In vitro transcription–translation using bacterial genome as a template to reconstitute intracellular profile

Kei Fujiwara[1],[*],[†], Tsunehito Sawamura[1],[†], Tatsuya Niwa[2], Tatsuki Deyama[1], Shin-ichiro M. Nomura[3], Hideki Taguchi[2] and Nobuhide Doi[1]

[1]Department of Biosciences and Informatics, Faculty of Science and Technology, Keio University, Yokohama 223-8522, Japan, [2]Cell Biology Center, Institute of Innovative Research, Tokyo Institute of Technology, Yokohama 226-8503, Japan and [3]Department of Robotics, School of Engineering, Tohoku University, Sendai 980-8579, Japan

ABSTRACT

***In vitro* transcription–translation systems (TX–TL) can synthesize most of individual genes encoded in genomes by using strong promoters and translation initiation sequences. This fact raises a possibility that TX–TL using genome as a template can reconstitute the profile of RNA and proteins in living cells. By using cell extracts and genome prepared from different organisms, here we developed a system for *in vitro* genome transcription–translation (iGeTT) using bacterial genome and cell extracts, and surveyed *de novo* synthesis of RNA and proteins. Two-dimensional electrophoresis and nano LC–MS/MS showed that proteins were actually expressed by iGeTT. Quantitation of transcription levels of 50 genes for intracellular homeostasis revealed that the levels of RNA synthesis by iGeTT are highly correlated with those in growth phase cells. Furthermore, activity of iGeTT was influenced by transcription derived from genome structure and gene location in genome. These results suggest that intracellular profiles and characters of genome can be emulated by TX–TL using genome as a template.**

INTRODUCTION

Construction of biological systems *in vitro* is a remarkable subject in synthetic biology (1–3). Studies aiming at this subject have elucidated essential factors and detailed mechanism for biological systems, and contributed to illuminating a road toward building an artificial cell behaving like living cells (1,3–5). Up to date, several biological systems have been reconstructed in test tubes. A central metabolism, glycolysis, is the first biological system reconstituted *in vitro*

nearly 100 years ago (6). To date, systems for transcription, translation and DNA replication from more than 10 factors has been successfully reconstituted *in vitro* (7–11). Very recently, fatty acid synthesis coupled with glycolysis from 30–40 proteins was also achieved by defined factors *in vitro* (12). These successes indicate the possibility that reconstitution of living cells from a mixture of biological factors may be achieved in near future.

To achieve the reconstitution of living cells, mimicking systems of living cells *in vitro* is unavoidable. In this point, *in vitro* transcription–translation system (TX–TL) is the best platform for mimicking living cells. TX–TL is a method that synthesizes proteins from DNA or mRNA in test tubes like the central dogma of molecular biology (13). Previous researches have shown that TX–TL using multiple DNA as templates is able to reconstitute many biological systems. For examples, RNA polymerase (14), transcriptional cascades (15), a simplified DNA replication system (16), lipid synthesis (17,18) and translocon (19) are able to be reconstituted by TX–TL. Moreover, TX–TL using genomic DNA of virus and phages as the template is able to produces φx174 phage, T7 phage and encephalomyocarditis virus (20,21).

These successes of bio-system reconstitution by TX–TL will focus on the next challenge: *in vitro* synthesis of autonomous growth organisms like bacteria from their genomes (3). Toward the challenge, TX–TL using bacteria genomic DNA should be analysed. Due to its large size, genomic DNA is known to show different behaviour from the conventional TX–TL using short DNA. For examples, it has been shown that DNA larger than 100 kb shows specific physicochemical characters (22–24), and topology of genomic DNA influences transcription in living cells (25,26). Thus, the differences and similarities between TX–TL using plasmid and genomic DNA should be examined.

---

*To whom correspondence should be addressed. Tel: +81 45 566 1533; Email: fujiwara@bio.keio.ac.jp
†These authors are contributed equal to this work as first authors.

In this paper, we report *in vitro* genomic DNA transcription–translation system (iGeTT), which uses bacterial genome as a template for TX–TL. To identify products from iGeTT, we used a heterologous iGeTT that uses cell extract of *Escherichia coli* as transcription–translation machinery and genomic DNA of *Thermus thermophilus* HB27. Proteome analysis confirmed iGeTT expresses proteins encoded in the genome. Quantitative analysis revealed that transcriptional levels by iGeTT are highly correlated with those in growth phase of *T. thermophilus* cells and are influenced by genome structure and gene locus in genome. These results indicate that iGeTT is able to recapitulate intracellular profile of genome, and simultaneously, indicate the possibility of *in vitro* reconstitution of living cells from genome and cell extract.

MATERIALS AND METHODS

Materials

Chemicals used in this study were mainly purchased from Nacalai tesque (Kyoto, Japan). Creatine kinase and lysozyme chloride was purchased from Nacalai tesque. Folinic acid and CTP were purchased from Tokyo Chemical Industry (Tokyo, Japan). Potassium glutamate and magnesium glutamate were purchased from Sigma-Aldrich (St. Louis, MO, USA). tRNA, creatine kinase, and Proteinase K were purchased from Roche (Basel, Switzerland). ATP, GTP and chemicals for LC–MS/MS were purchased from Wako Chemical (Osaka, Japan). CTP was purchased from Affymetrix (Santa Clara, CA, USA).

Plasmids

A control plasmid, pOR2OR1-sfGFP was constructed in the previous study (33,34). For the compatibility assay, three promoter-SD sequences sets (TTC1532, TTC0834, TTC1172) were amplified from *Thermus thermophilus* HB27 genome (Tth genome) by PCR. Then, the amplified fragment was fused with sfGFP gene by overlap PCR. The PCR product was cloned into pUC19 by Gibson Assembly mix (New England Biolabs, Ipswich, MA, USA). The resultant plasmids were purified from transformants of *E. coli* HST08 (TAKARA, Shiga, Japan) by Mini prep (Qiagen, Venlo, Netherlands). DNA sequences were verified by a Sanger DNA sequencer (ABI PRISM 3100 Genetic Analyzer, Applied Biosystems, Foster City, CA, USA).

For genome manipulation, each 600 bp of upstream and downstream at insertion sites was amplified from Tth genomic DNA by PCR. $P_{TTC1532}$-sfGFP was amplified from the plasmid described above. These three fragments were fused by two rounds overlap PCR. The amplified fragment was cloned into pUC19 by Gibson Assembly mix. The resultant plasmids were purified by the same way as above. After verification of DNA sequence, whole plasmid was amplified by PrimeSTAR max to fuse with thermostable kanamycin resistant gene amplified from pBGL73 (58). The PCR products were connected by Gibson Assembly. After the transformation and selection by kanamycin, plasmids were purified as the same way.

Calculation of molar concentration of DNA

The length of Tth genome and pOR2OR1-sfGFP are 1 894 877 bp and 3610 bp, respectively. To calculate molar concentrations, molecular masses were calculated from the length of DNA (1.9 Mbp for Tth genome and 3.6 kb for pOR2OR1-sfGFP) and by counting 1 bp = 660 Da. DNA concentrations (weight per volume) were determined by OD_{260}.

Preparation of cell extracts for TX–TL system

The protocol for cell extract preparation from *E. coli* BL21-CodonPlus(DE3)-RIPL (Agilent Technologies, Santa Clara, CA, USA) was according to the previous report (34). Briefly, *E. coli* was cultivated in LB medium and harvested by centrifugation at $OD_{600} = 1.0$. After washing with 400 mM sucrose, 400 μg/ml lysozyme chloride at final was added in the suspension, and incubated on ice for 1 h. Then, the cells were washed with 400 mM sucrose twice, and were suspended in 1.5 times volume of S30A buffer [10 mM Tris–HCl (pH 7.6), 60 mM potassium glutamate, 14 mM magnesium glutamate, 2 mM DTT]. The cells were frozen by liquid nitrogen for 15 min and thawed in ice water. Supernatant after centrifugation at 30 000 g for 1 h was obtained as EcoS30. EcoS30 was washed with 10 times volume of S30B buffer [5 mM Tris–HCl (pH 8.0), 60 mM potassium glutamate, 14 mM magnesium glutamate, 1 mM DTT] by Amicon Ultra-15 Centrifugal Filter Units 10 kDa (Merck Millipore, Billerica, MA, USA). By the washing, typically final volume of EcoS30 became one-fourth of before washing, and final protein concentrations were 60–75 mg/ml. EcoS30 was stocked in -30°C.

TX–TL reactions

For TX–TL reactions, the mixture containing 2.0 mM ATP, 2.0 mM GTP, 1.3 mM UTP, 1.3 mM CTP, 0.6 mM each amino acids, 1 mM spermidine, 0.2 mg/ml tRNA, 100 mM potassium glutamate, 10 mM magnesium glutamate, 50 mM Hepes–KOH (pH 8.0), 68 μM folinic acid, 80 mM creatine phosphate, 250 μg/ml creatine kinase (CK) at final was mixed with EcoS30. Final concentration of EcoS30 was set at 15 mg/ml. Genome DNA or plasmid DNA was added at various concentrations as indicated. Reaction temperature was usually at 30°C for protein expression. For maximizing transcription levels, transcription reactions were performed at 25°C for 3 h using a modified reaction mixture in which spermidine was omitted and 5.6 mM magnesium glutamate was used. For dialysis mode, TX–TL was performed using slide-A-Lyzer MINI Dialysis units (10 000 MWCO, Thermo Fisher Scientific, Waltham, MA, USA). Thirty micro liter of reaction mixtures were incubated at 30°C with dialysis against 400 μl of outer solution in which EcoS30, template DNA, CK were removed from the reaction mixture. During the dialysis mode, feeding solution was stirred for prompting diffusion of molecules.

To detect sfGFP synthesized by TX–TL, the reaction mixtures were mixed with equal amount of sample buffer [0.125 M Tris–HCl (pH 8.0), 4% sodium lauryl sulfate (SDS), 20% glycerol, 0.01% bromophenol blue] and examined by 12.5% SDS-polyacrylamide gel without boiling

prior to loading. The fluorescence of sfGFP was detected by Molecular Imager FX (Bio-Rad).

Silver staining and 2D electrophoresis

After heat treatment at 70°C for 3 h, supernatant of the reaction mixture of TX–TL system after centrifuged was collected, and pelleted by TCA/acetone. The pellet was examined by isoelectric focusing as following manufacture's instruction using Multiphor II Electrophoresis Unit (GE Healthcare, Chicago, IL, USA). Then, the gel was washed with distilled water and moved to equilibration buffer [50 mM Tris–HCl (pH 8.0), 6 M urea, 30% glycerol, 2% SDS]. After shaking for 30 min, the gel was put on 10% polyacrylamide gel and sealed by agarose. Proteins were separated by electrophoresis at 30 mA for 1 h, proteins were stained by Silver Stain II kit (Wako Chemical).

Western blotting

In the case of TX–TL system using sfGFP gene or Tth genome with sfGFP gene insertion, the solution after reactions were treated with DNase, and then were examined by SDS-PAGE, and blotted on PVDF membranes. Blotting was performed with Anti-GFP pAb (Thermo Fisher) and HRP-labelled anti-rabbit IgG (Cell Signaling Technology Japan, Tokyo, Japan). Signals from HRP activity was detected by a detection reagent with ultra-high sensitivity (Chemi-Lumi One Ultra, Nacalai) using C-DiGit Blot Scaner (LI-COR Wetzlar, Germany). The standard detection reagent in the manuscript was Chemi-Lumi One Super (Nacalai).

Proteome analysis by nano LC–MS/MS

Proteomics using nano LC–MS/MS was performed according to a previous report (59). Proteins synthesized by iGeTT were precipitated by acetone and re-suspended in 28.5 μl of 8 M urea in 100 mM NH_4HCO_3. Then, the solution was reduced with 1.5 μl of 100 mM NH_4HCO_3 with 1 mM DTT at 25°C for 30 min, and alkylated with 10 μl of 220 mM Iodoacetamide in 100 mM NH_4HCO_3 at 25°C for 30 min. After the reductive alkylation, the solution was incubated with 0.5 μg of Wako Lys-C at 25°C for 3 h and diluted with 100 mM NH_4HCO_3 to 180 μl. Proteins were further digested with 1 μg of Trypsin Gold (Promega, Madison, WI, USA) at 37°C for overnight. After trypsin digestion, salt and organic solvent in the solution were removed by GL-Tip SDB (GL science, Tokyo, Japan) and the solution was concentrated to 20 μl by a centrifugal evaporator (TOMY SEIKO, Japan). This solution was diluted with 0.3% trifluoroacetic acid to 30 μl and put in Q-Exactive & EasynLC 1000 (Thermo fisher) equipped with a 12.5 cm $\times$ 75 μm C-18 separation column (Nikkyo Technos, Japan). Peptides in the solutions were analyzed by Proteome Discoverer (Thermo fisher). Below 1% of false discovery rate and observation of more than two unique peptides were criteria that we used to determine as identified proteins. All proteins listed were identified in more than two runs of nano LC–MS/MS.

Purification of genomic DNA by agarose-trap method

Genome DNA in cells trapped by agarose was purified based on a previous report (35). Cells at growth phase (OD_{600} = 0.6–1.0) and at stationary phase were collected by centrifugation. The pellet was washed with 10 ml NT buffer [5 mM Tris–HCl (pH 7.6), 1 M NaCl], and resuspended with 0.75 ml NT buffer. The solution was incubated at 4°C for 1 h after addition of 0.75 ml of 1% low melting agarose. The agarose plug was immersed in EC buffer [5 mM Tris–HCl (pH 7.6), 1M NaCl, 0.1 mM EDTA, 0.5% Brij58, 0.1% deoxycholate, 0.5% SDS] containing 1 mg/ml lysozyme chloride, and incubated at 37°C for 80 rpm shaking for overnight. Then, the agarose plug was transferred into ES buffer (0.5 M EDTA, 1% SDS) containing 0.1 mg/ml Proteinase K, and incubated at 37°C under 80 rpm shaking for overnight. The agarose plug was washed with TE buffer [10 mM Tris–HCl (pH 8.0), 1 mM EDTA], and then transferred into TE buffer containing 1 mM Phenylmethylsulfonyl fluoride and incubated at 43°C for 2 h under 80 rpm shaking. After two sets of this washing step, agarose plug was moved to TE buffer and incubated at 25°C for 2 h under 80 rpm shaking. After 4 sets of this washing treatment, agarose gel plug in TE buffer was stored at 4°C.

Purification of genomic DNA by CTAB method

Cells at OD_{600} = 1.0 were collected by centrifugation. The pellet was suspended with 500 μl TE buffer, and then, 20 μl of 10% SDS and 1 μl of Proteinase K (Roche) was added in the suspension. After 10 min incubation at 37°C, 90 μl of 5 M NaCl and 70 μl of 10% cetyltrimethylammonium bromide (CTAB) and 0.7 M NaCl were added in the solution. After 10 min incubation at 60°C, genomic DNA was purified by phenol/chloroform treatment and ethanol precipitation.

Quantification of transcriptional levels by RT-qPCR

In the case of TX–TL, total RNA was purified from the TX–TL mixtures after 3 h reactions at 25°C by using RNeasy kit or acid guanidinium thiocyanate-phenol-chloroform extraction (AGPC) method. AGPC method was as follows. TX–TL reaction mixtures were mixed with 300 μl solution D (6 M Urea, 25 mM sodium citrate, 1% SDS, 0.1 M 2-melcaptoethanol) and 10 μl of 1.5 M sodium acetate (pH 4.0). Then, the solution was mixed with 400 μl phenol/chloroform by vortex. After incubation at 4°C for 10 min, the solution was centrifuged and the upper layer was collected. After this treatment, total RNA was precipitated by ethanol. In the case of living cells, *T. thermophilus* cells at OD_{600} = 0.6 or at stationary growth were harvested by centrifugation. The cell pellet was suspended with buffer RLT in RNeasy kit (Qiagen) and was disrupted with microbeads (GB-01, Tomy, Tokyo, Japan) using Mini-BeadBeater (Biospec Bartlesville, OK, USA). After the disruption, total RNA was purified by using RNeasy kit (Qiagen).

All steps of RT-qPCR were followed as manufactures instructions of One Step SYBR PrimeScript RT-PCR kit (TAKARA) and Light cycler (Roche). Briefly, total RNA and primers (0.2 μM each at final) were mixed with the kit components (thermostable DNA polymerase, reverse

transcriptase, and buffer containing SYBR Green I, dNTP Mixture, and magnesium ion). This mixture was set in LightCycler® Capillaries (Roche), and analysed by the following protocol in Light cycler. The protocol for temperature shifts was 5 min at 42°C, 10 s at 95°C (for reverse transcription and denaturation of reverse transcriptase), a 40 cycles of 5 s at 95°C and 20 s at 60°C (for PCR), and 0 s at 95°C, 15 s at 65°C, and gradual increase to 95°C (0.1°C/s for obtaining melting curves). Heating and cooling were set at 20°C/s except for the case of obtaining melting curves. Ct values were calculated from the amplification curves by second derivative maximum method using the software included in Light Cycler. Concentrations of target RNAs were calculated from Ct values using standard curves obtained from Ct values of DNA with corresponding sequences at known concentrations. For obtaining relative concentrations, concentration was normalized by that of 16S rRNA derived from EcoS30 in total RNA. As a negative control, amino acids and NTPs were omitted from the reaction mixtures, and total RNA in them were purified just after the preparation. We verified reaction samples showed significantly higher levels of RT-qPCR signals than negative controls in every case.

For the absolute quantification, standard mRNAs were synthesized from DNA using T7 RNA polymerase. Template DNAs for the absolute quantification were prepared by fusing corresponding partial sequences of genes from genomic DNA with T7 promoter (TAATACGACTCACT ATAGGGAGA) by PCR. RNA was transcribed from PCR products by T7 RiboMAX Express Large Scale RNA Production System (Promega).

Modification of Tth genome by homologous recombination

General Protocol for Tth genome modification was as follows in previous studies (42,43). Five hundred microliter of cell culture at $OD_{600} = 0.4$ was mixed with 1 μl of the cassette DNA with 0.33 mM $CaCl_2$ and 0.4 mM $MgCl_2$. The cell mixture was cultivated with 180 rpm for 2 h at 70°C, and then, was spread on TY agarose containing 500 μg/ml kanamycin. Recombination of genomic DNA was checked by PCR using purified genome prepared by CTAB method.

RESULTS

Compatibility between cell extract from *Escherichia coli* and genome of *Thermus thermophilus*

In our system, *E. coli* cell extract (EcoS30) was used for the transcription–translation machinery and genome DNA of *Thermus thermophilus* (Tth genome, 1.9 Mbp) was used as the template. The merit of *E. coli* cell extract system is its high productivity of protein expression and its versatile usage (15,27). The reason for using Tth genome was because proteins of *T. thermophilus* are able to be separated from components in the *E. coli* cell extract by heat treatment, and a wide range of information on this organism has been accumulated from various omics studies (28–30).

First, compatibility of transcription and translation between EcoS30 and Tth genome was quantitatively analysed. Similarities of consensus sequences for primary sigma factor σ^{70} (31) and Shine-Dalgarno (SD) sequences (32) sug-
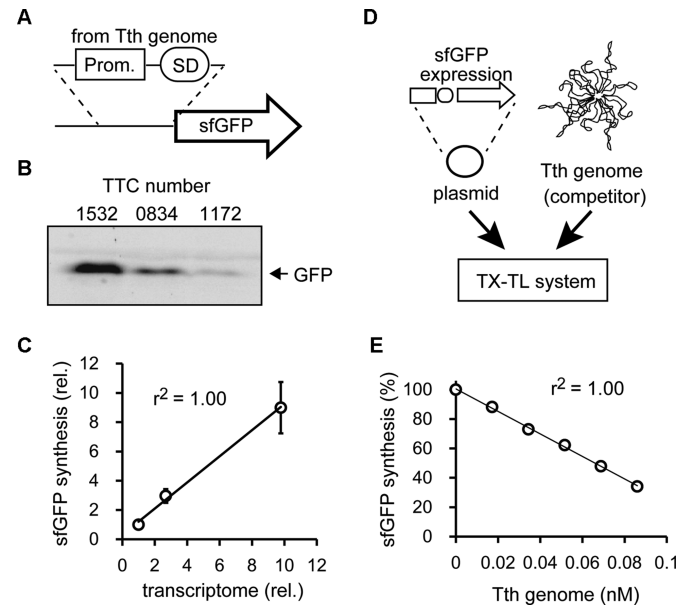

Figure 1. Compatibility of transcription and translation between EcoS30 and Tth genome. (**A**) Plasmid harbouring sets of promoter and SD sequence of three *Thermus* genes (TTC1532, TTC0834, TTC1172) fused with upstream of the reporter, sfGFP gene, were constructed for the compatibility assay. (**B**) GFP synthesized by EcoS30 and the reporter plasmids were detected from fluorescent signals after non-boil SDS-PAGE. (**C**) Levels of GFP synthesis by EcoS30 were plotted as a function of the expression levels in living cells of the corresponding genes. The levels of expression were referred to transcriptome analysis in a previous report (41). Synthesis levels of sfGFP and expression levels were both normalized by its value of TTC1172 as 1 (rel.). Error bars indicate standard deviations from three independent experiments. (**D**) Representative illustration of the competition assay. Tth genome was supplied as competitor to 0.5 nM OR2OR1-sfGFP plasmid. (**E**) Competition of TX–TL reaction between the OR2OR1-sfGFP plasmid and Tth genome. The expression level of sfGFP from the OR2OR1-sfGFP plasmid in the absence of Tth genome was determined as 100%.

gested the compatibility of promoters and SD sequences of housekeeping genes among bacteria. Thus, we expected that EcoS30 expresses genes under σ^{70} and SD sequence of Tth genome as similar levels of those in living cells. To examine this possibility, upstream region of three genes with different expression levels in *T. thermophilus* cells were fused at the upstream of a reporter gene, sfGFP gene (Figure 1A–C). This transplantation of promoter and SD sequences showed that the levels of sfGFP synthesis by EcoS30 significantly correlated with their expression levels in *T. thermophilus* cells ($P < 0.05$), and indicated that EcoS30 and Tth genome are quantitatively compatible at least on transcription.

The compatibility between EcoS30 and Tth genome was further surveyed by a competition assay (Figure 1D). We used a plasmid DNA as a standard and Tth genome as a competitor. The standard plasmid encodes OR2OR1-sfGFP (33,34), in which a pair of strong promoter and SD sequence for EcoS30 was fused with the upstream of sfGFP. Tth genome was prepared from cells by agarose-trap method (35,36). Because of its intactness, this method has been used for estimating genome sizes by pulse-field electrophoresis (37), and recently applied for genome transplantation among bacteria (38–40). If EcoS30 efficiently tran-
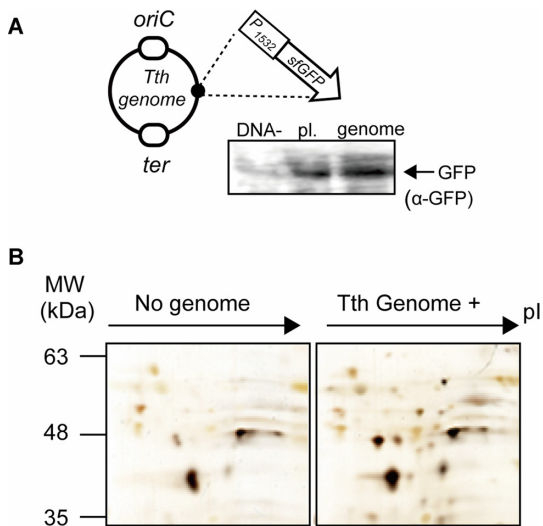
Figure 2. Proteins synthesized by the TX–TL system using EcoS30 and Tth genome. (**A**, top) The reporter sfGFP under promoter and SD sequence of TTC1532 was inserted in the genome of *T. thermophilus* HB27. (bottom) sfGFP proteins detected by western blotting after SDS-PAGE. pl. indicated sfGFP expression from the plasmids encoding the same reporter at 0.3 nM. We should note the proteins of the lane 'pl.' was 10 times diluted to make the signal comparable to the case of sfGFP synthesis from the modified genome. (**B**) Gels after 2D-PAGE of the TX–TL sample using 0.3 nM Tth genome. Proteins were detected by silver staining. Partial area of the gels was shown. Figures for total gels were shown in Supplementary Figure S3.

scribes Tth genome, sfGFP synthesis levels decrease by dose dependent manner because of competition. Actually, mixing various amounts of Tth genome with 0.5 nM of the OR2OR1-sfGFP plasmid showed that Tth genome supplementation linearly decreased the sfGFP expression from the plasmid (Figure 1E). The concentration of the Tth genome for 50% inhibition was 0.066 nM versus 0.5 nM of the plasmid, supporting that EcoS30 actually transcribes the Tth genome. This inhibition by Tth genome was not observed when concentration of the plasmid was increased to 9.6 nM (Supplementary Figure S1), supporting that the competition was not derived from secondary inhibitory effect such as crowding or ion absorption by the large-sized genome.

Protein expression by TX–TL using genome as a template

Next, we showed EcoS30 is able to translate genes encoded in Tth genome. First, translation was detected by western blotting. For the aim, sfGFP fused with the promoter and the SD sequence of TTC1532 was inserted in the genome of *T. thermophilus* HB27 by a conventional *in vivo* method (42,43). A noncoding region at the middle of *oriC* and *ter* of the genome were selected as the location of the gene insertion. The modified genome prepared by the agarose-trap method was added in the TX–TL system using EcoS30. However, we failed to obtain signals of GFP by western blotting when typical detection reagents were used. Then, we changed the detection reagent with ultra-high sensitivity, and observed that EcoS30 synthesized sfGFP protein encoded in the genome (Figure 2A). The expression level is about 4 nM GFP from 0.3 nM Tth genome, which was 10

times lower than that using 0.3 nM of the OR2OR1-sfGFP plasmid (Figure 2A).

To increase the amount of synthesis levels, second, we changed batch reaction to dialysis mode reaction which increases expression levels 10 times higher (13,44). In this mode, energy and materials like amino acids and NTPs are supplied through a 10 kDa cut-off membrane. However, gains of total protein concentrations before and after TX–TL reaction were not significantly high to detect even in the case of the dialysis mode. We therefore concentrated proteins synthesized from Tth genome by heat treatment. Heat treatment was performed by incubating samples at 70°C for 2 h, and supernatant after centrifugation was used for further analysis. By the heat treatment, 95% proteins in EcoS30 were denatured, but almost all of Tth proteins remained (Supplementary Figure S2). After the heat treatment, *E. coli* proteins were reduced to 5% but almost all of Tth proteins were remained, and thus, the treatment raised relative amounts of Tth proteins 20-fold higher to *E. coli* proteins than those without the treatment. After the heat treatment, proteins were further concentrated by TCA precipitation, and then, were analysed by 2D-PAGE. Although remarkable numbers of protein were appeared by the addition of Tth genome, the amounts were relatively low compared with residual proteins from EcoS30 (Figure 2B).

Third, proteins synthesized by the system were identified by nano LC–MS/MS. Recent progresses of nano LC–MS/MS-based proteomics enable us to identify from hundreds to thousands proteins in a solution by a single run (45,46). Actually, ∼800 proteins were identified when cell extracts of *T. thermophilus* HB27 were examined (Supplementary Table S1). By the method, proteins specifically expressed from Tth genome were surveyed. However, most of the identified proteins were derived from EcoS30, and only four proteins encoded in Tth genome were repeatedly identified. To overcome the point, raising temperature at the heat treatment and fractionation by cut-off filters were tested. Although these treatments prior to nano LC–MS/MS did not bring remarkable increase the number of proteins identified, additional four proteins were identified as a product of Tth genome in TX–TL system using EcoS30 (Table 1). These proteins were not identified in the heat-treated negative control sample (mixture of EcoS30 and Tth genome, but lacks NTPs and amino acids). All proteins identified were abundant proteins in *T. thermophilus* cells as suggested by the quantitative compatibility between EcoS30 and Tth genome (Figure 1C).

Quantitative comparison of transcriptional levels by iGeTT and in cells

The results above clearly indicated that proteins encoded in genome are able to be expressed in TX–TL system. Therefore, we termed this system as *in vitro* genome transcription–translation (iGeTT). Next, we tried to the quantitative profile of the products by the iGeTT. However, it was difficult to quantify of the products of translation by the present iGeTT system. Therefore, we analysed, because transcription levels were correlated with translation levels (Figure 1C) (47) and can be specifically quantified even from small amount of samples.

Table 1. Proteins derived from iGeTT identified by nano LC–MS/MS

Gene ID	Function	MW (kDa)	pI	Unique peptides	PEP score	Fractionation conditions prior to MS
TTC1714	60 kDa chaperonin groEL	57.9	5.2	6	35.3	70°C for 2 h[*1]
TTC1330	elongation factor Tu	44.8	5.4	4	43.2	70°C for 2 h[*1]
TTC1172	isocitrate dehydrogenase (NADP)	54.4	6.3	4	17.5	70°C for 2 h[*1]
TTC0666	thiosulfate sulfurtransferase	32.9	5.3	3	8.9	70°C for 2 h[*1]
TTC1754	dihydrolipoamide acetyltransferase	50.2	5.7	5	33.6	Sup. of 100 kDa filter[*2]
TTC1872	Pseudocatalase	33.3	5.5	6	24.1	Sup. of 100 kDa filter[*2]
TTC1776	LSU ribosomal protein L12P (L7/L12)	13.1	5.1	2	13.6	Sup. of 100 kDa filter[*2]
TTC1123	endopeptidase clp ATP-binding chain B, clpB	96.1	5.6	2	8.5	90°C for 2 h[*2]

[*1] and [*2]: Samples were fractionated from different reaction mixtures. Samples marked as the same asterisk number were derived from the same reaction mixture.

We firstly intended to perform RNA-seq analysis using next generation sequencers. However, the levels of TTC1532 mRNA, which is one of the most expressed genes, was 100 000 times smaller than that of rRNA derived from EcoS30. Although the efficiency of commercially available rRNA removal kits enable ∼1000-fold enrichment, we judged RNA-seq is inadequate to analyse expression profile by iGeTT. Thus, we performed RT-qPCR, which show a very high specificity and sensitivity instead of RNA-seq.

By referring the previous transcriptome dataset (41), we chose six genes with various expression levels (from 1 copy to 1000 copies in single cells) from various positions of genome. Their absolute expression levels were quantified by RT-qPCR (Supplementary Table S2). As a result, the expression levels by iGeTT and the amount of intracellular RNA in cells at growth phase showed a very high correlation ($P < 0.001$, Supplementary Figure S4). To widen the analysis, we picked up 50 homeostasis genes which are involved in glycolysis, TCA cycles, chaperones, DNA synthesis, RNA synthesis and protein synthesis. By comparing transcription levels by iGeTT with those in growth phase cells revealed that the transcription profile by iGeTT show high correlation with that in living cells (Figure 3A, Supplementary Table S3).

In our iGeTT system, genomes were prepared from cells at stationary phase, and cell extract was prepared from cells at growth phase. Thus, we investigated whether mRNA levels synthesized by iGeTT is similar to either of the phases or shows no correlation among them. As a result, the correlation of RNA synthesis by iGeTT and in cells became lower when compared with RNA in stationary phase cells (Supplementary Figure S5). However, the lowering was small (correlation coefficients were 0.62 for growth phase, and 0.53 for stationary phase). To more clarify the correlation, mRNA levels of genes for glycolysis pathway, which is the most fundamental pathway in life, were analysed by RT-qPCR. Plotting transcription levels of glycolysis genes by iGeTT as a function of intracellular levels showed considerably high correlation with those in cells at growth phase (Figure 3B, left). Such high correlation was not observed by comparing with mRNA levels at stationary phase (Figure 3B, right). This difference of the correlation among cell phases was remained when potential outliers were omitted from the plot (Supplementary Figure S6).
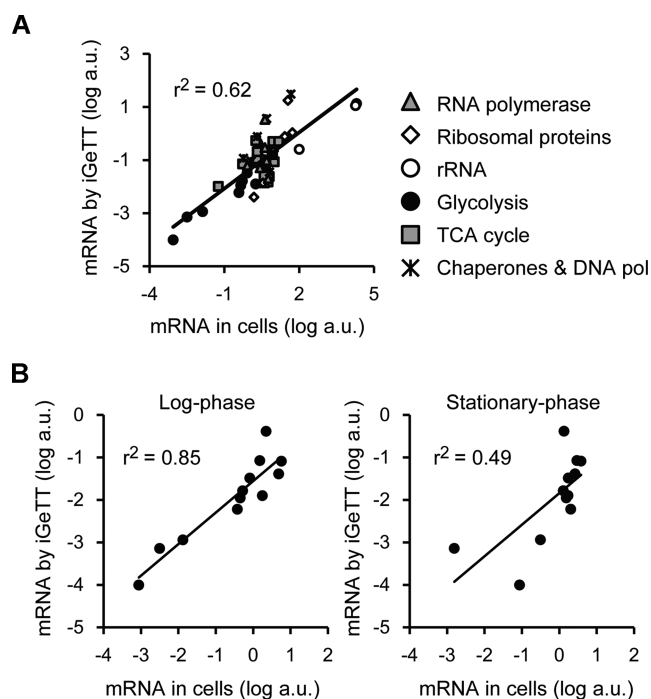

Figure 3. Correlation of transcription levels between by iGeTT and in growth phase cells. (**A**) Transcription levels of 50 homeostasis genes expressed by iGeTT using 0.2 nM Tth genome was plotted as a function of those in growth phase cells. (**B**) Transcription levels of glycolysis genes expressed by iGeTT was plotted as functions of those in growth phase cells (left) and in stationary phase cells (right). a.u.: arbitrary units.

Influences of intactness of genome on transcription by iGeTT

Since iGeTT reflected the transcriptional profile within living cells, we next analyzed dependence of transcriptional levels on states and structures of Tth genome. Large size DNA like genome is very fragile, and thus, is easily degraded by shear forces during purification (48). Genomes prepared by the conventional method are known to be fragmented as 100 kbp (48). Furthermore, genomes in growth phase cells are not necessarily intact, because the lagging strand of genome are synthesized as fragments and ligated together during replication. Fragmentation of genomes could affect transcription rates in iGeTT, and thus, we investigated the effect of genome conditions for iGeTT.
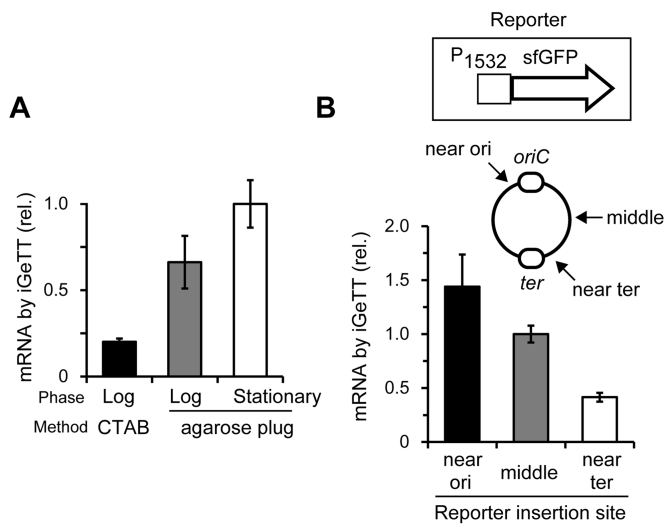
Figure 4. Effect of Intactness of genomes and location of genes on transcription levels of iGeTT. (**A**) Comparison of transcription levels from genomes prepared by two methods and two growth phases. Expression of TTC1532 gene was quantified and the values were normalized by the value of genome prepared from stationary-growth cells by agarose-trap method. (**B**) Expression levels of the same gene construct located in various location of the genome. $P_{TTC1532}$-sfGFP was inserted in the genome at near *oriC*, middle point, and near *ter*. Expression of sfGFP gene was quantified by RT-qPCR (Supplementary Table S4) and the values were normalized by the value of the middle. Genome concentration was 0.2 nM at final. The y-axis (rel.) indicates relative levels of transcription. Error bars indicate standard deviations from three trials.

We had used genomes of stationary-phase cells prepared by the agarose-trap in experiments for Figures 1-3. To investigate the effect of intactness of the genome, two additional genomes was prepared: genomes of growth phase cells at prepared by conventional method, and by the agarose-trap method. Transcription levels of TTC1532 in the genomes estimated by RT-qPCR were compared among the three samples. The expression levels were normalized by the amount of DNA encoding TTC1532 to adjust expression levels per DNA. Surprisingly, iGeTT using genome prepared by the conventional method showed the lowest expression levels, and genomes of stationary-phase cells prepared by the agarose-trap showed the highest (Figure 4A). We should note that transcription levels of TTC1532 by iGeTT were independent on concentration of Tth genome at the ranges for this experiment (0.18–0.33 nM) (Supplementary Figure S7).

Since a bacterial exonuclease, RecBCD, degrades linearized DNA, one may anticipate that degradation of fragmented DNA is the reason of the difference of the expression levels. To check this possibility, the same experiments were performed in the presence of a specific inhibitor of RecBCD, GamS (49). Consequently, supplementation of GamS did not improve the low expression from genomes from the fragmented genomes (Supplementary Figure S8). These results suggested that structure of genome rather than degradation during iGeTT is the reason of the difference of the expression levels.

Influences of gene locus in genome on transcription by iGeTT

The influence of genomic structures suggests that location of gene affect transcription levels. This effect was evaluated by iGeTT using genomes with the same gene cassette $P_{TTC1532}$-sfGFP at distinct genomic positions (Figure 4B). The cassette was inserted near replication origin (*oriC*), near replication terminator (*ter*), and the middle point of those by homologous recombination (Supplementary Figure S9A). Unfortunately, we failed to purify cells with the cassette near *oriC* even after four times single colony isolation (Supplementary Figure S9B). However, the host strain does not have sfGFP gene. Thus, we ignored the contaminant of wild type, and quantified transcription levels of sfGFP by iGeTT using these modified genomes. To avoid the effect of variable copy number of the gene, the levels were normalized by the DNA amount of the corresponding cassette quantified by qPCR. These tests suggest that expression levels were high in the case of at near *oriC*, but low in the case of at near *ter* (Figure 4B). Although leakage of transcription from outside the cassette or other factors can bring the results, it was shown that gene location in genome is an important factor to determine transcription levels by iGeTT.

Effect of genome DNA condensation on iGeTT

Finally, the effect of high-ordered structure on iGeTT was investigated. Previous studies using longer than 100 kb DNA showed that multivalent cations like spermidine and spermine drastically changes DNA structures from a random coil to a packed globule (23). It has been shown that this globule transition switching transcription in the case of T7 phage genome (24). We investigated whether this switching is also observed in the case of bacterial genomes. Spermidine concentration in the iGeTT mixture was set from zero to 2 mM spermidine, and transcriptional levels were quantified by RT-qPCR. In this case, transcription levels from the modified genome with $P_{TTC1532}$-sfGFP were analysed. As a control, the plasmid encoding $P_{TTC1532}$-sfGFP (total plasmid length: 4 kb) was also examined in the same condition. In the case of the plasmid, low concentration of spermidine positively affected on transcription, and high concentration of spermidine negatively affected (Figure 5A, white circles). By contrast, spermidine strongly inhibited iGeTT at the lower concentration (Figure 5A, black circles). These results demonstrated that globule transition by multivalent cation also inhibit transcription of mega-sized bacterial genome.

Spermidine enhances translation levels (50), and therefore, it was considered this positive effect can cancel the negative effect of spermidine. Raising concentration of spermidine in the system increased expression levels of sfGFP in the case of the plasmid (Figure 5B, white circles), but decreased those by iGeTT (Figure 5B, black circles). However, the levels of decrease were gentle in comparison with strong inhibition on transcription. This gentle decrease may be the reason we were able to detect the expression of several proteins by 2D-PAGE and nano LC–MS/MS.
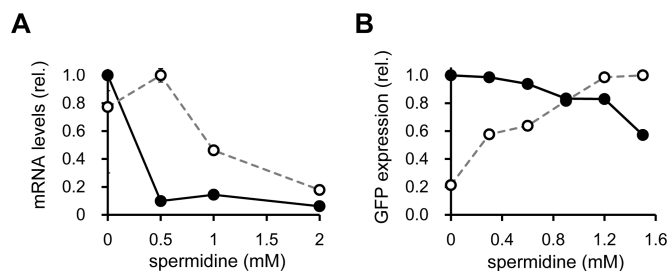
Figure 5. Effect of spermidine on iGeTT and plasmid TX–TL. Spermidine effects on transcription (**A**) and on TX–TL (**B**). Transcription was performed at 30°C for 3 h by using 0.2 nM OR2OR1-sfGFP plasmid or Tth genome as the template. TX–TL was performed at 30°C for overnight by using 0.3 nM OR2OR1-sfGFP plasmid or Tth genome as the template. Each value was normalized by the maximum value of each plot. This means that the value 1 is different between the plasmid TX–TL and iGeTT. Black circles with solid line: iGeTT, white circles with dashed line: plasmid TX–TL. In these experiments, 10.2 mM magnesium glutamate was used. Error bars indicate standard deviations from three trials.

DISCUSSION

TX–TL systems have played important roles in understanding of biological systems such as the importance of metabolic balance (50) and dependence of chaperones on folding (51,52). In this regard, iGeTT was able to recapitulate the regulation of expression levels by genome properties. In living cells, topology of DNA like supercoil structures and distance of genes are important factor to determine transcriptional levels (25,26). In agreement with the facts, genome structures and position of genes influenced transcriptional levels of iGeTT (Figures 4 and 5).

We anticipate that iGeTT is also applied for *in vitro* genome-wide analysis. To date, genome-wide analysis by *in vitro* expression of individual genes under a strong promoter, called as *in vitro* proteome, has revealed the solubility and stability of proteins under various conditions and has identified obligate substrates of chaperones (53–55). However, process on preparing and expressing thousands of genes with strong promoter and less reflection of intracellular expression levels are obstacles to perform the *in vitro* proteome. By contrast, iGeTT do not need gene constructs but reflect intracellular expression profile at least in transcription. Thus, *in vitro* proteome approaches using iGeTT will contribute to elucidating the behaviour of biological components at genome-wide scales. For the purpose, methods to realize specific identification of newly synthesized proteins by iGeTT should be developed.

Despite the remarkable results on recapitulation of the expression profile and the dependency of genome structures, the present iGeTT has several problems. First problem is lack of cell membranes which are not only important for intracellular homeostasis but also are shown to be important for stable expression and avoiding aggregation of membrane proteins (56). However, previous studies have developed methods to entrap TX–TL system in cell-sized liposomes, and have successfully expressed proteins in them (4,15,57). Thus, this membrane problem seems to be solved in near future by developing a method to encapsulate very large genome inside liposome.

Second problem is the heterologous of the present iGeTT in which genome is derived from different organisms from the source for TX–TL system. Only two σ factors for housekeeping (σ^{70}) and heat shock response (σ^{32}) are encoded in *Thermus* genome, and their compatibility among bacteria are a plausible explanation of the success of the present iGeTT. Similarly, a possible reason of the correlation of the transcriptome by the present iGeTT using Eco30 with growth-phase *T. thermophilus* cells is that σ^{70} works as a main sigma factor for promoters in growth-phase and recognizes very similar sequences among bacteria (31). As another possibility, expression of Tth transcription factors can be associated with the reconstitution of transcription profile. Although we did not reach to detect any sign of this kind of automatic formation of gene network by the present iGeTT, we believe that it will be a valuable consideration for future iGeTT. Thus, it is still elusive whether the conclusion obtained in this study is able to be expanded to other iGeTT systems. Expanding the possibility of iGeTT is an indispensable research to clarify the blueprint to create living cells. Especially, analysing iGeTT using TX–TL system and genome from the same organisms will be necessary in order to pave the way to the reconstitution of living cells *in vitro*. Preliminary experiments suggested that homologous iGeTT (using EcoS30 and *E. coli* genome) was able to synthesize remarkable amount of proteins, but also showed that synthesized RNA remaining after 3-h reaction were relatively low (Supplementary Figure S10). Further tuning of the system to establish homologous iGeTT is remained as the next challenge.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Forster,A.C. and Church,G.M. (2007) Synthetic biology projects *in vitro*. *Genome Res.*, **17**, 1–6.
2. Liu,A.P. and Fletcher,D.A. (2009) Biology under construction: *in vitro* reconstitution of cellular function. *Nat. Rev. Mol. Cell Biol.*, **10**, 644–650.

3. Fujiwara,K., Yanagisawa,M. and Nomura,S.M. (2014) Reconstitution of intracellular environments *in vitro* and in artificial cells. *Biophysics*, **10**, 43–48.

4. Noireaux,V. and Libchaber,A. (2004) A vesicle bioreactor as a step toward an artificial cell assembly. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 17669–17674.

5. Caschera,F. and Noireaux,V. (2014) Integration of biological parts toward the synthesis of a minimal cell. *Curr. Opin. Chem. Biol.*, **22**, 85–91.

6. Kresge,N., Simoni,R.D. and Hill,R.L. (2005) Otto Fritz Meyerhof and the elucidation of the glycolytic pathway. *J. Biol. Chem.*, **280**, e3.

7. Kaguni,J.M. and Kornberg,A. (1984) Replication initiated at the origin (oriC) of the *E. coli* chromosome reconstituted with purified enzymes. *Cell*, **38**, 183–190.

8. Shimizu,Y., Inoue,A., Tomari,Y., Suzuki,T., Yokogawa,T., Nishikawa,K. and Ueda,T. (2001) Cell-free translation reconstituted with purified components. *Nat. Biotechnol.*, **19**, 751–755.

9. Zhou,Y., Asahara,H., Gaucher,E.A. and Chong,S. (2012) Reconstitution of translation from *Thermus thermophilus* reveals a minimal set of components sufficient for protein synthesis at high temperatures and functional conservation of modern and ancient translation components. *Nucleic Acids Res.*, **40**, 7932–7945.

10. Yeeles,J.T., Deegan,T.D., Janska,A., Early,A. and Diffley,J.F. (2015) Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature*, **519**, 431–435.

11. Srivastava,A., Asahara,H., Zhang,M., Zhang,W., Liu,H., Cui,S., Jin,Q. and Chong,S. (2016) Reconstitution of protein translation of Mycobacterium reveals functional conservation and divergence with the gram-negative bacterium *Escherichia coli*. *PLoS One*, **11**, e0162020.

12. Liu,Z., Zhang,Y., Jia,X., Hu,M., Deng,Z., Xu,Y. and Liu,T. (2017) *In vitro* reconstitution and optimization of the entire pathway to convert glucose into fatty acid. *ACS Synth. Biol.*, **6**, 701–709.

13. Spirin,A.S. and Swartz,J.R. (2008) *Cell-free Protein Synthesis: Methods and Protocols*. John Wiley & Sons.

14. Asahara,H. and Chong,S. (2010) *In vitro* genetic reconstruction of bacterial transcription initiation by coupled synthesis and detection of RNA polymerase holoenzyme. *Nucleic Acids Res.*, **38**, e141.

15. Garamella,J., Marshall,R., Rustad,M. and Noireaux,V. (2016) The all *E. coli* TX–TL Toolbox 2.0: A platform for cell-free synthetic biology. *ACS Synth. Biol.*, **5**, 344–355.

16. Fujiwara,K., Katayama,T. and Nomura,S.M. (2013) Cooperative working of bacterial chromosome replication proteins generated by a reconstituted protein expression system. *Nucleic Acids Res.*, **41**, 7176–7183.

17. Kuruma,Y., Stano,P., Ueda,T. and Luisi,P.L. (2009) A synthetic biology approach to the construction of membrane proteins in semi-synthetic minimal cells. *Biochim. Biophys. Acta*, **1788**, 567–574.

18. Scott,A., Noga,M.J., de Graaf,P., Westerlaken,I., Yildirim,E. and Danelon,C. (2016) Cell-free phospholipid biosynthesis by gene-encoded enzymes reconstituted in liposomes. *PLoS One*, **11**, e0163058.

19. Matsubayashi,H., Kuruma,Y. and Ueda,T. (2014) *In vitro* synthesis of the *E. coli* Sec translocon from DNA. *Angew. Chem.*, **53**, 7535–7538.

20. Kobayashi,T., Nakamura,Y., Mikami,S., Masutani,M., Machida,K. and Imataka,H. (2012) Synthesis of encephalomyocarditis virus in a cell-free system: from DNA to RNA virus in one tube. *Biotechnol. Lett.*, **34**, 67–73.

21. Shin,J., Jardine,P. and Noireaux,V. (2012) Genome replication, synthesis, and assembly of the bacteriophage T7 in a single cell-free reaction. *ACS Synth. Biol.*, **1**, 408–413.

22. Murphy,L.D. and Zimmerman,S.B. (1995) Condensation and cohesion of lambda DNA in cell extracts and other media: implications for the structure and function of DNA in prokaryotes. *Biophys. Chem.*, **57**, 71–92.

23. Baigl,D. and Yoshikawa,K. (2005) Dielectric control of counterion-induced single-chain folding transition of DNA. *Biophys. J.*, **88**, 3486–3493.

24. Tsuji,A. and Yoshikawa,K. (2010) ON-OFF switching of transcriptional activity of large DNA through a conformational transition in cooperation with phospholipid membrane. *J. Am. Chem. Soc.*, **132**, 12464–12471.

25. Pruss,G.J. and Drlica,K. (1989) DNA supercoiling and prokaryotic transcription. *Cell*, **56**, 521–523.

26. Cavalli,G. and Misteli,T. (2013) Functional implications of genome topology. *Nat. Struct. Mol. Biol.*, **20**, 290–299.

27. Harris,D.C. and Jewett,M.C. (2012) Cell-free biology: exploiting the interface between synthetic biology and synthetic chemistry. *Curr. Opin. Biotechnol.*, **23**, 672–678.

28. Henne,A., Bruggemann,H., Raasch,C., Wiezer,A., Hartsch,T., Liesegang,H., Johann,A., Lienard,T., Gohl,O., Martinez-Arias,R. *et al.* (2004) The genome sequence of the extreme thermophile *Thermus thermophilus*. *Nat. Biotechnol.*, **22**, 547–553.

29. Kim,K., Okanishi,H., Masui,R., Harada,A., Ueyama,N. and Kuramitsu,S. (2012) Whole-cell proteome reference maps of an extreme thermophile, *Thermusthermophilus* HB8. *Proteomics*, **12**, 3063–3068.

30. Iwanaga,N., Ide,K., Nagashima,T., Tomita,T., Agari,Y., Shinkai,A., Kuramitsu,S., Okada-Hatakeyema,M., Kuzuyama,T. and Nishiyama,M. (2014) Genome-wide comprehensive analysis of transcriptional regulation by ArgR in *Thermus thermophilus*. *Extremophiles*, **18**, 995–1008.

31. Paget,M.S. and Helmann,J.D. (2003) The sigma70 family of sigma factors. *Genome Biol.*, **4**, 203.

32. Shine,J. and Dalgarno,L. (1975) Determinant of cistron specificity in bacterial ribosomes. *Nature*, **254**, 34–38.

33. Shin,J. and Noireaux,V. (2010) Efficient cell-free expression with the endogenous *E. coli* RNA polymerase and sigma factor 70. *J. Biol. Eng.*, **4**, 8.

34. Fujiwara,K. and Doi,N. (2016) Biochemical preparation of cell extract for cell-free protein synthesis without physical disruption. *PLoS One*, **11**, e0154614.

35. McClelland,M., Jones,R., Patel,Y. and Nelson,M. (1987) Restriction endonucleases for pulsed field mapping of bacterial genomes. *Nucleic Acids Res.*, **15**, 5985–6005.

36. Zhang,M., Zhang,Y., Scheuring,C.F., Wu,C.C., Dong,J.J. and Zhang,H.B. (2012) Preparation of megabase-sized DNA from a variety of organisms using the nuclei method for advanced genomics research. *Nat. Protoc.*, **7**, 467–478.

37. Overhauser,J. and Radic,M. (1987) Encapsulation of cells in agarose beads for use with pulsed-field gel electrophoresis. *Focus*, **9**, 8–9.

38. Lartigue,C., Glass,J.I., Alperovich,N., Pieper,R., Parmar,P.P., Hutchison,C.A. 3rd, Smith,H.O. and Venter,J.C. (2007) Genome transplantation in bacteria: changing one species to another. *Science*, **317**, 632–638.

39. Gibson,D.G., Glass,J.I., Lartigue,C., Noskov,V.N., Chuang,R.Y., Algire,M.A., Benders,G.A., Montague,M.G., Ma,L., Moodie,M.M. *et al.* (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, **329**, 52–56.

40. Hutchison,C.A. 3rd, Chuang,R.Y., Noskov,V.N., Assad-Garcia,N., Deerinck,T.J., Ellisman,M.H., Gill,J., Kannan,K., Karas,B.J., Ma,L. *et al.* (2016) Design and synthesis of a minimal bacterial genome. *Science*, **351**, aad6253.

41. Swarts,D.C., Jore,M.M., Westra,E.R., Zhu,Y., Janssen,J.H., Snijders,A.P., Wang,Y., Patel,D.J., Berenguer,J., Brouns,S.J. *et al.* (2014) DNA-guided DNA interference by a prokaryotic Argonaute. *Nature*, **507**, 258–261.

42. Koyama,Y., Hoshino,T., Tomizuka,N. and Furukawa,K. (1986) Genetic transformation of the extreme thermophile *Thermus thermophilus* and of other *Thermus* spp. *J. Bacteriol.*, **166**, 338–340.

43. Fujiwara,K., Tsubouchi,T., Kuzuyama,T. and Nishiyama,M. (2006) Involvement of the arginine repressor in lysine biosynthesis of *Thermus thermophilus*. *Microbiology*, **152**, 3585–3594.

44. Matsuda,T., Kigawa,T., Koshiba,S., Inoue,M., Aoki,M., Yamasaki,K., Seki,M., Shinozaki,K. and Yokoyama,S. (2006) Cell-free synthesis of zinc-binding proteins. *J. Struct. Funct. Genomics*, **7**, 93–100.

45. Iwasaki,M., Miwa,S., Ikegami,T., Tomita,M., Tanaka,N. and Ishihama,Y. (2010) One-dimensional capillary liquid chromatographic separation coupled with tandem mass spectrometry unveils the *Escherichia coli* proteome on a microarray scale. *Anal. Chem.*, **82**, 2616–2620.

46. Fujiwara,K., Ishihama,Y., Nakahigashi,K., Soga,T. and Taguchi,H. (2010) A systematic survey of in vivo obligate chaperonin-dependent substrates. *EMBO J.*, **29**, 1552–1564.

47. Ingolia,N.T., Ghaemmaghami,S., Newman,J.R. and Weissman,J.S. (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, **324**, 218–223.

48. Jones,E.P. and Chi,E. (1997) DNA isolation in agarose plugs for sizing and characterization of bacterial artificial chromosome (BAC) clones. *Tech. Tips Online*, **2**, 174–176.

49. Sitaraman,K., Esposito,D., Klarmann,G., Le Grice,S.F., Hartley,J.L. and Chatterjee,D.K. (2004) A novel cell-free protein synthesis system. *J. Biotechnol.*, **110**, 257–263.

50. Jewett,M.C. and Swartz,J.R. (2004) Mimicking the *Escherichia coli* cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnol. Bioeng.*, **86**, 19–26.

51. Ying,B.W., Taguchi,H., Kondo,M. and Ueda,T. (2005) Co-translational involvement of the chaperonin GroEL in the folding of newly translated polypeptides. *J. Biol. Chem.*, **280**, 12035–12040.

52. Niwa,T., Kanamori,T., Ueda,T. and Taguchi,H. (2012) Global analysis of chaperone effects using a reconstituted cell-free translation system. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 8937–8942.

53. Niwa,T., Ying,B.W., Saito,K., Jin,W., Takada,S., Ueda,T. and Taguchi,H. (2009) Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of *Escherichia coli* proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 4201–4206.

54. Niwa,T., Sugimoto,R., Watanabe,L., Nakamura,S., Ueda,T. and Taguchi,H. (2015) Large-scale analysis of macromolecular crowding effects on protein aggregation using a reconstituted cell-free translation system. *Front. Microbiol.*, **6**, 1113.

55. Niwa,T., Fujiwara,K. and Taguchi,H. (2016) Identification of novel in vivo obligate GroEL/ES substrates based on data from a cell-free proteomics approach. *FEBS Lett.*, **590**, 251–257.

56. Niwa,T., Sasaki,Y., Uemura,E., Nakamura,S., Akiyama,M., Ando,M., Sawada,S., Mukai,S.A., Ueda,T., Taguchi,H. *et al.* (2015) Comprehensive study of liposome-assisted synthesis of membrane proteins using a reconstituted cell-free translation system. *Sci. Rep.*, **5**, 18025.

57. Nomura,S.M., Tsumoto,K., Hamada,T., Akiyoshi,K., Nakatani,Y. and Yoshikawa,K. (2003) Gene expression within cell-sized lipid vesicles. *Chembiochem*, **4**, 1172–1175.

58. Ohta,T., Tokishita,S., Imazuka,R., Mori,I., Okamura,J. and Yamagata,H. (2006) beta-Glucosidase as a reporter for the gene expression studies in *Thermus thermophilus* and constitutive expression of DNA repair genes. *Mutagenesis*, **21**, 255–260.

59. Iwasaki,M., Masuda,T., Tomita,M. and Ishihama,Y. (2009) Chemical cleavage-assisted tryptic digestion for membrane proteome analysis. *J. Proteome Res.*, **8**, 3169–3175.