US 20040033564A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2004/0033564 A1**

Seong et al. (43) **Pub. Date: Feb. 19, 2004**

(54) **METHOD FOR INCREASING SOLUBILITY OF TARGET PROTEIN USING RNA-BINDING PROTEIN AS FUSION PARTNER**

(76) Inventors: **Balk Lin Seong**, Seoul (KR); **Seong-Il Choi**, Seoul (KR); **Hang Cheol Shin**, Seoul (KR)

Correspondence Address:
MUSERLIAN AND LUCAS AND MERCANTI, LLP
475 PARK AVENUE SOUTH
NEW YORK, NY 10016 (US)

(21) Appl. No.: **10/371,771**

(22) Filed: **Feb. 21, 2003**

(30) **Foreign Application Priority Data**

Aug. 19, 2002 (KR) KR/2002-0048929

Publication Classification

(51) **Int. Cl.**[7] **C12P 21/02**; C12N 5/06; C07H 21/02; C12N 15/86; C12N 5/02; C12N 5/00; C12N 15/74; C12N 15/70; C12N 15/63; C12N 15/09; C12N 15/00; C12P 21/06; C07H 21/04; C12P 21/04

(52) **U.S. Cl.** **435/69.1**; 435/456; 435/320.1; 435/325; 536/23.1; 435/69.7; 536/23.4

(57) **ABSTRACT**

Disclosed is a method of producing a target protein having improved solubility and folding efficiency, characterized by expressing a target protein as a fusion protein using an RNA-binding protein as a fusion partner, and binding the fusion partner fused to the target protein to an RNA molecule. More particularly, the present invention discloses a method of producing a target protein having improved solubility and folding efficiency, comprising constructing an expression vector encoding a fusion protein using a target protein and an RNA-binding protein as a fusion partner of the target protein; constructing an expression vector expressing an RNA molecule capable of binding to the RNA-binding protein; and cotransforming a host cell with the expression vectors and then culturing the resulting transformant, wherein the expressed fusion protein binds to the overexpressed RNA molecule. The method of the present invention is very useful for production of proteins for medical and industrial applications.

FIGURES

FIG. 1

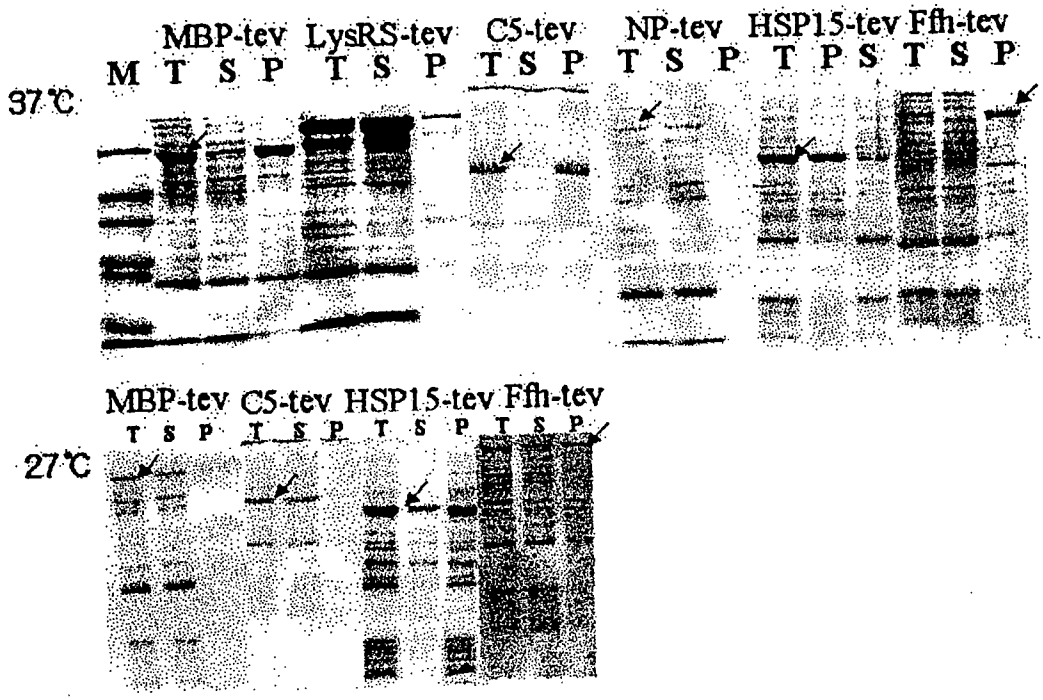


FIG. 2

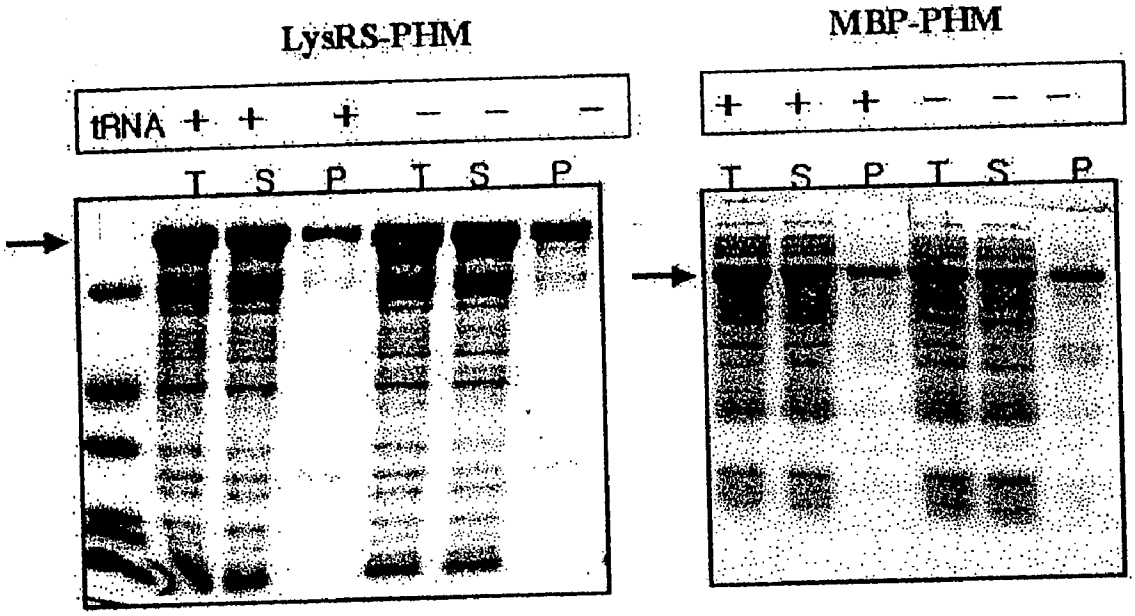


FIG. 3

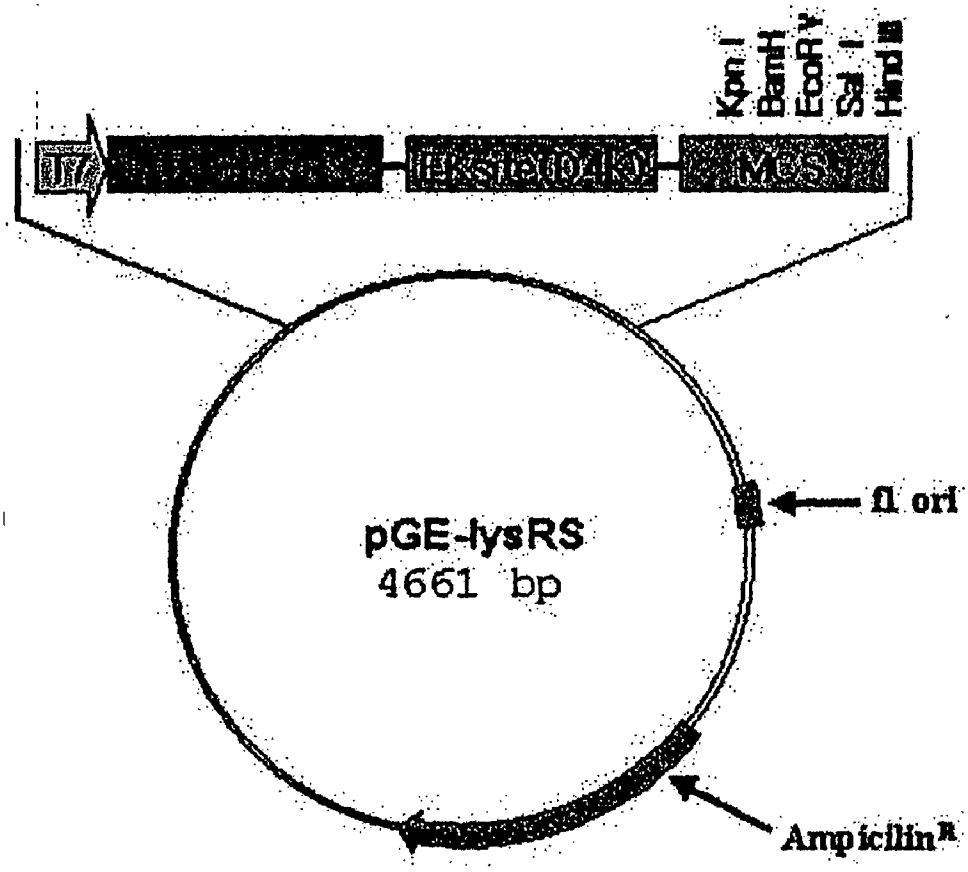

Kpn I
BamH
EcoR I
Sal I
Hind III
T7
MCS
pGE-lysRS
4661 bp
f1 ori
Ampicillin R

FIG. 4

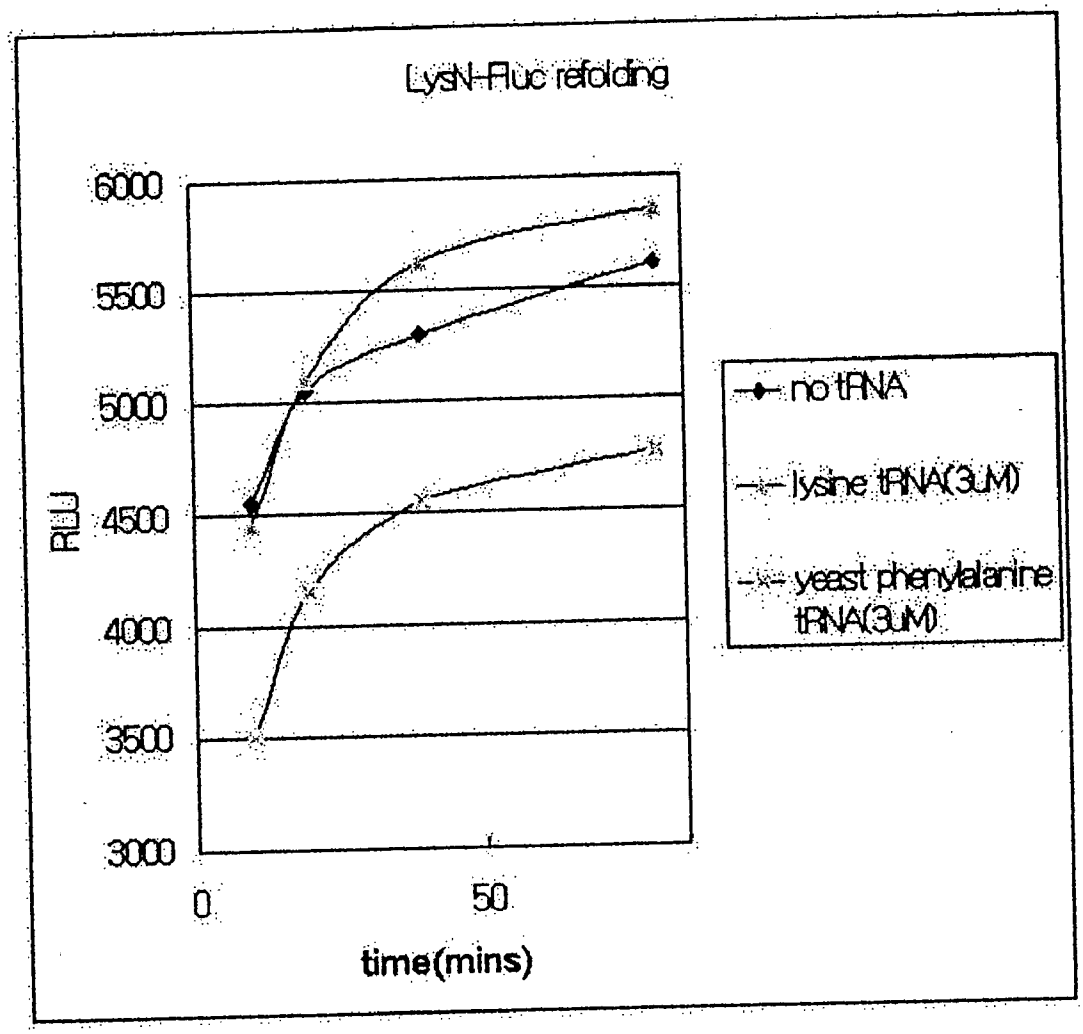
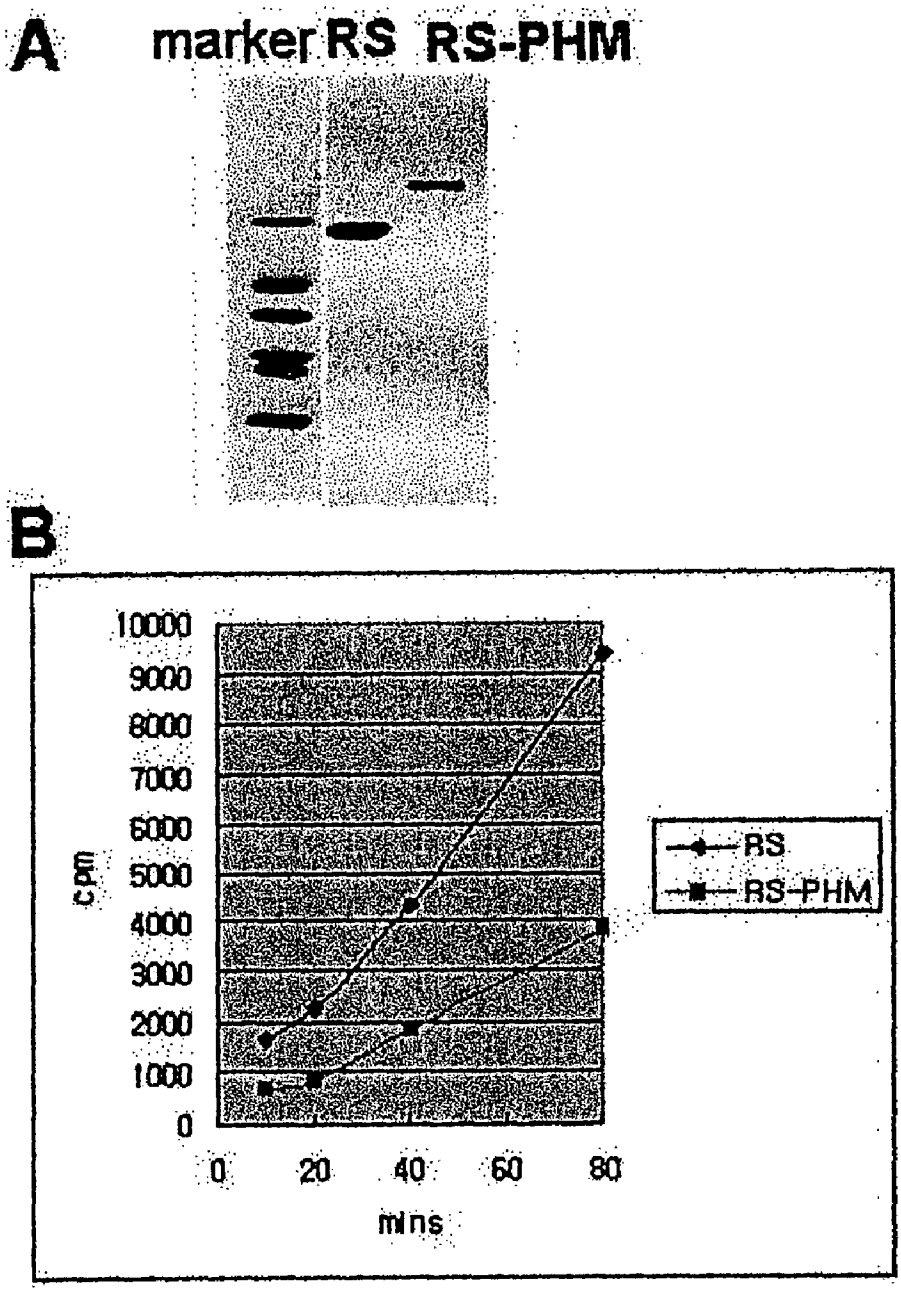

LysN-Fluc refolding
6000
5500
5000
4500
4000
3500
3000
RLU
0
50
time(mins)
no tRNA
lysine tRNA(3uM)
yeast phenylalanine tRNA(3uM)

FIG. 5


A
marker
RS
RS-PHM
B
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
cpm
0
20
40
60
80
mins
RS
RS-PHM

METHOD FOR INCREASING SOLUBILITY OF TARGET PROTEIN USING RNA-BINDING PROTEIN AS FUSION PARTNER

TECHNICAL FIELD

[0001] The present invention, in general, relates to a method of increasing solubility of a target protein, and more particularly, to a method of producing a target protein having improved solubility and folding efficiency, based on expression of a target protein as a fusion protein using a fusion partner, and binding of the fusion protein with an RNA molecule.

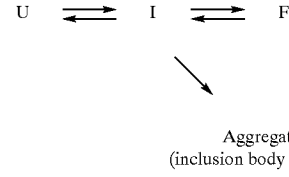
BACKGROUND ART

[0002] With the development of genetic recombination techniques, numerous target proteins are produced using animal cells, yeasts and prokaryotic systems including *E. coli*, and such proteins are widely used in the bioengineering industry, including medical fields. In particular, owing to its high growth rate and its relatively well identified genetic structure compared to other organisms, the bacterium *E. coli* is routinely used as a host cell for production of target proteins using genetic recombinant techniques.

[0003] However, *E. coli* has a severe disadvantage in terms of not having a variety of intracellular elements required for maturation of proteins in comparison with eukaryotic cells. In detail, post-translational modification, disulfide bond formation, glycosylation and compartmentation of proteins, which are achieved in eukaryotic cells, are not performed in *E. coli*. In addition, when a target protein is expressed in a large scale in *E. coli*, the expressed proteins frequently accumulate in the cytoplasm, forming insoluble protein aggregates referred to as inclusion bodies. Although being easily isolated and resistant to proteinase digestion, in order to obtain active proteins from the inclusion bodies, the inclusion bodies should be solubilized using a high concentration of urea or guanidium HCl to unfold proteins contained in the inclusion bodies into their primary structure, and then the resulting proteins must be refolded into biologically active conformation during or after removal of the chemical reagent. Since mechanisms associated in protein refolding are still not accurately identified, and refolding conditions vary according to proteins, finding effective refolding conditions requires much time and high cost. Because of recombinant proteins having low refolding rates, high-cost apparatuses are necessary for scaling up their industrial production, and most proteins having a high molecular weight are hard or impossible to refold, thereby creating difficulty in industrialization of such proteins.

[0004] Although biologically active proteins are stable thermodynamically, inclusion bodies are often formed during their expression in the *E. coli* system, formation of which is driven by intermolecular aggregation between folding intermediates during folding processes of proteins (Mitraki, A and King, J., *Bio/Technology*, 1989, 7:690-697) (Reaction Formula 1).

[Reaction Formula 1]



[0005] wherein, U is a protein in an unfolded state, F is a protein in a folded state, and I is a folding intermediate.

[0006] Typically, refolding a protein into an active form is accomplished experimentally, and is not always successfully achieved, thereby making large-scale production of a recombinant protein difficult. In addition, by the above-mentioned refolding process, it is difficult to obtain antibodies having a high molecular weight, tissue plasminogen activator and factor VIII in active forms.

[0007] To overcome the problems encountered when expressing target proteins as inclusion bodies, it is meaningful to express a target protein in a soluble form in *E. coli*. Until now, the following three methods have been used in effectively expressing a target protein.

[0008] First, a target protein can be obtained in a soluble form by linking a signal sequence to the N-terminus of the target protein to allow its secretion to the periplasm of *E. coli* (Stader, J. A. and Silhavy, T. J., *Methods in Enzymol.*, 1970, 165:166-187). However, such a method is not industrially available owing to low expression rate of the target protein.

[0009] Second, a target protein can be produced in a soluble form by co-expression with a chaperone gene, such as gro ES, gro El or dna K genes (Goloubinoff, P., Gatenby, A. A. and Lorimer, G. H., *Nature*, 1989, 337: 44-47). But this method is effective for specific proteins, and so is not for general use to prevent formation of inclusion bodies.

[0010] Third, a soluble target protein can be obtained by selecting a protein highly expressed in *E. coli* and then fusing a target protein to the C-terminus of the selected protein. Such fusion of the target protein with the C-terminus of a fusion partner protein allows effective use of translation initiation signals of the fusion partner, as well as increasing solubility of the target protein linked to the fusion partner, thereby leading to large-scale expression of the target protein in a soluble form in *E. coli*.

[0011] Among the methods of the prior arts for expressing a recombinant protein in a soluble form, the most successful one is to express the recombinant protein as a fusion protein using a highly soluble protein as a fusion partner. To produce a fusion protein in *E. coli*, Lac Z or Trp E protein is conventionally used as a fusion partner protein. However, fusion proteins with the Lac Z or Trp E protein are mostly produced as inclusion bodies, and thus it is hard to obtain a protein of interest in an active form. In this regard, many attempts to find new fusion partner proteins have been performed. As a result, several proteins or peptides were developed as fusion partner proteins: glutathion-S-transferase (Smith, D. B. and Johnson, K. S., *Gene*, 1988, 67: 31-40), maltose-binding protein (Bedouelle, H. and Duplay,

P., *Euro. J. Biochem.*, 1988, 171: 541-549), protein A (Nilsson, B. et al., *Nucleic Acid Res.*, 1985, 13: 1151-1162), Z domain of protein A (Nilsson, B. et al., *Prot. Eng.*, 1987, 1: 107-113), protein Z (Nygren, P. A. et al., *J. Mol. Recog.*, 1988, 1: 69-74), and thioredoxin (Lavallie, E. R. et al., *Bio/Technology*, 1993, 11: 187-193).

[0012] It has been reported that factors determining solubility of proteins include, in order of importance, average charge, fraction of turn-forming residues, cysteine fraction, proline fraction, hydrophilicity and total numbers of residues. And it also has been reported that average net charge and fractions of turn-forming residues are especially important (Wilkinson, DL and Harrison, RG., *Bio/Technology*, 1991, 9:443-448). Using the two very important parameters, model formula for solubility of a protein is defined as follows (Davis et al., *Biotechnol. Bioeng.*, 1999, 65: 382-388):

[0013] <Model Foumula>

$$CV=\lambda 1(\ (N+G+P+S)/n)+\lambda 2|((R+K)-(D+E))/n-0.03)|$$

[0014] wherein, CV is a canonical variable; n is the number of amino acids in the protein; N, G, P and S are numbers of residues of asparagine (N), glysine (G), proline (P) and serine (S), respectively; R, K, D and E are numbers of residues of arginine (R), lysine (K), asparaginic acid (D), glutamic acid (E), respectively; and λ1 and λ2 are coefficients of 15.43 and −29.56, respectively. If CV−CV' is positive, a protein is predicted to be insoluble. If CV−CV' is negative, a protein is predicted to soluble.

[0015] In the above formula, probability of solubility or insolubility is designated as 0.4934+0.276| CV−CV'|−0.0392(CV−CV')2, where CV' is a discriminant number of 1.71. That is, solubility of protein is determined by average charge and folding rate, where the higher the content of turn-forming residues including Asn, Gly, Pro and Ser is, the lower the folding rate is. Using the above formula, the *E. coli* protein Nus A was developed as a fusion partner (Davis et al., Biotechnol. Bioeng. 65, 382-388, 1999).

[0016] As described above, among the methods of the prior arts for expressing a recombinant protein as a soluble form, the most successful one is to express the recombinant protein as a fusion protein using a protein having high solubility as a fusion partner. The conventional fusion partner proteins include maltose binding protein, thioredoxin, glutathione-S-transferase, NusA, LysN (N-terminal domain of *E. coli* lysine tRNA synthetase), and lyss (Korean Pat. Application No. 1996-044010). A fusion partner protein improves solubility of a target protein according to Reaction Formula 2, below.

[Reaction Formula 2]

$$U_{(p,t)} \rightleftharpoons F_{(p)}—U_{(t)} \rightleftharpoons F_{(p,t)}$$

[0017] wherein, U is an unfolded state; F is a folded state; p is a fusion partner; and t is a target protein.

[0018] As apparent in the above Reaction Formula 1, the fusion protein increases overall solubility of the target protein by stabilizing intermediates using its high soluble property.

[0019] Molecular chaperones are known to help folding of proteins by temporarily binding to partially folded proteins and thus preventing their aggregation. Referring to the above Reaction Formula 2, a fusion partner is considered to serve as a chaperone. Because of being linked to a target protein, the fusion partner can be referred to a molecular chaperone. In the conventional concept of the molecular chaperones, a prosequence of a protein, for example, that of subtilisin, which is cleaved after assisting folding of a protein, is called a molecular chaperone (Shinde, U., and Inouye, M. 1995, *J.Mol.Biol.*). There is a difference between the prosequence and the fusion partner. The former has a limitation of acting to assist folding of only one protein, while the latter helps folding of a broad range of target proteins. Also, it has been reported that ribosome or the ribosomal component 23 S RNA help refolding of proteins (Das, B., Chattopadhyay, D. B., Bera, A. K., and Dasgupta, C. 1996, *Eur. J. Biochem.* 235, 623-621; Chattopadhyay. S., Das, B., and Dasgupta, C. (1996) Proc. Natl. Acad. Sci. 93, 8284-8287) In this case, the ribosome or 23 S RNA induces protein refolding in a trans-acting manner, not in a fused form with a target protein.

[0020] Performance ability of the fusion partner proteins fused to target proteins may basically depend on a rapid folding rate and high average net charge. The most urgent prior problem to be solved in the post-genome era is to identify the function of proteins. To solve the above problem, proteins are first produced in a soluble active form. In this regard, development of fusion partner proteins having excellent properties is very important in basic research and industrial processes. Fusion partner proteins have been discovered by experimental experiences or an aforementioned simple method, like the discovery of NusA.

DISCLOSURE OF THE INVENTION

[0021] Leading to the present invention, the intensive and thorough research into methods of improving solubility of a target protein, conducted by the present inventors, resulted in the finding that, when being expressed as a fusion protein using an RNA-binding protein as a fusion partner, a target protein has improved solubility and folding efficiency, wherein the RNA-binding protein fused to the target protein binds an RNA molecule, and thus an average net charge of the RNA molecule affects the target protein, increasing a net charge of the target protein.

[0022] It is therefore an object of the present invention to provide a method of producing a target protein having improved solubility and folding efficiency, comprising expressing a target protein as a fusion protein with an RNA-binding protein, and binding the RNA-binding protein fused to the target protein to an RNA molecule.

BRIEF DESCRIPTION OF THE DRAWINGS

[0023] The above and other objects, features and other advantages of the present invention will be more clearly understood from the following detailed description taken in conjunction with the accompanying drawings, in which:

[0024] **FIG. 1** is a photograph showing results of SDS-PAGE analysis for solubility of fusion proteins obtained by expressing a target protein as a fusion protein at 37° C. or 27° C. using various RNA-binding proteins to construct the

fusion proteins, (M: protein size marker; T: whole cell lysate; S: supernatant; and P: pellet);

[0025] **FIG. 2** is a photograph showing results of SDS-PAGE analyzing the effect of co-expression of lysine-tRNA on solubility of a LysRS-PHM fusion protein when expressing a target protein PHM using LysRS capable of binding to lysine tRNA as a fusion partner (T: whole cell lysate; S: supernatant; and P: pellet);

[0026] **FIG. 3** is a diagrammatic representation of a pGE-lysRS vector;

[0027] **FIG. 4** is a graph showing results of luciferase activity analysis for determining the effect of lysine tRNA on refolding of a LysRS-firefly luciferase fusion protein; and

[0028] **FIG. 5A** is a photograph showing results of SDS-PAGE of isolated LysRS and LysRS-PHM proteins, and **FIG. 5B** is a graph showing results of tRNA synthetase activity analysis for the LysRS and LysRS-PHM proteins to identify binding of the LysRS-PHM protein to lysine tRNA.

BEST MODES FOR CARRYING OUT THE INVENTION

[0029] To achieve the aforementioned object, the present invention provides a method of producing a target protein having improved solubility and folding efficiency, comprising expressing a target protein as a fusion protein using an RNA-binding protein as a fusion partner, and binding the RNA-binding protein fused to the target protein to an RNA molecule.

[0030] Since solubility of proteins is largely dependent on their average net charges and folding speeds, it is hard to increase solubility of proteins only by artificially modifying the folding speeds of proteins. Based on the fact that RNA molecules having high solubility in vivo and high net charges, in the present invention, a target protein is expressed as a fusion protein employing an RNA-binding protein as a fusion partner. And the RNA-binding protein fused to the target protein is allowed to bind an RNA molecule, where the RNA molecule present at a stable state supplies a strong negative charge, thereby increasing an average net charge of the target protein and leading to increased solubility of the target protein.

[0031] The fusion partner fused with the target protein may bind to an RNA molecule naturally present in cells, or an artificially co-expressed RNA molecule capable of binding to the fusion partner. Such co-expression of the RNA molecule may be achieved by constructing a vector expressing the RNA molecule and then introducing the vector into a host cell to overexpress the RNA molecule.

[0032] The method of producing a target protein having improved solubility and folding efficiency comprises the steps of:

[0033] 1) constructing an expression vector encoding a fusion protein using a target protein and an RNA-binding protein as a fusion partner of the target protein;

[0034] 2) constructing an expression vector expressing an RNA molecule capable of binding to the RNA-binding protein fused with the target protein; and

[0035] 3) cotransforming a host cell with the expression vectors prepared in Steps 1 and 2.

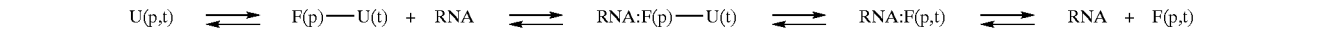
[0036] The RNA molecule binding to the fusion partner of the target protein may be selected from the group consisting of tRNA, mRNA, rRNA, nuclear RNA, and ribo-polynucleotides artificially prepared by genetic recombination techniques.

[0037] The fusion partner of the target protein may be a protein capable of binding to an RNA molecule, that is, a domain or polypeptide binding to an RNA molecule, or a derivative of a protein binding to an RNA molecule. Preferably, the fusion partner is a protein selected from the group consisting of aminoacyl-tRNA synthetases, ribosomal proteins, mRNA binding proteins, viral proteins having RNA-binding ability and proteins associated with cellular RNA processing and turnover, or a polypeptide corresponding to an RNA-binding region (domain) of the aforementioned proteins. More preferably, fusion partner is a protein selected from the group consisting of C5 protein of ribonuclease P (RNase P), Ffh protein of signal recognition particle, NP protein of influenza virus, ribosomal S1 protien, ribosomal S4 protein, ribosomal S17 protein, *E. coli* DbpA and *E. coli* Hsp15.

[0038] DNA or RNA molecules have a net negative charge due to oxygen molecules of phosphate groups in their backbone, and are thus highly soluble polymers. RNA molecules, in vivo, interact with a large number of proteins because of participating in replication, transcription and translation. When a protein composed of 200 neutral amino acid residues forms a complex with an RNA molecule composed of 100 nucleotides, an average net charge of the protein is calculated to increase by about −0.2, where molecular weight of the RNA molecule is converted to number of amino acid residues by dividing the molecular weight of the RNA molecule by an average molecular weight of amino acids (100×330/110=300) on the assumption that solubility of nucleotides and amino acid residues per unit mass excluding charges is identical, and an average number of the RNP complex converted to amino acid number is 500. If such a negative net charge of the RNP complex is converted to amino acid number, the RNP complex is composed of 200 amino acid residues and 50 residues consisting of aspartate and glutamate.

[0039] When expressed employing a protein having high affinity to DNA or RNA molecules as a fusion partner of a target protein, a fusion protein forms a nucleic acid-protein complex by binding of the DNA or RNA molecule to the fusion partner, and the strong negative charge of the DNA or RNA molecule changes an average net charge of the target protein, thereby increasing solubility of the target protein (Reaction Formula 3). RNA molecules are more effective than DNA molecules in improving solubility of a target protein in terms of being present in cells at a much higher amount and being more widely diffused than DNA.

[Reaction Formula 3]



[0040] wherein, U is an unfolded state; F is a folded state; p is a fusion partner; and T is a target protein.

[0041] According to affinity of a protein to an RNA molecule, the protein binds to the RNA molecule in an irreversible manner in which the protein strongly binds to the RNA molecule, for example, ribonucleoprotein (hereinafter, referred to as "RNP"), or in a reversible manner in which the proteins weakly bind to the RNA molecules. It is known that *E. coli* RNP molecules include ribosome, ribonulease P (hereinafter, referred to as "RNase P") and signal recognition particle (hereinafter, referred to as "SRP") Ribosomes are the largest RNP complexes among RNA-protein complexes identified until now.

[0042] Non-limiting examples of the protein binding to a target protein and having a binding affinity to an RNA molecule include the following proteins. RNase P, which is an endonuclease catalyzing cleavage of the 5' end of a tRNA precursor, consists of a catalytic RNA subunit (Ml RNA, 377 nucleotides) and C5 protein (119 amino acid residues) affecting stability and activity of RNase P (Gopalan et al., *J.Mol.Biol.*, 1997, 267:818-829). The C5 protein may be used as a fusion partner of a target protein having valuable biological activity. SRP is known to induce translocation of proteins to the endoplasmic reticulum (ER) membrane in eukaryotic cells, and to target proteins to nascent inner membrane proteins to transport sites the inner membrane in *E. coli. E. coli* SRP consists of 4.5 S RNA and Ffh protein, and the Ffh protein contains an RNA-binding domain ranging from 296 to 453 amino acid residues (Barty RT et al., *J.Mol.Biol.*, 2001, 307: 229-246).

[0043] The conventional fusion proteins using pro-sequences as fusion partners consist of only proteins. In contrast, the fusion partner of a target protein of the present invention differs from the conventional fusion partners in terms of being a ribonucleoprotein (RNP) complex consisting of an RNA molecule and a protein. The RNA molecule in the RNP complex serves as a molecular chaperone in a cis-acting manner, whereas the conventional refolding process uses ribosome or 23 S RNA.

[0044] In case of the conventional protein expression system using trans-acting chaperone proteins, in order to express a target protein in an active form, the chaperone should be additionally introduced into a host cell transformed with the target protein, or expressed along with the target protein. In contrast, the method of the present invention is advantageous in that a chaperone molecule is co-expressed in a fused form with a target protein, and thus the target protein is produced in an active form.

[0045] In addition, in the case of the conventional protein expression system using ribosome or 23 S RNA, ribosome or 23 S RNA is used in converting a target protein expressed in an inactive form to an active form in vitro. In contrast, the method of the present invention is distinguishable from the conventional method in terms of expressing a target protein in an active form, not requiring an additional refolding process.

[0046] AN RNA-protein complex is stably formed, in which a protein binds to an RNA molecule with a strong association constant, and the binding is in equilibrium. The RNA-protein complex is advantageous in terms of enabling solubility and folding efficiency of the protein fused to an RNA-binding protein to increase by the RNA molecule's properties of having a net negative charge and high solubility. That is, intermolecular aggregation of the fusion proteins is inhibited by repulsion of negative charges of the fusion proteins, resulting in that each of the fusion proteins is present in a separate form, and the fusion proteins are highly soluble in an aqueous environment by interaction between the net charges and water molecules. The RNA molecule binding to the RNA-binding protein serves as a molecular chaperone by inducing the protein to fold into its active form. With respect to the function and concept of the molecular chaperone, such function of the RNA molecule as a molecular chaperone, mediated by the charge-charge repulsion, is distinguished from the conventional fusion proteins fused with chaperone proteins that mediate protein folding by protein-protein interaction.

[0047] The present invention will be explained in more detail with reference to the following examples. However, the following examples are provided only to illustrate the present invention, and the present invention is not limited to them.

EXAMPLE 1

[0048] Preparation of Expression Vectors Encoding Proteins to be Fused to Target Proteins

[0049] Expression vectors expressing proteins binding to tRNA, rRNA or mRNA, proteins forming RNP complexes, and viral proteins binding to RNA molecules were constructed, in which the proteins will be used as fusion partners of target proteins. In detail, the following proteins were selected as fusion partners of target proteins: *E. coli* lysyl tRNA synthetase (hereinafter, referred to as "lysRS"), tyrosyl tRNA synthetase (hereinafter, referred to as "tyrRS"), tryptophane tRNA synthetase (hereinafter, referred to as "trpRS"), *E. coli* rRNA binding proteins S1, S4 and S17, *E. coli* Hsp 15 and DbpA, Ffh protein of *E. coli* signal recognition particle (SRP), C5 protein of RNase P, and NP protein of influenza A virus.

[0050] First, PCR was carried out using genomic DNA obtained from JM109 cells (*Gene*, 1985, 33, 103-119) as a template and primers designated SEQ ID NO.: 1 and SEQ ID NO.: 2. The resulting PCR product 'lysRS' was cloned to a pGEMEX-ΔNdeI vector, which was prepared by removing a NdeI site at the 3251 position among two NdeI sites present in the pGEMEX-1 vector (Promega), giving a pGE-lysRS expression vector (**FIG. 3**). In addition, expression vectors carrying genes encoding other aforementioned proteins except NP protein were prepared by PCR using primers designated with the SEQ ID NOs listed in Table 1, below, according to the same method as in constructing the expres-

sion vector carrying the lysRS gene. An expression vector expressing NP protein of influenza virus was prepared by PCR using plasmid DNA from a PIVA-NP vector having a gene encoding NP protein, in which the gene is positioned at an EcoRI site of pUC19 vector (*Gene*, 1985, 33, 103-119) and under the regulation of T7 promoter, as a template, and primers designated SEQ ID NO.: 21 and SEQ ID 5 NO.: 22, and then inserting the resulting amplified product into NdeI/KpnI sites of the pGE-lysRS vector, giving a pGE-NP vector. In order to compare effect of the fusion partners on solubility and folding efficiency of target proteins to proteins not binding to RNA molecules, a vector expressing *E. coli* maltose binding protein (MBP) that is known to not bind to RNA molecules was prepared according to the same method as described above, thus yielding a PGE-MBP vector.

TABLE 1

Genes	Primers	Expression vectors
tyrRS	SEQ ID NOs: 3 and 4	pGE-tyrRS
trpRS	SEQ ID NOs: 5 and 6	pGE-trpRS
S1	SEQ ID NOs: 7 and 8	pGE-S1
S4	SEQ ID NOs: 9 and 10	pGE-S4
S17	SEQ ID NOs: 11 and 12	pGE-S17
Hsp15	SEQ ID NOs: 13 and 14	pGE-Hsp15
DbpA	SEQ ID NOs: 15 and 16	pGE-DbpA
Ffh	SEQ ID NOs: 17 and 18	pGE-Ffh
C5	SEQ ID NOs: 19 and 20	pGE-C5
NP	SEQ ID NOs: 21 and 22	pGE-NP
MBP	SEQ ID NOs: 23 and 24	pGE-MBP

EXAMPLE 2

[0051] Evaluation of Effect of Fusion Partners on Solubility of a Target Protein

[0052] In order to investigate effect of the fusion partners on solubility of a target protein, protease of tobacco etch virus (hereinafter, referred to as "TEV") was used as a target protein, and fused with each of the fusion partners. In detail, PCR was carried out using pRK793 plasmid (*Protein Engineering*, 2001, 14, 993-1000) as a template and primers designated SEQ ID NO.: 25 and SEQ ID NO.: 26. the amplified PCR product was inserted into each of the vectors prepared in Example 1, pGE-lysRS, pGE-Hsp15, pGE-Ffh, pGE-C5, pGE-NP and pGE-MBP. The resulting expression vectors were designated as "plysRS-TEV", "pHsp15-TEV", "pFfh-TEV", "pC5-TEV", "pNP-TEV" and "pMBP-TEV", respectively.

[0053] Then, each of the expression vectors was introduced into *E. coli* HMS174(DE3)plysE (Novagen). Single colonies were inoculated in 2 ml of LB medium containing ampicillin of 50 μg/ml and chloramphenicol of 30 μg/ml, followed by incubation at 37° C. overnight. The cultured cells were diluted in 20 ml of LB medium, and cultured until OD_{600} reached 0.5. Thereafter, 1 mM IPTG was added to the culture medium, and the transformed cells were incubated at 37° C. or 27° C. for 5 hrs to express the recombinant proteins. After collecting 10 ml from the resulting cultured medium, the harvested cell pellet was supplemented with 0.3 ml of PBS and sonicated using a sonifier. 50 μl of the total cell lysates was mixed with 2×SDS buffer, and the remainder of cell lysate was centrifuged at 13,000 rpm for 12 min, thus yielding a supernatant. Also, the pellet was suspended in 250 μl of PBS. 50 μl of each of the supernatant and the pellet was mixed with 50 μl of 2×SDS buffer. After being boiled at 100° C., the mixtures were electrophoresed on a SDS-PAGE gel, and the separated proteins were stained with Coommassie blue.

[0054] As a result, when expressed at 37° C., the target protein fused with LysRS or NP was found to have higher solubility than the control fused with MBP, and the other fusion proteins were found to being expressed in an insoluble form. In addition, when expressed at 27° C., the fusion proteins fused with C5 and Hsp15 showed solubility similar to that of the fusion protein fused with MBP, and the fusion protein fused with Ffh showed solubility lower than that of the fusion protein fused with MBP (**FIG. 1**).

EXAMPLE 3

[0055] Analysis of Effect of Co-expression of RNA Molecules with Fusion Partners on Solubility of a Target Protein

[0056] When co-expressing RNA molecules known to bind the fusion partners, effect of such co-expression on solubility of a target protein was evaluated, as follows. A vector expressing lysine tRNA binding to lysRS and a vector expressing a target protein PHM (42-384 amino acid residues of peptidylglysine alpha-monooxygenase) fused with lysRS were prepared. After co-expressing the two vectors, solubility of the fusion protein was analyzed.

[0057] First, the vector expressing lysine tRNA was constructed, as follows. tRNA gene was amplified by PCR using genomic DNA from JM109 cells as a template and primers designated SEQ ID NO.: 27 and SEQ ID NO.: 28. Separately, PCR was carried out using primers designated SEQ ID NO.: 29 and SEQ ID NO.: 30 to amplify T7 terminator region gene. The amplified tRNA gene and T7 terminator region gene were digested with SalI/NcoI and NcoI/SphI, respectively, and then ligated to a pLysE vector (Novagen) digested with SalI/SphI. The resulting vector was designated as "pT71ys-tRNA ".

[0058] Next, the vector expressing PHM fused to lysRS was constructed, as follows. A gene corresponding to a region ranging from 42 to 384 amino acid residues of rat peptidylglysine alpha-monooxygenase was amplified by PCR using PBSkrPHM(E) (Sean et al, Nature, 1999, 6, 976-983) as a template and primers designated SEQ ID NO.: 31 and SEQ ID NO.: 32. The PCR product was inserted into BamHI/HindIII sites at the multi cloning site (MCS) of pGE-lysRS vector, and the resulting vector was designated as "plysRS-PHM".

[0059] HMS174(DE3) cells were cotransfected with the pT71ys-tRNA vector and plysRS-PHM vector, and the resulting transformant was designated as "HMS174(plysRS-PHM+pT71ys-tRNA) ". Separately, HMS174(DE3) cells were cotransfected with the pMBP-PHM vector and pT71ys-tRNA vector, and the resulting transformant was designated as "HMS174(pMBP-PHM +pT71ys-tRNA)", which was used as a control. After incubating the transformants HMS174(plysRS-PHM+pT71ys-tRNA) and HMS174(pMBP-PHM+pT71ys-tRNA) at 37° C. and 30° C., respectively, protein expression was induced according to the same method as in Example 1, and then solubility of fusion proteins was evaluated. Herein, because MBP-PHM fusion protein was found to be mainly expressed in an

insoluble form at 37° C., to provide an expression environment similar to that of LysRS-PHM, the MBP-PHM fusion protein was expressed at 30° C. Also, HMS174(pMBP-PHM) and HMS174(pMBP-PHM) transformants were used as negative controls, which were not cotransfected with the pT71ys-tRNA vector.

[0060] As a result, when being co-expressed with lys-tRNA, LysRS-PHM fusion protein was found to have 10% higher solubility than when expressed with no co-expression of lys-tRNA, while MPB-PHM fusion protein showed similar solubility when expressed with or without co-expression of lys-tRNA (**FIG. 2**). These results indicate that lysine tRNA increases solubility of a fusion protein fused to LysRS protein by specifically binding to the LysRS protein, in agreement with the Reaction Formula 3.

EXAMPLE 4

[0061] Analysis of Effect of an RNA Molecule on Folding Efficiency of a Target Protein when a Fusion Partner is Allowed to Bind to the RNA Molecule

[0062] It was demonstrated in Example 3 that binding of RNA molecules to the fusion partners increases solubility of target proteins. In this test, the effect of such binding of RNA molecules on protein refolding to an active form was investigated, as follows. A fusion protein was used, which was prepared by linking a luciferase gene to a gene encoding Lys N, which is the the N teminal domain of LysRS specifically binding the anticodon of lysine tRNA. The luciferase gene was amplified by PCR using pGL2-Basic vector (Promega) as a template and primers designated SEQ ID NO.: 33 and SEQ ID NO.: 34. The amplified luciferase gene was inserted into BamHI/HindIII sites of pGE-lysN vector (Korean Pat. No. 203919), the resulting vector was designated as "pLysN-firefly luciferase". Thereafter, the pLysN-firefly luciferase vector was introduced into HMS174(DE3)plysE cells to express LysN-luciferase fusion protein, where the fusion protein was expressed as inclusion bodies. After being washed with PBS containing 200 mM NaCl, 1 mM EDTA and 1% triton X-100 three times, and with distilled water three times, the inclusion bodies were solubilized in PBS containing 6 M guanidium HCl and 2 mM DTT, and then diluted with 100 times of a refolding buffer containing 20 mM KCl, 3 mM $MgCl_2$, 2 mM DTT and 0.1 mg/ml BSA. The diluted solution was analyzed for luciferase activity using a fireflyluciferase assay kit (Promega) at 30° C. at 10, 20, 40 and 80 min in the presence of lysine tRNA or phenylalanine tRNA. Herein, phenylalanine tRNA was used as a control because the anticodon of phenylalanine tRNA is opposite to the lysine codon, and the anticodon of lysine tRNA is required for recognition by LysN.

[0063] As a result, when the LysN-luciferase fusion protein was present with lysine tRNA, luciferase activity was higher than in the presence of phenylalanine tRNA (**FIG. 4**), indicating that refolding of the luciferase enzyme to an active form takes place. These results demonstrate that binding of RNA molecules to the fusion partners leads to increased folding efficiency and solubility of target proteins.

[0064] In addition, despite the fact that LysRS is known to bind to lysine tRNA, in order to investigate whether the LysRS-PHM fusion protein actually binds to lysine tRNA, and such binding induces refolding of the target protein into an active form, binding of LysRS-PHM to lysine tRNA was estimated. RNA binding was analyzed by the method for analyzing tRNA synthetase activity, that is, the aminoacylation charging assay. After isolating LysRS protein and LysRS-PHM fusion protein from cells (**FIG. 5A**), their RNA binding activity was evaluated, where each of 0.4 μM LysRS and 0.4 μM LysRS-PHM was added to 100 μl of a buffer containing 150 mM KCl, 2 mM ATP, 0.1 mM EDTA, 7 mM $MgCl_2$ and ^{14}C-labelled lysine, and incubated at 37° C. 10 μl samples were collected from the reaction solution at 10, 20, 40 and 80 min, and spotted on a Whatman filter. Radioactivity was measured using a Beckman scintillation counter.

[0065] As a result, when compared to the LysRS protein, the LysRS-PHM fusion protein showed high tRNA synthetase activity (**FIG. 5B**). This result indicates that the LysRS-PHM fusion protein, like the LysRS protein, has an ability to bind to lysine tRNA, and that binding of the LysRS-PHM fusion protein to lysine tRNA increases solubility and refolding efficiency of the target protein.

[0066] As described hereinbefore, when a target protein is fused with an RNA-binding protein, and expressed with an RNA molecule specifically recognized by the RNA-binding protein, the fusion protein binds to the RNA molecule, and the net negative charge of the RNA molecule improves solubilization of the target protein and its refolding to an active form, leading to increased recovery of the target protein. Therefore, the method of producing a target protein as a fused protein with an RNA-binding protein is very useful for production of proteins for medical and industrial applications.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 34

<210> SEQ ID NO 1
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Forward primer for lysyl tRNA synthetase

<400> SEQUENCE: 1

gactaccata tgtctgaaaa cacgcacagg gcgct 35

-continued

<210> SEQ ID NO 2
<211> LENGTH: 78
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Reverse primer for lysyl tRNA synthetase

<400> SEQUENCE: 2

gactccaagc ttgtcgacga tatcggatcc ggtacccttg tcatcgtcat cttttaccgg 60

acgcatcgcc gggaacag 78

<210> SEQ ID NO 3
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Forward primer for tyrosyl tRNA synthetase

<400> SEQUENCE: 3

gtcatccata tggcaagcag taacttgatt aaacaattgc 40

<210> SEQ ID NO 4
<211> LENGTH: 46
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Reverse primer for tyrosyl tRNA synthetase

<400> SEQUENCE: 4

gactacggta cctttccagc aaatcagaca gtaattcttt ttaccg 46

<210> SEQ ID NO 5
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Forward primer for trptophane tRNA synthetase

<400> SEQUENCE: 5

gtcatccata tgactaagcc catcgttttt agtggc 36

<210> SEQ ID NO 6
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Reverse primer for trptophane tRNA synthetase

<400> SEQUENCE: 6

gtcatcggat ccacgcttcg ccacaaaacc aatcgc 36

<210> SEQ ID NO 7
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Forward primer for S1

-continued

<400> SEQUENCE: 7

gtcatccata tgactgaatc ttttgctcaa ctc 33

<210> SEQ ID NO 8
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Reverse primer for S1

<400> SEQUENCE: 8

gtcatcggat ccctcgcctt tagctgcttt gaaagc 36

<210> SEQ ID NO 9
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Forward primer for S4

<400> SEQUENCE: 9

gtcatccata tgcagggttc tgtgacagag 30

<210> SEQ ID NO 10
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Reverse primer for S4

<400> SEQUENCE: 10

gtcatcggat ccaattcggg tagaagccgg cac 33

<210> SEQ ID NO 11
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Forward primer for S17

<400> SEQUENCE: 11

gtcatccata tgaccgataa aatccgtact ctg 33

<210> SEQ ID NO 12
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Reverse primer for S17

<400> SEQUENCE: 12

gtcatcggat cccagaaccg ctttctctac aac 33

<210> SEQ ID NO 13
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: Description of Artificial Sequence: Forward primer for Hsp15

<400> SEQUENCE: 13

gtcatccata tgaaagagaa a 21

<210> SEQ ID NO 14
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Reverse primer for Hsp15

<400> SEQUENCE: 14

gtcatcggat ccttcactgt cgccgtgttt aaatcg 36

<210> SEQ ID NO 15
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Forward primer for DbpA

<400> SEQUENCE: 15

gtcatccata tgacgccggt gcaggccgcc 30

<210> SEQ ID NO 16
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Reverse primer for DbpA

<400> SEQUENCE: 16

gtcatcggat ccttttaata accgcacccg gc 32

<210> SEQ ID NO 17
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Forward primer for Ffh

<400> SEQUENCE: 17

gtcatccata tgtttgataa tttaaccgat cg 32

<210> SEQ ID NO 18
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Reverse primer for Ffh

<400> SEQUENCE: 18

gtcatcggat ccgcgaccag ggaagcctgg 30

<210> SEQ ID NO 19
<211> LENGTH: 36
<212> TYPE: DNA

-continued

<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Forward primer for c5

<400> SEQUENCE: 19

gtcatccata tggttaagct cgcatttccc agggag 36

<210> SEQ ID NO 20
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Reverse primer for C5

<400> SEQUENCE: 20

gtcatcggat ccggacccgc gagccaggcg ac 32

<210> SEQ ID NO 21
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Forward primer for NP

<400> SEQUENCE: 21

gtcatcgtca tccatatggc gtctcaaggc accaaacgat c 41

<210> SEQ ID NO 22
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Reverse primer for NP

<400> SEQUENCE: 22

gtcatcggta ccattgtcgt actcctctgc attgtctcc 39

<210> SEQ ID NO 23
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Forward primer for MBP

<400> SEQUENCE: 23

gtcatgcata tggaagaagg taaactggta atctgg 36

<210> SEQ ID NO 24
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Reverse primer for MBP

<400> SEQUENCE: 24

gtcatgggta cccttggtga tacgagtctg cgc 33

<210> SEQ ID NO 25

-continued

<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Forward primer for tobacco etch virus protease

<400> SEQUENCE: 25

gtcataggat ccggagaaag cttgtttaag ggg 33

<210> SEQ ID NO 26
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Reverse primer for tobacco etch virus protease

<400> SEQUENCE: 26

gtcatcgtcg acttattaat tcatgagttg agtcgcttcc 40

<210> SEQ ID NO 27
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Forward primer for lysine tRNA gene

<400> SEQUENCE: 27

gtcatcgtcg actaatacga ctcactatag ggtcgttagc tcagttggta gag 53

<210> SEQ ID NO 28
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Reverse primer for lysine tRNA gene

<400> SEQUENCE: 28

gtcatcccat ggtggtgggt cgtgcaggat tcg 33

<210> SEQ ID NO 29
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Forward primer for T7 terminator region

<400> SEQUENCE: 29

gtcatcccat ggctagcata acccccttggg gcctctaaac gggtc 45

<210> SEQ ID NO 30
<211> LENGTH: 48
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Reverse primer for T7 terminator region

-continued

<400> SEQUENCE: 30

gtcatcgcat gccaaaaaac ccctcaagac ccgtttagag gccccaag 48

<210> SEQ ID NO 31
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Forward primer for rat peptidylglycine alpha monooxygenase

<400> SEQUENCE: 31

gatatcggat cctcattttc caatgaatgc cttggtacc 39

<210> SEQ ID NO 32
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Reverse primer for rat peptidylglycine alpha monooxygenase

<400> SEQUENCE: 32

atactcaagc ttctagacag gaattgggat attggcctc 39

<210> SEQ ID NO 33
<211> LENGTH: 60
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Forward primer for luciferase

<400> SEQUENCE: 33

gtcacgggat cccaccacca ccaccacatg gaagacgcca aaaacataaa gaaaggcccg 60

<210> SEQ ID NO 34
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Reverse primer for luciferase

<400> SEQUENCE: 34

gtcacgaagc ttttacacgg cgatctttcc gcc 33

1. A method of producing a target protein having improved solubility and folding efficiency, comprising expressing a target protein as a fusion protein using an RNA-binding protein as a fusion partner, and binding the fusion partner fused to the target protein to an RNA molecule.

2. The method as set forth in claim 1, wherein the fusion partner of the target protein is an RNA-binding protein or polypeptide selected from the group consisting of mRNA, tRNA, rRNA, nuclear RNA, viral RNA and ribo-polynucleotides prepared by genetic recombination techniques.

3. The method as set forth in claim 1, wherein the fusion partner of the target protein is a protein selected from the group consisting of aminoacyl-tRNA synthetases, ribosomal proteins, mRNA-binding proteins, viral proteins having RNA-binding ability and proteins associated with cellular RNA processing and turnover, or a polypeptide corresponding to an RNA-binding region of the aforementioned proteins.

4. The method as set forth in claim 3, wherein the fusion partner of the target protein is a protein selected from the group consisting of C5 protein of ribonuclease P (RNase P), Ffh protein of signal recognition particle, NP protein of influenza virus, ribosomal S1 protien, ribosomal S4 protein, ribosomal S17 protien, *E. coli* DbpA and *E. coli* Hsp15, or a polypeptide corresponding to an RNA-binding region of the aforementioned proteins.

5. The method as set forth in claim 1, wherein the fusion partner of the target protein binds to an RNA molecule naturally present in cells, or an artificially co-expressed RNA molecule.

6. The method as set forth in claim 5, wherein the RNA molecule is artificially overexpressed by constructing a vector expressing said RNA molecule and then introducing the vector into a host cell.

7. The method as set forth in any of claims 1 to 6, wherein the method comprises the steps of:

1) constructing an expression vector encoding a fusion protein using a target protein and an RNA-binding protein as a fusion partner of the target protein;

2) constructing an expression vector expressing an RNA molecule capable of binding to the RNA-binding protein fused with the target protein; and

3) cotransforming a host cell with the expression vectors prepared in Steps 1 and 2.

* * * * *