Cell-free protein production and labeling protocol for NMR-based structural proteomics

Dmitriy A Vinarov[1,2], Betsy L Lytle[1,3], Francis C Peterson[1,3], Ejan M Tyler[1,2], Brian F Volkman[1,3] & John L Markley[1,2]

Structural proteomics requires robust, scalable methods. Here we describe a wheat germ cell-free platform for protein production that supports efficient NMR structural studies of eukaryotic proteins and offers advantages over cell-based methods. To illustrate this platform, we describe its application to a specific target (At3g01050.1) from *Arabidopsis thaliana*. After cloning the target gene into a specialized plasmid, we carry out a small-scale (50 μl) *in vitro* sequential transcription and translation trial to ascertain the level of protein production and solubility. Next, we prepare mRNA for use in a 4-ml semicontinuous cell-free translation reaction to incorporate ^{15}N-labeled amino acids into a protein sample that we purify and test for suitability for NMR structural analysis. We then repeat the cell-free approach with ^{13}C,^{15}N-labeled amino acids to prepare a doubly labeled sample. The three-dimensional (3D) structure of At3g01050.1 shows that this protein is an unusual member of the β-grasp protein family.

A chief bottleneck in high-throughput protein structure determinations is the production of soluble folded protein. Although most structural proteomics centers rely on the overproduction of proteins from *Escherichia coli* cells, this default approach fails for many individual proteins. The failure rate is about a third for prokaryotic proteins and much higher for eukaryotic proteins (see the Protein Structure Initiative Target Database, http://www.rcsb.org/pdb/strucgen.html). For this reason, the development of improved platforms for the production of proteins or protein domains remains an important goal. Optimal platforms for NMR spectroscopy must support the production of proteins labeled with stable isotopes (^{15}N, ^{13}C and/or ^{2}H), and those for X-ray crystallography must facilitate the incorporation of selenomethionine or other appropriately labeled amino acids.

Cell-free methods for protein synthesis with extracts from prokaryotic[1] or eukaryotic[2] sources offer an alternative to *E. coli* cell-based platforms. Stable-isotope or selenomethionine labeling is easier with cell-free systems than with fungal or insect cell systems[3–5]. Cell-free systems permit the production of proteins that undergo proteolysis in cells[6,7] or that aggregate in inclusion bodies[8]. Cell-free systems support selective labeling strategies that cannot be achieved in bacterial whole-cell systems[9–12] and have been used to produce various types of protein, including membrane proteins[13] and proteins that are toxic to cells[8,14]. It is possible to collect NMR spectra of ^{15}N-labeled proteins without isolating them from the cell-free protein synthesis mixture[15]. Successful implementation of cell-free protein expression obviates the need for cell collection, cell lysis and some lengthy steps in protein purification (such as protein concentration).

The RIKEN Structural Genomics Center, in collaboration with Roche Applied Sciences, has pioneered the use of cell-free protein production through a coupled transcription-translation system that uses *E. coli* extracts[16–20]. Most of the proteins that are produced well in *E. coli* cell-free systems are ones that are produced successfully in *E. coli* cell culture (M. Kainosho, personal communication). Thus, despite its other potential advantages, this approach may not greatly expand the range of proteins that can be produced in a soluble folded state, although it may be possible to overcome this limitation by redesigning the gene sequence (see below), either by adding chaperones or other factors[21,22] or by re-engineering ribosomal proteins[23].

The Center for Eukaryotic Structural Genomics (CESG) in cooperation with Ehime University, Japan, and the CellFree Sciences Co., Ltd., Yokohama, Japan, is investigating the potential of wheat germ cell-free protein production (refs. 24–28 and Endo, Y., personal communication; 3rd ORCS International Symposium on Ribosome Engineering, January 22–23, 2001, Tsukuba, Japan, 2001) as an enabling technology for NMR-based structural proteomics. CESG has developed the platform described here and has found that it offers potential advantages over *E. coli* cell culture and *E. coli* cell-free approaches. A larger proportion of targets are produced in a folded soluble form by the wheat germ cell-free approach than by the *E. coli* cell-based approach. In a comparison of 96 randomly chosen *A. thaliana* targets taken through the wheat germ cell-free and *E. coli* cell pipelines of CESG to produce ^{15}N-labeled proteins, 8 proteins from the cell-free pipeline described here, as compared with 5 proteins from the cell-based pipeline,

[1]Center for Eukaryotic Structural Genomics, and [2]Department of Biochemistry, University of Wisconsin–Madison, 433 Babcock Drive, Madison, Wisconsin 53706, USA. [3]Department of Biochemistry, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, Wisconsin 53226, USA. Correspondence should be addressed to J.L.M. (markley@nmrfam.wisc.edu).

PUBLISHED ONLINE 21 OCTOBER 2004; DOI:10.1038/NMETH716

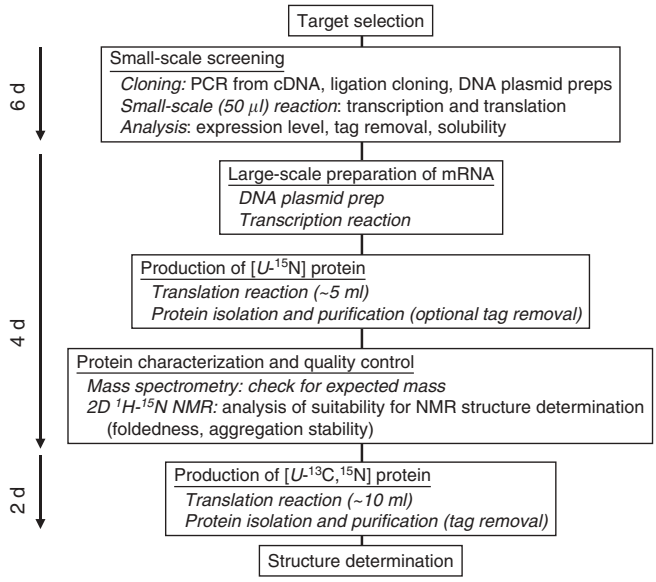


Figure 1 | Workflow diagram for the production of isotopically enriched NMR samples by the wheat germ cell-free expression system.

were found suitable for NMR structural analysis on the basis of ^{1}H-^{15}N correlation NMR spectra (R.C. Tyler *et al.*, unpublished data).

The platform uses a single construct for all targets without any redesign of the DNA or RNA; thus, it offers advantages over commercial cell-free methods that use *E. coli* extracts[29] and that require several constructs or redesign of the open reading frame (for example, see the Roche Protein Expression Group website, http://www.rpeg.com/rpeg/content/custom_protein.htmlcustom, or the ProteoExpert RTS *E. coli* HY website, http://www.roche-applied-science.com/sis/proteoexpert/). The wheat germ cell-free protocol needs no additives such as polyethylene glycol, which is frequently used to improve yields in *E. coli* S30 cell-free synthesis[21]. Because of the small volume requirements for screening (25–50 μl) and protein production for structural studies (4–12 ml), the method is amenable to automation.

Here we describe the wheat germ cell-free pipeline protocol developed at CESG for NMR structure determinations. We illustrate the approach through its application to the hypothetical *A. thaliana* protein At3g01050.1. We describe the methods that we used to evaluate the expression and solubility of this target and the procedures that we used to prepare samples labeled with ^{15}N and with ^{13}C and ^{15}N for NMR structural investigations leading to the determination of the 3D structure of this protein.

RESULTS

Overview

The approach consists of four steps (**Fig. 1**): first, creation of a plasmid for *in vitro* transcription; second, small-scale (50 μl) screening to assay protein production and solubility; third, larger-scale (4–12 ml) production of uniform ^{15}N (*U*-^{15}N)-labeled protein to evaluate the protein as a structural target; and fourth, production of sufficient *U*-^{13}C,^{15}N-labeled protein for multidimensional, multinuclear NMR data collection.

A well-defined series of cloning steps is used to create a DNA plasmid containing the target gene and 5′ and 3′ extensions that promote efficient transcription and translation. Small-scale protein expression and purification trials are generally done in a 96-well format. Successful candidates from these screens (those estimated to yield $\geq$0.5 mg/ml of target protein with $>$50% solubility under larger-scale production) are selected for larger-scale protein expression on *U*-^{15}N-labeled amino acids. Purified *U*-^{15}N-labeled proteins are then assayed by ^{1}H-^{15}N correlation spectroscopy for their suitability as structural candidates (they must be folded, monodisperse and stable at room temperature for at least 14 d). Targets that pass these tests are prepared as *U*-^{13}C,^{15}N-labeled protein samples.

Screening for expression and solubility

We inserted the target protein gene into plasmid peU-N_{tag}-At3g01050.1 (in which N_{tag} encodes the purification tag $MG(H)_6LE$). We used this plasmid as a template for producing mRNA. The small-scale (50 μl) translation reaction yielded 5–10 μg of the target protein. The protein product (N_{tag}-At3g01050.1) was more than 75% soluble, as assessed by a comparison of the total and soluble fractions.

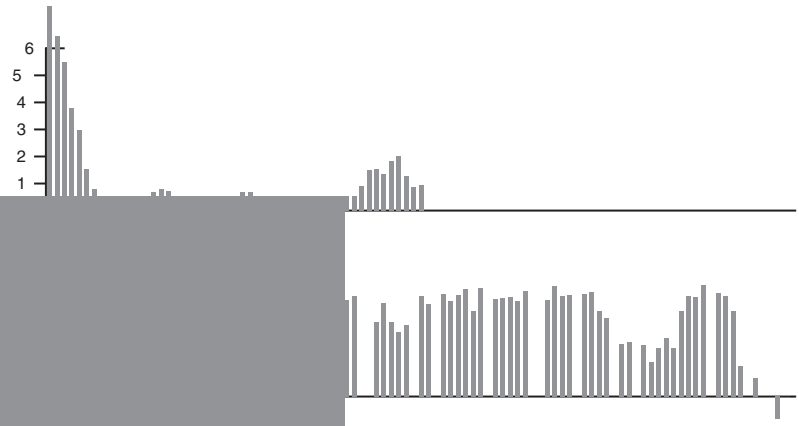
Production of labeled protein for NMR investigations

On the basis of the screening results, we adjusted the volumes of each of the reactions used to produce *U*-^{15}N- and *U*-^{13}C,^{15}N-labeled protein to 8 ml. The yields ($>$84% pure) of [*U*-^{15}N]N_{tag}-At3g01050.1 and [*U*-^{13}C,^{15}N]N_{tag}-At3g01050.1 were each about 2.5 mg. The protein eluted as a single peak from a nickel nitrilotriacetic acid (Ni-NTA) column (**Fig. 2a**) and

was observed as a band on SDS-PAGE corresponding to about 14 kDa (**Fig. 2b**).

NMR analysis

The solution conditions chosen for NMR spectroscopy were 10 mM potassium phosphate buffer (pH 6.5) containing 50 mM NaCl and 10% ^2H_2O for the frequency lock of the NMR spectrometer. Under these conditions, the number, dispersion and line widths of the observed signals from a ^1H-^{15}N heteronuclear single quantum correlation (HSQC) spectrum (**Fig. 2c**) were consistent with a homogeneous folded protein. The translational self-diffusion coeffi-



ermined five protein structures, the coordinates
NMR data of which have been deposited in the
(accession numbers 1SE9 (At3g01050.1), 1T0G,
1XFL) and BioMagResBank (accession numbers
), 6138, 6213, 6240 and 6318), respectively.
he labeled amino acids present in the reaction
nally incorporated into the purified protein.
is more than 95%, as determined by mass
approach is rapid and amenable to automation.
ddition of a GeneDecoder1000 robotic system
, CESG can carry out as many as 1,052 small-
ning reactions per week and, by using a Prote-
tem (CellFree Sciences), CESG can perform 24
-scale (4-ml) reactions per week. The cell-free
less labor intensive than is the corresponding
ipeline.
d out a detailed analysis of its protein production
sidered the costs (labor and supplies) of indivi-
, expression testing and preparation of *U*-^{15}N-
eled protein) and took into account the yields at
luded that the cell-free approach is no more
cell-based approach for preparing samples for
nvestigations. The costs of producing labeled
ree systems could be lowered by using less
of labeled amino acids, by tailoring the mixture

Table 1 | Structural statistics for 20 NMR structures

Constraints	
Dihedral angles from TALOS[39]:	
ϕ	60
ψ	62
NOE:	
Long	367
Medium	138
Short	302
Intraresidue	330
Total	1,137
Ramachandran from PROCHECK[40] (%)	
Most favored	86.3
Additionally allowed	12.3
Generously allowed	1.1
Disallowed	0.3
Constraint violations (no. $>$ 0.5 Å)	0 ± 0
NOE distance r.m.s. deviation (Å)	0.026 ± 0.001
Torsion angle violations	
No. $> 5^\circ$	0 ± 0
r.m.s. deviation ($^\circ$)	0.57 ± 0.09
Atomic r.m.s. deviation to mean structure (Å)	
Residues 1–101:	
Backbone	1.98 ± 0.30
Heavy atom	2.54 ± 0.29
Residues 7–43, 50–73, 85–93:	
Backbone	0.60 ± 0.10
Heavy atom	1.03 ± 0.09
WHATCHECK[41] quality indicators	
Z-score	−1.65
r.m.s. *Z*-score:	
Bond lengths	0.80
Bond angles	0.70
Bumps	0

of amino acids to match the amino acid composition of the target protein, or by recovering and recycling the unused amino acids. Small-scale screening for protein yield and solubility is nearly three times less expensive by the cell-free protocol than by the corresponding *E. coli* cell-based approach. The most important advantage of the wheat germ cell-free method is that it supports the production of a larger fraction of targets as folded soluble proteins; the chief advantage of the *E. coli* approach, when successful, is that larger quantities of protein are produced at a given cost.

Fold analysis of the At3g01050.1 protein

The level of sequence identity required to retain the β-grasp fold is very low (~10%)[36]. A bioinformatics analysis done at the time of target selection did not identify any likely structural homologs for the At3g01050.1 protein; however, newer methods have been successful in predicting ubiquitin-like (UBL) domains in unexpected protein families, such as tubulin-binding cofactors[37].

The common β-grasp fold, which At3g01050.1 adopts, is a domain superfamily that includes ubiquitin and UBL proteins such as SUMO and NEDD8 that function as covalent modifiers of protein function. In addition to their well-known role in directing proteasomal degradation, modification with polyubiquitin or a single SUMO or NEDD8 moiety can markedly alter the intracellular trafficking and biochemical activity of modified proteins. Although At3g01050.1 has a β-grasp fold and shares some sequence similarity (a maximum of 25% identity) with other proteins in this family, it seems unlikely that it functions in this manner because it lacks the conserved C-terminal Gly-Gly sequence required for covalent conjugation of ubiquitin and other UBL proteins. Instead, on the basis of the C-terminal CAAX box motif present in its amino acid sequence, we propose that At3g01050.1 may act as a substrate for posttranslational lipid modifications and may function as a membrane-associated protein (see **Supplementary Methods** online).

METHODS

DNA manipulations. The At3g01050.1 gene was derived from total mRNA isolated from an *A. thaliana* T87 tissue culture by a procedure involving reverse transcriptase. We cloned this gene into the pEU-N_{ext} vector, a gift of Y. Endo (Ehime University, Matsuyama, Japan). Complete details are given in the **Supplementary Methods** online. In brief, we designed PCR primers to incorporate *Xho*I and *Not*I restriction sites at the 5′ and 3′ ends of the gene, respectively. We used the ligation product to transform competent Top10 cells. We picked colonies for growth in 2× YT medium and used this culture for a small-scale DNA preparation done manually with a MiniprepSpin kit (Qiagen). A 5-ml cell culture typically resulted in about 20 μg of purified DNA plasmid. We sequenced the whole coding region of the plasmid to verify the absence of mutations.

Small-scale expression and purification. The plasmid (4 μg) was used as the template for small-scale reactions. The transcription reactions used published protocols[25]. In brief, mRNA encoding the N_{tag}-At3g01050.1 protein was synthesized by *in vitro* transcription of the peU-N_{tag}-At3g01050.1 plasmid carrying the gene under control of the SP6 RNA polymerase promoter. We purified the resulting mRNA by ethanol precipitation.

The small-scale translation reaction was done overnight in a dialysis cup (molecular weight cutoff 12,000; Biotech International) by published protocols[27] with modifications. The reaction mixtures (50 μl) contained 15 μl of wheat germ extract (final optical density at 600 nm (OD_{600}) = 60) purchased from CellFree Sciences, 35 μl of translation buffer solution, 450 μg/ml of creatine kinase and N_{tag}-At3g01050.1 mRNA. The translation buffer contained 24 mM HEPES/KOH (pH 7.8), 1.2 mM ATP, 0.25 mM GTP, 16 mM creatine phosphate, 250 units of RNasin (ribonuclease inhibitor), 1 mM dithiothreitol, 0.4 mM spermidine, 0.3 mM each of the 20 amino acids, 2.7 mM magnesium acetate and 100 mM potassium acetate. We immersed the dialysis cup containing the reaction mixture in 2.5 ml of a solution containing all of the ingredients of the translation buffer plus creatine kinase. Incubations were carried out overnight at 26 °C. SDS-PAGE with Coomassie staining was used to determine target protein yields ('total', 'soluble' and 'pellet' fractions). Ni-HiTrap spin columns (Amersham) were used for small-scale purifications.

Large-scale expression and purification. Labeled amino acids (each of the 20 common amino acids as a single pure species) were obtained from Cambridge Isotope Laboratories. Separate mixtures of *U*-^{15}N- and *U*-^{13}C,^{15}N-labeled amino acids were prepared such that each of the 20 amino acids was present at 0.3 mM in the translation buffer. We produced labeled protein samples as follows. The 1-ml reaction mixture contained 0.3 ml of

wheat germ extract (final OD_{600} = 60), 450 μg/ml of creatine kinase and about 10 mg of At3g01050.1 mRNA. A dialysis bag containing the reaction mixture was immersed in 5 ml of translation buffer solution supplemented with *U*-^{15}N- or *U*-^{13}C, ^{15}N-labeled amino acids to achieve the desired labeling pattern. The reaction was carried out for 48 h at 26 °C and the substrate solution was replaced every 12 h. As a control, mRNA encoding green fluorescent protein (GFP) was substituted for the mRNA of the target protein in an aliquot of the reaction mixture. We determined protein production and solubility by SDS-PAGE analysis of the total, soluble and pellet fractions. Proteins were purified on an ÅKTA Prime system (Amersham) by using 1-ml HiTrap Chelating HP columns with adaptors (Amersham). Before loading the protein onto the column, the buffer was exchanged for one that was appropriate for affinity binding and the solution was centrifuged. The flow rate was 1 ml/min. After the column was washed, the target protein was eluted with 250 mM imidazole. We determined the purity of the column fractions by SDS-PAGE.

NMR spectroscopy. All NMR data were acquired at 25 °C on a Bruker 600-MHz spectrometer equipped with a triple-resonance CryoProbe, and were processed with NMRPipe software[38] (see the **Supplementary Methods** online for details).

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

We thank Y. Endo and CellFree Sciences Co., Ltd., for advice and encouragement in using the wheat germ cell-free system that they have developed; and M. Kainosho for sharing the unpublished information cited here. This research was supported by a grant from the National Institutes of Health Protein Structure Initiative (1 P50 GM64598 to J.L.M.).

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests; see the *Nature Methods* website for details.

Received 26 May; accepted 21 September 2004
Published online at http://www.nature.com/naturemethods/

1. Kramer, G., Kudlicki, W. & Hardesty, B. Cell-free coupled transcription-translation systems from *Escherichia coli*. in *Protein Expression. A Practical Approach* (eds. Higgens, S.J. & Hames, B.D.) 201–223 (Oxford Univ. Press, Oxford, UK, 1999).
2. Clemens, M.M. & Prujin, G.J. Protein synthesis in eukaryotic cell-free systems. in *Protein Expression. A Practical Approach* (eds. Higgens, S.J. & Hames, B.D.) 129–165 (Oxford Univ. Press, Oxford, UK, 1999).
3. Cubeddu, L. *et al. Dictyostelium discoideum* as expression host: isotopic labeling of a recombinant glycoprotein for NMR studies. *Protein Expr. Purif.* **19**, 335–342 (2000).
4. Strauss, A. *et al.* Amino-acid-type selective isotope labeling of proteins expressed in Baculovirus-infected insect cells useful for NMR studies. *J. Biomol. NMR* **26**, 367–372 (2003).
5. Bruggert, M., Rehm, T., Shanker, S., Georgescu, J. & Holak, T.A. A novel medium for expression of proteins selectively labeled with ^{15}N-amino acids in *Spodoptera frugiperda* (Sf9) insect cells. *J. Biomol. NMR* **25**, 335–348 (2003).
6. Goff, S.A. & Goldberg, A.L. An increased content of protease LA, the Lon gene product, increases protein degradation and blocks growth in *Escherichia coli*. *J. Biol. Chem.* **262**, 4508–4515 (1987).
7. Maurizi, M.R. Degradation *in vitro* of bacteriophage λ N protein by Lon protease from *Escherichia coli*. *J. Biol. Chem.* **262**, 2696–2703 (1987).
8. Chrunyk, B.A., Evans, J., Lillquist, J., Young, P. & Wetzel, R. Inclusion-body formation and protein stability in sequence variants of interleukin-1β. *J. Biol. Chem.* **268**, 18053–18061 (1993).
9. Shi, J., Pelton, J.G., Cho, H.S. & Wemmer, D.E. Protein signal assignments using specific labeling and cell-free synthesis. *J. Biomol. NMR* **28**, 235–247 (2004).
10. Torizawa, T., Terauchi, T. & Kainosho, M. [Recent developments in NMR methods for structural biology]. *Seikagaku* **74**, 1279–1284 (2002).
11. Kigawa, T., Muto, Y. & Yokoyama, S. Cell-free synthesis and amino acid-selective stable isotope labeling of proteins for NMR analysis. *J. Biomol. NMR* **6**, 129–134 (1995).
12. Yabuki, T. *et al.* 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 (1998).
13. Klammt, C. *et al.* High level cell-free expression and specific labeling of integral membrane proteins. *Eur. J. Biochem.* **271**, 568–580 (2004).
14. Henrich, B., Lubitz, W. & Plapp, R. Lysis of *Escherichia coli* by induction of cloned ϕX174 genes. *Mol. Gen. Genet.* **185**, 493–497 (1982).
15. Guignard, L., Ozawa, K., Pursglove, S.E., Otting, G. & Dixon, N.E. NMR analysis of *in vitro*–synthesized proteins without purification: a high-throughput approach. *FEBS Lett.* **524**, 159–162 (2002).
16. Kigawa, T. & Yokoyama, S. [High-throughput cell-free protein expression system for structural genomics and proteomics studies]. *Tanpakushitsu Kakusan Koso* **47**, 1014–1019 (2002).
17. Kigawa, T. *et al.* Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett.* **442**, 15–19 (1999).
18. Kim, D.M., Kigawa, T., Choi, C.Y. & Yokoyama, S. A highly efficient cell-free protein synthesis system from *Escherichia coli*. *Eur. J. Biochem.* **239**, 881–886 (1996).
19. Yokoyama, S. Protein expression systems for structural genomics and proteomics. *Curr. Opin. Chem. Biol.* **7**, 39–43 (2003).
20. Yokoyama, S. *et al.* Structural genomics projects in Japan. *Nat. Struct. Biol.* **7**, 943–945 (2000).
21. Kim, D.M. & Swartz, J.R. Prolonging cell-free protein synthesis by selective reagent additions. *Biotechnol. Prog.* **16**, 385–390 (2000).
22. Yin, G. & Swartz, J.R. Enhancing multiple disulfide bonded protein folding in a cell-free system. *Biotechnol. Bioeng.* **86**, 188–195 (2004).
23. Chumpolkulwong, N. *et al.* Effects of *Escherichia coli* ribosomal protein S12 mutations on cell-free protein synthesis. *Eur. J. Biochem.* **271**, 1127–1134 (2004).
24. Kawasaki, T., Gouda, M.D., Sawasaki, T., Takai, K. & Endo, Y. Efficient synthesis of a disulfide-containing protein through a batch cell-free system from wheat germ. *Eur. J. Biochem.* **270**, 4780–4786 (2003).
25. Madin, K., Sawasaki, T., Ogasawara, T. & Endo, Y. 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. USA* **97**, 559–564 (2000).
26. Morita, E.H., Sawasaki, T., Tanaka, R., Endo, Y. & Kohno, T. A wheat germ cell-free system is a novel way to screen protein folding and function. *Protein Sci.* **12**, 1216–1221 (2003).
27. Sawasaki, T. *et al.* A bilayer cell-free protein synthesis system for high-throughput screening of gene products. *FEBS Lett.* **514**, 102–105 (2002).
28. Sawasaki, T., Ogasawara, T., Morishita, R. & Endo, Y. A cell-free protein synthesis system for high-throughput proteomics. *Proc. Natl. Acad. Sci. USA* **99**, 14652–14657 (2002).
29. Betton, J.M. Rapid translation system (RTS): a promising alternative for recombinant protein production. *Curr. Protein Pept. Sci.* **4**, 73–80 (2003).
30. Altieri, A.S., Hinton, D.P. & Byrd, R.A. Association of biomolecular systems via pulsed field gradient NMR self-diffusion measurements. *J. Am. Chem. Soc.* **117**, 7566–7567 (1995).
31. Bartels, C., Billeter, M., Güntert, P. & Wüthrich, K. Automated sequence-specific NMR assignment of homologous proteins using the program GARANT. *J. Biomol. NMR* **7**, 207–213 (1996).
32. Herrmann, T., Guntert, P. & Wuthrich, K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319**, 209–227 (2002).
33. Linge, J.P., Williams, M.A., Spronk, C.A., Bonvin, A.M. & Nilges, M. Refinement of protein structures in explicit solvent. *Proteins* **50**, 496–506 (2003).
34. Palmer III, A.G. NMR probes of molecular dynamics: overview and comparison with other techniques. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 129–155 (2001).
35. Terauchi, T., Ohki, S.Y. & Kainosho, M. [Developing a new approach for high-throughput, high-accuracy NMR structural analyses of genomic proteins]. *Tanpakushitsu Kakusan Koso* **47**, 1045–1051 (2002).
36. Larsen, C.N. & Wang, H. The ubiquitin superfamily: members, features, and phylogenies. *J. Proteome Res.* **1**, 411–419 (2002).
37. Grynberg, M., Jaroszewski, L. & Godzik, A. Domain analysis of the tubulin cofactor system: a model for tubulin folding and dimerization. *BMC Bioinformatics* **4**, 46 (2003).
38. Delaglio, F. *et al.* NMRPIPE—a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
39. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289–302 (1999).
40. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
41. Hooft, R.W.W., Vriend, G., Sander, C. & Abola, E.E. Errors in protein structures. *Nature* **381**, 272 (1996).