

Construction of Recombinant Expression Vectors to Study the Effect of Thioredoxin on Heterologous Protein Solubility

Liz Geum, Reed Huber, Nicole Leung, Matthew Lowe
Department of Microbiology and Immunology, University of British Columbia

The pET-32 vector series is designed for high-level expression of recombinant thioredoxin fusion proteins. Previous studies have shown that linkage to thioredoxin increases the yield of biologically active proteins expressed in *Escherichia coli*. How thioredoxin promotes protein solubility is not known, although a catalytic contribution involving the formation of disulphide bonds in the *E. coli* cytoplasm has been proposed. In this study, we have cloned the proteinase inhibitor II (PI2) gene from potato into wild type pET-32a and previously created pET-32a Δ trx vectors. PI2 was chosen as a model protein to study the oxido-reductase activity of thioredoxin since it contains eight disulphide bonds. These new vectors allow comparison of the solubility of overexpressed PI2 fused to thioredoxin to apo-PI2. To our knowledge, this was the first attempt to express PI2 in a prokaryotic system. SDS-PAGE analysis did not detect induction of the thioredoxin-PI2 fusion protein or the apo-PI2 protein, leading us to speculate that PI2 may not be a suitable protein for high level protein expression studies using the pET-32 vector series in *E. coli* strain BL21.

High-level expression of recombinant proteins in *Escherichia coli* has many useful applications in biological research and is a hallmark of the biotechnology industry. However, efficient expression of recombinant proteins is difficult to achieve as a number of obstacles are frequently encountered when expressing eukaryotic genes in *E. coli*. Additional problems beyond expression may arise when the physical characteristics of heterologous proteins are not compatible with the prokaryotic system. For example, many proteins require post-translational modification such as glycosylation and disulfide cross-linking in order to remain soluble and to fold properly into their native state. Unfortunately, the cytoplasm of *E. coli* is not conducive to these stabilizing influences, often leading to formation of insoluble aggregates of heterologous proteins called inclusion bodies (1).

Recent strategies devised to circumvent challenges associated with inclusion bodies have centered around the production of a fusion protein by linking a gene which is known to be expressed well in *E. coli* to the gene of interest. This method however is not infallible as fusion proteins are frequently expressed as insoluble inclusion bodies, and when soluble isolation is possible, an additional purification step is necessary in order to separate the protein of interest from its fusion partner. One such example is the pET-32 vector series commercialized by Novagen for high-level expression of peptide sequences fused with the thioredoxin (*trxA*) protein (2).

Although the mechanism by which thioredoxin is able to contribute to the increased solubility of certain foreign proteins is not well understood, the prevailing belief is that two cysteine residues in the active site function as a redox switch catalyzing the transfer of a disulfide bond between a substrate protein and thioredoxin (3). In the present study, we have constructed two modified pET-32a(+) vectors that are designed to assess the utility of the Trx□TagTM as a protein solubilizer. We have capitalized on previous work done by Wong (4) using the SW057 vector in which the *trxA* coding sequence has been removed, and have cloned the

proteinase inhibitor II (PI2) gene into both SW057 and wild type pET-32 vectors.

PI2 from potatoes is an important protein from an agricultural standpoint since it has been shown to protect plants from pathogens and ultraviolet radiation (5, 6). For the purpose of our study, PI2 is useful for investigating thioredoxin activity due to its relatively small size, 148 amino acids (18 kDa) and because it contains eight disulphide bonds in its native state. Importantly, we were able to resolve some difficulties encountered by previous groups (7, 8) in creating recombinant vectors containing the PI2 gene. Thus, we have provided a useful platform for directly assessing the utility of thioredoxin as a fusion partner for over-expression of a heterologous protein requiring disulfide cross-linking for biological activity.

MATERIALS AND METHODS

Bacterial growth conditions and plasmid isolation. DH5 α *E. coli* strains used in this study were chosen according to plasmid identity as outlined in Table 1 and BL21 (DE3) *E. coli* were used for protein expression. All *E. coli* strains were obtained from the culture collection of the Department of Microbiology and Immunology at the University of British Columbia (UBC). Cultures were grown overnight in Luria-Bertani (LB) broth (10 g/L tryptone, 5 g/L yeast extract, 10 g/L sodium chloride) containing 100 μ g/ml Ampicillin (Gibco) at 37°C in quarter-full 5 ml glass tubes. Plasmid isolation was performed using the PureLink® Quick Plasmid Miniprep Kit (Life Technologies, Cat. No. K2100-10). DNA concentrations were determined using the NanoDropTM spectrophotometer (Thermo Scientific).

TABLE 1 Ampicillin-resistant DH5 α used in this study

Strain Name	Plasmid	Source
<i>E. coli</i> DH5 α	pET-32a(+)	Novagen (2)
SW057	pET-32a(+) Δ trx	UBC (4)
<i>E. coli</i> DH5 α	pE32-PI2	UBC
<i>E. coli</i> DH5 α	pEmoA	UBC
MRLN03	pET-32a(+) Δ trx/PI2+	This study
MRLN04	pET-32a(+)/PI2+	This study

Amplification of PI2. A 590bp fragment containing the PI2 coding sequence was amplified from the pE32-PI2 plasmid in *E. coli* (Table 1). MRLNPI2-F and MRLNPI2-R forward and reverse primers were designed for in frame ligation into NcoI and EcoRI restriction sites, respectively within the multiple cloning site (MCS) in pET-32a(+) and SW057 vectors. The PCR conditions were set up according to (7), though we found that use of Platinum® Pfx Polymerase (Life Technologies) resulted in a cleaner PCR product when visualized by gel electrophoresis. The touchdown PCR used involved an initial 4 minute denaturation step at 94°C, followed by annealing and elongation at 68°C, annealing temperatures in steps 1 to 7 decreased by 1°C from 58°C to 51°C, and then the last was repeated 17 times. PCR products were visualized using a 1.5% (w/v) agarose gel dissolved in 0.5x TBE (54 g of Tris base, 27.5 g of boric acid, and 20 ml of 0.5 M EDTA made up to 500 ml in distilled water and diluted 1 in 10 for working concentration). PCR products were loaded by mixing 5 or 10 µl of sample with 1 or 2 µl of loading dye (Thermo Scientific), respectively. High-mass DNA ladder (Invitrogen), low-mass DNA ladder (Invitrogen) and O'GeneRuler DNA ladder (Thermo Scientific) were used for determining amplicon sizes. Gels were run at 200 V for 40 minutes and stained in ethidium bromide (0.5 µg/ml) for 10 minutes. Samples containing the predicted 590 bp product were purified using the PureLink® PCR purification kit from Invitrogen (Cat. No. K3100-01) and stored at -20°C.

Preparation of CaCl₂ competent cells. DH5α and BL21 (DE3) *E. coli* were cultured overnight in LB broth and then diluted 1 in 10 and grown to an OD₅₅₀ of 0.4. Cultures were centrifuged at 5000 x g using the Beckman JA-10 rotor at 4°C for 15 minutes to pellet cells. The supernatant was decanted and pellets were resuspended in 5 ml of ice-cold MgCl₂ (0.1 M). Second and third washes with MgCl₂ (0.1 M) were done following a 15 minute 3000 x g centrifugation and a 10 minute 2500 x g centrifugation, respectively. If competent cells were to be stored for later use, then 15% glycerol (v/v) was added and 50 µl aliquots were stored at -80°C.

Double digests and ligations to generate MRLN03 and MRLN04 constructs. PI2 PCR products and pET-32a(+) and SW057 plasmid backbones were digested with EcoRI and NcoI restriction enzymes (both New England Biolabs) in 1x EcoRI Buffer (New England Biolabs), in which both enzymes retain 100% enzymatic activity. Double digest samples were set up with 1 µg of DNA, 5 µl of 10x EcoRI Buffer, 0.5 µl of EcoRI, 1 µl of NcoI, and distilled water to 50 µl reaction volumes. Following the double digest, enzymes were heat inactivated for 20 minutes at 65°C on a heating block. Digested products were ligated with T4 DNA ligase (New England Biolabs) using a 5:1 insert DNA to vector DNA ratio. 20 µl reaction volumes included 1 µl T4 DNA ligase, 2 µl 10x Ligase Buffer (New England Biolabs), 55 ng plasmid DNA, and 27 ng PI2 insert DNA in distilled water. Tubes were incubated at room temperature for 16 to 24 hours, then enzyme was heat inactivated as per above and products were either stored at -20°C or visualized by gel electrophoresis.

Transformation, screening and colony PCR. Competent DH5α and BL21 (DE3) *E. coli* were transformed with plasmid DNA using the Hanahan protocol as outlined in Green and Sambrook (9). Dilute and concentrated 100 µl aliquots of the transformation reaction was added to LB-Amp agar and the spread plate method was used for screening of isolated colonies. After an overnight incubation at 37°C colonies were picked and plasmid DNA was analyzed for the presence of the PI2 gene using the same conditions and method as described above for the initial amplification of the coding sequence. Samples from individual colonies were run on a 1.5% (w/v) agarose gel for detection of a 590 bp band.

SDS-PAGE for detection of apo-PI2 and PI-thioredoxin proteins. BL21 (DE3) *E. coli* were transformed with our putative MRLN03 and MRLN04 vectors by the Hanahan protocol (9) and isolated colonies were obtained using the same method as for colony PCR. Isolated colonies were used to inoculate tubes containing 5 ml of LB broth and following an overnight incubation period samples were diluted 1 in 10 and grown to an OD₅₅₀ of 0.4. The culture was then split into two equal volumes and isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a concentration of 100 mM in one of the samples. IPTG-induced and uninduced samples were then incubated overnight. Sample concentrations were normalized by measuring optical density (OD_{600 nm}), then two to three centrifugation steps were done to remove all LB media and cell pellets were resuspended in 1 ml of distilled water. Whole cell lysate and soluble fractions were prepared for all samples. For the whole cell lysates, cells were resuspended in 4x SDS Loading Buffer (2 ml 1 M Tris-HCl pH 6.8, 0.8 g SDS, 4 ml 100% glycerol, 0.4 ml 14.7 M β-mercaptoethanol, 1 ml 0.5 M EDTA, and 8 mg of Bromophenol Blue), and soluble protein samples were prepared using the FastPrep-24™ 5G instrument (MP Biomedicals) by adding the sample to 2 ml Lysing Matrix tubes and running the machine for 30 seconds. The supernatant containing the soluble fraction was collected from the tubes and 4x SDS Loading Buffer was added.

A 14% SDS-PAGE gel (see Table S1) was prepared for analysis of PI2 constructs. Prior to loading the wells, the samples were heated for 10 minutes at 80°C in a water bath. The gel was run at 50 V for 45 minutes, or until the samples reached the end of the stacking gel. The gel was then run at 150 V for 1 to 2 hours, or until the loading dye ran off the resolving gel and the SDS-PAGE MW ladder (Life Technologies) visibly separated.

RESULTS

Construction and characteristics of MRLN vectors. Our vectors contain the PI2 coding region cloned into wild type and modified pET-32aΔ*trx* vectors. The amino acid sequence of the PI2 gene is shown in Figure 1A with 16 cysteine residues highlighted which in the functional state of the protein fold together to form eight disulphide bonds. We have directionally cloned the PI2 coding region into unique NcoI and EcoRI sites in wild type pET-32a and SW057 vectors for downstream protein expression studies.

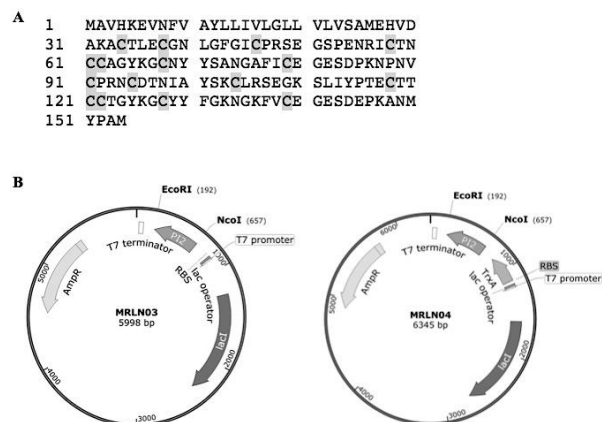


FIG 1 Plasmid maps and important sequence elements of our MRLN vectors. (A) PI2 amino acid sequence with 16 cysteine residues highlighted. (B) Left, MRLN03 (pET-32aΔ*trx*/PI2+) and right, MRLN04 (pET-32a/PI2+).

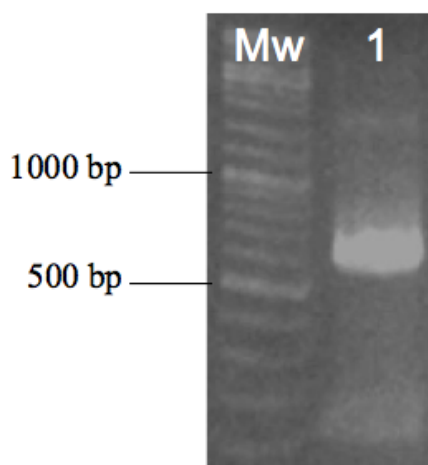


FIG 2 PCR amplification of PI2 from pE32-PI2 plasmid. Lane 1 shows a large amount of the expected 590 bp PCR product using the optimized protocol which included Platinum® *Pfx* polymerase, no DMSO and a touchdown protocol. Mw is GeneRuler DNA ladder. Data are representative of three experiments (n = 3).

We have named our pET-32aΔ*trx*/PI2+ and pET-32a/PI2+ vectors MRLN03 and MRLN04, respectively, and Figure 1B shows the plasmid map of both vectors as well as important sequence information. Expression of apo-PI2 and PI2-thioredoxin fusion proteins are under control of T7 promoters whose activity can be modulated by IPTG induction. Both vectors also contain the *bla* coding sequence for selection of transformed cells on ampicillin and the PI2 gene was cloned in frame of N terminal 6xHis- and S-Tags for production of proteins containing cleavable His□Tag® and S□Tag® sequences for immunodetection and purification (Fig S1 and S2).

Amplification of PI2 from pE32-PI2 vector DNA. The PI2 coding region was amplified using PCR for cloning of the gene into bacterial overexpression vectors. We designed primers with 5' and 3' extra bases encoding *Nco*I and *Eco*RI restriction sites, respectively (Fig S3). Our initial attempt using *Taq* polymerase at an annealing temperature of 51°C did not produce any specific product, but when 5% DMSO (v/v) was included as a PCR additive a significant 590 bp band could be observed on an agarose gel as well as a couple larger non-specific bands (Fig S4). Using touchdown PCR and Platinum® *Pfx* polymerase which provides an automatic “hot start”, we were able to generate large bands corresponding to the size of our desired product (Fig 2). Furthermore, use of *Pfx* polymerase and a touchdown protocol eliminated the need for PCR additive (data not shown). The purpose of the amplification step was for expression cloning to yield recombinant vectors, which is considered to be a demanding PCR application thus it is not surprising that using reagents and a protocol aimed at providing high yields of specific product gave the best results.

Colony PCR confirms the identity of MRLN03 and MRLN04 clones. Following restriction digestion of the 590 bp PCR product and pET-32 and SW057 vectors and

ligation to generate our desired MRLN constructs, we transformed DH5a *E. coli* and performed colony PCR to assay for the presence of the PI2 gene. We screened 12 putative MRLN03 clones and 13 putative MRLN04 clones and a 590 bp product amplified using our PI2 primer set was detected in both corresponding to pET-32aΔ*trx*/PI2+ and pET-32a/PI2+ vector DNA (Fig 3). It appeared that our ligation was highly efficient since 10/12 (MRLN03) and 13/13 (MRLN04) of the colonies that were screened produced a clear 590 bp band (Figure S5). However this is not surprising since a double digest was performed using *Nco*I and *Eco*RI which prevented the vector backbones from re-circularizing.

Expression of MRLN03 and MRLN04 constructs for comparison of PI2 solubility. Given that our vector constructs contained recombinant PI2 DNA we wanted to see if the protein could be expressed in a prokaryotic system. BL21 (DE3) *E. coli* were transformed with MRLN03 and MRLN04 vectors and regulation of transcription was under control of the T7 promoter which was inducible by the addition of IPTG. We were expecting to see dark bands around 18 and 30 kDa corresponding to apo-PI2 and PI2-thioredoxin fusion proteins, respectively. Though, neither band appeared on the gel, and furthermore, there were no observable differences between induced and uninduced conditions for either of our transformed BL21 (DE3) *E. coli* constructs. These results indicate that the PI2 gene on both vectors was not expressed *in vivo*. Previous studies involving PI2 expression were done in mammalian cells so differences in codon usage between *E. coli* and eukaryotic genes or the requirement for translation initiation to occur in the context of rough endoplasmic reticulum which is not available in prokaryotic cells, might explain the lack of detectable PI2 in our BL21 constructs. Alternatively, a point mutation may have occurred in our selected MRLN03 and MRLN04 vectors which could be confirmed by sequencing the putative PI2 coding region on both plasmids.

DISCUSSION

Despite the widespread commercial application of the pET-32 vector series for high-level expression of heterologous proteins as thioredoxin fusions in the *E. coli* cytoplasm, a clear biological answer that explains why thioredoxin is such a good fusion partner has not been provided. Nonetheless, a few non-mutually exclusive suggestions have garnered attention. It is possible that the physical linkage of a heterologous protein to a highly soluble fusion partner may prevent insoluble aggregates from forming, allowing proper folding to occur. The N-terminal Trx□Tag™ may restrict inclusion body formation since it has been shown that thioredoxin remains soluble even when it is expressed at unnaturally high levels (10). Another idea is that overexpression of thioredoxin fusions can change the redox environment of the *E. coli* cytoplasm since it is well known that the balance between thioredoxin reductase (*trx*B) and thioredoxin is such that thioredoxin

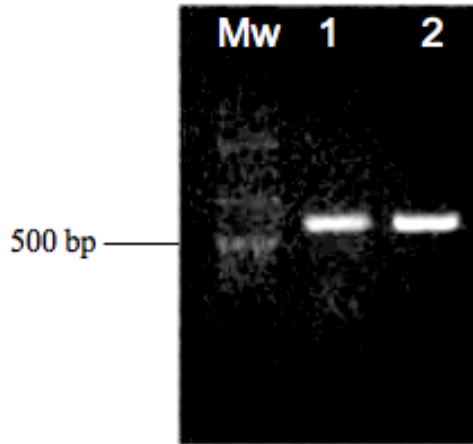


FIG 3 Colony PCR to confirm the presence of the PI2 gene from putative MRLN03 and MRLN04 constructs. In lane 1 the expected ligation product was pET-32a Δ trx/PI2+ (MRLN03) and the plasmid DNA was probed using our PI2 primers to amplify the expected 590 bp fragment. In lane 2 the expected ligation product was pET-32a/PI2+ (MRLN04) and a similar 590 bp fragment was detected using the same approach. Mw is GeneRuler DNA ladder. Data are from one experiment (n = 1).

is almost exclusively found in the reduced form and thus unable to catalyze the oxidation of cysteine residues (11). Thereby overexpression of thioredoxin may interfere with the natural balance to facilitate protein oxidation that would otherwise be non-permissive in a reducing environment. Furthermore, covalent linkage to the Trx□TagTM has been proposed to play a direct role in enhancing protein solubility by acting as an oxidoreductase enzyme to catalyze the formation of energetically favorable disulfide bonds in the fusion partner protein.

The aim of this study was to construct two recombinant pET-32a(+) plasmids to investigate the potential catalytic activity of thioredoxin. Previous groups have attempted to create plasmids containing the PI2 gene for assessing protein expression based on the proteins apparent lack of solubility (7, 8), however, these putative constructs were never confirmed and our sequencing results showed that the PI2 gene was missing (data not shown). We were able to create pET-32a Δ trx/PI2+ (MRLN03) and pET-32a/PI2+ (MRLN04) vectors which have been verified using colony PCR (Fig 3). Furthermore, PI2 insert DNA has been cloned in frame of N terminal 6xHis- and S-Tags which may be useful for measuring levels of specific protein either by SDS-PAGE, Western blot, or flow cytometry analysis.

We originally hoped to construct four MRLN vectors under the premise that thioredoxin would have different effects on the solubility of PI2 and EDTA monooxygenase (*emoA*) which contains no cysteine residues so disulphide cross-linking is not a factor in polypeptide folding. The idea was that by comparing differences in solubility of PI2 and *EmoA* thioredoxin

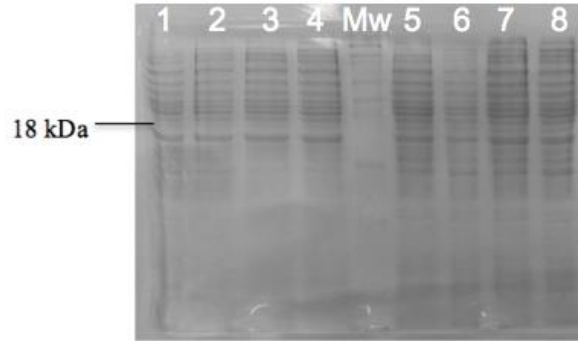


FIG 4 SDS-PAGE to compare solubility of apo-PI2 and PI2-thioredoxin fusion proteins. Both genes of interest on MRLN03 and MRLN04 vectors are under control of a T7 promoter that is regulated by LacI and thus are inducible by IPTG. Lanes 1 to 4 correspond to MRLN03; lane 1 uninduced whole cell lysate, lane 2 induced whole cell lysate, lane 3 uninduced soluble fraction, and lane 4 induced soluble fraction. Mw is protein molecular weight marker. Lanes 5 to 8 correspond to MRLN04; lane 5 uninduced whole cell lysate, lane 6 induced whole cell lysate, lane 7 uninduced soluble fraction, and lane 8 induced soluble fraction. Apo-PI2 and PI2-thioredoxin proteins are 18 kDa and 30 kDa respectively, however, there were no apparent difference between induced and uninduced conditions signifying that expression of PI2 was not inducible from either vector. Data are from one experiment (n = 1).

fusions to corresponding apo-Trx□TagTM proteins, then the possible oxido-reductase contribution to enhanced protein solubility could be inferred. Though, upon closer inspection of our results we determined that our *emoA*-containing vectors are not suitable for downstream protein expression studies. It appears that the *emoA* fragment was ligated out of frame of the Trx□TagTM which typically can be corrected by using either pET-32b or pET-32c vectors which yield thioredoxin fusions in different reading frames. However, we made the error of designing our *emoA* fragment for directional ligation into the NcoI and EcoRI regions of the MCS and to our dismay the single base pair difference between the pET-32a-c vectors is between these two unique restriction sites. Therefore pET-32a-c vectors double digested with NcoI and EcoRI yield identical linearized vector products and corresponding 24 bp, 23 bp, and 25 bp excised fragments. In follow-up studies we do not recommend ligating *emoA* DNA into the NcoI restriction site and for protein expression purposes we would advise using a software tool for planning insertion of recombinant DNA into the correct reading frame.

The vector cartoon in Figure 1 shows the unique NcoI and EcoRI restriction sites in our putative vectors. To validate the expected ligation products, a double digest of both vectors was performed, however the expected 590 bp product was missing from both of our plasmid DNA samples (Figure S6). As a control, single digests of pET-32a and SW057 vectors were performed and the expected 5.9 kb and 5.6 kb products were observed, though a control for the double digest condition was not done because the expected single digest and double

digest products could not be differentiated using gel electrophoresis. Thus the apparent discrepancy between the results of the colony PCR and the double digest experiments has led us to conclude that a sequential digest may be necessary to excise the 590 bp PI2 fragment from our MRLN constructs. Furthermore, a previous study reported similar issues when attempting to excise the PI2 gene from a different vector by a double digest (8).

FUTURE DIRECTIONS

Follow-up studies intending to use MRLN03 and MRLN04 vectors should begin by troubleshooting the sequencing of both vectors to screen for any mutations. Sequencing should be attempted using DMSO as an additive given that it was useful for PCR amplification of PI2. It would also be worthwhile to attempt sequential single digests of our vectors. Future research looking at protein expression levels may consider Western blot or flow cytometry using an anti-6x His□Tag® antibody to aid in detection of PI2. Furthermore, creating the corresponding *emoA* vectors to complement our PI2 vectors would be useful for uncovering the mechanism of thioredoxin.

ACKNOWLEDGEMENTS

This work was financially supported by the Department of Microbiology and Immunology at the University of British Columbia. We would like to thank Dr. David Oliver and Chris Deeg for their knowledge and support over the course of this project. We also wish to thank the Wesbrook media room staff for supplying the equipment needed to complete this experiment.

REFERENCES

1. **Mitraki A, King, J.** 1989. Protein folding intermediates and inclusion body formation. *Nat. Biotechnol.* **7**:690-697.
2. **Novy B, Berg J, Yaeger K, Mierendorf R.** 1995. Novagen in Novations. **3**:4
<http://www.emdbiosciences.com/docs/NDIS/inno03-003.pdf>.
3. **Collet J, Messens J.** 2010. Structure, function, and mechanism of thioredoxin proteins. *Antioxid. Redox Signaling.* **13**:1205-1216.
4. **Wong SH.** 2005. Cloning of flavin reductase into pET32a(+) expression vector lacking the thioredoxin A tag to study solubility of EDTA monooxygenase A in overexpression systems. *J. Exp. Microbiol. Immunol.* **8**:59-66.
5. **Bergey DR, Howe GA, Ryan CA.** 1996. Polypeptide signaling for plant defensive genes exhibits analogies to defense signaling in animals. *PNAS.* **93**:12053-12058.
6. **Huang C, Ma W, Ryan CA, Dong Z.** 1997. Proteinase inhibitors I and II from potatoes specifically block UV-induced activator protein-1 activation through a pathway that is independent of extracellular signal regulated kinases, c-Jun N-terminal kinases, and P38 kinase. *PNAS.* **94**:11957-11962.
7. **Park JE.** 2006. Generation of recombinant plasmids constructs to assess the ability of NADH:flavin oxidoreductase to solubilize proteinase inhibitor 2 in bacterial protein overexpression systems. *J. Exp. Microbiol. Immunol.* **10**:27-33.

8. **Duronio C.** 2012. Production of a recombinant vector to the study of thioredoxin function as a bound or detached solubilizer or proteinase inhibitor 2 in a bacterial protein overexpression system. *J. Exp. Microbiol. Immunol.* **16**:79-84.
9. **Green MR, Sambrook J.** 2012. *Molecular cloning: a laboratory manual*, 4th ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
10. **LaVallie ER, DiBlasio EA, Kovacic S, Grant KL, Schendel PF, McCoy JM.** 1993. A thioredoxin gene fusion system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Nat. Biotechnol.* **11**:187-193.
11. **Prinz WA, Åslund F, Holmgren A, Beckwith J.** 1997. The role of thioredoxin and glutaredoxin pathways in reducing protein disulphide bonds in *Escherichia coli* cytoplasm. *J. Biol. Chem.* **272**:15661-15667.