen, Promega, etc.).

2.3 *E. coli* Cell-Free Protein Synthesis System

The S30 fraction (the supernatant fraction obtained after cell disruption and centrifugation at $30,000 \times g$) of *E. coli* cells is used as the cell extract for cell-free protein synthesis. We usually use the S30 extract of *E. coli* strain BL21 CodonPlus-RIL (Agilent Technologies), containing extra copies of the genes encoding minor tRNAs [10, 43]. Various kits for cell-free protein synthesis with *E. coli* cell extracts are commercially available, and those suitable for structural biology sample preparation should be chosen. For structural biology purposes, the cell-free protein expression kit “Musaibo-Kun” (Taiyo Nippon Sanso, Japan), “iPE Kit” (Sigma-Aldrich, USA), and the Remarkable Yield Translation System (RYTS) Kit (Protein Express, Japan) are useful and based on the method by Kigawa et al. [9]. The RTS 100 *E. coli* HY Kit (Biotech rabbit GmbH, Germany), the EasyXpress Protein Synthesis Kit (QIAGEN, The Netherlands), and the S30 T7 High-Yield System (Promega, USA) are also suitable. Some products are optimized for special purposes, such as the use of a linear template DNA and disulfide bond formation. To facilitate proper folding, we prepare the S30 extract from *E. coli* BL21 cells expressing a set of *E. coli* chaperones (DnaK/DnaJ/GrpE and/or GroEL/GroES) in addition to the minor tRNAs for rare codons, such as AGA/AGG, AUA, and CUA. Notably, non-natural amino acids can be incorporated into proteins in response to UAG codons much more efficiently by using the S30 extract of the *E. coli* RFzero strain, which lacks the release factor 1 gene [35, 36]. The detailed protocols for *E. coli* cell extract preparation have been published [9, 10, 29].

2.4 Cell-Free Protein Synthesis Reaction Solution

The reaction solution for *E. coli* cell-free coupled transcription-translation contains the *E. coli* S30 extract, the DNA template, the T7 RNA polymerase, and the substrates for transcription and translation. The components of the standard reaction solution are listed in Table 3. The order of the components in Table 3 roughly corresponds to that used to set up the reaction solution. For transcription, T7 RNA polymerase, prepared as reported in [44], is used. For translation, the S30 extract is prepared in 10 mM Tris-acetate buffer (pH 8.2), containing 60 mM potassium acetate, 16 mM magnesium acetate, and 1 mM DTT, and used at a final concentration of 30 % (v/v) in the reaction solution. The S30 extract contains the endogenous tRNAs from *E. coli* cells, but is supplemented with *E. coli* MRE600-derived tRNA (Roche Applied Science, 109550). The low-molecular-mass component mixture solution, *low-molecular-weight creatine phosphate tyrosine* (LMCPY), contains 160 mM HEPES-KOH buffer (pH 7.5), 4.13 mM L-tyrosine, 534 mM potassium L-glutamate, 5 mM DTT, 3.47 mM ATP, 2.40 mM GTP, 2.40 mM CTP, 2.40 mM UTP, 0.217 mM folic acid, 1.78 mM cAMP, 74 mM ammonium acetate, and 214 mM creatine phosphate. The other amino acids besides L-tyrosine, which is included in LMCPY, are provided as “A.A.(-Y)”, containing 10 mM DTT and 20 mM each of the 19 amino acids, as shown in Table 3.

As DTT is used in the standard reaction solution, protein synthesis is fundamentally performed under reducing conditions, whereas the use of DTT-free LMCPY is recommended for the production of disulfide bond-forming proteins, such as secreted

Table 3
Standard composition of the *E. coli* cell-free protein synthesis reaction solution

Reagent	Stock conc	Final conc
LMCP(Y)		37.33 % (v/v)
NaN ₃	5 % (w/v)	0.05 % (w/v)
Mg(OAc) ₂	1.6 M	9.28 mM
A.A.(-Y)	20 mM	1.5 mM each
tRNA	17.5 mg/ml	0.175 mg/ml
Creatine kinase	3.75 mg/ml	0.25 mg/ml
S30 extract		30 % (v/v)
T7 RNA polymerase	10 mg/ml	66.7 μg/ml
Other factors		
Milli-Q water		To the desired volume

proteins and membrane proteins, as described in the next section (Sect. 2.5). The optimal magnesium concentration depends to some extent on the target proteins, and it should therefore be optimized for each target protein, in the range of 5–20 mM. For ATP regeneration, creatine kinase and its substrate, creatine phosphate, are used. The optimal DNA template concentration for coupled transcription-translation should be determined by a preliminary small-scale cell-free experiment.

2.5 Protein Folding

Metal Ligation, Ligand Binding, and Complex Formation: For Zn-binding proteins, an appropriate concentration (usually around 50 μ M) of ZnCl_2 or ZnSO_4 should be added [30, 31]. Ligand-binding proteins are synthesized in the presence of the ligand (cofactor, substrate, inhibitor, etc.) in the reaction solution, since the ligand is expected to help the protein fold properly. For protein complex formation, two or more DNA templates are simultaneously used. The ratio of these templates should be adjusted prior to the large-scale cell-free production [30].

Molecular Chaperones: To facilitate correct folding, appropriate molecular chaperones [45] are prepared separately, and their mixture is added to the cell-free reaction. Otherwise, the S30 extract for the correct folding of the target protein(s) and protein complex(es) should be added. Among the *E. coli* chaperones [45], DnaK/DnaJ/GrpE and GroEL/GroES may function in the early and late stages, respectively, of chaperone-assisted protein folding. Therefore, single and/or dual uses of the two sets of chaperones in the cell-free protein synthesis are usually tested for precipitating or aggregating proteins.

Disulfide Bonds: For disulfide bond-containing proteins, cell-free synthesis is performed under more oxidative redox conditions than the standard conditions. The ratio between reduced glutathione (GSH) and oxidized glutathione (GSSG) may be optimized, by testing ratios between 1:9 and 9:1. A disulfide isomerase [46], such as *E. coli* DsbC, is usually added to facilitate proper protein folding. *E. coli* Skp [47, 48] may be used as a chaperone in addition to DsbC.

Reaction Temperature: For proper protein folding, the incubation temperature may be selected according to the efficiency of folding in the range of 15–37 °C, while the standard temperature is about 25 °C.

2.6 Amino Acid Labeling for Structure Determination

Selenomethionine Incorporation for X-ray Crystallography: Selenomethionine-substituted proteins for MAD phasing can be obtained by cell-free protein synthesis, in which the L-methionine in the reaction and external solutions is simply replaced by L-selenomethionine. The amino acid mixture lacking L-methionine

and the 20 mM selenomethionine solution with 10 mM DTT are prepared separately and used in place of the standard amino acid mixture. Selenocysteines may be used instead of selenomethionine in the reaction solution and incorporated in place of cysteine in the protein for the MAD method. Iodine-/bromine-substituted amino acids such as tyrosine can be incorporated into specified site(s) of the protein, by using the “expanded genetic code system,” and may also be used for MAD phasing [49].

Stable Isotope Labeling for NMR Spectroscopy: The production of stable isotope (SI)-labeled protein samples for multinuclear NMR spectroscopy is performed by replacing the amino acid(s) to be labeled in the cell-free reaction solution with SI-labeled ones. The mixture solution containing 10 mM DTT and 20 mM each of the SI-labeled amino acids should be used. Uniform SI labeling of proteins is accomplished with mixtures of the 20 amino acids uniformly labeled with ^{15}N , ^{13}C , and/or ^2H . Amino acid-selective SI labeling with respect to one or several kinds of amino acids can be performed more easily by the cell-free method than by the conventional recombinant method, because the SI scrambling between amino acids is minimized in the cell-free reaction. The cell-free protein synthesis method is quite useful for the stereo-array isotope labeling (SAIL) method [50]. We developed a cell-free system that utilizes potassium D-glutamate in place of L-glutamate, for efficient SI labeling [21, 34].

2.7 Reaction Modes of Cell-Free Protein Synthesis

The Batch and Dialysis Modes: The cell-free coupled transcription-translation may be performed in either the batch or dialysis mode (Fig. 1). The batch mode of cell-free protein synthesis is the simplest: the reaction is performed by incubating the reaction solution in a container, such as a test tube. In the reaction solution, the low-molecular-mass substrates for coupled transcription-translation and ATP regeneration become exhausted and by-products accumulate. Thus, the batch reaction reaches a plateau in a few hours. In order to achieve higher yields, the dialysis-mode cell-free synthesis reaction is performed by placing the reaction solution in a compartment with a dialysis membrane, such as a dialysis bag, and incubating it with the external solution, containing the same low-molecular-mass components as those in the reaction solution. In this mode, the substrates and the by-products are continuously provided and removed, respectively, by the external solution through the dialysis membrane. In the standard conditions, the molecular weight cutoff of the dialysis membrane is 10–15 kDa, and the ratio of the volume of the external solution to that of the reaction solution is equal to or greater than 10. Therefore, the protein synthesis reaction continues much longer in the dialysis mode than in the batch mode. The synthesis yield at 25–30 °C may reach 1–5 mg/ml reaction in 3–4 h. In practice, we usually stop the reaction at 3–4 h to avoid denaturation of the products, although it may continue longer. For synthesis at 15 °C, the reaction may be continued up to overnight.

For large-scale structural biology sample preparation, the cell-free synthesis reaction is performed in the dialysis mode, usually with a 1–10 ml reaction solution, while difficult targets such as large complexes and membrane proteins may be synthesized with a 30 ml or larger reaction solution. We recommend optimizing the construct and the conditions of the cell-free synthesis reaction, by performing small-scale reactions (5–30 μ l) prior to the large-scale synthesis. For example, multiple PCR-amplified linear template DNAs encoding protein constructs with different terminal deletions may be generated and tested, with no cloning steps, by small-scale cell-free synthesis in multi-well plates, in either the batch or dialysis mode. Typically, multiple dialysis-mode cell-free reactions are performed in 96-well plates equipped with a dialysis membrane, and the volumes of the reaction solutions are 5 μ l per well. The optimal construct/conditions are selected with respect to the yield, the solubility, etc., toward the larger-scale cell-free production, as described above, for structure determination.

2.8 Purification of Synthesized Proteins

After the large-scale protein synthesis reaction, the product is purified by affinity chromatography, incubated with the specific protease to cleave the affinity tag, and then purified again with the affinity column to remove the affinity-tag peptides. When the protease is fused with the same affinity tag without the cleavage site, the affinity-tagged protease can be removed together with the affinity-tag peptides in one step. The resultant fraction is subjected to further ion-exchange chromatography and gel-filtration chromatography for X-ray analysis.

3 Examples of Heteromultimeric Complexes Produced by the Cell-Free Method for Structure Determination

The cell-free protein synthesis method is highly advantageous for the production of heteromultimeric complexes consisting of two or more different component proteins [30]. Here, we describe several examples of cell-free heteromultimeric proteins produced for structural biology.

3.1 DOCK2•ELM01

DOCK2 (dedicator of cytokinesis 2), which is specifically expressed in hematopoietic cells, activates the small GTP-binding protein Rac and thereby plays a critical role in cellular signaling events. The formation of a complex between DOCK2 and ELM01 (engulfment and cell motility 1) is required for DOCK2-mediated Rac signaling. In 2012, we identified the regions of DOCK2 and ELM01 required for their association and determined the complex structure by the following experimental strategies [51].

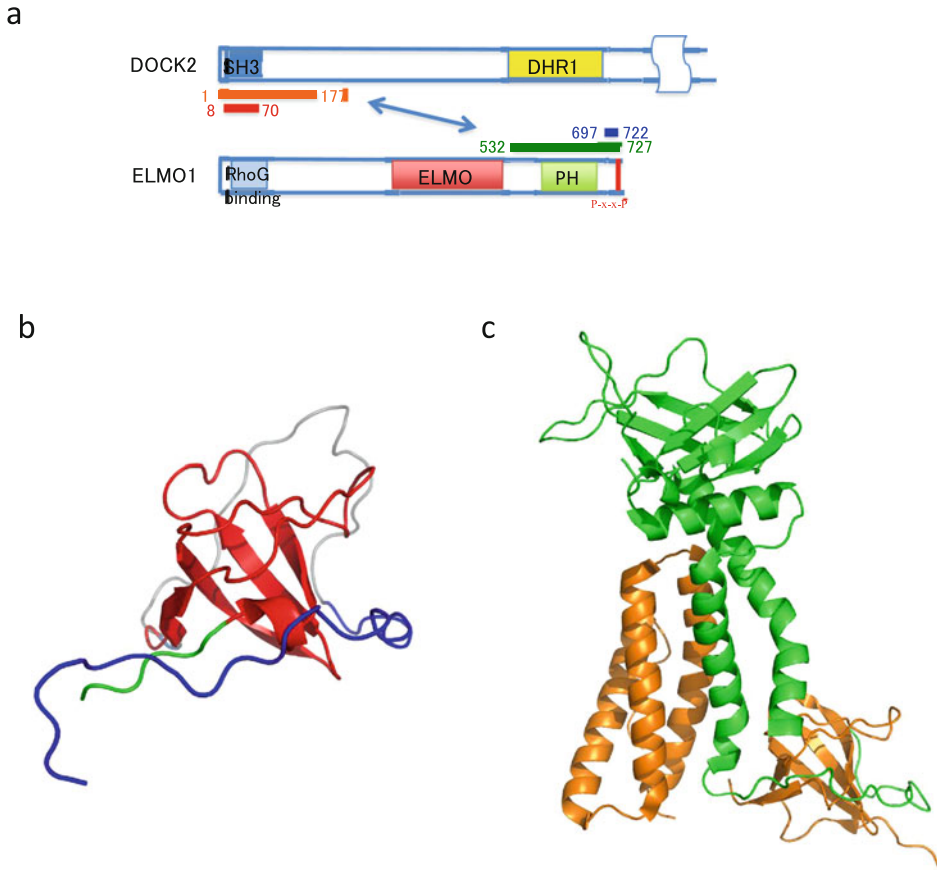


Fig. 4 Structures of the interactive regions of DOCK2 and ELMO1. **(a)** The domain organizations of DOCK2 and ELMO1. The *red* and *blue bars* indicate the DOCK2 and ELMO1 regions included in the fusion construct for NMR. The *orange* and *green bars* indicate the regions co-expressed for crystallization. **(b)** The NMR structure of the DOCK2 SH3-ELMO1 peptide fusion protein (PDB ID: 2RQR) (*ribbon* representation). The DOCK2 SH3 domain and the ELMO1 peptide are colored *red* and *blue*, respectively. **(c)** The crystal structure of the DOCK2 (1–177)•ELMO1(532–727) complex (PDB ID: 3A98) (*ribbon* representation). The DOCK2(1–177) and ELMO1 (532–727) proteins are colored *orange* and *green*, respectively

First, the N-terminal SH3 domain of human DOCK2 was found to bind to the C-terminal Pro-rich sequence of human ELMO1. Therefore, 87 differently designed DNA fragments encoding the human DOCK2 SH3 domain, with a human ELMO1 Pro-rich sequence peptide fused to its N- or C-terminus (Fig. 4a), were generated by PCR. The fragments were cloned into the pCR2.1 vector (Invitrogen) as fusions with an N-terminal histidine tag (a modified HAT tag) and a tobacco etch virus (TEV) protease cleavage site. Among these constructs, one fusion construct including an ELMO1 peptide (residues 697–722) fused to the N-terminus of the DOCK2 SH3 domain (residues 8–70) (designated as the DOCK2 SH3-ELMO1 peptide fusion protein) was selected as a suitable construct for NMR analysis, after checking

the productivity and solubility of the constructs by the small-scale dialysis-mode of cell-free protein synthesis.

For NMR structure determination, the $^{13}\text{C}/^{15}\text{N}$ -labeled DOCK2-ELMO1 peptide fusion protein was prepared by the large-scale dialysis-mode cell-free method. The solution structure determined by NMR (Fig. 4b, PDB ID: 2RQR) confirmed that the C-terminal Pro-rich region, especially P_{714-x-x}-P₇₁₇, of ELMO1 interacts with the SH3 domain of DOCK2, and prompted us to investigate the more detailed interactions between DOCK2 and ELMO1 by X-ray crystallography.

To identify the precise interacting regions of DOCK2 and ELMO1, a variety of N-terminal fragments of DOCK2 (residues 1–160, 1–177, 1–190, 9–160, 9–177, 9–190, 21–160, 21–177, and 21–190) and C-terminal fragments of ELMO1 (residues 532–717, 541–717, 550–717, 532–727, 541–727, and 550–727), with the N-terminal histidine tag (a modified HAT tag) sequence and the TEV cleavage site sequence, were generated by the two-step PCR method [27]. Using these PCR products as the templates for the small-scale dialysis-mode cell-free synthesis reactions, co- and separate protein expression studies were conducted. Among the above fragments, the DOCK2(1–177) fragment, consisting of the SH3 domain and the flanking region, and the ELMO1(532–727) fragment, consisting of the PH domain and the Pro-rich sequence, were selected as suitable fragments for crystallographic analyses and separately cloned into the pCR2.1 vector. By co-expression of the DOCK2(1–177) and ELMO1(532–727) fragments by the large-scale dialysis-mode cell-free synthesis method, the DOCK2(1–177)•ELMO1(532–727) complex protein was obtained in a soluble, selenomethionine-labeled form, whereas the DOCK2(1–177) fragment alone precipitated during the cell-free synthesis. The DOCK2(1–177)•ELMO1(532–727) complex protein, purified by histidine-tag affinity chromatography, histidine-tag cleavage with TEV protease, ion-exchange chromatography, and gel-filtration chromatography, was crystallized and the structure was determined at 2.1-Å resolution, as shown in Fig. 4c (PDB ID: 3A98, Structure Weight: 89,622.52). The complex structure revealed the structural basis for the mutual relief of DOCK2 and ELMO1 from their autoinhibited forms.

3.2 Rab27B•Slac2-a

Rab27A is required for actin-based melanosome transport in mammalian skin melanocytes. Rab27A (221 residues) and its isoform Rab27B (218 residues) bind to several effectors in common, including their specific effector, Slac2-a/melanophilin (590 residues).

We chose a C-terminally truncated form of the GTPase-deficient mutant Rab27B(Q78L) (residues 1–201; designated simply as Rab27B(1–201) hereafter) and the minimum effector region of Slac2-a that specifically binds to the GTP-bound form of Rab27

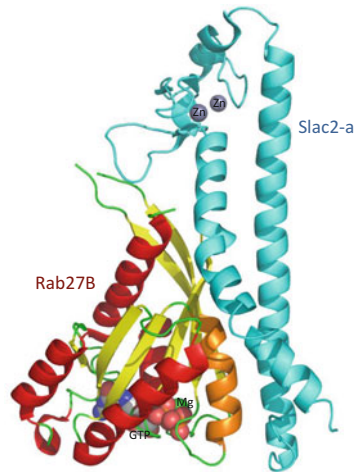


Fig. 5 Crystal structure of the Rab27B•Slac2-a complex. *Ribbon* representation of the Rab27B•Slac2-a complex structure (PDB ID: 2ZET). Rab27B is colored *red, orange, and yellow*. Slac2-a is colored cyan. Zn²⁺, GTP, and Mg²⁺ are represented by spheres

(residues 1–146; designated as Slac2-a(1–146)) and produced their complex, Rab27B•Slac2-a, by the *E. coli* cell-free production method. First, two PCR-amplified DNA fragments encoding the proteins were independently cloned into the pCR2.1 vector (Invitrogen), as fusions with an N-terminal histidine tag (a modified HAT tag) and TEV protease cleavage site. The selenomethionine-labeled Rab27B•Slac2-a complex was obtained in a soluble form by the cell-free co-expression synthesis method, with 50 μ M ZnCl₂ present in the reaction solution. The Rab27B•Slac2-a complex was stable and monomeric with 1:1 stoichiometry, as determined by gel filtration. The purified Rab27B•Slac2-a complex was crystallized and the structure was determined at 3.0-Å resolution, as shown in Fig. 5 (PDB ID: 2ZET, Structure Weight: 83,464.15). The crystal structure revealed the residues involved in the specific Rab27B•Slac2-a interaction [52].

3.3 V-ATPase

By using cell-free synthesized protein complex samples, high-quality structures of the *Enterococcus hirae* V1-ATPase A₃B₃ [53], DF [54, 55], and A₃B₃DF [53–55] complexes were determined by X-ray crystallography.

The PCR-amplified DNA fragments encoding the *E. hirae* V1-ATPase subunits A, B, D, and F (Eh-A, -B, -D, -F) were independently subcloned into the pCR2.1 vector (Invitrogen). These subunit proteins could only be expressed in the soluble forms by co-expression, and they formed the stoichiometric complexes by the cell-free protein synthesis method. To form the stable subcomplex and whole complex, the optimum concentrations of the plasmid

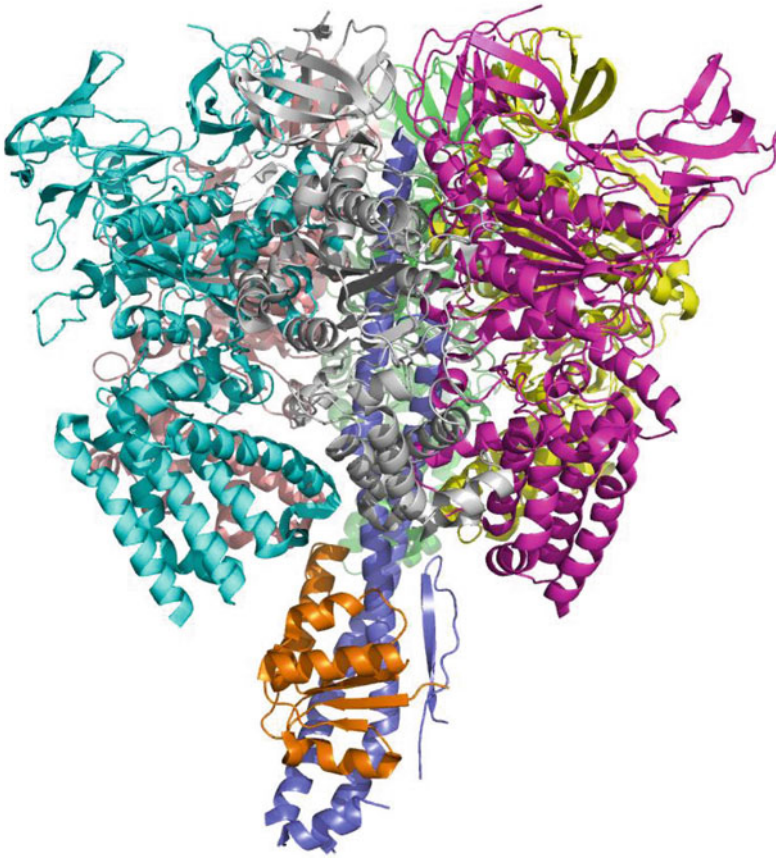


Fig. 6 Crystal structure of the *E. hirae* V-ATPase A₃B₃DF complex. *Ribbon* representation of the *E. hirae* V-ATPase A₃B₃DF complex (PDB ID: 3VR4)

DNA templates were determined by small-scale cell-free expression.

To promote X-ray crystallographic analyses, the selenomethionine-substituted Eh-A₃B₃ [53] and Eh-DF [54, 55] proteins were synthesized by the large-scale dialysis-mode *E. coli* cell-free method. The Eh-A₃B₃DF complex (Fig. 5.6, PDB ID: 3VR4, Structure Weight: 399,405.1) [53] was reconstituted from the Eh-A₃B₃ [53] and Eh-DF [54, 55] subcomplexes.

Using 27 mL of the cell-free reaction solution, more than 15 mg of the purified complex proteins were produced [54, 55].

3.4 Complexes of Disulfide-Bonded Proteins

Disulfide bond formation is required for the correct folding and structural stabilization of secreted and membrane proteins [56]. Cells have protein-folding catalysts to ensure that the correct pairs of cysteine residues interact during the folding process [57]. These enzymatic systems are located in the endoplasmic reticulum (ER) of eukaryotes and the periplasm of Gram-negative bacteria [58]. In bacteria, electron transfer occurs through cascades of disulfide

bond formation/reduction between a series of proteins (DsbA, DsbB, DsbC, and DsbD) [46]. However, the overproduction of disulfide-bonded proteins by *E. coli* cells tends to result in precipitation, aggregation, or inclusion body formation, thus requiring protein solubilization and refolding.

We applied the cell-free synthesis method to the large-scale preparation of a variety of heterodimeric complexes of disulfide-bonded proteins and determined their crystal structures, including the complexes such as (1) the extracellular domains (ECDs) of the calcitonin receptor-like receptor (CLR) and the receptor activity-modifying protein 2 (RAMP2) [the adrenomedullin 1 (AM₁) receptor] [59] and (2) the secreted homodimeric interleukin-5 (IL-5) and the IL-5 receptor α -subunit (IL-5RA) ECDs [60]. Human CLR (residues 23–136, including three pairs of Cys residues forming disulfide bonds) and human RAMP2 (residues 56–139, including two pairs of Cys residues forming disulfide bonds) were cloned into the TA vector pCR2.1TOPO (Life Technologies). The CLR and RAMP2 ECDs were produced as fusions with an N-terminal histidine tag and a TEV cleavage site. The selenomethionine-labeled proteins were synthesized by the *E. coli* cell-free method, using the large-scale dialysis mode [9, 22]. The CLR and RAMP2 ECDs both precipitated during synthesis. The precipitated proteins were denatured with 50 mM Tris-HCl buffer (pH 8.3), containing 8 M guanidine hydrochloride and 20 mM DTT, and were refolded together (co-refolded) by rapid dilution into 50 mM Tris-HCl buffer (pH 8.3), containing 1 M arginine hydrochloride, 5 mM reduced glutathione, and 0.5 mM oxidized glutathione. The co-refolded CLR•RAMP2 ECD complex was successfully purified to homogeneity by chromatography, after the affinity tags were enzymatically removed by TEV protease. The disulfide bonds were properly formed during the co-refolding process. By a similar method, human IL-5 (residues 23–134, including two pairs of Cys residues for disulfide bonding per subunit) and human IL-5RA (residues 21–335, including three pairs of Cys residues for disulfide bonding) were synthesized and co-refolded. The co-refolded IL-5•IL-5RA ECD complex was also successfully purified [60].

For larger proteins and more difficult complexes with numerous disulfide bonds, co-translational disulfide bonding is necessary. The openness of the cell-free system offers direct and flexible control of the reaction environment to promote proper disulfide-bond formation. Several groups have developed cell-free synthesis methods for disulfide-bonded proteins, based on crude extracts from *E. coli*, wheat germ, or insect cells [9, 22, 61–65]. To facilitate disulfide-bond formation, glutathione buffer is used to control the relatively oxidative environment. Usually, glutathione buffer is composed of 0–5 mM oxidized glutathione (glutathione-S-S-glutathione, GSSG) or a mixture of various ratios of oxidized

glutathione (GSSG) and reduced glutathione (GSH). In addition, incorrectly formed disulfide bonds are reshuffled by the addition of 0.2–0.8 mg/ml disulfide isomerase, DsbC, or another protein disulfide isomerase (PDI). To favor disulfide bond formation, the reducing agent should be removed to maintain the oxidizing conditions. In fact, by the cell-free co-expression method, the above-mentioned CLR•RAMP2 ECD complex can be produced in the disulfide-bonded and soluble form without refolding (the final yield is 0.3 mg purified complex protein/ml reaction solution).

The cell-free synthesis method enables the efficient synthesis of antibody fragments by co-expression of the heavy chain (Hc) and light chain (Lc) genes encoding the Fv or Fab fragment. Several hundred micrograms of functional anti-human IL-23 single-chain Fv and anti-human IL-13 α 1R Fab fragment were produced from a 1 ml batch reaction [66, 67]. Intact mouse IgG1 against human creatine kinase was successfully produced, although the productivity was relatively low (0.5 μ g/ml reaction) even with the dialysis-mode cell-free system [68]. Structural analyses require milligram quantities of protein samples. The batch-mode cell-free synthesis method is unable to produce sufficient amounts for this purpose. For large-scale disulfide-bonded protein production, the dialysis-mode cell-free synthesis method has been improved. For example, 3.3 mg/ml of human lysozyme-C was obtained from 1 ml reaction solution in 6 h [69]. This method can be applied to produce antibody fragments, including Fv, scFv, and Fab, for structural analysis.

Acknowledgments

We thank Mr. K. Ake, Ms. T. Imada, and Ms. T. Nakayama for their assistance in the manuscript preparation. This work was supported by the RIKEN Structural Genomics/Proteomics Initiative (RSGI), the National Project on Protein Structural and Functional Analyses, the Targeted Proteins Research Program, and the Platform for Drug Discovery, Informatics, and Structural Life Science, of the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan (to S.Y.).

References

1. Zubay G (1973) In vitro synthesis of protein in microbial systems. *Annu Rev Genet* 7:267–287
2. Pratt JM (1984) Coupled transcription-translation in prokaryotic cell-free system. In: Hames BD, Higgins SJ (eds) *Transcription and translation*. IRL Press, Washington, DC, pp 179–209
3. Spirin AS, Baranov VI, Ryabova LA et al (1988) A continuous cell-free translation system capable of producing polypeptides in high yield. *Science* 242:1162–1164
4. Kigawa T, Yokoyama S (1991) A continuous cell-free protein synthesis system for coupled transcription-translation. *J Biochem* 110:166–168
5. Kigawa T, Muto Y, Yokoyama S (1995) Cell-free synthesis and amino acid-selective stable

- isotope labeling of proteins for NMR analysis. *J Biomol NMR* 6:129–134
6. Kim DM, Kigawa T, Choi C-Y et al (1996) A highly efficient cell-free protein synthesis system from *Escherichia coli*. *Eur J Biochem* 239:881–886
 7. Yabuki T, Kigawa T, Dohmae N et al (1998) Dual amino acid-selective and site-directed stable-isotope labeling of the human c-Ha-Ras protein by cell-free synthesis. *J Biomol NMR* 11:295–306
 8. Kigawa T, Yabuki T, Yoshida Y et al (1999) Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett* 442:15–19
 9. Kigawa T, Yabuki T, Matsuda N et al (2004) Preparation of *Escherichia coli* cell extract for highly productive cell-free protein expression. *J Struct Funct Genomics* 5:63–68
 10. Kigawa T (2010) Cell-free protein preparation through prokaryotic transcription-translation methods. *Methods Mol Biol* 607:1–10
 11. Madin K, Sawasaki T, Ogasawara T et al (2000) A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes. *Proc Natl Acad Sci U S A* 97:559–564
 12. Takai K, Endo Y (2010) The cell-free protein synthesis system from wheat germ. *Methods Mol Biol* 607:23–30
 13. Takai K, Sawasaki T, Endo Y (2010) Practical cell-free protein synthesis system using purified wheat embryos. *Nat Protoc* 5(2):227–238
 14. Tarui H, Imanishi S, Hara T (2000) A novel cell-free translation/glycosylation system prepared from insect cells. *J Biosci Bioeng* 90:508–514
 15. Wakiyama M, Kaitsu Y, Yokoyama S (2006) Cell-free translation system from *Drosophila* S2 cells that recapitulates RNAi. *Biochem Biophys Res Commun* 343:1067–1071
 16. Suzuki T, Ezure T, Ito M et al (2009) An insect cell-free system for recombinant protein expression using cDNA resources. *Methods Mol Biol* 577:97–108
 17. Mikami S, Masutani M, Sonenberg N et al (2006) An efficient mammalian cell-free translation system supplemented with translation factors. *Protein Expr Purif* 46:348–357
 18. Mikami S, Kobayashi T, Masutani M et al (2008) A human cell-derived *in vitro* coupled transcription/translation system optimized for production of recombinant proteins. *Protein Expr Purif* 62:190–198
 19. Kim DM, Choi CH (1996) A semicontinuous prokaryotic coupled transcription/translation system using a dialysis membrane. *Biotechnol Prog* 12:645–649
 20. Kigawa T, Yamaguchi-Nunokawa E, Kodama K et al (2002) Selenomethionine incorporation into a protein by cell-free synthesis. *J Struct Funct Genomics* 2:29–35
 21. Matsuda T, Koshiba S, Tochio N et al (2007) Improving cell-free protein synthesis for stable-isotope labeling. *J Biomol NMR* 37:225–229
 22. Kigawa T, Matsuda T, Yabuki T, et al (2008) Bacterial cell-free system for highly efficient protein synthesis. In: Spirin AS, Swartz JR (eds) *Cell-free protein synthesis*. Wiley-VCH, pp 83–97
 23. Kigawa T, Inoue M, Aoki M, et al (2008) The use of the *Escherichia coli* cell-free protein synthesis for structural biology and structural proteomics. In: Spirin AS, Swartz JR (eds) *Cell-free protein synthesis*. Wiley-VCH, pp 99–109
 24. Kigawa T (2010) Cell-free protein production system with the *E. coli* crude extract for determination of protein folds. *Methods Mol Biol* 607:101–111
 25. Jackson AM, Boutell J, Cooley N et al (2003) Cell-free protein synthesis for proteomics. *Brief Funct Genomic Proteomic* 2:308–319
 26. Carlson ED, Gan R, Hodgman CE et al (2012) Cell-free protein synthesis: applications come of age. *Biootechnol Adv* 30:1185–1194
 27. Yabuki T, Motoda Y, Hanada K et al (2007) A robust two-step PCR method of template DNA production for high-throughput cell-free protein synthesis. *J Struct Funct Genomics* 8:173–191
 28. Aoki M, Matsuda T, Tomo Y et al (2009) Automated system for high-throughput protein production using the dialysis cell-free method. *Protein Expr Purif* 68:128–136
 29. Seki E, Matsuda N, Yokoyama S et al (2008) Cell-free protein synthesis system from *Escherichia coli* cells cultured at decreased temperatures improves productivity by decreasing DNA template degradation. *Anal Biochem* 377:156–161
 30. Terada T, Murata T, Shirouzu M et al (2014) Cell-free expression of protein complexes for structural biology. *Methods Mol Biol* 1091:151–159
 31. Matsuda T, Kigawa T, Koshiba S et al (2006) Cell-free synthesis of zinc-binding proteins. *J Struct Funct Genomics* 7:93–100
 32. Wada T, Shirouzu M, Terada T et al (2003) Structure of a conserved CoA-binding protein synthesized by a cell-free system. *Acta Crystallogr D Biol Crystallogr* 59:1213–1218
 33. Yokoyama J, Matsuda T, Koshiba S et al (2010) An economical method for producing stable-

- isotope labeled proteins by the *E. coli* cell-free system. *J Biomol NMR* 48(4):193–201
34. Yokoyama J, Matsuda T, Koshihara S et al (2011) A practical method for cell-free protein synthesis to avoid stable isotope scrambling and dilution. *Anal Biochem* 411(2):223–229
 35. Hirao I, Ohtsuki T, Fujiwara T et al (2002) An unnatural base pair for incorporating amino acid analogs into proteins. *Nat Biotechnol* 20:177–182
 36. Kiga D, Sakamoto K, Kodama K et al (2002) An engineered *Escherichia coli* tyrosyl-tRNA synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukaryotic translation and its application in a wheat germ cell-free system. *Proc Natl Acad Sci U S A* 99:9715–9720
 37. Kodama K, Fukuzawa S, Nakayama H et al (2006) Regioselective carbon-carbon bond formation in proteins with palladium catalysis; new protein chemistry by organometallic chemistry. *ChemBiochem* 7:134–139
 38. Mukai T, Yanagisawa T, Ohtake K et al (2011) Genetic-code evolution for protein synthesis with non-natural amino acids. *Biochem Biophys Res Commun* 411:757–761
 39. Mukai T, Hayashi A, Iraha F et al (2010) Codon reassignment in the *Escherichia coli* genetic code. *Nucleic Acids Res* 38:8188–8195
 40. Yokoyama S, Hirota H, Kigawa T et al (2000) Structural genomics project in Japan. *Nat Struct Biol* 7(Suppl):943–945
 41. Yokoyama S (2003) Protein expression systems for structural genomics and proteomics. *Curr Opin Chem Biol* 7:39–43
 42. Yokoyama S, Terwilliger TC, Kuramitsu S et al (2007) RIKEN aids international structural genomics efforts. *Nature* 445:21
 43. URL: <http://www.genomics.agilent.com/article.jsp?pageId=484>
 44. Davanloo P, Rosenberg AH, Dunn JJ et al (1984) Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proc Natl Acad Sci U S A* 81:2035–2039
 45. Thomas JG, Ayling A, Baneyx F (1997) Molecular chaperones, folding catalysts, and the recovery of active recombinant proteins from *E. coli*. To fold or to refold. *Appl Biochem Biotechnol* 66:197–238
 46. Kadokura H, Katzen F, Beckwith J (2003) Protein disulfide bond formation in prokaryotes. *Annu Rev Biochem* 72:111–135
 47. Muller M, Koch HG, Beck K et al (2001) Protein traffic in bacteria: multiple routes from the ribosome to and across the membrane. *Prog Nucleic Acid Res Mol Biol* 66:107–157
 48. Weski J, Ehrmann M (2012) Genetic analysis of 15 protein folding factors and proteases of the *Escherichia coli* cell envelope. *J Bacteriol* 194:3225–3233
 49. Sakamoto K, Murayama K, Oki K et al (2009) Genetic encoding of 3-iodo-L-tyrosine in *Escherichia coli* for single-wavelength anomalous dispersion phasing in protein crystallography. *Structure* 17:335–344
 50. Kainosho M, Torizawa T, Iwashita Y et al (2006) Optimal isotope labelling for NMR protein structure determinations. *Nature* 440:52–57
 51. Hanawa-Suetsugu K, Kukimoto-Niino M, Mishima-Tsumagari C et al (2012) Structural basis for mutual relief of the Rac guanine nucleotide exchange factor DOCK2 and its partner ELMO1 from their autoinhibited forms. *Proc Natl Acad Sci U S A* 109:3305–3310
 52. Kukimoto-Niino M, Sakamoto A, Kanno E et al (2008) Structural basis for the exclusive specificity of Slac2-a/melanophilin for the Rab27 GTPases. *Structure* 16:1478–1490
 53. Arai S, Saijo S, Suzuki K et al (2013) Rotation mechanism of *Enterococcus hirae* V1-ATPase based on asymmetric crystal structures. *Nature* 493:703–707
 54. Arai S, Yamato I, Shiokawa A et al (2009) Reconstitution *in vitro* of the catalytic portion (NtpA₃-B₃-D-G complex) of *Enterococcus hirae* V-type Na⁺-ATPase. *Biochem Biophys Res Commun* 390:698–702
 55. Saijo S, Arai S, Hossain KM et al (2011) Crystal structure of the central axis DF complex of the prokaryotic V-ATPase. *Proc Natl Acad Sci U S A* 108:19955–19960
 56. Creighton TE (1988) Toward a better understanding of protein folding pathways. *Proc Natl Acad Sci U S A* 85:5082–5086
 57. Paget MS, Buttner MJ (2003) Thiol-based regulatory switches. *Annu Rev Genet* 37:91–121
 58. Sevier CS, Kaiser CA (2002) Formation and transfer of disulphide bonds in living cells. *Nat Rev Mol Cell Biol* 3:836–847
 59. Kusano S, Kukimoto-Niino M, Hino N et al (2012) Structural basis for extracellular interactions between calcitonin receptor-like receptor and receptor activity-modifying protein 2 for adrenomedullin-specific binding. *Protein Sci* 21:199–210
 60. Kusano S, Kukimoto-Niino M, Hino N et al (2012) Structural basis of interleukin-5 dimer recognition by its α receptor. *Protein Sci* 21:850–864
 61. Goerke AR, Swartz JR (2008) Development of cell-free protein synthesis platforms for disulfide bonded proteins. *Biotechnol Bioeng* 99:351–367

62. Michel E, Wüthrich K (2012) Cell-free expression of disulfide-containing eukaryotic proteins for structural biology. *FEBS J* 279:3176–3184
63. Kawasaki T, Gouda MD, Sawasaki T et al (2003) Efficient synthesis of a disulfide-containing protein through a batch cell-free system from wheat germ. *Eur J Biochem* 270:4780–4786
64. Ezure T, Suzuki T, Shikata M et al (2007) Expression of proteins containing disulfide bonds in an insect cell-free system and confirmation of their arrangements by MALDI-TOF MS. *Proteomics* 7:4424–4434
65. Stech M, Merk H, Schenk JA et al (2012) Production of functional antibody fragments in a vesicle-based eukaryotic cell-free translation system. *J Biotechnol* 164:220–231
66. Yin G, Garces ED, Yang J, et al (2012) Aglycosylated antibodies and antibody fragments produced in a scalable *in vitro* transcription-translation system. *MAbs* 4(2)
67. Matsuda T, Furumoto S, Higuchi K et al (2012) Rapid biochemical synthesis of ¹¹C-labeled single chain variable fragment antibody for immuno-PET by cell-free protein synthesis. *Bioorg Med Chem* 20(22):6579–6582
68. Frey S, Haslbeck M, Hainzl O et al (2008) Synthesis and characterization of a functional intact IgG in a prokaryotic cell-free expression system. *Biol Chem* 389:37–45
69. Matsuda T, Watanabe S, Kigawa T (2013) Cell-free synthesis system suitable for disulfide-containing proteins. *Biochem Biophys Res Commun* 431:296–301