

A robust two-step PCR method of template DNA production for high-throughput cell-free protein synthesis

Takashi Yabuki · Yoko Motoda · Kazuharu Hanada ·
Emi Nunokawa · Miyuki Saito · Eiko Seki · Makoto Inoue ·
Takanori Kigawa · Shigeyuki Yokoyama

Received: 16 August 2007 / Accepted: 29 October 2007 / Published online: 1 January 2008
© Springer Science+Business Media B.V. 2007

Abstract A two-step PCR method has been developed for the robust, high-throughput production of linear templates ready for cell-free protein synthesis. The construct made from the cDNA expresses a target protein region with N- and/or C-terminal tags. The procedure consists only of mixing, dilution, and PCR steps, and is free from cloning and purification steps. In the first step of the two-step PCR, a target region within the coding sequence is amplified using two gene-specific forward and reverse primers, which contain the linker sequences and the terminal sequences of the target region. The second PCR concatenates the first PCR product with the N- and C-terminal double-stranded fragments, which contain the linker sequences as well as the sequences for the tag(s) and the initiation and termination, respectively, for T7 transcription and ribosomal translation, and amplifies it with the universal primer. Proteins can be fused with a variety of tags,

such as natural poly-histidine, glutathione-*S*-transferase, maltose-binding protein, and/or streptavidin-binding peptide. The two-step PCR method was successfully applied to 42 human target protein regions with various GC contents (38–77%). The robustness of the two-step PCR method against possible fluctuations of experimental conditions in practical use was explored. The second PCR product was obtained at 60–120 µg/ml, and was used without purification as a template at a concentration of 2–4 µg/ml in an *Escherichia coli* coupled transcription-translation system. This combination of two-step PCR with cell-free protein synthesis is suitable for the rapid production of proteins in milligram quantities for genome-scale studies.

Keywords Cell-free protein synthesis · Linear template construction · Two-step PCR · Tagged proteins · Human proteins

Electronic supplementary material The online version of this article (doi:10.1007/s10969-007-9038-z) contains supplementary material, which is available to authorized users.

T. Yabuki · Y. Motoda · K. Hanada · E. Nunokawa · M. Saito ·
E. Seki · M. Inoue · T. Kigawa · S. Yokoyama
Protein Research Group, RIKEN Genomic Sciences Center,
1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan

T. Kigawa
Department of Computational Intelligence and Systems Science,
Interdisciplinary Graduate School of Science and Engineering,
Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku,
Yokohama 226-8502, Japan

S. Yokoyama (✉)
Department of Biophysics and Biochemistry, Graduate School
of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku,
Tokyo 113-0033, Japan
e-mail: yokoyama@biochem.s.u-tokyo.ac.jp

Abbreviations

GST	Glutathione- <i>S</i> -transferase
MBP	<i>E. coli</i> maltose binding protein
SBP	Streptavidin binding peptide
NHis	Natural poly-histidine
TEV	Tobacco etch virus
DMSO	Dimethyl sulfoxide
HA	Hemagglutinin

Introduction

The demand for high-throughput and flexible protein expression has increased in contemporary genome-scale research. For massive, high-throughput experiments using

robotics, the simplicity and robustness of the experimental procedures are especially important. Greater complexity of the experimental protocol usually increases the rates of human error, other experimental errors, and contamination, and thus reduces the overall success rate of the experiment.

In the case of a protein expression experiment, one of the keys is the protein expression method itself, and the other is the template DNA preparation method for protein expression. For the protein expression method, cell-free protein synthesis is most suitable for high-throughput use. First, the proteins can be produced from a linear template DNA. Second, the reaction conditions can be optimized for a target protein by adding or deleting components. For example, $^{15}\text{N}/^{13}\text{C}$ -labeled proteins were produced and subjected to structural analyses by NMR [1, 2, 3, 4]. Selenomethionine-substituted protein for X-ray crystallography were synthesized, and the structures were solved [5, 6, 7]. The addition of chaperones or protein disulfide isomerases allowed some proteins to be expressed in soluble and active forms [8, 9]. Functional membrane proteins were produced in the presence of detergents [10]. Furthermore, the protocol is simple and suitable for automation. Parallel protein expression can easily be achieved in a multi-well format. The reaction mixtures can then be directly subjected to protein analysis or purification. Complicated procedures, such as cell fermentation and cell disruption, are not necessary for producing proteins by cell-free protein synthesis. Thus, cell-free protein synthesis is suitable for massive and high-throughput applications [11, 12]. For the preparation of a template DNA for protein expression, complicated procedures such as cloning, including ligation, transformation and culture, are used in many cases. Gateway cloning or Ligation Independent Cloning can simplify these procedures; however, these methods still require transformation and time-consuming culture steps. Some template DNA preparation methods for cell-free protein synthesis without cloning steps were reported [13].

Recently, another method for linear DNA template production by two-step PCR was published [14] and successful protein synthesis was obtained with the two-step PCR product. In the first PCR step, a linear DNA containing the coding sequence was amplified by PCR using gene-specific primers, and was purified. In the first stage of the second PCR step, the product DNA was combined with the dsDNA for the T7 promoter and terminator elements. In the second stage of the second PCR step, the extended template DNA was amplified with an additional primer. The resultant PCR product was used as the template DNA for cell-free protein synthesis. This method was rapid, and generated high yields and high success rates for many coding sequences. However, there were still some

complicated steps, such as the purification step and the separate PCR stages, which are not suitable for automation.

In the present study, we developed a simpler and more robust two-step PCR protocol to prepare template DNA for high-throughput cell-free protein synthesis. Our two-step PCR protocol does not involve DNA purification. The second PCR step does not require a break for primer addition, and therefore is not separated into two stages. This protocol provided a high yield and a high success rate, and the protocol was applicable to target DNAs with various GC contents. In the second PCR step, the 5'- and 3'-termini of the target coding sequence can be connected to additional sequences, such as N- and/or C-terminal tags and transcription and translation elements, e.g., a promoter, a ribosome binding site, and a terminator. We explored the robustness of the two-step PCR against possible fluctuations in experimental conditions in practical use. We also expressed proteins fused with various tags by cell-free synthesis. Both of the present methods of two-step PCR and cell-free protein synthesis are suitable for robotics, because they involve only simple procedures and are free from any cloning process. Thus, we have established a practical platform to produce and analyze proteins on a genome scale in a high-throughput manner.

Materials and methods

Materials

Oligonucleotides were purchased from Invitrogen and SIGMA Genosys. The purification grades of the oligonucleotides were 'desalted' for Invitrogen and 'cartridge' for SIGMA Genosys, respectively. The Expand Hi-Fi PCR kit was obtained from Roche. The iProof Hi-Fi PCR kit was obtained from Bio-Rad. The pCR2.1-TOPO cloning vector was from Invitrogen.

In our experience, it is important to check the quality of the primers and to select the appropriate primer manufacturer and grade carefully, especially when working with many samples, because the quality of the primers was strongly dependent on the manufacturers and grades. The concentrations of the primers in some lots from some manufacturers were far less than the specified ones. Some primer lots contained a large amount of defective product. The two-step PCR sometimes failed when using these primers. The two-step PCR product obtained using some primer lots contained more nucleotide deletions in the primer region than in other regions, as confirmed by sequencing after cloning into a vector followed by single colony isolation, although the products seemed to have the proper length, as confirmed by agarose gel electrophoresis.

Target clones

Human cDNA clones (Ultimate ORF clones, Invitrogen), which varied in GC contents, were used as test clones (Table 1). A total of 18 clones were for excised domain expression, and 24 clones were for full-length expression. The pk7-Ras plasmid [15, 7], which encodes the human Ha-Ras protein, was used as the standard. For full-length expression, the clones for which the localization of the translated proteins was predicted as cytoplasmic or nuclear by PSort2 [16] were selected. The cDNA vectors were used to transform *E. coli* strain DH10B, and the cells were cultured in a 96-well plate in LB medium with 50 µg/ml ampicillin and 7% glycerol to the steady-state phase without shaking. For the culture of cells with the pk7-Ras plasmid, kanamycin was used as the antibiotic.

In our experience, the quality of the cDNA clones differed widely among the manufacturers. Some clones from some manufacturers had a mismatch between the actual sequence and the provided sequence information. Quite a few samples that failed in the two-step PCR had this mismatch, and the unique primers were designed according to the mismatch region. For trouble-shooting of problems with the two-step PCR, reconfirming the sequence of the target cDNA clone is strongly recommended.

T7 promoter (T7P) fragment

A fragment was excised by PCR using the U2T7PL primer (GCTCTTGTCATTGTGCTTCG CATGATTACGAATTC AGATCTCGATCCCG) and the c(NL1) primer (CCCGA GGAGCCGCTGG) from pk7b2-NHisRas, a derivative of pk7-Ras, which has a natural poly-histidine affinity tag (NHis), a tobacco etch virus (TEV) protease recognition sequence [17], and the NL1-linker sequence (CCCGAG-GAGCCGCTGG) upstream of the c-Ha-Ras coding sequence (NHis-TEV-TV2 fragment, Fig. 1a). The NHis tag is a modified version of the HAT tag [18], which is part of the chicken lactate dehydrogenase-A gene [19]. In addition to the NHis tag fragment, fragments with glutathione-S-transferase (GST), maltose-binding protein (MBP) or streptavidin-binding peptide (SBP) [20] tags were also constructed (the GST-TEV-NL1, MBP-TEV-NL1, and SBP-TEV-NL1 fragments, respectively). In a similar way, a derivative tag that contains the TV2-linker sequence (ACTGAGAACCTGTACTTCCAGGG), instead of the TEV recognition site and the NL1-linker, was also constructed (the NHis-TEV-TV2 fragment). Some fragments were cloned into the pCR2.1-TOPO vector

(Invitrogen), sequence-verified, and amplified from the cloned vector by PCR, using Pyrobest polymerase (Takara) with the U2 primer (GCTCTTGTCATTGTGCTTCG) and the c(NL1) primer (CCCGAGGAGCCGCTGG) or c(TV2) primer (CCCTGGAAGTACAGGTTCTCAGTAGTTGGG ATATCG) for the fragments with the NL1-linker or TV2-linker, respectively. The resultant fragment was purified by agarose electrophoresis. The gels were stained with SYBR-Gold and were visualized by excitation with blue-light. Gel slices with the appropriate DNA bands were excised, and the DNA was purified by absorption to a glass-surface under chaotropic conditions, using a GFX column (Amersham Biosciences) or a QIAGEN Plasmid Midi Kit (QIAGEN).

T7 terminator (T7T) fragment

The CL1-Term T7T fragment was excised by PCR using the c(CL1-Term-LN1) primer (GCGGTGGCAGCAGCC AACTCAGCATCAATCAATTATTATCCTGACGAGGG CCCCC; the sequences complementary to the termination codons are underlined) and the U2T7TL2 primer (GCTCTTGTCATTGTGCTTCG CCAAGCTTGCATGC CTGCAGCTC) from the pk7b2-NHisRas vector. The sequence of the fragment was confirmed, and the fragment was purified in a similar manner as the T7P fragment (Fig. 1b). The CL1-Term fragment contains the CL1-linker sequence (CCTGACGAGGGCCCCG in the anti-sense strand) followed by the tandem repeat of termination codons (TAATAATTGATTGAT). Derivative fragments with the SBP- or MBP-coding sequence with the TEV protease recognition sequence, instead of the termination codon repeat (the CL1-TEV-SBP and CL1-TEV-MBP fragments, respectively), and a derivative fragment in which the CL1-linker sequence was replaced by the DT2-linker (GGGCGGGGATCAATCAATCATT in the anti-sense strand) were also constructed (the DT2-Term fragment). The linkers are shown in Fig. 2, and the entire sequences of the fragments are shown in Supplementary Table 1.

Unique primers for two-step PCR

The forward (FW) unique primer for the NL1 linker consisted of the NL1-linker sequence and the unique sequence: 5'-CCAGCGGCTCCTCGGGA-X_n-3'. The sequence X_n was identical to the 5'-terminal sequence of the target coding sequence. The reverse (RV) unique primer for the CL1 linker consisted of the CL1-linker sequence and the unique sequence: 5'-CCTGACGAGGGCCCCG-Y_n-3'.

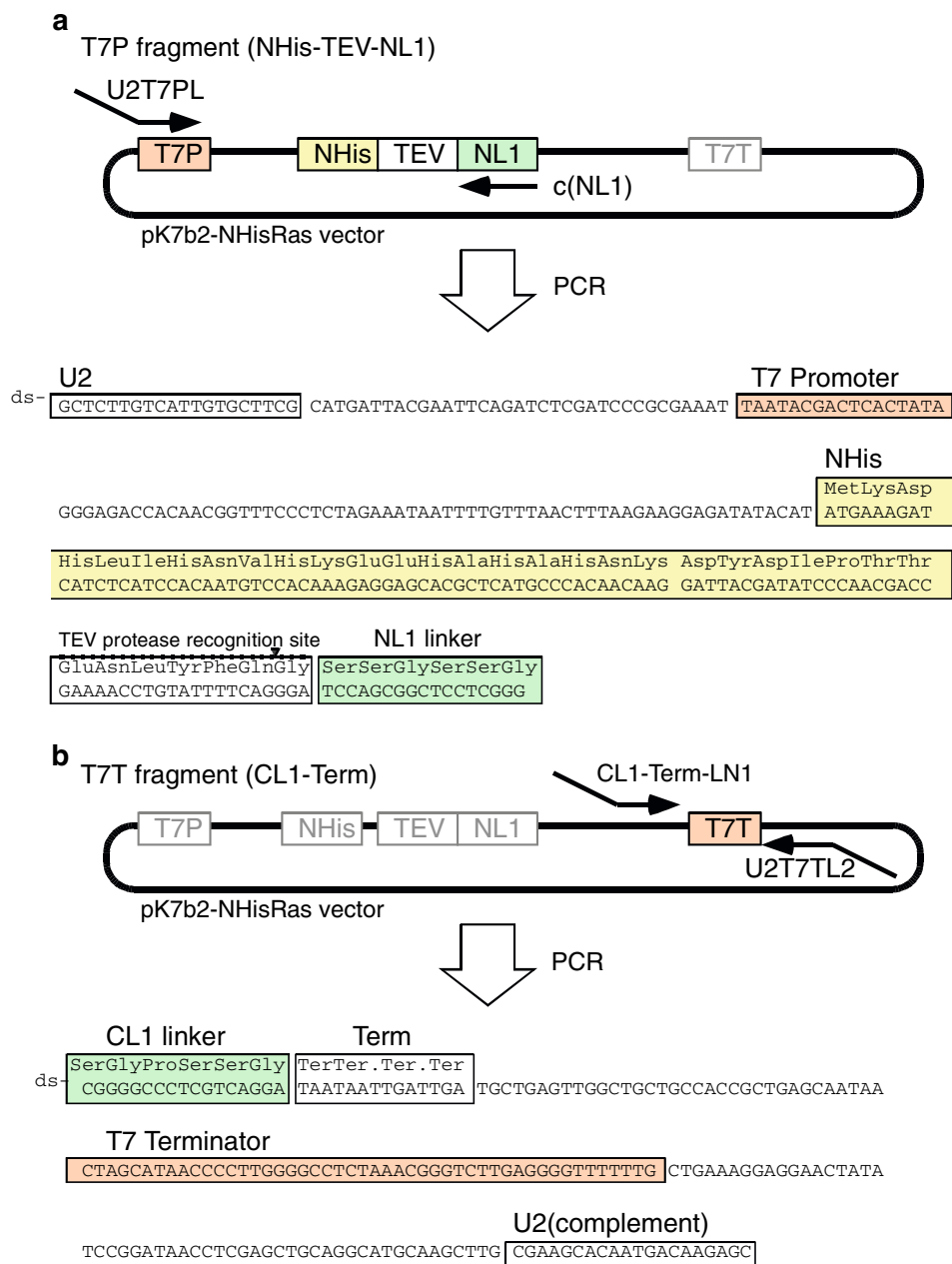
Table 1 continued^d Annotation of full-length sequence of the target clone in GenBank or SwissProt^e GC content of whole sequence of the target^f Maximum GC content scanned over whole sequence of the target with 150-bp window^g Results of the second PCR under the 'Std' conditions were about 60–120 ng/μl of proper length product (2), less amount of proper length product or associated with improper length product (1), no proper length product detected (0) or no experiment performed (blank)^h Results of the second PCR under the '+DMSO' conditions. Value definitions are the same as in the caption ^g

The sequence Y_n was complementary to the 3'-terminal sequence of the target coding sequence. The lengths of the unique sequences (X_n and Y_n) were designed to be 14 nt or longer, to provide a T_m of at least 46°C (Fig. 3a). The unique sequences used for test construction are shown in Supplementary Table 2. For the forward TV2-linker, the FW unique primer sequence was 5'-ACTGAGAACCTG TACTTCCAGGGA- X_n -3'. For the reverse DT2-linker, the RV unique primer sequence was 5'-GGGCGGGGATCA ATCAATCATT- Y_n -3'.

Two-step PCR

We defined the 'Standard (Std)' two-step PCR conditions as follows. The first PCR was carried out in a reaction mixture (20 μl) with 3 μl of 50-fold diluted culture medium of the cDNA clone as the template, 50 nM each of the FW and RV unique primers, 0.2 mM each of dNTPs, 1× Expand-Hi-Fi buffer and 0.5 U Expand-Hi-Fi Enzyme (Roche) with hot start (Fig. 3a). The PCR program began with a 2 min denaturation step at 94°C. This step was followed by 40 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 30 s and extension at 72°C for 1 min (after the 20th cycle, the extension duration was prolonged for 5 s per cycle). The last step was an incubation at 72°C for 7 min. The resultant product was immediately cooled to 10°C. The second PCR was carried out in a reaction mixture (20 μl) with 5 μl of 5-fold diluted first PCR product, 50 pM T7P fragment, 50 pM T7T fragment, 1 μM U2 universal primer (GCTCTTGTCATTGTGCTTCG), 0.2 mM each of dNTPs, 1× Expand-Hi-Fi buffer and 0.5 U Expand-Hi-Fi Enzyme (Roche) with hot start (Fig. 3b). The PCR program began with a 2 min denaturation step at 94°C. This step was followed by 30 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 30 s and extension at 72°C for 2 min for the N-NHis tag (the NHis-TEV-NL1 and CL1-Term fragments), 3 min for the N-SBP tag (the SBP-TEV-NL1 and CL1-Term fragments) and the N-NHis/C-SBP tag (the NHis-TEV-NL1 and CL1-TEV-SBP fragments) and 4 min for the N-MBP tag (the MBP-TEV-NL1 and CL1-Term fragments), the N-GST tag (the GST-TEV-NL1 and CL1-Term fragments) and the N-NHis/C-MBP tag (the NHis-TEV-NL1 and CL1-TEV-MBP fragments). After the 10th cycle, the annealing temperature was changed to 64°C, and the extension duration was prolonged for 5 s per cycle. The last step was an incubation at 72°C for 7 min. The resultant product was immediately cooled to 10°C. The concentration of the resultant product was determined with a PicoGreen dsDNA quantification kit (Invitrogen). All of the dilution steps in the two-step PCR protocol were carried out using the dilution buffer (1 mM Tris-HCl, 0.01 mM EDTA, pH 8.0).

Fig. 1 Fragments for the second PCR. **(a)** Preparation of the T7P fragment (NHis-TEV-NL1). **(b)** Preparation of the T7T fragment (CL1-Term). Arrows represent primers to excise the fragment



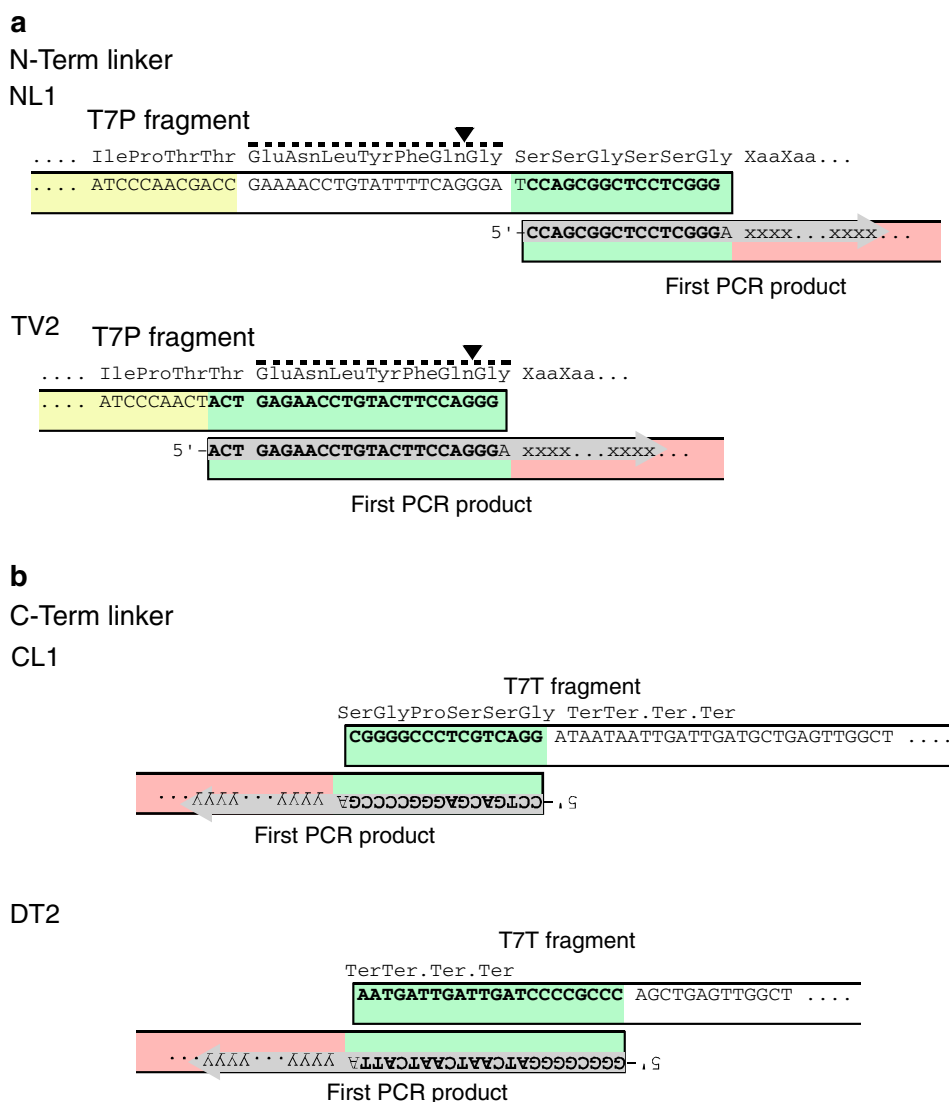
Two-step PCR with DMSO conditions

The ‘Std’ PCR conditions were modified as follows for the ‘+DMSO’ conditions. DMSO (5% v/v) was added to both the first and second PCR reaction mixtures. The first denaturation temperature was 95°C. This step was followed by 30 cycles of denaturation at 95°C for 30 s, annealing at 60°C for 30 s and extension at 72°C for 3 min (after the 10th cycle, the extension duration was prolonged for 5 s per cycle). The last step was an incubation step at 72°C for 7 min. The resultant product was immediately cooled to 10°C. The same cycle protocol was used for both the first and second PCRs.

High-fidelity and fast two step PCR

We defined the ‘High fidelity and Fast (HF)’ two-step PCR conditions as follows. The first PCR was carried out a reaction mixture (20 µl) with 10 ng of purified cDNA vector as the template, 50 nM each of the FW and RV unique primers, 0.2 mM each of dNTPs, 1× iProof-HF buffer and 0.4 U iProof enzyme (BioRad) with hot start. The PCR program began with a 30 s denaturation step at 98°C. This step was followed by 25 cycles of denaturation at 98°C for 5 s, annealing at 60°C for 10 s and extension at 72°C for 30 s. The last step was an incubation at 72°C for 5 min. The resultant product was immediately cooled to 10°C. The second PCR

Fig. 2 Linker sequence variation. **(a)** The T7P fragments. **(b)** The T7T fragments. Sequences in bold letters represent linkers. Shaded arrows indicate the unique primer-binding regions of the first PCR product. Amino acid residues under the broken line represent the TEV recognition sequence. Arrowheads represent the TEV cleavage site



was carried out in a reaction mixture (20 μ l) with 5 μ l of 5-fold diluted first PCR product, 50 pM T7P fragment, 50 pM T7T fragment, 1 μ M U2 universal primer, 0.2 mM each of dNTPs, 1 \times iProof-HF buffer and 0.4 U iProof enzyme (BioRad) with hot start. The PCR program began with a 30 s denaturation step at 98 $^{\circ}$ C. This step was followed by 25 cycles of denaturation at 98 $^{\circ}$ C for 5 s, annealing at 60 $^{\circ}$ C for 10 s and extension at 72 $^{\circ}$ C for 45 s. The last step was an incubation at 72 $^{\circ}$ C for 5 min. The resultant product was immediately cooled to 10 $^{\circ}$ C.

Robustness of two-step PCR

Dilution factor of template culture

To explore the effects of the dilution factor of the template culture on the two-step PCR, a culture of *E. coli* cells with the pk7Ras plasmid, which was grown to 0.8 OD₆₀₀, was

diluted to various relative culture concentrations (from 1/16- to 4-fold). One-fold of the relative culture concentration corresponds to a 50-fold culture dilution, which is used in the 'Std' PCR conditions. The two-step PCR was carried out with the diluted cultures as the template for the first PCR.

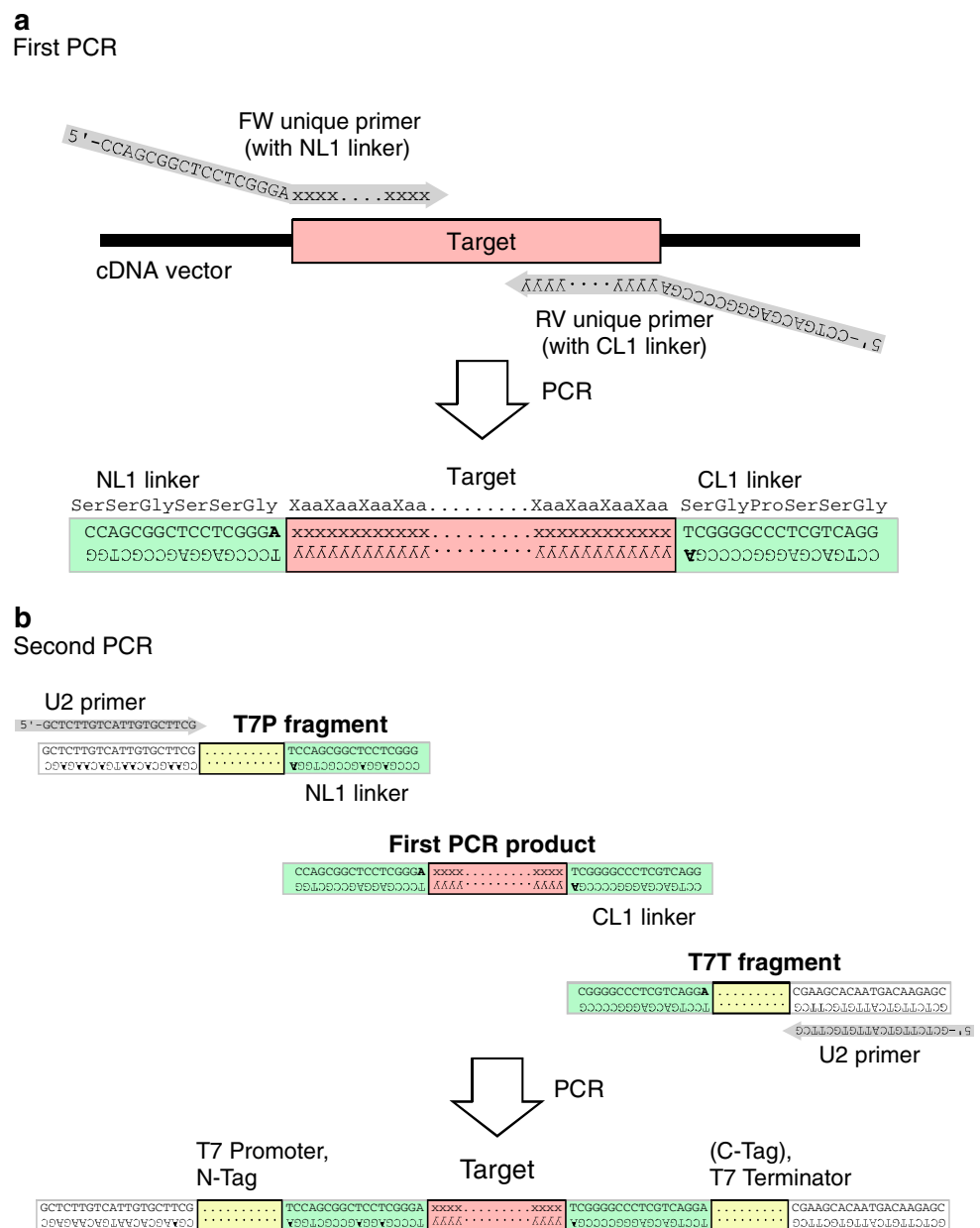
Cell density in culture

To clarify the effects of the cell-density in the culture, cells harboring the pk7Ras plasmid were grown to various densities (from 0.05 to 2 OD₆₀₀), and the resultant culture was subjected to the two-step PCR with the 'Std' conditions.

Plasmid content in a unit quantity of cells

To determine the effect of plasmid contents in a unit quantity of cells, cultures of cells with the pk7Ras plasmid

Fig. 3 Two-step PCR. **(a)** Design of the unique primers and the first PCR to amplify the target domain-coding region. Shaded arrows represent the unique primers. **(b)** Design of the second PCR. Shaded arrows represent the universal primer. Bold bases should be 'A' in order to compensate for the 3'-adenine overhang in the fragments and the first PCR products produced by Taq polymerase and its variations



(pk7Ras culture) and without the plasmid (blank culture) were grown to 0.8 OD₆₀₀. The pk7Ras culture was diluted by the blank culture to various extents (from 2⁻¹⁵- to 1-fold), and the resulting cultures were subjected to the two-step PCR with the 'Std' conditions.

Ramp rate of heating and cooling in PCR program

To explore the effects of the ramp rate of heating and cooling in the PCR programs, two-step PCR under the 'Std' conditions with restricted ramp rates for heating and cooling (1, 2 and 3°C/s) were carried out for test constructs (Nos. 1–10).

Primer concentration

To determine the effects of FW and RV primer concentration variations, the concentration of each primer was calibrated by its A₂₆₀ value, using the molar absorbance coefficient for each primer, which was provided by the primer manufacturer, and two-step PCR was carried out with various primer concentrations (from 1/4- to 8-fold concentrations for the 'Std' conditions for both the FW and RV primers) for the test constructs (Nos. 1–10). To determine the effect of an imbalance between the FW and RV primer concentrations, FW and RV primers with 1/2-, 1- and 2-fold of the 'Std' concentration were mixed with each other, and two-step PCR was carried out using the primer mix for the first PCR.

Cell-free protein synthesis

The cell-free protein synthesis reaction was carried out at 30°C overnight, with the 30- μ l scale dialysis method [21]. The second PCR product (0.5 μ l), was used without purification as the template for cell-free protein synthesis. An aliquot of the resultant product was reserved (total fraction), and then the remainder was centrifuged at 15,000 g at 4°C for 5 min, and the supernatant was reserved (sup. fraction). The fractions were analyzed by SDS-PAGE (Perfect NT Gel, DRC) and were stained by Quick-CBB (Wako Pure Chemicals). The gel images were acquired by LAS3000 (Fuji) or FAS III (Toyobo) imagers. The target protein was quantified from the density of the band in the image, using BSA as the standard. The quantification error was estimated from multiple sets of assays of a subset of the clones. For the tag cleavage assay, the cell-free reaction mixture was incubated with 15 μ g/ml of TEV protease at 30°C for 3 h after protein synthesis.

PCR error rate analysis and assessment of effects on HSQC spectra

In order to quantify the error rate for the two-step PCR, the second PCR products of construct Nos. 7 and 35 were cloned into the pCR2.1-TOPO vector, and clones were picked and sequenced. The error rate was calculated from the resultant sequences.

The ^1H - ^{15}N -HSQC spectrum provides information about protein folding and may be used to determine the viability of the structure determination of the protein. We examined the effect of template error rates higher than those in the normal two-step PCR products on the ^1H - ^{15}N -HSQC spectrum. The two-step PCR product with the c-Ha-Ras coding sequence, NHis-TEV-NL1, and CL1-Term fragments were cloned into the pCR2.1-TOPO vector. The clones were sequenced and a clone with no errors was selected. The selected vector was used as the error-free template for cell-free protein synthesis. Linear templates with a higher error rate than the typical two-step PCR product were produced by error-prone PCR, which is a repetition of the PCR or by mutagenic PCR [22], from a two-step PCR product. Cell-free protein synthesis reactions were carried out with ^{15}N -labeled amino acids, based on the previously described protocol [23] with the error-free vector template as the standard and linear templates with increased error rates of 9.2×10^{-4} and 3.9×10^{-3} mutation/bp. The NHis-Ras proteins were purified with Ni-affinity resin. The NHis tag was cleaved in a reaction mixture with 7.6 μ g/ml TEV and 0.4–0.6 mg/ml of the tagged protein at 30°C overnight, and then the ^1H - ^{15}N -HSQC spectrum was acquired using the tag and Ras protein mixture.

Simulation of effect of template error rate on HSQC spectra

About 1000 randomly mutated DNA sequences with a given error rate (E) were generated from the NHis-Ras open reading frame sequence (656 bp, including tag, linker sequences and a termination codon). Each sequence was translated and the mutations in the amino acid sequence were analyzed. The average number of point mutations per sequence ($Np(E)$ and $Ap(E)$ for nucleotide and amino acid mutations, respectively) and the relative amount of non-sense mutants ($Rn(E)$) in the set were calculated.

The relative peak height ($H(E)$) of the ^1N - ^{15}N -HSQC amide cross peaks of the random mutant pool relative to that from an error-free ensemble can be roughly estimated as:

$$H(E) = 1 - Rn(E) - Ap(E) * M/N,$$

where M is the number of peaks with chemical shifts that change by more than the peak width at half height of the residue of the wild type, and N is the total number of amino acid residues of the translation product from the open reading frame. Here, it was assumed that neither the shifted peaks nor the peaks from nonsense mutants were observed in the ensemble.

N is 217 (amino acid residues) for the NHis-Ras protein. M was estimated to be 26 (a.a.), which was derived from a comparison of the ^1N - ^{15}N -HSQC spectra of the wild type and the Y32W mutant Ras protein (personal communication from *T. Matsuda*).

All of the calculations were carried out using Microsoft Excel and Visual Basic. Standard deviations of the values were determined from three sets of simulations.

Results

Design of two-step PCR protocols

In the first step of the two-step PCR (Fig. 3a), a linear DNA fragment containing the sequence encoding the target protein is amplified by PCR using two “unique primers”, consisting of the gene-specific sequences and the N- and C-terminal linker sequences (“NL1 and CL1 linkers”). In the second PCR step (Fig. 3b), the product of the first PCR step is treated with two dsDNA fragments (“T7P and T7T fragments”) and a single PCR primer (“U2 primer”). The T7P and T7T fragments have the promoter and terminator sequences for T7 RNA polymerase, the optional N- and/or C-terminal tag-coding sequences, and the NL1 and CL1 linker sequences. The NL1 and CL1 linkers were designed based on the following concept. These linkers encode six small hydrophilic amino acid residues, Ser-Ser-Gly-Ser-Ser-Gly and Ser-Gly-Pro-Ser-Ser-Gly, respectively, to be

connected to the N- and C-termini of the target protein (Figs. 1 and 3a), so that their influence on the structure and other properties of the target protein is minimized. The GC contents of the linkers were set to be as high as about 75%, in order to shorten the length of the unique primers and thus reduce their costs. The NL1 and CL1 linkers are the standard linkers for our high-throughput protein expression system for structural analysis. In addition to them, we designed the TV2 and DT2 linkers in order to minimize the number of residual amino acids after tag cleavage with the TEV protease (Fig. 2). The TV2 N-terminal linker encodes the TEV protease recognition sequence, and only one glycine remains at the N-terminus of the target protein region after TEV protease cleavage. The DT2 C-terminal linker provides four termination codons, UAA-UAA-U-UGA-U-UGA, where the tandem in-frame UAA codons are used to avoid read-through, and the two out-of-frame UGA codons are used to prevent read-through by frame shifting in the target protein region. As compared to the NL1 and CL1 linkers, the TV2 and DT2 linkers require longer unique primers because of their lower GC contents. The DT2 linker is designed to make the same C-terminal amino acid residue as that of the target protein. Generally, an additional adenine overhang is often attached to the 3'-terminus of the amplified dsDNA product, in PCR using the family of Taq polymerases. Thus, the bases around the linkers were designed in order to completely match the fragments with the first PCR product, even if the adenine overhangs were attached (Fig. 3b, bold 'A' bases).

The U2 primer, designed with an artificial sequence to avoid mispriming to popular vectors, was used to amplify the final product in the second PCR step (Figs. 1 and 3b). The U2 primer-binding site is incorporated at both the 5'-terminus of the sense strand of the T7P fragment and the 5'-terminus of the anti-sense strand of the T7T fragment. Therefore, the second PCR amplification is performed with the single U2 universal primer. Amplification with the single primer could also inhibit dimerization of the primer and increase the yield of the proper PCR product [24].

During the preparation of the T7P and T7T fragments, UV-light excitation for visualizing the DNA band for extraction from the agarose gel should be avoided, in order to minimize the damage to the DNA. When the fragment band was visualized by UV-light excitation, the following second PCR step tended to fail (data not shown).

The concentrations of the FW and RV "unique primers" in the first PCR step are important. The primer concentrations should be as low as 50 nM each, in order to increase the priming rate per primer for amplification of the target-coding region. This is to reduce the production of "primer dimers", in which the two primers are connected head-to-head, without the target-coding region, but with an

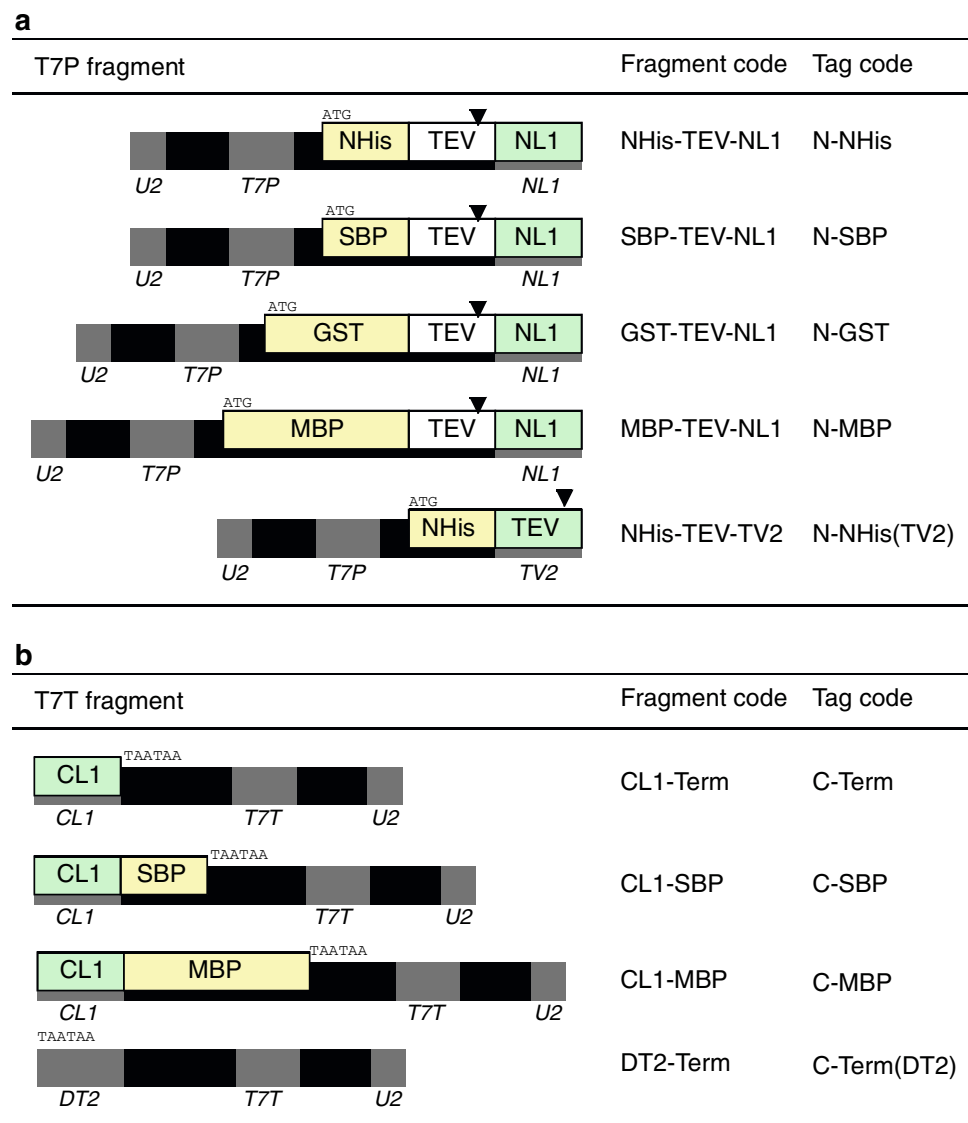
insertion or deletion of several nucleotides. The primer dimers generate a byproduct, the direct concatemer of the T7P and T7T fragments, in the second PCR step. Therefore, the low concentrations of the unique primers enable the direct use of the first PCR product, without any purification, as the template in the second PCR step. The concentrations of the four materials in the second PCR step should be in the following order: the U2 universal primer \gg the first PCR product \gg the T7P and T7T fragments, which is to avoid direct concatenation of the T7P and T7T fragments and to obtain mostly a single PCR product, by consuming the first PCR product, the T7P fragment, and the T7T fragment. Thus, the second PCR product can be used as the template for protein synthesis without any purification step. About 40 cycles of amplification were used in the first PCR step, to ensure amplification even from cell cultures with poor growth as the template of the two-step PCR.

Performance and robustness of the two-step PCR protocols

The fragments used for two-step PCR are shown in Fig. 4. Two-step PCR experiments were performed for a test set of human cDNA clones, with various GC contents (Table 1). The results for the N-NHis and C-Term (no tag) fragments are summarized in Table 1 and Fig. 5. For constructs of the proper length, the concentration of the second PCR product was about 60–120 $\mu\text{g}/\text{ml}$. Under the standard PCR conditions (Fig. 5a, 'Std' condition), construction for target Nos. 37, 40, 43 and 44, which have relatively high GC contents, failed and a 400-bp byproduct was observed in the second PCR products. The byproduct was the direct concatemer of the N-NHis and C-Term fragments, as confirmed by sequencing (data not shown). Construction for a high GC content target tended to fail in the two-step PCR under the 'Std' conditions. Moreover, a stronger correlation was observed between the two-step PCR results and the maximum values of GC content, scanned over the whole sequence with a 150-bp window (GCMax150) (Table 1). These failed constructs were successfully recovered by the use of the two-step PCR under the '+DMSO' conditions (Fig. 5a, '+DMSO' conditions). This indicates that PCR under the '+DMSO' conditions would be effective for target protein regions with a high GCMax150 value.

Construction by two-step PCR was successful at least for target protein regions with 100–2200-bp lengths and with 45–84% GCMax150 values or 37–75% GC contents. Irrespective of the tag size, the N-NHis, N-SBP, N-GST, and N-MBP tagged constructs were successfully obtained (Table 1, Fig. 5b). The N-NHis/C-SBP tagged constructs and the N-NHis tagged constructs with the TV2 and DT2

Fig. 4 Schematic representations of fragments, fragment codes and tag codes. **(a)** The T7P fragments. **(b)** The T7T fragments. The closed boxes represent dsDNA. Element names in the fragment DNA are written in italic letters. The open boxes represent elements that are coding sequences, and the letters in the box indicate element names in the translated polypeptide. The element ‘TEV’ represents the TEV protease recognition sequence, and the arrowheads show the TEV cleavage site



linkers were also successfully obtained (Table 1). The two-step PCR construction could cover a large variety of target protein regions and tags with a few modifications of the reaction conditions. Moreover, the experimental procedure is simple, because it contains neither a purification step nor a separation step.

Experimental protocols for high-throughput use, i.e., parallel processing of many samples, should be tolerant of fluctuations in experimental conditions, because it is difficult to equalize the conditions precisely for all of the samples. For example, cumulative liquid-dispensing on the micro-liter scale using robotic systems may sometimes cause an accumulated error in the total concentration of two-fold or more, and the cell density and the plasmid contents of cultures for PCR templates may vary. Therefore, we investigated the tolerance of the two-step PCR protocols against some possible variations in practical use.

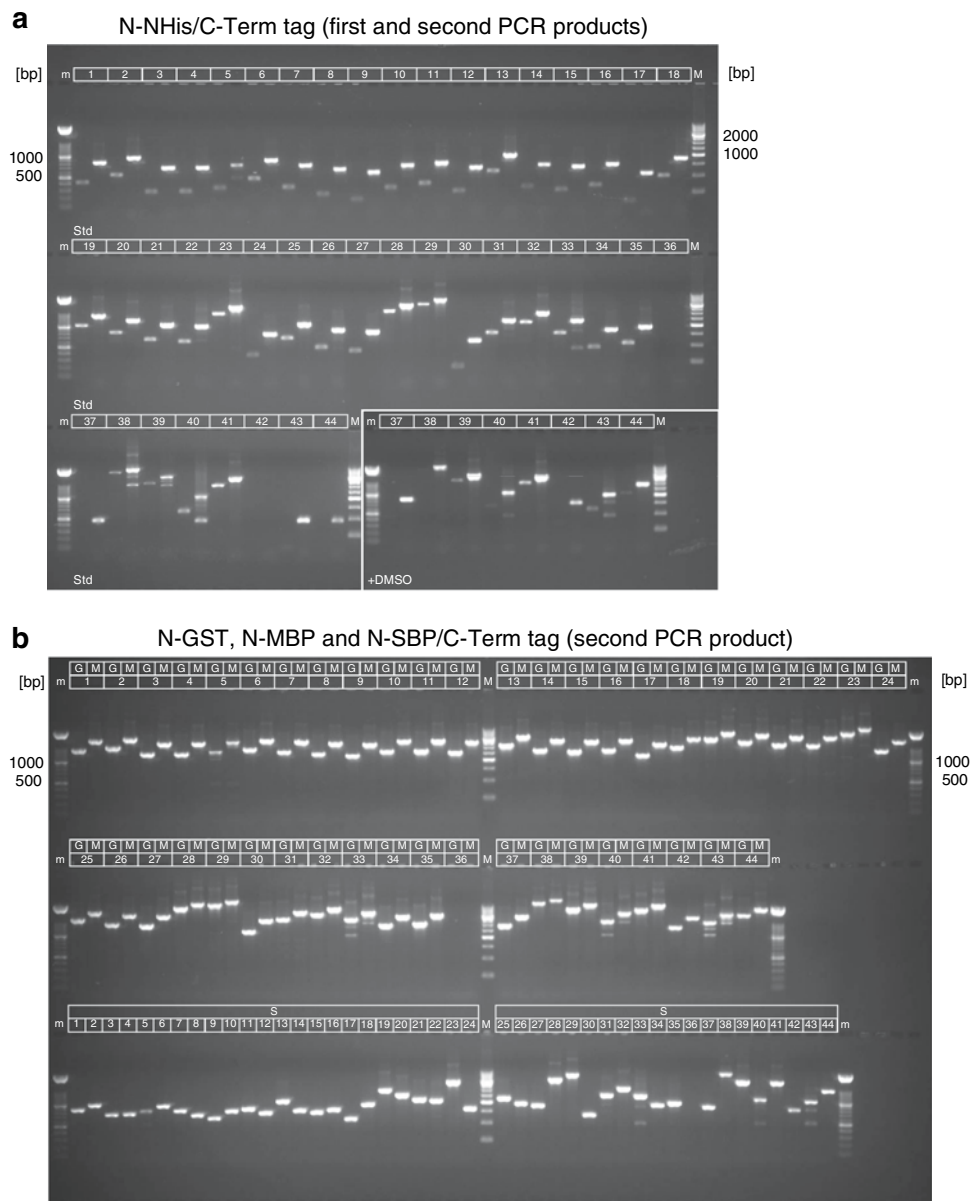
Dilution factor of template culture

We examined the effect of template dilution on the two-step PCR, and found that the PCR was successful when the concentration relative to that under the ‘Std’ PCR conditions was from 1/16- to 4-fold (Supplementary Fig. 2a). The results indicated that the two-step PCR would be tolerant of actual liquid-dispensing errors.

Cell density in culture

Even if the template cultures were incubated for a certain time at a certain temperature, the cell densities may differ, due to the various vectors and inserts. The two-step PCR from template cultures with various cell densities from 0.05 to 2 OD₆₀₀ was performed, and in all of the tested

Fig. 5 Agarose electrophoresis of the PCR product of the two-step PCR. **(a)** The N-NHis/C-Term tag construct in the 'Std' conditions (without DMSO) and the '+DMSO' conditions (with DMSO). Numbers (1–44) represent the construct number in Table 1. For constructs 37–44, PCR results under the '+DMSO' conditions are also shown. Each construct was run in two lanes (the first PCR product, left lane; the second PCR product, right lane). Markers are 100-bp and 250-bp DNA ladders. The PCR product (2 μ l) was electrophoresed in each lane. **(b)** N-GST/C-Term (G), N-MBP/C-Term (M), and N-SBP/C-Term (S) tag constructs in the 'Std' conditions for constructs 1–36 and in the '+DMSO' conditions for constructs 37–44



conditions, the PCR was successful (Supplementary Fig. 2b). Therefore, the two-step PCR is tolerant of variations in cell density resulting from practical template cultivation procedures.

Plasmid content in unit quantity of cells

Target plasmid contents per unit cell weight may vary because (1) unstable inserts may become deleted from the plasmids during cultivation, (2) plasmids without the proper insert may contaminate the plasmid-construction step by some procedures, (3) cells without plasmids may contaminate some transformation procedures and grow in cultivation after the antibiotics have been exhausted, (4) in

the case of the simultaneous culture of many clones in a multi-well plate, the culture may be contaminated with neighboring cultures, and (5) the copy number of the plasmids depends on the type of replication origin. Therefore, we have investigated the tolerance to fluctuations in the plasmid content in a unit amount of cells. A culture of cells harboring the pk7Ras plasmid was diluted with a culture of cells without plasmid at various dilution factors, and the two-step PCR was carried out. The PCR was successful for the 1- to 2^{-10} (=1/1024)-fold concentrations, and a byproduct band was observed for conditions diluted more than 2^{-10} -fold (Supplementary Fig. 2c). Therefore, the tolerance to variation in the plasmid content is sufficient for practical template preparation procedures. The pk7Ras plasmid is a derivative of a high copy-number

vector with the ColE1 origin, and therefore, lower copy-number plasmids and single copy-number vectors, such as pBR322, pACYC, and BAC derivatives, may also be used as templates for the two-step PCR under the 'Std' conditions.

Ramp rate of heating and cooling in the PCR program

The ramp rate of heating and cooling is dependent on the PCR manufacturers and machines. The ramp rate of one PCR machine may also be affected by the deterioration of the Peltier device over time. In order to assess the durability of the PCR program against the ramp rate, the two-step PCR experiments were performed with forced maximum ramp rates of 3, 2 and 1°C/s. The rate of 3°C/s corresponds to that in the 'Std' conditions. The PCR was successful for all of the ramp rates (Supplementary Fig. 2d).

Primer concentration

We examined the tolerance of the two-step PCR against variations in the unique primer concentration. When the concentration of the unique primer was varied from 1/8- to 8-fold of that in the 'Std' conditions in the first PCR mixture, the two-step PCR was successful at least with 1/2- to 8-fold concentrations (Supplementary Fig. 3a). The best quality of the second PCR product was obtained from 1- to 2-fold concentrations. The tolerance of an imbalance between the FW and RV unique primer concentrations was also examined. A two-fold imbalance of the primer concentrations relative to those in the 'Std' conditions did not affect the two-step PCR result (Supplementary Fig. 3b).

The two-step PCR is tolerant to possible variations in practical use, and therefore, it is robust and suitable for high-throughput use with a robotic system.

Cell-free protein synthesis and tag variations

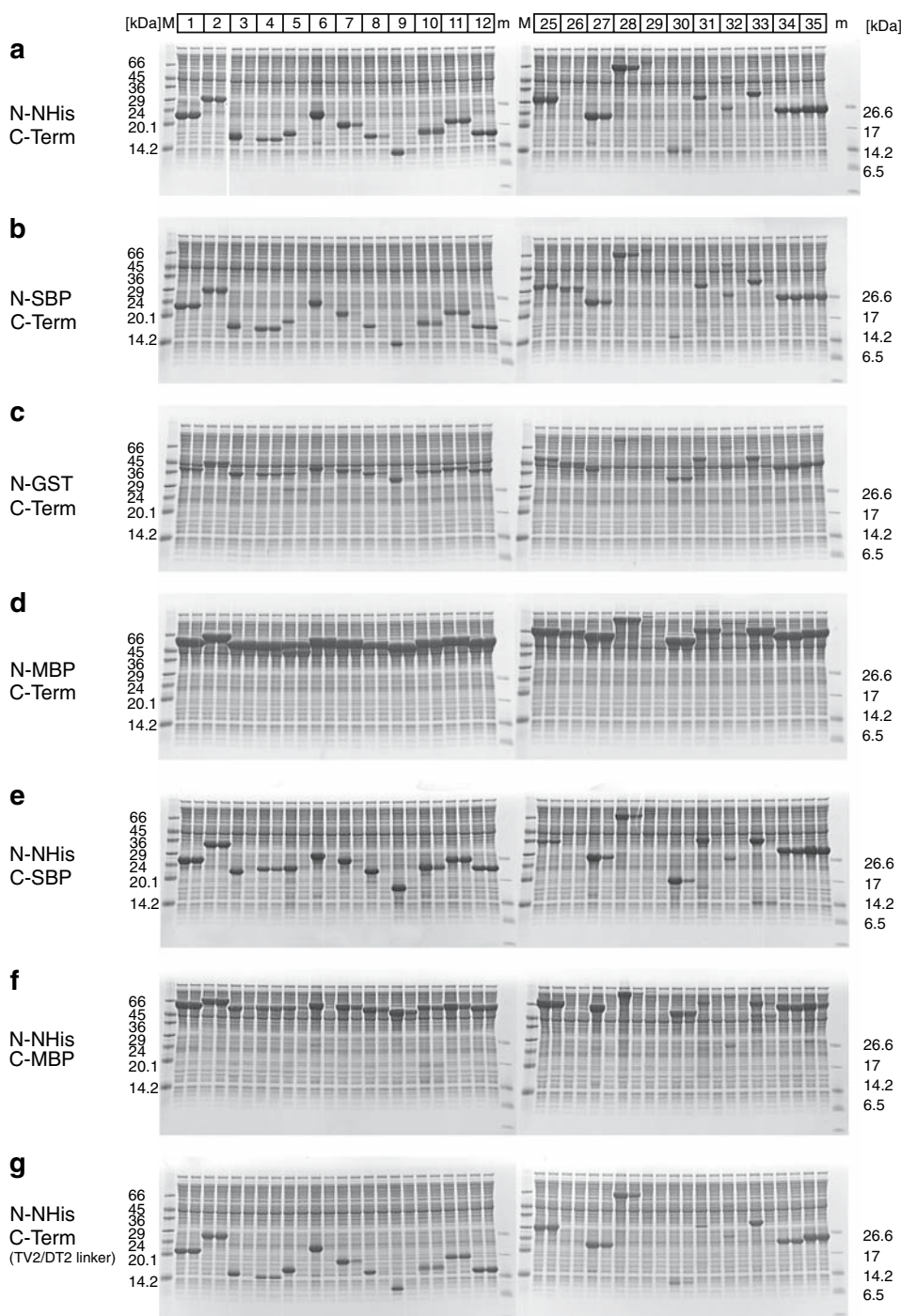
Cell-free synthesized proteins from template DNAs constructed by the two-step PCR were analyzed by SDS-PAGE (Fig. 6 for some constructs, Supplementary Fig. 1 for all constructs). The productivity and the solubility with various fusion tags are summarized in Table 2. The productivity and the solubility were similar between the N-NHis and N-SBP tagged proteins for most of the test targets (Fig. 6a and b), although some targets showed variations in the solubility with different tags. Although the production of the N-GST tagged proteins was less abundant

than that of the N-NHis tagged proteins on average, they were slightly more soluble than the N-NHis tagged proteins (Fig. 6c). MBP is a well-studied solubility enhancer, and the fused target proteins do not remain soluble in some cases after removal of the MBP tag [25]. In our study, the N-MBP tag increased the productivity and the solubility (Fig. 6d); out of the 44 target protein regions, 17 target proteins were soluble with the N-MBP tag but insoluble with the N-NHis tag, and 6 out of the 17 MBP-solubilized protein regions precipitated after N-MBP tag cleavage (data not shown). This result suggests that some proteins were solubilized with the MBP tag, but were not folded properly. The C-MBP tag also raised the solubility; however, it was less effective than the N-MBP tag. The N-NHis/C-SBP doubly tagged proteins and most of the N-NHis tagged proteins with the TV2 and DT2 linkers showed solubilities similar to those of the N-NHis tagged proteins (Fig. 6e and g). The production levels of the N-NHis tagged proteins with the TV2 and DT2 linkers were slightly less than those with the N-NHis tag with the NL1 and CL1 linkers.

PCR error rate analysis and assessment of effects on HSQC spectra

The error rate of the two-step PCR was examined using constructs No. 7 and No. 35, and was determined to be 3-bp mutations per 13,300-bp total read ($=2.3 \times 10^{-4}$ mutation/bp) and 16-bp mutations per 21,032-bp total read ($=7.6 \times 10^{-4}$ mutation/bp). This rate is equivalent to the rough estimation that 0.1–0.5 point mutation exists in one molecule, in the case of a target region with a 651-bp length, which corresponds to the size of the NHis-tagged Ras protein. This estimation includes silent mutations; therefore, the error rate at the amino acid sequence level would be lower. We tried to examine the effects of the PCR error rate on our ^1H - ^{15}N -HSQC protein-folding assay, but the effects were too small to detect (data not shown). Thus, in addition to the error-free vector template (a), two kinds of linear templates (b) and (c) with purposely-increased error rates, 9.2×10^{-4} and 3.9×10^{-3} mutation/bp, respectively, were produced using error-prone PCR. The mutation contents of templates (a), (b), and (c) were estimated to be 0, 0.6 and 2.5 point mutations per DNA molecule, respectively. The error rate of template (b) was slightly higher than that of the two-step PCR product for NHis-Ras, whereas the error rate of template (c) was about 5-times higher than that of the two-step PCR product. Using these three templates in cell-free protein synthesis, we synthesized ^{15}N -labeled Ras proteins and measured their HSQC spectra (Fig. 7). Protein sample (b), with an error rate slightly higher than that of the standard

Fig. 6 SDS-PAGE of cell-free protein synthesis reaction mixtures for some of the test constructs with various tags and linkers. The tags were N-NHis/C-Term (a), which indicates N-terminal NHis tag and C-terminal no tag, N-SBP/C-Term (b), N-GST/C-Term (c), N-MBP/C-Term (d), N-NHis/C-SBP (e), N-NHis/C-MBP (f) and N-NHis/C-Term (TV2/DT2 linker) (g). The linker types were NL1/CL1 (a to f) and TV2/DT2 (g). Numbers represent construct numbers in Table 1. Each construct was run in two lanes (total fraction, left lane; supernatant fraction, right lane). The reaction mixture (1.7- μ l aliquots) was electrophoresed for each lane and was stained with CBB



two-step PCR product, provided almost the same spectrum as that of the error-free sample (a). Moreover, the sample (c), with an error rate 5-times higher than that of the two-step PCR product, also provided almost the same spectrum as that of the error-free sample (a). We also found that the protein productivity of sample (b) was the same as that of sample (a), while sample (c) was somewhat less productive than samples (a) and (b) (Fig. 7b). The efficiency of TEV cleavage was more than 80% for

samples (b) and (c) and slightly less than that of the error-free sample (a). The solubility of the target proteins after tag cleavage was the same for all of the samples. This result indicated that even sample (c) exhibits unaffected solubility of the ensemble of randomly mutated protein molecules. Therefore, we concluded that the error rate of the two-step PCR does not disturb the HSQC protein-folding assay. For experiments that may require higher homogeneity, e.g., for crystallization of proteins, it may

Table 2 Results of cell-free protein synthesis for test clones

No. ^a	MW ^b (kDa)	PCR Conditions ^c	Cell-free protein synthesis summary													
			N-NHIs C-Term +5.0 kDa	N-SBP C-Term +6.1 kDa	N-GST C-Term +27.8 kDa	N-MBP C-Term +44.8 kDa	N-NHIs C-SBP +10.1 kDa	N-NHIs C-MBP +47.6 kDa	N-NHIs(TV2) C-Term(DT2) +4.1 kDa							
			Pro. ^d (mg/ml)	Sol. ^e	Pro. ^d (mg/ml)	Sol. ^e	Pro. ^d (mg/ml)	Sol. ^e	Pro. ^d (mg/ml)	Sol. ^e	Pro. ^d (mg/ml)	Sol. ^e	Pro. ^d (mg/ml)	Sol. ^e		
1	14.9	Std	2.9	+++	1.3	++	0.2	+++	2.6	+++	2.0	+++	2.4	+++	1.4	+++
2	18.7	Std	2.4	+++	0.9	+++	0.4	++	2.4	+++	1.5	+++	2.2	+++	1.3	+++
3	9.4	Std	2.3	-	1.1	-	0.5	-	2.1	+++	1.1	+++	1.4	+	0.6	-
4	9.2	Std	0.3	+++	0.9	+++	0.3	+	2.6	+++	0.8	+++	1.4	+++	0.3	+++
5	11.8	Std	0.7	-	0.4	+	0.1	-	1.2	+++	1.6	+++	1.4	+	1.1	-
6	16.6	Std	3.3	-	1.5	-	0.7	-	3.4	+++	3.0	+++	3.1	-	1.9	-
7	11.7	Std	1.9	+	0.4	+	0.3	++	3.0	+++	1.7	+++	2.7	++	0.7	+
8	7.9	Std	0.5	+	0.3	+	0.2	+	1.7	+++	1.6	+++	1.8	++	0.4	-
9	5.1	Std	1.5	-	0.7	+	0.9	-	3.1	+++	1.7	+++	3.4	+	0.3	-
10	10.0	Std	1.2	+++	0.8	++	0.3	+++	2.3	+++	1.9	+++	1.5	+++	0.3	+++
11	12.5	Std	1.2	+++	0.8	+++	0.3	+++	2.5	+++	1.1	+++	3.4	++	0.5	+++
12	9.2	Std	1.1	+++	0.5	+++	0.2	+++	3.2	+++	1.5	+++	2.3	++	0.9	+++
13	21.3	Std	2.8	+++	0.6	+++	0.1	++	2.3	+++	1.1	+++	1.2	+	1.1	+++
14	11.1	Std	1.4	-	0.7	-	0.6	-	2.1	+++	0.9	+++	2.6	-	1.0	-
15	9.7	Std	0.6	-	0.3	+	0.3	+	1.5	+++	0.6	+++	1.1	++	0.2	-
16	11.2	Std	1.2	++	0.5	+++	0.4	+++	2.3	+++	1.1	+++	1.8	+++	0.4	+++
17	4.2	Std	0.8	+++	0.4	+++	0.5	+++	0.1	+++	1.2	+++	3.1	+++	0.0	ND
18	15.8	Std	1.1	-	0.5	-	0.7	-	2.4	+++	1.3	+++	2.0	+++	1.1	-
19	36.5	Std	0.0	ND	0.1	+	0.1	+	0.0	ND	0.0	ND	0.0	ND	0.0	ND
20	30.0	Std	1.0	+++	0.6	+++	0.3	+++	3.0	+++	0.7	+++	1.4	+++	0.6	+++
21	23.1	Std	3.2	+++	1.6	+++	0.4	+	3.0	+++	1.7	+++	3.3	+	2.4	+++
22	21.2	Std	2.5	+	1.3	-	0.5	+	3.5	+++	1.9	+++	2.6	+	2.2	+
23	55.4	Std	0.2	+	0.1	+	0.0	ND	0.3	++	0.1	++	0.1	+	0.2	+
24	12.4	Std	0.0	ND	0.1	+++	0.3	+	1.3	+++	0.0	+++	0.1	+++	0.0	ND
25	23.7	Std	2.3	+++	0.7	+++	0.2	+++	2.6	+++	0.3	+++	3.4	++	1.4	+++
26	17.9	Std	0.1	++	0.2	+++	0.1	++	0.8	+++	0.1	++	0.1	++	0.0	ND
27	15.9	Std	2.9	++	1.5	+	0.2	-	3.0	+++	3.1	+++	3.9	-	1.7	++
28	60.6	Std	3.3	+	1.2	+	0.1	+	1.2	+++	1.7	+++	2.0	-	1.9	++
29	78.0	Std	0.0	ND	0.0	ND	0.0	ND	0.1	+	0.1	++	0.1	+	0.0	ND
30	6.0	Std	0.5	+	0.1	+	0.3	+++	3.1	+++	2.1	+++	1.7	++	0.2	++

Table 2 continued

No. ^a	MW ^b (kDa)	PCR Conditions ^c	Cell-free protein synthesis summary											
			N-NHIs C-Term +5.0 kDa	N-SBP C-Term +6.1 kDa	N-GST C-Term +27.8 kDa	N-MBP C-Term +44.8 kDa	N-NHIs C-SBP +10.1 kDa	N-NHIs C-MBP +47.6 kDa	N-NHIs(TV2) C-Term(DT2) +4.1 kDa	Pro. ^d (mg/ml)	Sol. ^e	Pro. ^d (mg/ml)	Sol. ^e	Pro. ^d (mg/ml)
31	26.8	Std	0.8	0.7	0.6	1.1	0.9	0.2	0.2	0.2	0.2	0.2	0.2	+
32	38.0	Std	0.2	0.2	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	+
33	26.4	Std	1.4	0.9	0.9	2.9	1.9	1.2	1.2	1.1	1.1	1.1	1.1	-
34	15.9	Std	2.0	1.2	0.7	3.2	1.7	2.0	2.0	1.0	1.0	1.0	1.0	+++
35	19.5	Std	3.4	1.0	0.7	2.6	3.1	2.8	2.8	1.6	1.6	1.6	1.6	+++
36														
37	13.8	+DMSO	1.1	0.7	0.6	2.5	0.8	1.1	1.1	0.5	0.5	0.5	0.5	-
38	85.1	+DMSO	1.2	1.1	0.3	0.6	1.2	0.3	0.3	1.2	1.2	1.2	1.2	-
39	57.6	+DMSO	0.1	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	+
40	21.4	+DMSO	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	ND
41	51.6	+DMSO	1.3	0.7	0.6	1.3	0.9	0.4	0.4	0.3	0.3	0.3	0.3	-
42	10.8	+DMSO	0.3	0.1	0.2	1.3	0.3	0.7	0.7	0.0	0.0	0.0	0.0	ND
43	20.1	+DMSO	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	ND
44	35.1	+DMSO	3.4	0.9	0.5	1.8	0.2	0.1	0.1	0.5	0.5	0.5	0.5	-

^a The local index for the test construct, which is defined in Table 1

^b Molecular weight of the target domain excluding tags

^c Template preparation conditions used for protein synthesis were under either the '+DMSO' or the 'Std' conditions

^d Productivity of the protein by cell-free protein synthesis, which was determined from band size and density on PAGE. A unit is defined as mg BSA equivalent per mL of cell-free protein reaction mixture. The standard deviations for the values were estimated to be 20% + 0.1 mg/ml

^e Solubility of the synthesized protein, which was determined from PAGE, was >70% (+++), >40% (++), >10% (+), ≤10% (-) or not determined (ND)

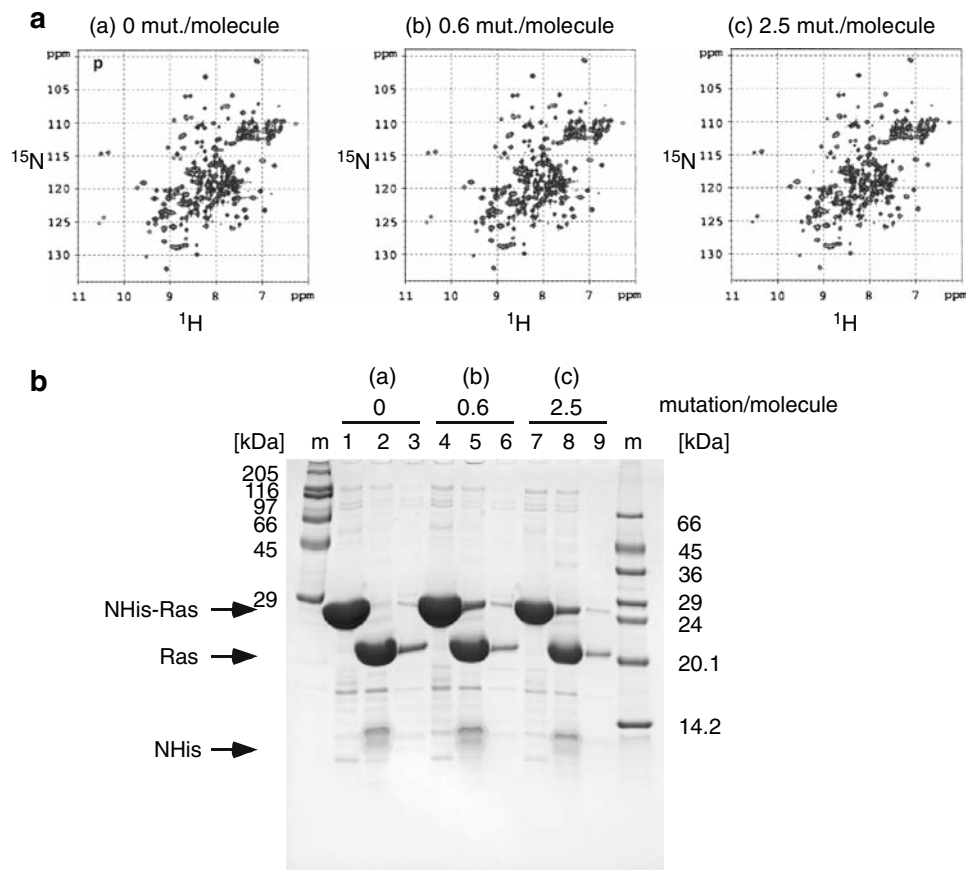


Fig. 7 Effects of the PCR error rate on protein properties. (a) ¹H-¹⁵N-HSQC spectra for the mixtures of the c-Ha-Ras protein and the cleaved NHis tag from an error-free vector (a, 0 mutation/molecule), and linear templates with intentionally-increased error rates of 9.2×10^{-4} (b, 0.6 mutation/molecule) and 3.9×10^{-3} (c, 2.5 mutation/molecule). The error rates of templates (b) and (c) were 1.1- and 5-times higher than that of the two-step PCR product. (b) SDS-PAGE of the proteins from

an error-free vector template (a), and linear templates with error rates of 9.2×10^{-4} (b) and 3.9×10^{-3} (c). The units of the error rates are mutation/bp. Proteins before cleavage by TEV (lanes 1, 4 and 7), supernatant after cleavage by TEV (lanes 2, 5 and 8) and precipitate after cleavage (lanes 3, 6 and 9). Arrows represent the positions of the NHis-tagged Ras protein (24.5 kDa), the Ras protein with the tag removed (20.5 kDa), and the NHis tag (4 kDa)

be necessary to incorporate the linear template into a vector, and after the sequence is confirmed, the vector may be used as the template for (cell-free) protein synthesis.

Simulation of effects of the template error rate on the HSQC spectrum

We also simulated the effects of errors at different rates in the DNA templates on the heights of the cross peaks, observed in the ¹H-¹⁵N-HSQC spectrum. When a point mutation is introduced into a protein, the chemical shifts of the peaks of amino acid residues near the mutated one may change, while the chemical shifts of the other peaks would be the same as those of the wild-type protein in most cases. Therefore, in the HSQC spectrum for the ensemble of a random point mutant protein pool, the shifted peaks of each mutant would be too small to detect,

because the cross-peak shifts are different among the mutants, and the other peaks remain at the same positions. As a consequence, the same HSQC spectrum as that of the wild-type protein is observed with somewhat weaker intensities, even for the mutant protein pool prepared from DNA templates with an increased error rate. We have roughly simulated the relative peak height (*H*) of the peaks observed at positions for the wild-type protein in the ¹H-¹⁵N-HSQC spectra for given error rates on templates (*E*), which corresponded to the spectra in Fig. 7a. As a result, the relative heights (*H*) were 93% and 71% for $E = 9.2 \times 10^{-4}$ (0.6 nucleotide mutation/molecule) and 3.9×10^{-3} (2.5 nucleotide mutation/molecule), respectively (Supplemental Table 3). An HSQC spectrum, with a pattern that is almost the same as that of the wild-type protein was predicted to be observed, even for the increased error rate (2.5 nucleotide mutation/molecule). These simulation results agree with the experimental results in Fig. 7a.

High-fidelity and fast two-step PCR

In order to reduce the error rate of the two-step PCR, we developed the 'HF' conditions of the two-step PCR, using a high-fidelity and fast-processing PCR enzyme, iProof DNA polymerase. In addition to the use of the high fidelity polymerase, the purified cDNA vector, instead of the culture, was used as the template in the first PCR, and the number of PCR cycles was reduced. The iProof polymerase has more than 17-times higher fidelity than the Expand-Hi-Fi enzyme mix, based on their characteristics value listed in their catalogs. For NHis-Ras, the total error rate in the two-step PCR product prepared under the HF conditions was 3.0×10^{-4} (0.2 mutation/molecule), which was 40% of that of the 'Std' condition, 7.6×10^{-4} (0.5 mutation/molecule). In more detail, the error rates in the fragments (T7P and T7T fragments, 409 bp), the unique primers (FW and RV primers for the first PCR, 71 bp) and the target domain regions (476 bp) were 5.3×10^{-4} , 9.6×10^{-4} , and 0 (no mutation in 41,888-bp total read, or $<2.4 \times 10^{-5}$), respectively, under the 'HF' conditions, and 1.0×10^{-3} , 1.3×10^{-3} and 4.8×10^{-4} , respectively, under the 'Std' conditions. The relatively high error rates within the primers are due to errors in the chemical synthesis, but not in the PCR amplification, indicating that the use of higher quality primers may lower the error rates in these regions. It is notable that the error rate in the target region under the 'HF' conditions was reduced to one-tenth or less of that of in the 'Std' conditions. This indicates that the total mutation content under the 'HF' conditions would remain low, regardless of the length of the target region. For example, the total mutant contents of a 1000-residue target protein were estimated to be 0.3 and 1.9 mutations per molecule under the 'HF' and 'Std' PCR conditions, respectively. It should be emphasized that the error rate in the linear template DNA prepared by two-step PCR under the 'HF' conditions is low enough to synthesize a 1000 amino acid residue protein in cell-free protein synthesis. The errors in the target regions could be further reduced by optimizing the preparation protocol. Another feature of the iProof polymerase is its high processing speed. The time required for the two-step PCR under the 'HF' conditions was about 2 h, which is about one-third of that under the 'Std' conditions. This feature would improve the throughput of processing with a number of samples.

Discussion

We have developed a robust two-step PCR method for generating linear templates, which can express the target regions of coding sequences with N- and C-terminal tags, for cell-free protein synthesis. This method is applicable to

targets with various GC contents (37–75%) or GC contents with a 150-bp window scan (45–84%) and with different tags, by few modifications of the PCR conditions. The main feature of the two-step PCR is (1) its simple protocol and (2) its high tolerance to fluctuations in experimental conditions and concentrations of materials. These features are essential for automation, in order to process many samples in parallel.

It is important to pay attention to the error rate of the two-step PCR. In the 'Std' PCR condition, the mutation content per molecule of the synthesized protein was at most 0.5 for a 200-residue protein and would be 1.9 for a 1000-residue protein, as estimated from the error rates of the two-step PCR product. The mutation was randomly incorporated into each molecule. Therefore, this mutant content did not disturb our HSQC protein-folding assay, which was experimentally determined and was also confirmed by the simulation. The mutation content could be reduced by a high-fidelity DNA polymerase. In the 'HF' PCR conditions, the mutation content was 0.2 for a 200-residue protein and would be 0.3 for a 1000-residue protein. This mutation content would not disturb most biochemical assays. However, for experiments that may require even higher homogeneity, e.g., for crystallization of proteins and single molecule assays, it may be necessary to incorporate the linear template into a vector, and after the sequence is confirmed, the vector may be used as the template for (cell-free) protein synthesis. Even in these cases, the two-step PCR is valuable for preliminary screening, especially for a large number of constructs.

By combining two-step PCR and cell-free protein synthesis, robust high-throughput protein synthesis using a robotic system for practical use can be achieved. We employed this system to select suitable protein samples for structural analysis by NMR in a project of the RIKEN Structural Genomics/Proteomics Initiative (RSGI), and generated more than 200,000 constructs for comprehensive screening of targets. The system can also be applied to functional analysis.

Another feature of the two-step PCR is its extensibility of construction. Tags can be replaced by simply switching fragments at the second PCR step; therefore, it is easy to screen suitable tags for the target and its application. Fragments with SBP or epitope tags, e.g., FLAG, Myc, or hemagglutinin (HA), may be useful for protein stabilization onto surfaces for binding assays by surface plasmon resonance, protein arrays [26] or bead displays [27], and could be used for producing a detection probe with antibodies. Fragments with the green fluorescent protein (GFP) will be useful as direct fluorescence detection probes or for solubility screening [28]. If mutagenic PCR [22] is performed during the first PCR step, then an expression template suitable for selection experiments may be obtained by the

two-step PCR. Moreover, this method may be applicable to produce expression templates for other systems than the *E. coli* cell-free systems, by simply changing the expression regulation elements in the fragments, e.g., the promoter, the ribosome binding site and the enhancer, to the appropriate sequences.

Applications of the two-step PCR are not restricted to the construction of protein expression constructs. Eventually, this method can robustly concatenate a dsDNA fragment and two other dsDNA fragments, and thus the two-step PCR method may serve as a useful DNA construction tool for many purposes.

Acknowledgements The authors thank Dr. Daisuke Kiga (Tokyo Institute of Technology) and Dr. Yasuaki Kawarasaki (University of Shizuoka) for helpful discussions; Dr. Takayoshi Matsuda (RIKEN) for providing HSQC spectra of Ras mutants; Natsuko Matsuda, Natsumi Suzuki, and Yikkiko Fujikura for their technical assistance; and Tomoko Nakayama and Azusa Ishii for expert secretarial assistance. This work was supported by the RIKEN Structural Genomics/Proteomics Initiative (RSGI), the National Project on Protein Structural and Functional Analysis, Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

- Nameki N, Tochio N, Koshiha S, Inoue M, Yabuki T, Aoki M, Seki E, Matsuda T, Fujikura Y, Saito M, Ikari M, Watanabe M, Terada T, Shirouzu M, Yoshida M, Hirota H, Tanaka A, Hayashizaki Y, Guntert P, Kigawa T, Yokoyama S (2005) *Protein Sci* 14:756–764
- Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T, Yabuki T, Aoki M, Seki E, Matsuda T, Tomo Y, Hayami N, Terada T, Shirouzu M, Tanaka A, Seki M, Shinozaki K, Yokoyama S (2005) *Plant Cell* 17:944–956
- Li H, Inoue M, Yabuki T, Aoki M, Seki E, Matsuda T, Nunokawa E, Motoda Y, Kobayashi A, Terada T, Shirouzu M, Koshiha S, Lin YJ, Guntert P, Suzuki H, Hayashizaki Y, Kigawa T, Yokoyama S (2005) *J Biomol NMR* 32:329–334
- Yabuki T, Kigawa T, Dohmae N, Takio K, Terada T, Ito Y, Laue ED, Cooper JA, Kainosho M, Yokoyama S (1998) *J Biomol NMR* 11:295–306
- Kurimoto K, Muto Y, Obayashi N, Terada T, Shirouzu M, Yabuki T, Aoki M, Seki E, Matsuda T, Kigawa T, Okumura H, Tanaka A, Shibata N, Kashikawa M, Agata K, Yokoyama S (2005) *J Struct Biol* 150:58–68
- Arai R, Kukimoto-Niino M, Uda-Tochio H, Morita S, Uchikubo-Kamo T, Akasaka R, Etou Y, Hayashizaki Y, Kigawa T, Terada T, Shirouzu M, Yokoyama S (2005) *Protein Sci* 14:1888–1893
- Kigawa T, Yamaguchi-Nunokawa E, Kodama K, Matsuda T, Yabuki T, Matsuda N, Ishitani R, Nureki O, Yokoyama S (2002) *J Struct Funct Genomics* 2:29–35
- Kang SH, Kim DM, Kim HJ, Jun SY, Lee KY, Kim HJ (2005) *Biotechnol Prog* 21:1412–1419
- Ryabova LA, Desplancq D, Spirin AS, Pluckthun A (1997) *Nat Biotechnol* 15:79–84
- Klammt C, Schwarz D, Lohr F, Schneider B, Dotsch V, Bernhard F (2006) *FEBS J* 273:4141–4153
- Yokoyama S, Matsuo Y, Hirota H, Kigawa T, Shirouzu M, Kuroda Y, Kurumizaka H, Kawaguchi S, Ito Y, Shibata T, Kainosho M, Nishimura Y, Inoue Y, Kuramitsu S (2000) *Prog Biophys Mol Biol* 73:363–376
- Endo Y, Sawasaki T (2004) *J Struct Funct Genomics* 5:45–57
- Sawasaki T, Ogasawara T, Morishita R, Endo Y (2002) *Proc Natl Acad Sci USA* 99:14652–14657
- Woodrow KA, Airen IO, Swartz JR (2006) *J Proteome Res* 5:3288–3300
- Kigawa T, Muto Y, Yokoyama S (1995) *J Biomol NMR* 6:129–134
- Nakai K, Horton P (1999) *Trends Biochem Sci* 24:34–36
- Dougherty WG, Cary SM, Parks TD (1989) *Virology* 171:356–364
- Chaga G, Bochkariov DE, Jokhadze GG, Hopp J, Nelson P (1999) *J Chromatogr A* 864:247–256
- Hirota Y, Katsumata A, Takeya T (1990) *Nucleic Acids Res* 18:6432
- Keefe AD, Wilson DS, Seelig B, Szostak JW (2001) *Protein Expr Purif* 23:440–446
- Matsuda T, Kigawa T, Koshiha S, Inoue M, Aoki M, Yamasaki K, Seki M, Shinozaki K, Yokoyama S (2006) *J Struct Funct Genomics* 7:93–100
- Lin-Goerke JL, Robbins DJ, Burczak JD (1997) *Biotechniques* 23:409–412
- Kigawa T, Yabuki T, Yoshida Y, Tsutsui M, Ito Y, Shibata T, Yokoyama S (1999) *FEBS Lett* 442:15–19
- Nakano H, Kobayashi K, Ohuchi S, Sekiguchi S, Yamane T (2000) *J Biosci Bioeng* 90:456–458
- Esposito D, Chatterjee DK (2006) *Curr Opin Biotechnol* 17:353–358
- Labaer J, Ramachandran N (2005) *Curr Opin Chem Biol* 9:14–19
- Griffiths AD, Tawfik DS (2006) *Trends Biotechnol* 24:395–402
- Waldo GS, Standish BM, Berendzen J, Terwilliger TC (1999) *Nat Biotechnol* 17:691–695