Expression of soluble heterologous proteins via fusion with NusA protein

Roger G. Harrison — School of Chemical Engineering and Materials Science, University of Oklahoma, Norman, Oklahoma

The production of proteins by genetically engineered *Escherichia coli* bacteria cells has become well established in biotechnology research and large-scale production. Cloning and expression of proteins in *E. coli* are favored in many instances because *E. coli* has relatively simple genetics, is well characterized, and has a relatively rapid growth rate. One disadvantage, however, of expressing heterologous proteins in *E. coli* is that the proteins are frequently expressed as insoluble aggregated folding intermediates, known as "inclusion bodies." In order to recover an active protein expressed in inclusion bodies, the protein that is insoluble must be solubilized with denaturants such as 8 M urea or 6 M guanidine hydrochloride, and then the denaturant must be removed under conditions that are optimal for protein folding. While progress has been made in the refolding of recombinant proteins (1, 2), the specific folding conditions regarding buffer composition, protein concentration, temperature, etc. must be optimized for every protein. Even for refolding processes that have been optimized, the yield of renatured protein may still be relatively low, requiring large volumes and significant cost for the preparation of large quantities of protein.

One approach to expressing heterologous proteins in soluble form has been to co-express molecular chaperones which aid in protein folding. Although there have been several instances of *E. coli* chaperones aiding the folding of heterologous proteins, unfortunately it is a trial and error process to determine the specific match between the target protein and the chaperone that will lead to correct folding (3, 4).

Another approach that has achieved success in recent years in producing soluble heterologous proteins in *E. coli* has been the use of gene fusions. The first fusion protein system specifically aimed at increasing solubility of a target protein was the one in which *E. coli* thioredoxin (TRX) was fused to the N-terminus of the target protein (5). Other carrier proteins that have been widely used in fusion proteins in *E. coli* are maltose binding protein (MBP) and glutathione S-transferase (GST). MBP and GST were chosen as carriers because they enable the fusion protein to be affinity purified: MBP binds to amylose, while GST binds to immobilized glutathione. While these carrier proteins have resulted in the successful overexpression of many heterologous proteins in *E. coli*, each was discovered empirically and certainly may not possess maximal solubilizing characteristics.

The new fusion protein systems that we describe here are based on a systematic evaluation to identify *E. coli* proteins that have the highest potential for solubility when overexpressed. A modified version of the solubility model of Wilkinson and Harrison (6) for recombinant proteins expressed in *E. coli* was used to determine the solubility potential of the more than 4000 *E. coli* proteins in the SwissProt protein database. Based on this solubility model, we identified three *E. coli* proteins--BFR, GrpE, and NusA--that gave a significantly higher level of solubility when expressed as a fusion with the target heterologous protein human interleukin-3 (hIL-3) compared to the fusion of TRX to hIL-3. NusA expressed as a fusion with hIL-3 gave the highest level of solubility and hIL-3 expression of the carrier proteins tested. NusA fused to human interferon-γ (hIFN-γ), bovine growth hormone (bGH), or tyrosinase from *Rhizobium meliloti* was also expressed in nearly completely soluble form and at high levels. This article elaborates on these new findings.

Recombinant protein solubility modeling

In the original model of Wilkinson and Harrison (6) for the solubility of recombinant proteins expressed in *E. coli*, five amino acid-based parameters were correlated with inclusion body formation. It has since been discovered, however, that only two of the parameters are critical in distinguishing soluble versus insoluble protein expression (7). The critical parameters are the approximate charge average, which accounts for the differences in the numbers of Asp plus Glu versus Lys plus Arg residues, and the turn-forming residue content, which accounts for the total number of Asn, Gly, Pro, and Ser residues. Incorporation of the two-parameter solubil-

Table 1. Predicted solubilities of carrier and target proteins

Protein	MW (kDa)	Amino Acid Length	Probability of Solubility or Insolubility[1]
NusA	54.8	495	95% soluble
BFR	18.5	158	95% soluble
GrpE	21.7	197	92% soluble
thioredoxin	11.7	109	73% soluble
hIL-3	15.1	133	73% insoluble
bGH	21.6	189	60% insoluble
hIFN-γ	17.1	146	58% insoluble
tyrosinase	54.1	494	51% soluble

1. The revised Wilkinson-Harrison solubility model involves calculating a canonical variable (CV) or composite parameter for the protein for which the solubility is being predicted. The canonical variable in the two-parameter model is defined as:

$$CV=\lambda_1\left(\frac{N+G+P+S}{n}\right)+\lambda_2\left|\left(\frac{(R+K)-(D+E)}{n}-0.03\right)\right|$$

n = number of amino acids in protein
N,G,P,S = number of Asn, Gly, Pro, or Ser residues, respectively.
R,K,D,E = number of Arg, Lys, Asp, or Glu residues, respectively.
λ_1, λ_2 = coefficients (15.43 and -29.56, respectively)
The probability of the protein solubility is based on the parameter CV-CV', where CV' is the discriminant equal to 1.71. If CV-CV' is positive, the protein is predicted to be insoluble, while if CV-CV' is negative, the protein is predicted to be soluble. The probability of solubility or insolubility can be predicted from the following equation:

$$\text{Probability of solubility or insolubility} = 0.4934 + 0.276\,|(CV-CV')| - 0.0392\,(CV-CV')^2$$

ity model into a custom C computer program allowed the rapid evaluation of all known *E. coli* protein sequences. As input, this program used a single text file created by the SwissProt server (http://www.expasy.ch), which contained approximately 4000 *E. coli* protein sequences.

E. coli proteins identified by the solubility model which possessed a solubility probability of greater than 90% were considered for fusion protein experiments. The list of possible carrier proteins was then reduced to those which were stable when expressed in *E. coli* as shown in previous studies. Next, choices were further restricted to those containing more than 100 amino acids to ensure that predicted solubility characteristics would not be simply limited by the molecular weight of the carrier protein relative to the target protein chosen.

Three *E. coli* proteins, NusA, BFR (bacterioferritin), and GrpE, were identified by the solubility model as having a very high probability (> 90%) of being soluble when expressed in *E. coli*. All three of these proteins were evaluated as N-terminal fusions with human interleukin-3 (hIL-3), which has been found in inclusion bodies both when expressed alone (8) and as a fusion protein with thioredoxin (5). Bovine growth hormone (bGH) and human interferon-γ (hIFN-γ) were selected for fusion with NusA at their N-terminus because they had previously been expressed alone in inclusion bodies (9, 10). Tyrosinase from *Rhizobium meliloti* was not able to be overexpressed by itself in *E. coli* but was expressed as a fusion protein with NusA connected to its N-terminus (11); this tyrosinase is interesting because of its large molecular weight (54 kDa) and large number of rare codons. The predicted solubilities of the carrier and target proteins studied, including the carrier thioredoxin as a control, are shown in Table 1.

Evaluation of solubility of *E. coli* fusion proteins

The results of overexpression of fusion proteins in *E. coli* at 37°C are shown in Figures 1, 2 and 3. The SDS-PAGE and Western blot results show that all fusion proteins containing hIL-3 were expressed at a high level with respect to the percent of the total cell protein, namely: NusA, 18.4%; GrpE, 7.4%; BFR, 13.7%; and thioredoxin, 8.5% (see fig. 1). The highest soluble expression level was achieved by the NusA/hIL-3 protein (97% soluble), while the thioredoxin/hIL-3 fusion protein was expressed almost completely in the inclusion body fraction (8% soluble). Thus, the percentage solubility of the NusA/hIL-3 fusion protein was over 12 times that of the thioredoxin/hIL-3 fusion protein.

The Western blot (see fig. 1, lower panel) shows, with more clarity, the distribution of hIL-3 among the soluble and insoluble fractions. BFR and GrpE are distributed between the soluble and insoluble fractions. Upon centrifugation of BFR/hIL-3 cultures at the end of induction periods, it was noticed that the cell pellet had a slightly red tint compared to uninduced cultures. This is presumably due to the iron binding properties of BFR (12) and serves as a convenient marker for confirmation of protein expression. One striking finding from the Western blot data (fig. 1) is that the level of expression of hIL-3 in the soluble fraction was higher in the NusA fusion than in the other fusions. Thus, the large size of NusA (55 kDa) was not at all limiting in being able to express high levels of hIL-3.

In order to determine if the hIL-3 present in each of the fusion proteins was biologically active, indicating that hIL-3 was properly folded, a cell proliferation assay was performed on each fusion protein in soluble cell lysate. hIL-3 activity was found to be present in all fusion proteins, with 67% of the hIL-3 determined to be active for the NusA/hIL-3 fusion protein, which was the highest of the fusion proteins tested (7).

For the expression of the NusA/bGH and NusA/hIFN-γ fusion proteins (see fig. 2), it is clear from SDS-PAGE and Western blot results that the vast majority of these fusion proteins are expressed in the soluble fraction (estimated as 89% for NusA/bGH and 87% for NusA/hIFN-γ). To our knowledge, this is the first evidence of soluble expression of bGH in *E. coli*.

The fusion of NusA with the relatively large tyrosinase (54 kDa) was expressed in almost completely soluble form (>90%) based on SDS-PAGE analysis (see fig. 3). Thus, NusA may be a very good carrier protein for solubilizing large target proteins. In

continued on page 6

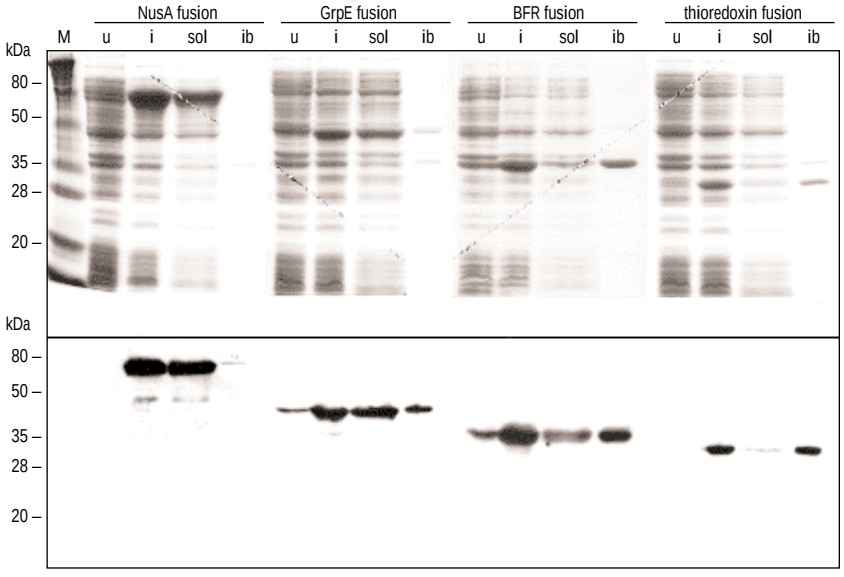


Figure 1. Analysis of fusion proteins containing hIL-3

SDS-PAGE (upper panel) and Western blot (lower panel) analyses show NusA, GrpE, BFR, and thioredoxin fusion proteins containing hIL-3. Equal portions of cell lysate, soluble fraction, and insoluble fraction were loaded. Fusion proteins were expressed from plasmid pKK223-3 under control of the *tac* promoter in *E. coli* JM105 at 37°C. Cells were induced with 1 mM IPTG and grown for 3 h post-induction. The Western blot was probed with mouse anti-hIL-3 monoclonal antibody and visualized using chemiluminescence. Percentage solubility based on the blots (density of soluble band divided by the density of the soluble plus insoluble bands): 97%, NusA/hIL-3; 71%, GrpE/hIL-3; 47% BFR/hIL-3; and 8% thioredoxin/hIL-3. u = uninduced whole cell lysate, i = induced whole cell lysate, sol = soluble fraction, ib = inclusion body fraction.

continued from page 5

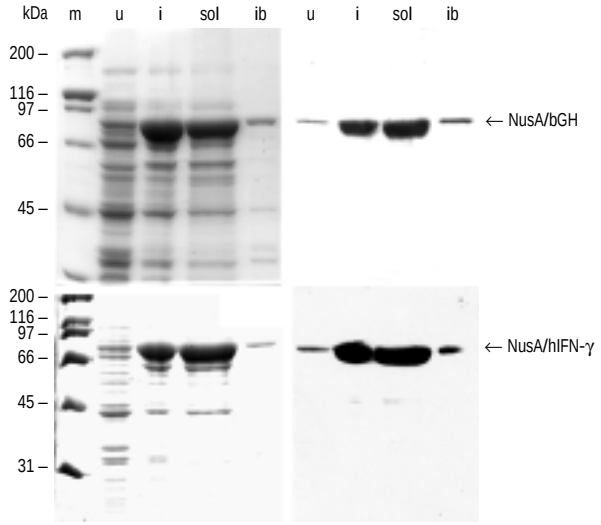


Figure 2. Analysis of bovine growth hormone (bGH) and human interferon-γ (hIFN-γ) expressed as fusions to NusA

Equal portions of cell lysate, soluble fraction, and insoluble fraction were analyzed by SDS-PAGE (upper panel) and Western blot (lower panel). The NusA/bGH and NusA/hIFN-γ fusion proteins were expressed from plasmid pKK223-3 under control of the *tac* promoter in *E. coli* JM105 at 37°C. Cells were induced with 1 mM IPTG and grown for 3 h post-induction. The Western blots were probed with rabbit anti-bGH polyclonal antibody, murine anti-hIFN-γ monoclonal antibody, and visualized using chemiluminescence. Percentage solubility based on blots: 89%, NusA/bGH; 87%, NusA/hIFN-γ. u = uninduced whole cell lysate, i = induced whole cell lysate, sol = soluble fraction, ib = inclusion body fraction.

a separate gene construction, we were not able to express the tyrosinase gene by itself at all (11); this is possibly due to the fact that the tyrosinase gene has a high content of rare codons (over three times as many as in the gene for NusA), or because of a 5'-mRNA structure that inhibits translation initiation.

Because the levels of soluble hIL-3 expression were the highest for the fusion protein with NusA, despite the fact that NusA was the largest of the carrier proteins investigated, we postulated that the biological activity of NusA might be a contributing factor to solubilization. One of the biological activities of NusA is that of pause-enhancement during transcription (13). To test whether NusA could independently influence the solubility of a target protein, we co-expressed NusA and human interferon-γ (hIFN-γ), normally insoluble when expressed in *E. coli* by itself, using dicistronic expression in *E. coli*. This strategy resulted in an increase in hIFN-γ overexpression (94% increase) when co-expressed with NusA in a dicistronic system, compared to when hIFN-γ was expressed by itself under the control of the same promoter and ribosome binding site (11). However, no increase in solubility was noted since hIFN-γ appeared to be almost completely insoluble either when expressed by itself or when co-expressed with NusA. Therefore, the excellent solubility of different proteins fused to NusA does not appear to be due to the biological activity of NusA.

Purification of hIL-3 expressed as a NusA fusion

Since the NusA/hIL-3 fusion protein exhibited the highest solubility and hIL-3 expression levels, we developed a purification scheme for this system (7). A His_6 tag was added to the N-terminus of NusA for purification by immobilized metal affinity chromatography (IMAC) using an Ni^{2+} chelate affinity column. A factor Xa protease cleavage sequence (Ile-Glu-Gly-Arg) connected NusA and hIL-3 in the original construction.

Following IMAC purification, the His_6NusA/hIL-3 fusion protein was digested with factor Xa protease. The protease digest was then passed over a Q Sepharose anion exchange chromatography column to remove hIL-3 from the digest mixture. Pure hIL-3 was found to elute in the flow through during loading of the Q Sepharose column. The hIL-3 was purified to homogeneity after the anion exchange chromatography at a yield of 4.5 mg/L of expression culture (0.49 mg hIL-3 per gram of cell paste) and a 13% recovery. The largest loss of hIL-3 occurred in the IMAC step, where the yield of hIL-3 was 32%. This yield in the IMAC step is consistent with the recovery results for the purification by IMAC of an upstream stimulatory factor (USF) with a His_6 at the N terminus (14). The percent recovery for the wild type and mutant forms of USF were reported to be 17% and 28%, respectively, for the IMAC step. The purified hIL-3 had full biological activity when compared to an hIL-3 reference standard.

After the incubation of the NusA/hIL-3 fusion protein with factor Xa protease, some non-specific cleavage of NusA was noted (7). The use of a linker sequence between the C-terminus of NusA and the N-terminus of the factor Xa sequence should increase specificity. It has previously been shown that a glycine-rich linker (Pro-Gly-Ile-Ser-Gly_5) placed immediately downstream of a thrombin cleavage site greatly increased the cleavage efficiency of several GST fusion proteins (15). Other proteases could also be used, including enterokinase and Kex2, which both cut at the C-terminus of their cleavage sequences (Asp_4-Lys and Leu-Phe-Lys-Arg, respectively); neither of these cleavage sequences are present in NusA.

Summary

This research has shown that selection of proteins with high solubility based on a statistical solubility model of solubility in *E. coli* has been a successful approach to expressing a target insoluble protein in soluble form as part of a fusion protein. Three *E. coli* proteins selected using this model-based approach, NusA, GrpE, and BFR, all were predicted to be more soluble than TRX when expressed in *E. coli*, and fusions of all three with hIL-3 had significantly higher solubility than the fusion of TRX and hIL-3. One of the carrier proteins studied, NusA, may be advantageous in solubilizing many target insoluble proteins, not only because of its good solubilizing characteristics, but because of its very high expression level. The good results of NusA in solubilizing a 54 kDa tyrosinase in a fusion protein may

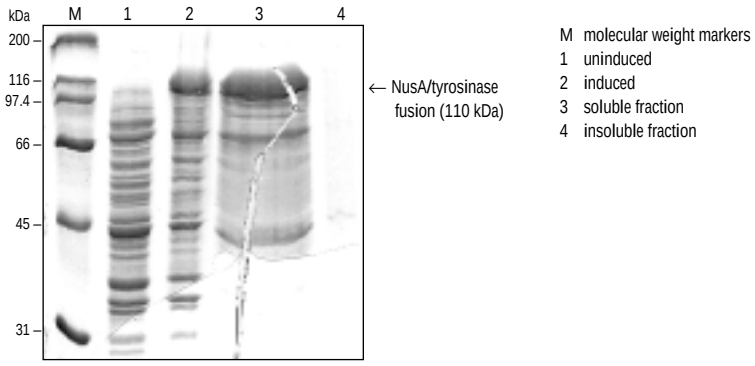


Figure 3. SDS-PAGE of *R. meliloti* tyrosinase expressed as a fusion to NusA

Equal portions of cell lysate, soluble fraction, and insoluble fraction were loaded and analyzed by SDS-PAGE. The NusA/tyrosinase fusion protein was expressed from plasmid pKK223-3 under control of the *tac* promoter in *E. coli* JM105 at 37°C. Cells were induced with 1 mM IPTG and grown for 3 h post-induction.

indicate that NusA may be a very good carrier for solubilizing large proteins. It was possible to obtain fully active hIL-3 starting from the NusA/hIL-3 fusion protein expressed in *E. coli* cells.

Acknowledgment

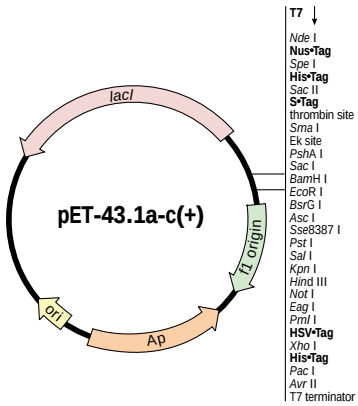
This work was supported in part by National Science Foundation grant CTS-9502235.

The NusA system is covered under University of Oklahoma U.S. patent number 5,989,868.

REFERENCES

1. Lilie, H., Schwarz, E. and Rainer, R. (1998) *Curr. Opin. Biotechnol.* **9**, 497–501.
2. De Bernadez Clark, E. (1998) *Curr. Opin. Biotechnol.* **9**, 157–163.
3. Georgiou, G. and Valax, P. (1996) *Curr. Opin. Biotechnol.* **7**, 190–197.
4. Hannig, G. and Makrides, S.C. (1998) *Trends in Biotech.* **16**, 54–60.
5. LaVallie E.R., DiBlasio E. A., Kovacic, S., Grant, K.L., Schendel, P.F. and McCoy, J.M. (1993) *Bio/Technology* **11**, 187–193.
6. Wilkinson, D.L. and Harrison, R.G. (1991) *Bio/Technology* **9**, 443–448.
7. Davis, G.D., Elisee, C., Newham, D.M. and Harrison, R.G. (1999) *Biotechnol. Bioeng.* **65**, 382–388.
8. van Leen, R.W., Bakhuis, J.G., van Beckhoven, R.G.W.C., Berger, H., Dorssers, L.C.J., Hommes, R.W.J., Lemson, P.J., Noordam, B., Persoon N.L.M. and Wagemaker G. (1991) *Bio/Technology* **9**, 47–52.
9. Langley, K.E., Berg, T.F., Strickland, T.W., Fenton, D.M., Boone, T.C. and Wypych J. (1987) *Eur. J. Biochem.* **163**, 313–321.
10. Haelewyn, J. and De Ley, M. (1995) *Biochem. Mol. Biol. Int.* **37**, 1163–1171.
11. Fayard, E.M.S. Thesis, University of Oklahoma, 1999.
12. Andrews, S.C., Harrison P.M. and Guest, J.R. (1989) *J. Bacteriol.* **171**, 3940–947.
13. Yager, T.D. and von Hippel, P.H. (1987) *in*: "*Escherichia coli* and *Salmonella typhimurium* Cellular and Molecular Biology", Ed. Neidhardt, F.C. American Soc. Microbiol., Washington, 1241–1275.
14. Van Dyke, M.W., Sirito, M. and Sawadogo, M. (1992) *Gene* **111**, 99–104.
15. Guan, K.L. and Dixon, J.E. (1991) *Anal. Biochem.* **192**, 262–267.

NusA gene now available as a fusion tag in pET-43.1 series



The NusA solubility promoting tag described in the accompanying article has now been incorporated into the pET vector system to create the pET43.1a-c(+) vector series. These vectors contain the following elements:

- Tightly controlled T7*lac* promoter and optimal ribosome binding site
- N-terminal Nus•Tag™ fusion sequence
- Internal His•Tag® and S•Tag™ sequences
- Thrombin and enterokinase cleavage sites
- Multiple cloning sites in all reading frames
- Optional C-terminal HSV•Tag® and His•Tag sequences
- *bla* (β-lactamase) and *lacI* (*lac* repressor) genes oriented in the opposite direction of the T7*lac* promoter
- ColE1 replication origin

Further enhancements in the generation of soluble, properly folded target protein may potentially be realized by combining the Nus•Tag fusion technology with expression in Origami™ host strains. Origami host strains are *trxB/gor* mutants (1) and are permissive to the formation of disulfide bonds in the cytoplasm. Therefore, target proteins that require disulfide bond formation may benefit from expression in these hosts.

1. Aslund, F., Bessette, P., Georgiou, G. and Beckwith, J. (1999) *inNovations* **10**, 11–12.

Product	Size	Cat. No.	Price
pET NusA Fusion System 43.1 (contains pET-43.1a, b, c(+) DNA, host strain glycerol stocks, and vector sequence CD)		70942-3	$310
pET NusA Fusion System 43.1 plus Competent Cells (contains pET-43.1a, b, c(+) DNA, host strain glycerol stocks, host strain competent cells, and vector sequence CD)		70943-3	$373
pET-43.1a(+) DNA	10 µg	70939-3	$131
pET-43.1b(+) DNA	10 µg	70940-3	$131
pET-43.1c(+) DNA	10 µg	70941-3	$131
His•Bind® Purification Kit (contains His•Bind Resin and buffers)		70239-3	$142