Journal of Biomolecular NMR **30:** 311–325, 2004.

Efficient production of isotopically labeled proteins by cell-free synthesis: A practical protocol

Takuya Torizawa[a], Masato Shimizu[a], Masato Taoka[b], Hiroshi Miyano[c] & Masatsune Kainosho*[,a,b]

[a]*CREST, JST;* [b]*Graduate School of Science, Tokyo Metropolitan University, 1-1 Minami-ohsawa, Hachioji 192-0397, Japan,* [c]*Ajinomoto Co., Inc. Institute of Life Sciences, 1-1 Suzuki-cho, Kwasaki, 210-8681, Japan*

Received 21 July 2004; Accepted 31 August 2004

Key words: cell-free extract from *E. coli*, cell-free protein synthesis, NMR assignments, protocol optimization, stable isotope labeling

Abstract

We provide detailed descriptions of our refined protocols for the cell-free production of labeled protein samples for NMR spectroscopy. These methods are efficient and overcome two critical problems associated with the use of conventional *Escherichia coli* extract systems. Endogenous amino acids normally present in *E. coli* S30 extracts dilute the added labeled amino acids and degrade the quality of NMR spectra of the target protein. This problem was solved by altering the protocol used in preparing the S30 extract so as to minimize the content of endogenous amino acids. The second problem encountered in conventional *E. coli* cell-free protein production is non-uniformity in the N-terminus of the target protein, which can complicate the NMR spectra. This problem was solved by adding a DNA sequence to the construct that codes for a cleavable N-terminal peptide tag. Addition of the tag serves to increase the yield of the protein as well as to ensure a homogeneous protein product following tag cleavage. We illustrate the method by describing its stepwise application to the production of calmodulin samples with different stable isotope labeling patterns for NMR analysis.

Introduction

Cell extracts from several organisms have been used for the purpose of cell-free protein synthesis (Clemens and Prujin, 1999; Kramer et al., 1999). Proteins that cannot be produced from cells because they are toxic to the host cells can be made from cell-free systems (Henrich et al., 1982; Chrunyk et al., 1993). In addition, proteins that tend to be digested by intracellular proteases (Goff and Goldberg, 1987; Maurizi, 1987) or aggregate in inclusion bodies (Chrunyk et al., 1993) in cells sometimes can be produced from cell-free extracts. Labeled amino acids can be incorporated efficiently and with minimal scrambling of the label, because the amino acid metabolic activity of cell-free extracts is low (Kigawa et al., 1995). Cell-free systems are now considered as alternatives to conventional in vivo expression. Commercial products are now available that provide improved yields of protein products and increased convenience (Betton, 2003). Most recently, it has been shown that protein synthesis can be carried out by a system, named the 'PURE' system, reconstituted from purified components of the translation machinery (Shimizu et al., 2001).

Cell-free systems offer particular advantages for preparing samples for NMR spectroscopy,

*To whom correspondence should be addressed. E-mail: kainosho@nmr.chem.metro-u.ac.jp

where one can take advantage of the inactive amino acid metabolism for stable isotope labeling, particularly for selective labeling (Kigawa et al., 1995, 1999; Guignard et al., 2002; Klammt et al., 2004; Shi et al., 2004). The ideal cell-free system for NMR spectroscopy would be one that is capable of producing milligram-scale samples of labeled protein in a robust, reproducible fashion at reasonable cost. Commercial cell-free systems, including the PURE system, however, are relatively expensive.

We present here an economical approach to preparing *E. coli* extracts for protein synthesis, along with a detailed description of the steps involved in producing labeled samples for NMR spectroscopic investigations. Our approach overcomes two major problems with conventional *E. coli* cell-free systems: the dilution of added labeled amino acids by endogenous, natural-abundance amino acids from the extract and the production of protein mixtures with heterogeneous N-termini. We describe methods for avoiding difficulties and for increasing protein yields. We illustrate the approach through a description of the preparation of calmodulin samples with different stable isotope labeling patterns.

Materials and methods

All chemicals were purchased from Sigma or Wako unless stated otherwise. All strains of *E. coli* were cultured at 37 °C. Other biochemical procedures were carried out at 4 °C, unless otherwise noted. Buffers used in preparing S30 extracts and in cell-free reactions contained 0.5% (v/v) diethylpyrocarbonate (DEPC) (Sigma) to inactivate RNases and were autoclaved. (The medium used for *E. coli* cell culture did not contain DEPC.) Instruments and glassware used in preparing S30 extracts and in cell-free synthesis reactions were washed with autoclaved water containing 0.5% (v/v) DEPC (denoted as DEPC-treated water) and subsequently autoclaved if possible.

E. coli *expression*

The gene for calmodulin (CaM) from *Xenopus laevis* (gift from M. Ikura) was cloned into pET3a (Novagen), pET15b (Novagen), and pET15b modified so as to remove the encoded $(His)_6$-tag. The modified pET15b vector was prepared by replacing the sequence between NcoI and NdeI in pET15b with a DNA oligomer (purchased from Proligo) that removed the sequence coding for the $(His)_6$-tag. These plasmids were transformed into *E. coli* BL21 (DE3) cells (Invitrogen) for expression. The gene for *E. coli* peptidyl-prolyl cis-trans isomerase b (EPPIb) (Kariya et al., 2000) was cloned into pET11a by means NdeI and BamHI restriction sites. This plasmid was transformed into BL21 (DE3) cells for expression. The strains were cultured in LB medium until OD_{600} reached 0.5 and incubated for 3 h after induction by adding isopropyl β-D-thiogalactopyranoside (IPTG) (Wako) to a level of 1 mM.

Preparation of S30 extract

The procedure followed was based on earlier work (Zubay, 1973; Davanloo et al., 1984; Pratt, 1984; Kigawa et al., 1999) and includes the improvements described here. *E. coli* cells, A19 (*met* B, *RNA*) or BL21 Star (DE3) (Invitrogen), were grown in 100 ml LB medium overnight. This served as the inoculum for 10 l of a medium containing 56 g KH_2PO_4, 289 g K_2HPO_4, 10 g Bacto yeast extract (DIFCO), 15 mg thiamine, 1 mM magnesium acetate and 2% (w/v) glucose, which was grown until OD_{650} reached 0.7. The cells from this growth were harvested by centrifugation, washed four times by suspending in 200 mL of S30 buffer (10 mM Tris–acetate (pH 8.2), 14 mM magnesium acetate, 60 mM potassium acetate and 1 mM dithiothreitol (DTT) containing 0.05% (v/v) 2-meraptoethanol (2-ME), and centrifuged again. Before the last wash, the pellets (18 g) can be stored at −80 °C. After the final wash step, the cells were resuspended in 23 ml S30 buffer and were disrupted with a French Press (Ohtake) at a constant pressure of 20,000 psi. DTT was added to the lysate to a concentration of 1 mM. The lysate was then centrifuged at $30{,}000 \times g$ for 30 min, after which the top 24 ml of the supernatant was withdrawn and recentrifuged. The top 18 ml of that supernatant was withdrawn carefully, and to it was added 5.4 ml of a solution composed of 293 mM Tris–acetate, 9.2 mM magnesium acetate, 4.4 mM DTT, 40 μM of each of the 20 common amino acids, 200 mM Tris (for

pH adjustment), 84 mM phospho(enol)pyruvate, 69.4 mM ATP, and 6.7 units/ml of pyruvate kinase; the resulting mixture was shaken at 80 rpm at 37 °C for 80 min. The resulting solution was dialyzed for 45 min against 2 l of S30 buffer three times using a Spectra/Por membrane with a molecular weight cut off (MWCO) of 6–8000 and a diameter of 20.4 mm (Spectrum). The retained extract was applied to a Sephadex G-25 medium gel-filtration column (2.5 cm × 18 cm; Amersham Biosciences) equilibrated with S30 buffer, and a fraction was collected equivalent to 1.4 times the applied volume. The collected fraction was concentrated to 0.86 times the original volume by dialysis (membrane described above) against 700 ml of an equal weight mixture of polyethylene glycol 8000 (PEG-8000) and S30 buffer. The concentrated extract was then dialyzed against 2 l of S30 buffer for 1 h. The extract was transferred into Eppendorf tubes, frozen with liquid nitrogen, and stored at −80 °C.

Preparation of T7 RNA polymerase

The method is based on published procedures (Zawadzki and Gross, 1991; Grodberg and Dunn, 1988). *E. coli* BL21 (DE3) transformed with pAR1219 (Davanloo et al., 1984) was grown overnight in 10 ml LB medium containing 50 μg/ml ampicillin. This culture served as the inoculum for 1 l of a medium containing 1 g of tryptone peptone (DIFCO), 5 g NaCl, 1 g NH_4Cl, 3 g KH_2PO_4, 6 g Na_2HPO_4, 1 mM $MgSO_4$, 0.4% (w/v) glucose, and 50 μg/ml ampicillin. This culture was grown until OD_{600} reached 0.5. IPTG was added to a final concentration of 0.5 mM to induce production of T7 RNA polymerase, and the culture was incubated for an additional 8 h. The harvested cells were suspended in a buffer composed of 50 mM Tris–HCl (pH 8.1), 20 mM NaCl, 2 mM EDTA, 1 mM DTT, 0.3 mg/ml lysozyme (Wako), 4 μg/ml PMSF (Wako), 2 μg/ml bacitracin (Wako), and 20 μM benzamidine (Wako) and stirred for 15 min. Following the addition of sodium deoxycholate (Wako) to a concentration of 0.8% (w/v), the solution was further stirred for 15 min and sonicated. After the addition of 15 ml of 2 M ammonium sulfate, the volume was brought to 150 ml by the addition of a buffer containing 50 mM Tris–HCl (pH 8.1), 20 mM NaCl, 2 mM EDTA, and 1 mM DTT. Then 15 ml of 10% (w/v) polyethyleneimine that had been adjusted to pH 7.0 with HCl was added gradually, and the mixture was stirred for 20 min. The product was centrifuged at $39{,}000 \times g$ for 15 min, and the supernatant was collected. To this was gradually added 0.82 volume of saturated ammonium sulfate that had been adjusted to pH 7.0 with Tris, and the mixture was stirred for 15 min. The pellets collected after centrifugation at $12{,}000 \times g$ for 15 min were suspended with 45 ml of dialysis buffer (20 mM sodium phosphate pH 7.7, 1 mM EDTA, 1 mM DTT, 5% glycerol) containing 100 mM NaCl and 20 μg/ml PMSF. The suspension was dialyzed three times, each time for 3 h against 2 l of the dialysis buffer containing 100 mM NaCl and 20 μg/ml PMSF, with use of a Spectra/Por membrane (MWCO, 6–8000; diameter, 20.4 mm). The retained fraction was diluted with 1 volume of dialysis buffer containing 20 μg/ml PMSF and loaded onto an SP Sepharose FF column (3 cm × 30 cm, Amersham Biosciences) equilibrated with dialysis buffer containing 50 mM NaCl and 20 mg/l PMSF. After washing the column with the same buffer used for equilibration, T7 RNA polymerase was eluted with dialysis buffer containing 200 mM NaCl and 20 μg/ml PMSF. The eluted material was dialyzed (using a membrane of the type described above) three times, the first two times for 6 h against dialysis buffer containing 100 mM NaCl and 20 μg/ml PMSF and once for 6 h against dialysis buffer containing 100 mM NaCl and 50% glycerol. The retained fraction was centrifuged, and the supernatant was collected. The amount of polymerase was determined spectroscopically ($1\ \text{mg/ml} = 98{,}000 \times OD_{280}/(1.4 \times 10^5)$), and the supernatant was stored at −20 °C.

Preparation of template DNA for cell free reaction

All plasmids for the cell free reactions, except that for EPPIb, were prepared from pIVEX2.3d (Roche); in the case of EPPIb, the pET vector used for *in vivo* expression was used also for the cell-free protein synthesis. Each target gene construct with its extension coding for the N-terminal tag sequence was transferred from the pET vectors used for *in vivo* expression into pIVEX2.3d using *Nco*I, *Bam*HI restriction sites. The sequence of each target gene with its tag was

submitted to the ProteoExpert (Roche) site (http://www.proteoexpert.com/) for prediction of silent mutations that would improve the efficiency of protein synthesis (the cost in Japan is JPY 15,000 per calculation). The original DNA sequence coding for the tag was ATGGGCAGCTCTTCCAGCGGCCTGGTGCCGCGCGGCAGCCATATG. The ten sequences predicted to change the N-terminal sequence of the construct used to produce calmodulin were

1: ATGGGTAGTAGTAGTAGTGGTCTGGTG CCGCGCGGCAGCCATATG
2: ATGGGTAGTAGTAGTTCAGGTCTGGTG CCGCGCGGCAGCCATATG
3: ATGGGCAGTAGTAGTAGTGGTCTGGT GC CGCGCGGCAGCCATATG
4: ATGGGTTCATCATCATCAGGTCTGGTG CCGCGCGGCAGCCATATG
5: ATGGGTTCAAGTAGTAGCGGTCTGGTG CCGCGCGGCAGCCATATG
6: ATGGGTAGTAGTTCATCAGGTCTGGTG CCGCGCGGCAGCCATATG
7: ATGGGCAGTTCATCATCAGGTCTGGTG CCGCGCGGCAGCCATATG
8: ATGGGTTCATCAAGTAGTGGTCTGGTG CCGCGCGGCAGCCATATG
9: ATGGGCTCATCATCATCAGGTCTGGTG CCGCGCGGCAGCCATATG
10:ATGGGCAGTTCATCATCAGGCCTGGT GCCGCGCGGCAGCCATATG

DNA oligomers, designed to include the desired mutations and NcoI and NdeI cleavage sites, were purchased (Proligo); they were introduced into the plasmid by restriction cloning. The plasmid was transformed into *E. coli* strain DH5α (Toyobo) for *in vitro* amplification. The template DNA was isolated and purified according to the instructions of the manufacturer either by a QIAGEN Plasmid Mega Kit (Qiagen) for milligram scale preparation or by a QIAprep Spin Miniprep Kit (Qiagen) for microgram scale. The purified DNA was dissolved in DEPC-treated water and stored at −20 °C. Of the predicted sequences 2 and 8 were found to improve the protein yield, and 8 was used to produce calmodulin.

Cell free protein synthesis reaction

Solution compositions are summarized in Tables 1 and 2 for convenience. Quantities shown in Table 1 are for a small-scale reaction; they are scaled up for a large-scale reaction. PEG-8000 (Table 2) was only added to LM mixtures used for large-scale reactions. A dializer chamber contained the 'reaction solution', and a test tube contained the 'dialysis solution' (Tables 1). The semi-permeable membrane used for the dializer chamber was Spectra/Por CE

Table 1. Reaction solution and dialysis solution

	Stock solution	Reaction solution	Dialysis solution
DEPC-treated water		112 µl	1160.8 µl
Ammonium acetate[a]	1.4 M	9.8 µl	39.2 µl
Magnesium acetate	0.5 M	15 µl	60 µl
Amino acid mixture	25 mM each	20 µl	80 µl
Creatine phosphate (Wako)	0.645 M	40 µl	160 µl
LM mixture		125 µl	500 µl
Template DNA	1 mg/ml	10 µl	–
T7 RNA polymerase	11 mg/ml	4.5 µl	–
RNase inhibitor (Takara)	40 units/µl	1.25 µl	–
Creatine kinase (Roche)	10 mg/ml	12.5 µl	–
S30 extract		150 µl	–
Final volume		0.5 ml	2 ml

[a]Replaced by [^{15}N]-ammonium acetate when preparing proteins labeled with nitrogen-15.

Table 2. LM mixture

	Amounts	Final concentration
2 M HEPES-KOH pH7.5	22 ml	210 mM
6 M potassium acetate	33.4 ml	1 M
DTT	210 mg	6.8 mM
ATP	530 mg	4.8 mM
CTP	338 mg	3.2 mM
GTP	335 mg	3.2 mM
UTP	310 mg	3.2 mM
cAMP	172 mg	2.56 mM
Folinic acid (Sigma)	28 mg	272 µM
tRNA (Roche)	140 mg	0.7 mg/ml
50% (w/v) PEG-8000[a]	64 ml	16 (w/v)%
DEPC-treated water up to 200 ml		

[a]Included only in LM mixtures used for large-scale protein production.

(MWCO of 50,000, Spectrum). For small-scale reactions, a dialyzer chamber with a diameter of 5 mm was placed in a test tube with an inner diameter of 10 mm. For large reactions a dialyzer chamber with a diameter of 15 mm was placed in a test tube with an inner diameter of 20 mm. The test tubes were set at an angle in a water shaker preheated to 37 °C and were shaken at 150 rpm during the dialysis. Following the 7 h reaction, the level of protein production was determined by SDS–PAGE of a sample withdrawn from the dializer chamber of the small-scale reaction apparatus. For large reactions, the synthesized proteins were retrieved from both the dialysis chamber and the test tube. The cell-free system was used to make CaM samples containing three different labelling patterns by appropriate substitution of the amino acid mixture (Table 1): (1) [*U*-^{15}N]-CaM was produced by using 4 mg/ml of ^{15}N-labeled algal hydrolysate (Shoko). (2) [*U*-^{2}H,^{13}C,^{15}N]-CaM was produced by using 4 mg/ml of ^{2}H,^{13}C,^{15}N-labeled algal hydrolysate (Shoko). With this sample, a ^{13}C-filtered ^{1}H pulse sequence was used to suppress signals from undeuterated carbons so that the level of the ^{1}H–^{12}C could be determined and used to assay the level of unlabeled amino acids incorporated. (3) [*U*-^{13}C,^{15}N]-CaM (prepared for analysis of amino acid incorporation levels by LC-MS/MS and CE-MS) was produced by using 1.7 mg/ml of a mixture of ^{13}C,^{15}N-labeled amino acids (purchased from Isotec or Cambridge Isotope Laboratories). Because the extract contains active transaminases, when ^{15}N-labeled samples are prepared, natural abundance ammonium acetate is substituted by [^{15}N]-ammonium acetate (Table 1).

Purification of CaM

Both native and tagged CaM were purified according to published methods (Ikura et al., 1990a, b). The tagged CaM was further treated after the purification as follows. About 0.5 mg/ml of the protein in a buffer composed of 25 mM Tris–HCl pH 7.5 and 2 mM $CaCl_2$ (Buffer A) was digested with thrombin (Amersham Biosciences) (17 units/mg of the protein) at 37 °C for 2 h. Immediately afterward, the solution was applied to a Phenyl Sepharose 6 FF column (1.5 cm × 3 cm; Amersham Biosciences) equilibrated with Buffer A and washed with Buffer A containing 0.5 M NaCl. Subsequently, the protein was eluted with a buffer composed of 25 mM Tris–HCl pH 7.5, 0.5 M NaCl and 2 mM EDTA (Buffer B) and loaded on a Benzamidine Sepharose 6B column (0.8 cm × 1.5 cm; Amersham Biosciences) equilibrated with Buffer B to trap the thrombin. The column was washed with Buffer B, and the CaM was collected in the flow-through. The purified protein was both concentrated and buffer-exchanged by centrifugation in Centriprep YM-3 (Amicon) and Centricon YM-3 (Amicon) tubes. Each NMR sample contained 0.5 mM CaM in 5 mM MES-d_{13} (Cambridge Isotope Laboratories), ca: 10 mM bis-Tris-d_{19} (Cambridge Isotope Laboratories) to adjust pH to 6.5 (Kelly et al., 2002), 5 mM $CaCl_2$, 0.1 mM NaN_3, and 10% D_2O.

NMR measurements

All NMR spectra of CaM were recorded at 37 °C with Bruker DRX800 or DRX600 spectrometers equipped with TXI xyz-gradient probes. Published procedures were used in acquiring ^{13}C-filtered ^{1}H spectra (Zwahlen et al., 1997) and ^{1}H–^{15}N HSQC spectra (Mori et al., 1995). All spectra were processed with XWINNMR version 3.5 software (Bruker). Sequence specific assignments of amide resonances were carried out by spectral analyses of HNCO, HN(CA)CO, HNCACB, and HN(CO)CACB datasets (Clore and Gronenborn, 1994). ^{1}H chemical shifts were referenced to external DSS and ^{15}N shift were indirectly referenced (Wishart et al., 1995; Markley et al., 1998).

LC-MS/MS, CE-MS and MALDI-TOF-MS measurements

An Agilent CE capillary electrophoresis system with an Agilent 1100 series MSD mass spectrometer and Agilent 1100 series isocratic HPLC pump were used in performing the CE-ESI-MS experiments. Amino acids were separated on a fused-silica capillary, 50 μm i.d. (365 μm o.d.) diameter × 100 cm length (Polymicro Technologies, LLC). The flushing buffer and running buffer were 1 M formic acid that had been filtered before use through a 0.22 μm pore-size nylon membrane (Millipore). Prior to each injection,

the capillary was preconditioned for 8 min by flushing with running electrolyte. The sample solution was inserted by pressure injection at 50 mbar for 3.0 sec (~3 nl). The applied voltage was increased linearly from 0 to 25 kV over the initial 1 min and then held at 25 kV for the next 19 min. The capillary temperature was thermostated at 25 °C. The Agilent 1100 series pump equipped with a 1:100 splitter was used to deliver 10 µl/min of 5 mM ammonium acetate in 50% (v/v) methanol–water to the CE interface used as a sheath liquid. ESI-MS was conducted in the positive ion mode, with the capillary voltage set at 3500 V. The mass distribution of each amino acid was detected in selective ion mode so as to monitor the protonated $[M + H]^+$ ion. For example, to investigate the ratio of labeled/total methionine, the spectrometer was scanned by 1 mass unit from m/z 150, which corresponds to $[^{12}C_5H_{12}{}^{14}NO_2S]^+$, to m/z 156, which corresponds to $[^{13}C_5H_{12}{}^{15}NO_2S]^+$. The quantity of protein synthesized was determined by a standard method (Bradford, 1976). For amino acid analysis, an aliquot of the $[^{13}C,^{15}N]$-CaM sample prepared for analysis by LC-MS/MS and CE-MS (corresponding to ~200 µg protein) was hydrolyzed in 6 N HCl at 110 °C for 18 h. Details of the LC-MS/MS procedure will be published elsewhere (Hiroshi Miyano et al., unpublished data).

For MALDI-TOF-MS analysis, 1 µl of matrix solution, 10 mg/ml alpha-cyano-4-hydroxycinamic acid in 50% acetonitrile/0.1% TFA, was mixed with a protein sample directly on a sample plate. Mass spectra were recorded using Voyager Elite DE-STR (Applied Biosystems) in positive reflector mode using myoglobin (MW: 16951.55) as an internal standard.

Results and discussion

Strategy for NMR sample preparation

Our strategy for the preparation of NMR samples is shown in Figure 1 as a simple scheme, and the individual steps are discussed in detail in the following paragraphs. Our approach was based on the pursuit of efficiency and low cost for the cell-free protein synthesis system along with high quality for the NMR sample.

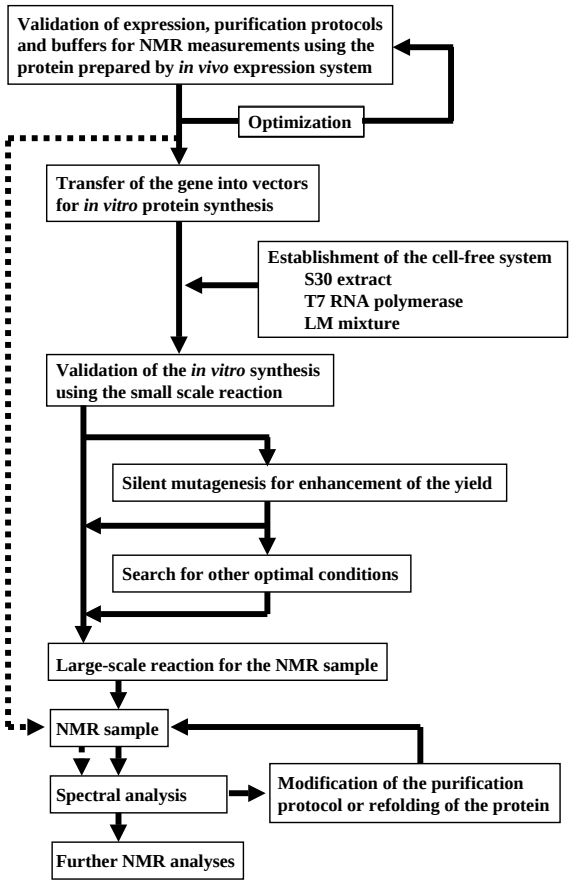

Figure 1. Flow chart showing our cell-free strategy for preparing labeled protein samples for NMR analysis.

Although, commercial cell-free systems are efficient and reproducible (Betton, 2003), they prove to be expensive for the preparation of the large amounts of protein required for NMR structural investigations, and they have not solved the problems specific to NMR samples described below. Our procedure incorporates some of the technology used in the commercial systems and combines it with savings realized by preparing most of the materials ourselves and with improvements that lead to higher quality NMR samples. Typically, it takes one person 2 days to prepare 28 ml of the lysate, which is sufficient for 15 large-scale (6-ml each) synthesis reactions. The approach has been used successfully for the cell-free production of a large number of proteins. SDS–PAGE analyses of many of the proteins synthesized by the cell-free system are shown in Figure 2.

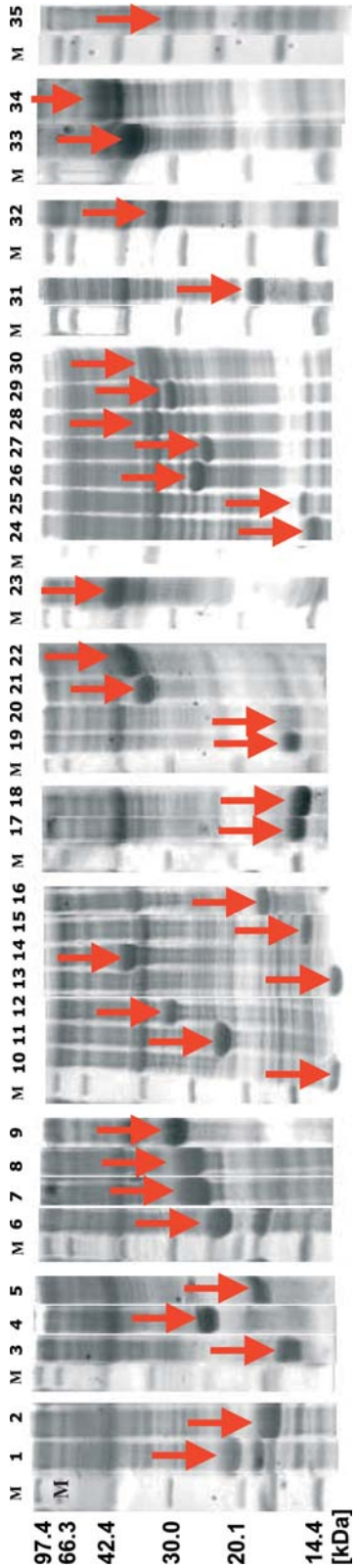

Figure 2. SDS–PAGE analyses of proteins synthesized by the small scale cell-free system: 1. EPPIa, 2. EPPIb, 3. E.coli Parvulin, 4. Human Parvulin, 5. Human Parvulin lacking WW domain, 6. DSBa, 7. DSBc, 8. DSBg, 9. DSBe, 10. drk SH3 domain, 11. CAT1, 12. Calcineurin B, 13. α-spectrin SH3 domain, 14. alkaline phosphatase, 15. SSI, 16. CPI17(35-120), 17.FKBP12.6, 18. GroES, 19. SecB, 20. SecE, 21. leader peptidase, 22. OmpA1, 23. IP3 receptor, 24. RuvC, 25. T. martina RuvC, 26. At5g45600, 27. At5g24650, 28. At4g20010, 29. At2g02570, 30. At1g18800, 31. Calmodulin, 32. At3g16450, 33. MBP, 34. β-Subunit, 35. GFPuv. 'M' denotes molecular weight markers. The synthesized proteins are indicated by an arrow.

Validation of E. coli *expression*

One of the advantages of cell-free systems is that proteins sometimes can be synthesized that cannot be expressed by *in vivo* methods (Henrich et al., 1982; Goff and Goldberg, 1987; Maurizi, 1987; Chrunyk et al., 1993). With the exception of such cases, we find it advantageous to carry out *E. coli* cellular expression of proteins prior to producing them by the cell-free system. The reasons for this are the following. (i) This makes it possible to compare the NMR spectra of the proteins expressed by *in-vivo* and *in-vitro* methods to determine if they are similar. (ii) It is less costly and requires less effort to produce proteins from *E. coli* cells than by the cell-free method. Thus protein prepared by the former method can be used to work out the purification protocols and search for optimal solution conditions prior to preparing the protein by cell-free methods. (iii) We have found that most proteins that express well in *E. coli* cells are produced successfully by the cell-free method. Of 44 proteins found to express in *E. coli* and randomly selected from a library in our laboratory, 35 (~80%) were found to work well in the cell-free synthesis system (Figure 2). Thus a protein that has been shown to express well in *E. coli* cells has a good probability of producing well in the cell-free system.

We have designed the expression vectors so that they are compatible with constructs used for expression from *E. coli* cells or for cell-free production (Figure 3). The *Nde*I and *Bam*HI restriction sites are used for cloning the target gene into the appropriate vector: native sequence (pET3a or pET11a), sequence with N-terminal thrombin-cleavable $(His)_6$-tag (pET15b), sequence with N-terminal thrombin-cleavable

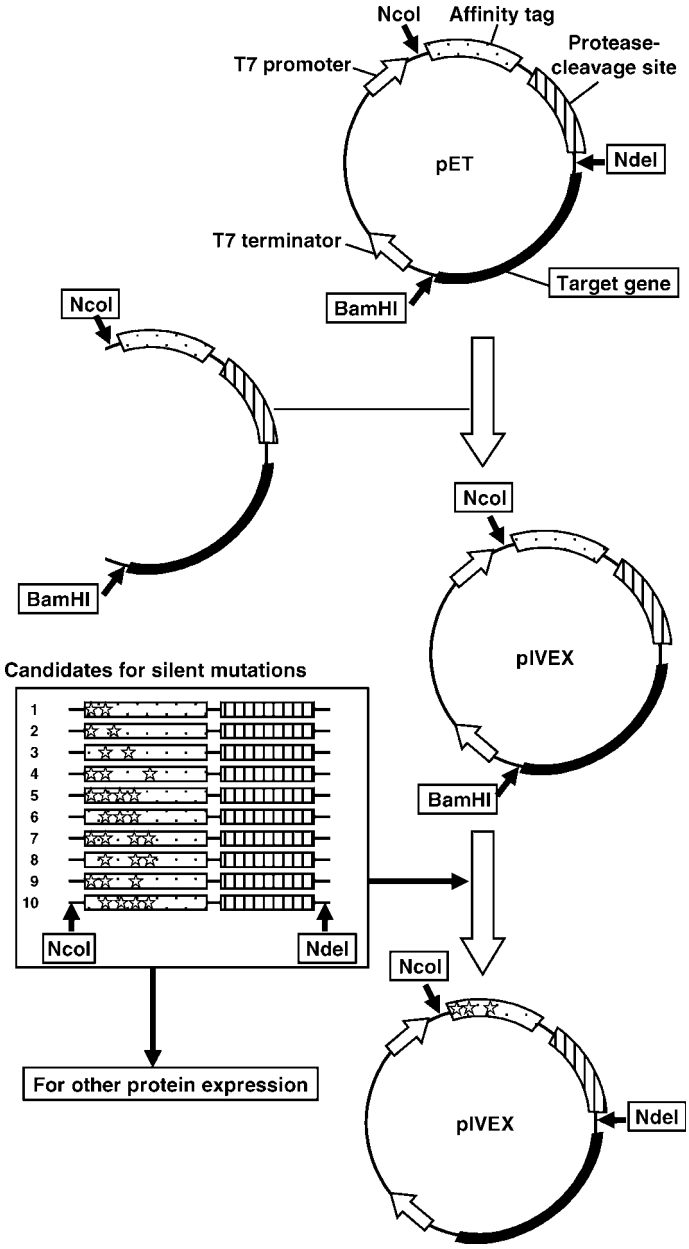

Figure 3. Schematic drawing illustrating the construction of plasmids for expression by *in vivo* and *in vitro* methods.

tag without $(His)_6$ (prepared from pET15b as described above). If the native protein is produced well but the tagged proteins are not (the tags hamper protein synthesis), it is possible that the yield can be increased by the silent mutagenesis method described below. Provided that the yield is adequate, use of a construct with a cleavable tag is preferable to one without, since it will ensure a product (post cleavage) with a homogeneous N-terminus. The affinity tagged version should be used if the purification of the native protein is difficult or time-consuming. Since the protocol for purifying CaM was easier for protein not containing the affinity tag, it was produced by the construct that did not code for the tag.

Construction of cell free protein synthesis system

The preparation of S30 extract is critical to the success of the system. As most investigators and commercial vendors of cell-free systems have emphasized, the most important issue is the level of activity of the extract (Spirin et al., 1988; Kim et al., 1996; Kigawa et al., 1999; Kim and Swartz, 2000; Madin et al., 2000). We first addressed the question of *E. coli* strain selection. Strain A19 has been found to yield a reliable extract because of its lack of RNase I (Kim and Swartz, 2000). Recently BL21 Star (DE3) lacking the activity of RNase E has become available. We found by comparing extracts from each of the two strains prepared by the same method, that the activity of the extract from BL21 Star (DE3) is 1.2–1.4 times higher than that of A19. This led to our choice of BL21 Star (DE3) as the strain for our extract. Furthermore, the activity is strongly dependent on how the extract is prepared and on its concentration. The concentration of the extract has been optimized to maximize the amount of protein produced (Kigawa et al., 1999). For labeled NMR samples, a higher concentration of extract leads to higher yields of protein, but also to dilution of the label by endogenous amino acids in the extract. We have analyzed this problem of label dilution. Figure 4 shows ^{13}C-filtered ^{1}H spectra of [^{15}N]-CaM and [^{2}H,^{13}C,^{15}N]-CaM prepared by the cell-free system. In the aliphatic region of the spectra, only ^{12}C-^{1}H resonances are detected. By comparing the spectrum of ^{15}N-labeled CaM (Figure 4a) with that of [^{2}H,^{13}C,^{15}N]-CaM produced from a conventional cell-free extract prepared without gel-filtration (Figure 4b), one sees that the labeling efficiency is only ~90%. By introducing a gel-filtration step into the protocol for extract production, the incorporation level can be increased to ~96% (Figure 4c) without loss of the activity of the extract.

The extract described here produces a maximum protein yield (EPPIb) of 2 mg/ml of the reaction solution under optimal concentrations of added amino acids (1 mM each). The yield of protein (the amount of the synthesized protein/amount of the added amino, acids) decreases as the levels of added amino acids are increased, although the labeling efficiency increases. Because the labeled amino acids represent the most important cost factor, we adjusted the mole ratios of the individual amino acids to match the amino acid composition of CaM, with exception that the mole ratio of Gly was increased by a factor of 4. The optimal concentration for the amino acid mixture was determined to be 1.7 mg/ml by analyzing protein yields as a function of concentration and by choosing a value slightly below the maximum. A sample of [^{13}C,^{15}N]-CaM prepared with the dialyzed extract and level of amino acids optimized for protein yield, was hydrolyzed to amino acids and analyzed by LC-MS/MS and CE-MS in order to investigate the ratios of unlabeled amino acids incorporated into the protein: unlabeled/(labeled + unlabeled). These were: Gly, 2.2%; Ala, 2.9%; Ser, 3.2%; Pro, 3.2%; Val, 3.7%; Thr, 2.3%; Leu, 3.6%; Asx, 7.5%; Lys, 4.5%; Glx, 0.8%; Met, 1.9%; His, 2.2%; Phe, 1.8%; Arg, 5.7%; and Tyr, 3.0%. When the added amount of Gly was a quarter, the level of incorporation of unlabeled Gly was 12.7%. This is the reason why the concentration of Gly was increased. If necessary, the levels of other unlabeled amino acids can be decreased by increasing the amounts of the labeled species.

Cell-free protein synthesis systems usually use a large amount of a plasmid containing sequences coding for the target protein, a promoter, and a terminator for T7 RNA polymerase. High-copy plasmids can reduce the costs and time required to produce the template DNA. We selected the pIVEX vector (Figure 3) provided by Roche over other high copy plasmids, because this vector enables the prediction of silent mutants that can improve cell-free synthesis yields as discussed below.

Validation of cell free synthesis of the protein

Small-scale reactions are useful for determining whether a target protein is synthesized or not. The recommended composition of reagents in the small scale reaction is the same as in the large-scale reaction used for preparation of NMR sample, with the important difference of PEG-8000. PEG plays an important role as a stabilizer of the system (Kim and Swartz, 2000). In the case of CaM, the ratio of protein yields with and without PEG-8000 was 1.6. The problem is that even a small amount of PEG can disturb the

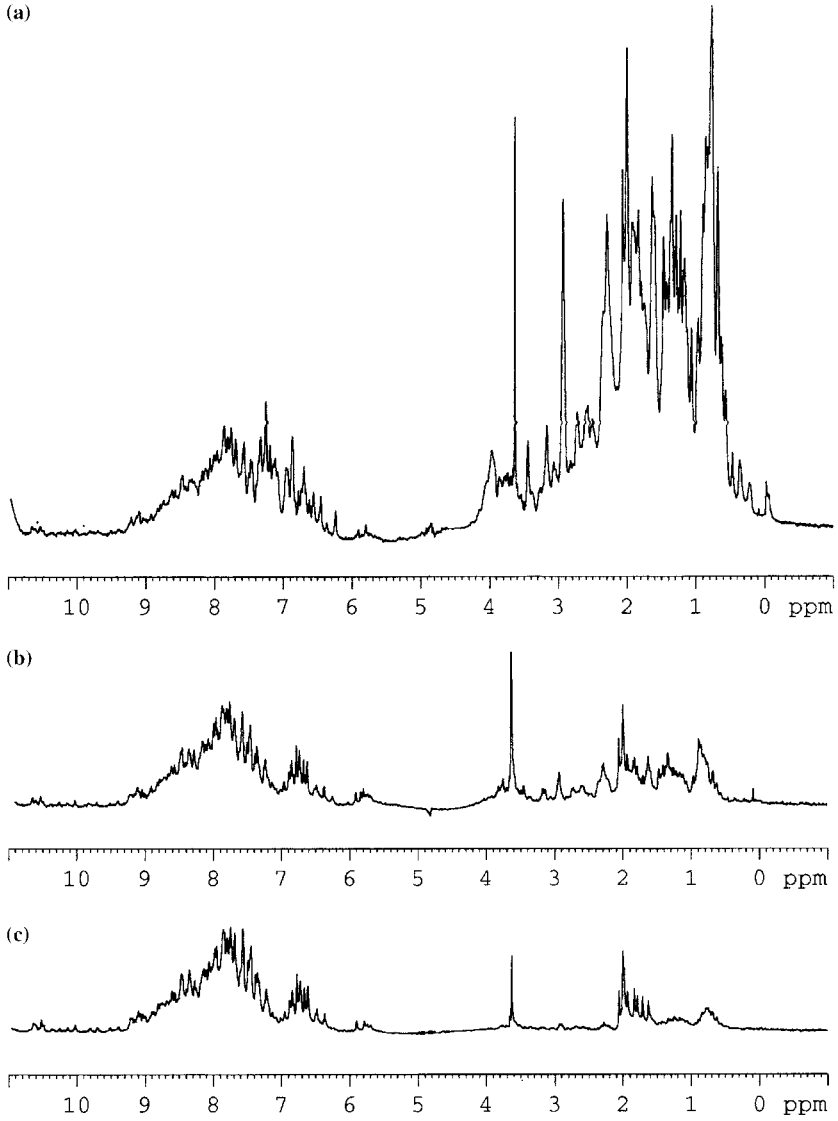

Figure 4. ^{13}C-filtered ^{1}H spectra of CaM. (a) ^{15}N-labeled CaM synthesized with the conventional extract. (b) ^{2}H, ^{13}C, ^{15}N-labeled CaM synthesized with the conventional extract. (c) ^{2}H, ^{13}C, ^{15}N-labeled CaM synthesized with our improved, dialyzed extract. All the spectra are normalized by intensities of the amide region. The resonances detected in the aliphatic region originate from ^{12}C-^{1}H. The water resonance was suppressed by a 3-9-19 pulse sequence.

SDS-PAGE quantitation of the level of production of a target protein, even if attempts are made to remove PEG by acetone or TCA precipitation of protein. For example, after cell-free synthesis and prior to purification, the SDS-PAGE band from acetone- or TCA-precipitated CaM was not observed, although it could be observed following purification. For this reason, we recommend that PEG be omitted from small-scale screening reactions.

Improvement of the yield of synthesized protein

It has been reported that the yield of protein from *E. coli* cell-free synthesis is dependent on the concentrations of magnesium ion and template DNA (Kramer et al., 1999). The concentrations in our protocol were set to be nearly optimal for the syntheses of many kinds of proteins. They could be improved to increase the yield of a particular protein. Another approach for improving the yield is

to optimize the sequence of the template DNA. Roche provides a commercial product that predicts silent mutants (those that do not alter the amino acid sequence of the target protein) that may improve the yield of a particular protein. This prediction is empirical and is based, in part, on calculations of the secondary structure of the mRNA (including the pIVEX and the target gene), although the full details of the algorithm are unavailable. One round of calculations provides ten suggested candidates for silent mutations in the region close to N-terminus of the protein. Roche recommends that these mutations be introduced by PCR into linear templates and that small-scale synthesis be used to check the relative yields. The mutated sequence that gives the highest yield is then transferred to a circular template and used for large-scale reactions.

Rather than PCR mutagenesis, our approach has been to design oligonucleotides that will introduce the best silent mutations; these are introduced into the circular plasmid (Figure 3) by restriction enzymes and ligase. This avoids possible difficulties and errors from the PCR reaction. Small-scale reactions are then carried out using the plasmids containing the silent mutations. Results for calmodulin are shown in Figure 5. Small scale protein yields were much higher with plasmids incorporating silent mutation candidates 2 and 8 than with the original plasmid. In the large-scale reaction carried out with the candidate 8 sequence and 51.1 mg of amino acid mixture (1.7 mg/ml), the amount of CaM synthesized was 5.2 mg (10 wt.%); this was twice the yield from the construct prior to the silent mutagenesis.

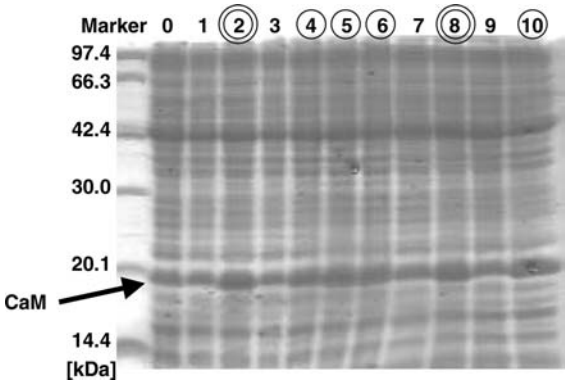

Figure 5. SDS–PAGE of CaM synthesized using plasmids without or with silent mutations predicted to enhance production. Lane 0: protein from the construct without silent mutations. Lanes 1–10: protein from constructs with 10 different silent mutations. In comparing the intensities of lanes 1–10 to lane 0, candidates with higher intensity are shown by a circle and those with much higher intensity by a doubled circle.

We have found for the gene targets analyzed thus far that many (but not all) of the silent mutations predicted to improve protein production map to the same positions in DNA coding for the N-terminal tag sequences. Thus previously optimized N-terminal tag sequences frequently can be reused by cutting and pasting (Figure 3) so as to reduce the number of mutations that need to be made to test the ten predicted improvements. This reduces the time and cost of making the complimentary DNA strands needed for mutagenesis. We have prepared two kinds of tags with a variety of the silent mutations. One tag has the same amino acid sequence as that of pET15b, namely, a $(His)_6$-tag followed by a thrombin cleavage site; the other is identical except for the absence of the $(His)_6$-tag. The former is used for easy purification of proteins by immobilized metal affinity chromatography, and the later is used in cases where the protein of interest can be purified easily for by other means. Both of the tags can be cleaved with thrombin. The cleavage can generate another important advantage as described in the next section.

Heterogeneous N-terminus of the protein product of cell-free synthesis

We collected the ^{1}H–^{15}N HSQC spectrum of CaM synthesized by our system as the native sequence (Figure 6a). More peaks than predicted from the sequence were observed in the spectrum, although the purified protein was observed as a single band by SDS–PAGE. Doubled peaks were found to arise from residues located close to the N-terminus (M0, A1, D2, Q3, L4 and T5) and in other regions (I9, K13, T70, M71, A73, R74, K77 and D78) (Figure 7). Some of the latter residues are spatially separated from the N-terminus in the X-ray structure of CaM, PDB code: 1CLL). The largest chemical shift differences between peaks assigned to the same atoms in corresponding residues in the two protein forms corresponded to M0, A1, D2, Q3 and L4. Intensities of the superfluous peaks were not always reproducible completely. In order to identify the source of the phenomenon, N-terminal

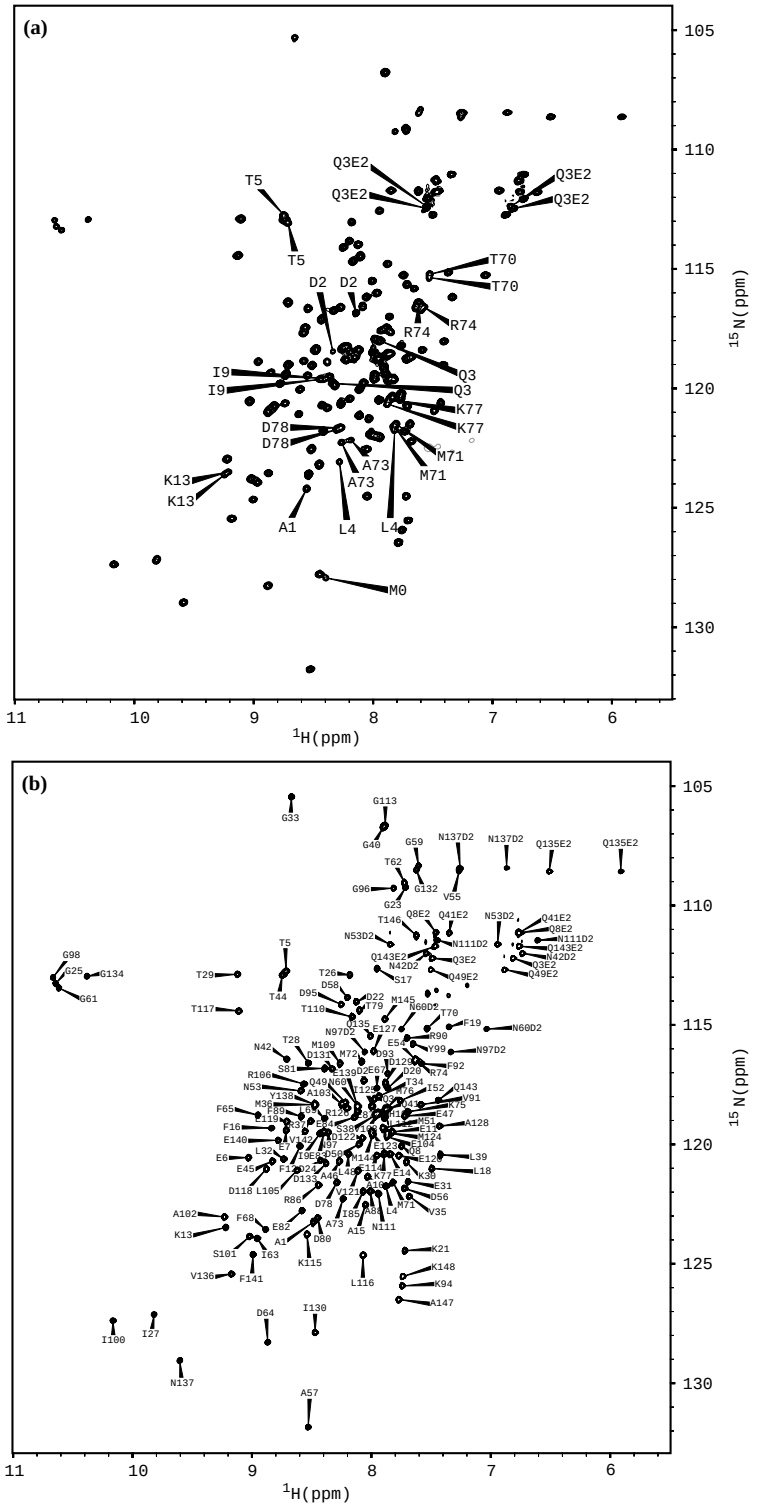

Figure 6. ^{1}H–^{15}N HSQC spectra of [*U*-^{13}C,^{15}N]-CaM. (a) CaM synthesized as the native amino acid sequence. (b) CaM synthesized with the N-terminal tag following its removal by thrombin digestion. In (a), the extra peaks are denoted with the assignment.

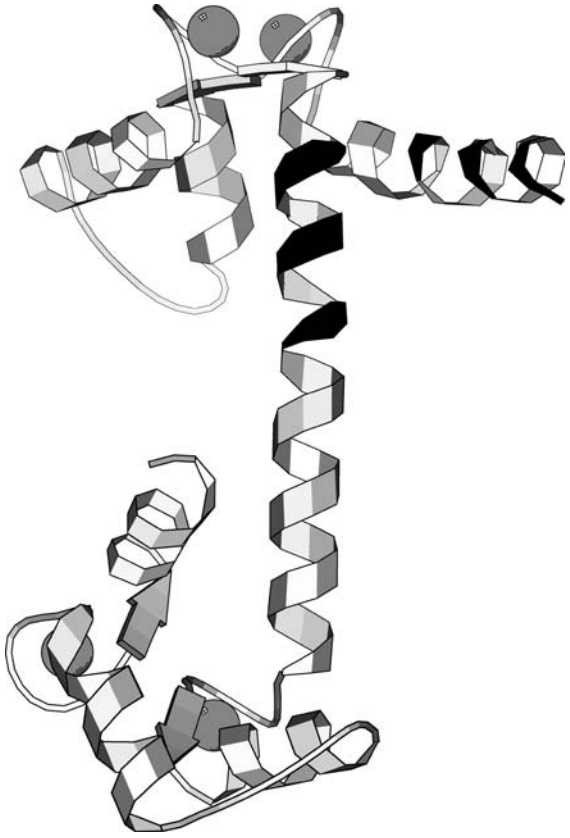

Figure 7. Residues of CaM whose chemical shifts are perturbed depending on whether the N-terminus of the molecule is fMet or not are shown in black in a ribbon model of the crystal structure of CaM (PDB id: 1CLL). Calcium ions bound to CaM are displayed by gray spheres. Residues M0, A1, D2, Q3 and L4, which were not resolved in the electron density, are not displayed in the structure.

analysis and mass spectroscopy were applied to the samples of CaM. The analyses showed that the sample consists of three molecular species: one with N-terminal formyl Met (f-Met_0) (60–90%), one with N-terminal Met_0 (~10%), and one with N-terminal Ala_1 (10–40%). On the other hand, CaM prepared by *in vivo* expression consisted of two molecular species: one starting with Met_0 (10–40%) and another starting with Ala_1 (60–90%). The results indicate that the components that lead to modification of the N-terminus are different in *E. coli* cells and in the *E. coli* cell-free extract. Heretofore, it has not been reported that heterogeneity of the N-terminus can be a critical problem for the structural analysis of a protein prepared from *E. coli* cells. In the original investigations of recombinant CaM, signals from the N terminal residues (M0, A1 D2, Q3 and L4) were not detected, and there was no indication of the presence of more peaks than predicted from the sequence (Ikura et al., 1990a). However, when the product contains molecules with and without an N-terminal f-Met, differences arising from the heterogeneity can become more apparent. A second case is EPPIb (164 amino acids residues) in which 33 extra peaks were observed in the ^{1}H–^{15}N HSQC spectrum of protein prepared by the cell free system than in the spectrum of protein produced from *E. coli* cells (data not shown).

This problem can be overcome by producing the protein with a cleavable N-terminal tag. The spectrum of CaM produced in this way shows no detectable extra peaks (Figure 6b). It is noteworthy that NMR signals from CaM, which has a flexible N-terminus, are affected by the heterogeneity and that the region affected is quite extensive (Figure 7). Proteins that have a more organized N-terminus could exhibit larger spectral complications arising from heterogeneity of the peptide chain. Since, as discussed above, the tag can be engineered to improve protein yields, our method of using a cleavable N-terminal tag to establish a homogeneous N-terminus following cleavage, serves two functions.

Large-scale preparation of labeled protein for NMR spectroscopy

In large-scale protein preparations it is critical to minimize losses of the product. A membrane with MWCO of 50 kDa was used for the reaction because it has been reported (Kigawa et al., 1999) and we have confirmed that membranes with smaller MWCO reduce the yield of protein. Because this cutoff is larger than most of the proteins we prepare for NMR spectroscopy, much of the synthesized protein goes into the outer dialysis solution during the reaction. For CaM (17 kDa) after the 7 h reaction, the amount of protein product in the inner reaction solution equaled that in the outer solution, although the protein in the outer chamber sometimes could not be detected well by SDS–PAGE owing to the dilution factor. We retrieved the synthesized protein from both solutions.

The purification protocol also must be optimized. It is frequently advantageous to dialyze the protein prior to a the first chromatography step,

either to remove salts prior to binding to an ion-exchange column or to remove nucleic acids or other contaminants that may interfere with chromatography. It may be better to dialyze the solutions against the buffer used for the first purification step. Although the protein with the $(His)_6$-tag is purified easily by Ni affinity chromatography, DTT in the system sometimes disturbs the purification if the chromatography is adopted as the first step. The presence of DTT in the S30 extract is necessary for preserving its activity over time. S30 extracts prepared with and without DTT have equivalent activity initially, but the activity of the extract without DTT is reduced to 80–90% after only 3 months of storage at −80 °C.

Conclusion

We have presented our strategy for the production of NMR samples by *E. coli* extract cell-free synthesis. The protocol described here incorporates our latest methods for improving yields and the quality of the labeled protein product. It includes solutions to two major problems encountered. The dilution of added stable isotope labeled amino acids by amino acids in the extract is solved by minimizing the levels of endogenous amino acids by a dialysis step that does not impair the activity of the S30 system. The problem of proteins produced with different N-terminal residues was solved by synthesizing proteins with a cleavable N-terminal stretch of amino acids engineered to increase the overall yield of the protein of interest. Yields were increased by optimizing sequences coding for the N-terminal fusions by incorporating silent mutations, selected from those predicted by ProteoExpert, that were found experimentally to increase protein production. For NMR spectroscopy, a major advantage of the cell-free approach is that it enables the incorporation of labeled amino acids without metabolic scrambling of isotope labeled amino acids. Applications of this technology will be presented in separate publications.

Acknowledgements

We thank Dr Toshiya Hayano for his contribution during the early stage of this work, Prof Mitsuhiko Ikura of Toronto University for providing the calmodulin cDNA, Prof Toshiaki Isobe of Tokyo Metropolitan University for the mass spectral analyses of calmodulin, and Prof John L. Markley of the University of Wisconsin-Madison for critical reading of the manuscript. This work was supported by CREST/JST and by grant P50 GM 64598 from the Protein Structure Initiative of the U.S. National Institutes of Health (J.L. Markley, PI).

References

Betton, J.M. (2003) *Curr. Protein Pept. Sci.*, **4**, 73–80.
Bradford, M.M. (1976) *Anal. Biochem.,* **72**, 248–254.
Chrunyk, B.A., Evans, J., Lillquist, J., Young, P. and Wetzel, R. (1993) *J. Biol. Chem.*, **268**, 18053–18061.
Clemens, M.J. and Prujin, G.J. (1999) *Protein Synthesis in Eukaryotic Cell-Free Systems*, Oxford University Press, New York, pp. 129–165.
Clore, G.M. and Gronenborn, A.M. (1994) *Meth. Enzymol.*, **239**, 349–363.
Davanloo, P., Rosenberg, A.H., Dunn, J.J. and Studier, F.W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 2035–2039.
Goff, S.A. and Goldberg, A.L. (1987) *J. Biol. Chem.*, **262**, 4508–4515.
Grodberg, J. and Dunn, J.J. (1988) *J. Bacteriol.*, **170**, 1245–1253.
Guignard, L., Ozawa, K., Pursglove, S.E. Otting, G. and Dixon, N.E. (2002) *FEBS Lett.*, **524**, 159–162.
Henrich, B., Lubitz, W. and Plapp, R. (1982) *Mol. Gen. Genet.*, **185**, 493–497.
Ikura, M., Kay, L.E. and Bax, A. (1990a) *Biochemistry*, **29**, 4659–4667.
Ikura, M., Marion, D., Kay, L.E., Shih, H., Krinks, M., Klee, C.B. and Bax, A. (1990b) *Biochem. Pharmacol.*, **40**, 153–160.
Kariya, E., Ohki, S., Hayano, T. and Kainosho, M. (2000) *J. Biomol. NMR.*, **18**, 75–76.
Kelly, A.E., Ou, H.D., Withers, R. and Dötsch, V. (2002) *J. Am. Chem. Soc.*, **124**, 12013–12019.
Kigawa, T., Muto, Y. and Yokoyama, S. (1995) *J. Biomol. NMR*, **6**, 129–134.
Kigawa, T., Yabuki, T., Yoshida, Y., Tsutsui, M., Ito, Y., Shibata, T. and Yokoyama, S. (1999) *FEBS Lett.*, **442**, 15–19.
Kim, D.M. and Swartz, J.R. (2000) *Biotechnol. Prog.*, **16**, 385–390.
Kim, D.M., Kigawa, T., Choi, C.Y. and Yokoyama, S. (1996) *Eur. J. Biochem.*, **239**, 881–886.
Klammt, C., Lohr, F., Schafer, B., Haase, W., Dötsch, V., Rüterjans, H., Glaubitz, C. and Bernhard, F. (2004) *Eur. J. Biochem.*, **271**, 568–580.
Kramer, G., Kudlicki, W. and Hardesty, B. (1999) *Cell-free Coupled Transcription–Translation Systems from* Escherichia coli, Oxford University Press, New York, pp. 129–165.
Madin, K., Sawasaki, T., Ogasawara, T. and Endo, Y. (2000) *Proc. Natl. Acad. Sci. USA*, **97**, 559–564.
Markley, J.L., Bax, A., Arata, Y., Hilbers, C.W., Kaptein, R., Sykes, B.D., Wright, P.E. and Wüthrich, K. (1998) *Eur. J. Biochem.*, **256**, 1–15.

Maurizi, M.R. (1987) *J. Biol. Chem.*, **262**, 2696–2703.
Mori, S., Abeygunawardana, C., Johnson, M.O. and van Zijl, P.C. (1995) *J. Magn. Reson. B*, **108**, 94–98.
Pratt, J.M. (1984) *Transcription and Translation: A Practical Approach*, IRL Press, New York, pp. 179–209.
Shi, J., Pelton, J.G., Cho, H.S. and Wemmer, D.E. (2004) *J. Biomol. NMR*, **28**, 235–247.
Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K. and Ueda, T. (2001) *Nat. Biotechnol.*, **19**, 751–755.
Spirin, A.S., Baranov, V.I., Ryabova, L.A., Ovodov, S.Y. and Alakhov, Y.B., (1988) *Science*, **242**, 1162–1164.
Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L. and Sykes, B.D. (1995) *J. Biomol. NMR*, **6**, 135–140.
Zawadzki, V. and Gross, H.J. (1991) *Nucl. Acids Res.*, **19**, 1948.
Zubay, G. (1973) *Annu. Rev. Genet.*, **7**, 267–287.
Zwahlen, C., Legault, P., Vincent, S.J.F. Greenblatt, J., Konrat, R. and Kay, L.E. (1997) *J. Am. Chem. Soc.*, **119**, 6711–6721.