Protein Expression Using Cotranslational Fusion and Cleavage of Ubiquitin

MUTAGENESIS OF THE GLUTATHIONE-BINDING SITE OF HUMAN Pi CLASS GLUTATHIONE S-TRANSFERASE*

(Received for publication, February 15, 1994, and in revised form, June 24, 1994)

Rohan T. Baker†, Stewart A. Smith, Robert Marano, Jeff McKee, and Philip G. Board

From the Molecular Genetics Group, The John Curtin School of Medical Research, The Australian National University, P. O. Box 334, Canberra, ACT 2601, Australia

Expression of cloned genes in prokaryotes such as Escherichia coli is a widely used technique in both basic research and biotechnology. Despite the availability of several E. coli expression vector systems, adequate levels of expression may not be achieved. Expressing proteins as fusions to the highly conserved eukaryotic protein ubiquitin has been reported by several investigators to enhance protein yield in both bacterial and eukaryotic systems. We have modified this technique by the co-expression in E. coli of a ubiquitin-fusion protein and the Saccharomyces cerevisiae ubiquitin-specific protease Ubp2. This allows the co-translational cleavage of engineered ubiquitin-fusion proteins expressed in E. coli.

This system was used to express a human Pi class glutathione S-transferase (GST) GSTP1 as well as two mutant GSTP1 derivatives, Trp39→Cys and Glu62→Glu. The yield of these enzymes was improved 40-fold by using the ubiquitin-fusion/co-translational cleavage technique, and no uncleaved product was detected. The Trp39→Cys mutant was totally devoid of GST activity, while the activity of the Glu62→Glu mutant was reduced to 6% of wild-type GSTP1-1. As both of the mutated residues map within the glutathione-binding site, the reduced GST activity is consistent with a marked reduction in glutathione binding ability.

Many studies of protein structure/function relationships depend upon the ability to express a cloned gene product in bacterial or eukaryotic cells, followed by the purification of that product for further study. Highly efficient heterologous expression systems are a prerequisite for such studies, especially for crystallographic and therapeutic applications, where large quantities are required. Despite the availability of several Escherichia coli expression vectors which utilize strong promoters and consensual transcription and translation signals, the final protein yield can vary widely under identical growth conditions and is often insufficient for the intended application. Low yields may result in spite of efficient synthesis due to subsequent rapid degradation of the product.

Several systems based on heterologous protein expression as protein fusions have been developed, which generally give reliably high yields. Examples include fusions to β-galactosidase and glutathione S-transferase (reviewed by Uhlén and Moks, 1990). One obvious disadvantage of this approach is that the desired product is fused to a large, unrelated protein. This has been circumvented by the engineering of protease recognition sites at the junction of the two proteins to allow release of the product after purification of the fusion protein (Uhlén and Moks, 1990). However, these proteases may also cleave at other sites within the protein, and authentic NH2 termini cannot be constructed (Butt et al., 1989).

More recently, several investigators have expressed heterologous proteins as fusions to the 76-amino acid protein ubiquitin. Ubiquitin is the most conserved eukaryotic protein known, but both it, and the enzymes of the ubiquitin pathway, are absent from bacteria (for review, see Finley and Chau, 1991; Hershko and Ciechanover, 1992). Ubiquitin-fusion protein genes occur naturally (Schlesinger and Bond, 1987), and expression of proteins as synthetic fusions to ubiquitin has been reported to enhance their yield in both E. coli (Butt et al., 1989; Yoo et al., 1989) and the yeast Saccharomyces cerevisiae (Ecker et al., 1989; Sabin et al., 1989). Expression of such fusion proteins in yeast results in the cleavage of ubiquitin from the fusion protein by endogenous ubiquitin-specific proteases (Ubps),1 whereas ubiquitin-fusion proteins remain intact in E. coli due to the absence of Ubps. Thus ubiquitin serves two functions in these fusions, first by stabilizing the downstream protein, and second as a recognition domain for Ubps. Cleavage of ubiquitin from fusion proteins purified from E. coli has been accomplished by the use of partially purified rabbit reticulocyte Ubps (ubiquitin-protein hydrolases) (Butt et al., 1989; Yoo et al., 1989). However, isolation of a pure product would require its repurification from this cleavage reaction mixture.

In this report we describe a heterologous protein expression system based on the co-expression in E. coli of a ubiquitin fusion protein and the yeast ubiquitin-specific protease Ubp2 (Baker et al., 1992). This system combines the ease of handling of bacterial expression systems, the high yields conferred by ubiquitin-fusion technology, and the eukaryotic convenience of co-translational cleavage of ubiquitin. This system was tested by expressing a human Pi class glutathione S-transferase (GST) GSTP1, as well as two GSTP1 mutants constructed by site-directed mutagenesis. Expression of these proteins by the ubiquitin-fusion/co-translational cleavage technique resulted in a dramatic improvement in yield of enzymatically active protein that was free of ubiquitin. This technique should be generally applicable to the expression of a broad range of proteins and peptides in bacteria.

Crystallographic studies of the rat and human Pi class GSTs have implicated Trp39 and Gln62 in glutathione binding (Reine-
Protein Expression using Ubiquitin Fusion and Cleavage

Plasmids—Recombinant proteins were expressed from a modified pKK223-3 E. coli expression vector (Brosius and Luspa, 1987) constructed as follows. The pKK223-3 polynuker (Fig. 1A) was replaced with an EcoRI-HindIII DNA fragment lacking BamHI and SalI restriction sites (Fig. 1B). This plasmid was then digested with BamHI and SalI, the ends made blunt with the Klenow fragment of DNA polymerase I, and self-ligated (Fig. 1C). The EcoRI-HindIII fragment was then replaced with the original pKK223-3 (pUC8) EcoRI-HindIII polynuker to yield the plasmid pKK261 (Fig. 1D), which has unique BamHI and SalI restriction sites in the polynuker and also lacks the SpIII site from the vector backbone. The plasmid pKK262 was constructed analogously except that the pUC18 EcoRI-HindIII polynuker was used to provide additional cloning sites (Fig. 1E).

The ubiquitin gene used for ubiquitin fusion constructions was obtained by the polymerase chain reaction (PCR) amplification of the first coding unit of the yeast UBI4 polyubiquitin gene contained in the plasmid pACUb-R-pgal (Baker et al., 1992) using the forward primer YUS (5′-d(GCCGAGATACTTATGATAGCCGTTCAAG)-3′) and reverse primer YUS (5′-d(GCCGATACACCCGGCGACCCGACACGCA)-3′). These primers result in the addition of EcoRI and BglII recognition sites at the 5′ end of the gene, and SacII and BamHI recognition sites at the 3′ end of the gene (Fig. 2). Amplification was performed using 10 ng of pACUb-R-fgal, 200 ng of each primer, and 1 unit of Taq DNA polymerase in a 100-μl reaction containing 50 mM KCl, 10 mM Tris-HCl, pH 8.8, 1.5 mM MgCl2, 0.1% Triton X-100, and 0.125 mM each dNTP. Reaction conditions were five cycles of 95 °C for 20 s, 40 °C for 20 s, and 72 °C for 20 s, followed by 25 cycles of 95 °C for 20 s, 68 °C for 20 s, and 72 °C for 20 s. The PCR product was digested with EcoRI and BamHI and the resulting 234-bp fragment was purified from an agarose gel using GeneClean (Bio 101). This fragment was ligated into EcoRI-BamHI-digested pKK262 (see above) to yield the plasmid pRB269.

A full-length human placental GSTPl cDNA clone pGT3EX (Board et al., 1993) was used as a template for PCR amplification using the forward primer UBGF (5′-d(GTGCAGGCTGTCGTTTCACTGCCGGATATAACG)-3′) and the reverse primer 3EX (5′-d(CTAAGCTTCCTCACTGTTCTCGG-3′). Amplification with these primers results in the addition of EcoRI and BglII restriction sites down the ends made blunt with the Klenow fragment of DNA polymerase I, and self-ligated (Fig. 2C) and ligated into the gel-purified SpIII-HindIII cut pRB307 vector to yield the plasmid pRB315. The presence of the correctly mutated insert was confirmed by sequencing the SpI-HindIII fragment from pRB315. Similarly, the SpI-HindIII fragment from a correctly mutagenized Q52E clone was ligated into SpI-HindIII cut pRB307 to yield the plasmid pRB316. In the latter case, the presence of the correctly mutated insert was detected by the presence of a SacI site diagnostic for the Gln-Glu mutation (Board et al., 1993).

Enzyme Purification and Assays—E. coli strain TG1 (Amersham) containing pRB307, pRB315, or pRB316 was grown under the conditions described by Board and Pierce (1987). Where pRB173 was also present as a source of the ubiquitin-specific protease Ubp2, chloramphenicol was added to the medium at a final concentration of 34 μg/ml to select for this plasmid. Cells were washed in 150 mM NaCl, 50 mM Tris-HCl, pH 7.5, and then resuspended in the same buffer containing 4 mg/ml lysozyme and 1 mM β-mercaptoethanol. After standing for 1 h on ice the cells were lysed by sonication (Branson Sonifier 250). After centrifugation at 10,000 × g for 20 min the supernatant was applied directly to a column of glutathione-agarose (Sigma), and the enzyme was processed as described previously by Simons and Vader Jagt (1977). In some instances, affinity chromatography was performed in batch mode as described by Smith and Johnson (1988). Glutathione S-transferase activity was measured spectrophotometrically with 1-chloro-2,4-dinitrobenzene and glutathione (GSH) as substrates as described previously by Habig and Jakoby (1981). In some instances, small-scale cell lysates were prepared by the lysozyme/Triton X-100 procedure described by Tobias and Varshavsky (1991). Extracts containing approximately equal amounts of protein were electrophoresed in 10% SDS-polyacrylamide gels (Laemmli, 1970) and detected by staining with Coomassie Brilliant blue R-250.

RESULTS

General Considerations of the Ubiquitin Fusion Expression System—A single coding unit from the yeast UBI4 polyubiquitin gene was modified to provide an EcoRI site 9 bp upstream of the initiation codon so as to place it in a favorable context and position with respect to the tac promoter in the modified pKK223-3 E. coli expression vector pKK262 (Fig. 2A). To facilitate the positioning of other genes downstream of the ubiquitin coding unit, a BamHI site was engineered spanning the last ( Gly20) codon of ubiquitin and the first codon thereafter (Fig. 2B). This BamHI site (made blunt with Klenow DNA polymerase) is necessary can be used for ligating other DNA fragments in cases where the amino acid exposed after cleavage of ubiquitin is not critical, or if by chance serine is desired in this position (the 3′ half of the BamHI site, TCC, is a serine codon; Fig. 2B). Likewise, any of the polynuker restriction sites downstream of BamHI can be used for cloning provided that the presence of extra residues at the NH2 terminus of the protein can be tolerated. Alternatively, a DNA fragment can be ligated to the BamHI site that has been made blunt with a single strand-specific nuclease (such as mung bean nuclease) provided that the first nucleotide of the DNA fragment to be joined is a “G,” so that the Gly20 codon of ubiquitin is reformed. In instances where the amino acid sequence is critical, a SacII site has been engineered near the 3′ end of the ubiquitin coding region, in the codons for Leu24, Arg18, and Gly20 (Fig. 2D). The Leu24 codon was changed from CTA to CTC, and the Arg18 codon was changed from AGA to CGC, using an appropriate oligonuclease and the polymerase chain reaction (see “Experimental Procedures”). The DNA sequence encoding the protein to be synthesized as a ubiquitin fusion is then extended at its 5′ end with the sequence 5′-d(GTG-GCCGCCGTTCTG)-3′ immediately preceding the codon for the residue to be at the NH2 terminus of the protein following ubiquitin cleavage. (The first three nucleotides are present to enhance subsequent SacII cleavage and may be any sequence.) This 5′ extension provides a SacII restriction site in the same reading frame as the engineered ubiquitin SacII site and completes the ubiquitin coding region. The only difficulty arises where proline is required at the protein’s NH2 terminus: ubiquitin-specific proteases are largely inhibited by the presence of proline immediately following the Gly20 of ubiquitin (Bachmair et al., 1986; Tobias and Varshavsky, 1991).

Expression of Wild-type and Mutant GSTPl in E. coli—We have previously attempted to express GSTPl in E. coli by cloning its cDNA downstream of the tac promoter in pKK223-3.
**Fig. 1. Construction of modified pKK223-3 plasmids.** pKK223-3 (A) was digested with EcoRI and HindIII and ligated with a 318-bp EcoRI-HindIII fragment (heavy black line) to yield pRB259 (B). pRB259 was digested with Sall and BamHI, the ends made blunt with the Klenow fragment of DNA polymerase I and dNTP's, and self-ligated to yield pRB260 (C). pRB260 was digested with EcoRI and HindIII, and the gel-purified vector was ligated with the EcoRI-HindIII polylinker from either pUC8 to yield pKK261 (D), or from pUC18 to yield pKK262 (E). In all plasmids, the tac promoter is shown as a black triangle and other regions as shaded boxes. Ori, colEl origin of replication; Amp', β-lactamase gene; ROP, regulator of plasmid copy number; term, transcriptional terminator region. Non-unique restriction enzyme sites are underlined for emphasis. Locations of restriction enzyme sites are given relative to EcoRI = 1. In pKK262 (E), 20 bp should be added to the positions of the Scal, PvuII, and Eagl sites.
When grown in the presence of isopropyl β-D-thiogalactoside, E. coli cells containing this plasmid produced very low levels of GST activity, and a protein of the expected molecular mass was not detectable in total protein extracts (data not shown). In order to enhance expression, the 5' end of the GSTPl cDNA was altered by oligonucleotide-directed mutagenesis to add a ribosome-binding site and to modify the codon usage of the first few codons to better suit the E. coli codon bias (Board et al., 1993). These changes resulted in slightly enhanced GSTPl expression and a detectable protein in total cell lysates (Fig. 3, lane 3), but the expression level was still low (~1 mg/liter of culture). For our mutagenesis studies, significantly higher expression levels were preferred. We therefore used the technique of protein expression as a ubiquitin fusion, which has been reported to enhance the yield of proteins expressed in E. coli (Bachmair et al., 1991; Baker et al., 1992). These crude lysates also resulted in the appearance of a predominant band on Coomassie-stained SDS gels with an apparent molecular mass of 30 kDa (Fig. 3, lane 4), approximately the size expected for a ubiquitin-GSTPl fusion protein (31.9 kDa). As noted by others, ubiquitin cleavage does not occur in E. coli because this organism lacks ubiquitin and enzymes of the ubiquitin system (Bachmair et al., 1986; Monia et al., 1989; Tobias and Varshavsky, 1991; Baker et al., 1992). These crude lysates also possessed abundant GST activity, indicating that the presence of ubiquitin at the NH2 terminus of GSTPl did not significantly interfere with its catalytic activity (not shown).

For the purposes of detailed kinetic analysis, however, an NH2-terminal extension on GSTPl was not desired. To remove the ubiquitin moiety from the fusion protein, pRB307 was transformed into competent E. coli containing the plasmid pRB173, which expresses the S. cerevisiae ubiquitin-specific protease Ubp2 (Baker et al., 1992). Crude lysates from these cells contained an abundant protein with an apparent molecular mass of 24 kDa (Fig. 3, lane 5), approximately the size

![Diagram of plasmid pRB269](image)

**Fig. 2. Structure and sequence of plasmid pRB269.** A, plasmid map of pRB269. The ubiquitin coding region is shown as an open box pointing in the direction of transcription/translation. Other designations are as in Fig. 1. B, restriction sites, nucleotide sequence, and protein translation of the ubiquitin coding unit. The tac promoter/ubiquitin coding region from A is expanded to show the DNA sequence at each end of the ubiquitin coding unit. Restriction enzymes are indicated above their recognition sequence. The protein translation is given below the sequence in the standard single letter code. Ubiquitin amino acids are numbered, and non-ubiquitin amino acids are in lower case. Dots indicate a gap in the sequence corresponding to position 13–216 of the first coding unit of the UBI4 polyubiquitin gene (Ozkaynak et al., 1987). C, ubiquitin-GSTPl fusion inserts in plasmids pRB307, 315, and 316. Ubiquitin and GSTPl coding regions are shown as boxes. Other designations are as in B. The locations of the SphI site used for subcloning and the Trp59 and Gin62 codons mutated in plasmids pRB315 and pRB316, respectively, are shown below the GSTPl coding region. ter is the termination codon.

![Expression and purification of GSTPl-1 using co-translational fusion and cleavage of ubiquitin](image)

**Fig. 3. Expression and purification of GSTPl-1 using co-translational fusion and cleavage of ubiquitin.** E. coli TG1 cultures containing various plasmids were induced with isopropyl β-D-thiogalactoside followed by lysis, SDS-polyacrylamide gel electrophoresis, and Coomassie Blue staining. Lane 1, molecular weight markers; lane 2, E. coli TG1 lysate; lane 3, TG1 containing pGTSEX (Board et al., 1993); lane 4, TG1 containing pRB307 (Ub-GSTPl); lane 5, TG1 containing pRB307 and pRB173 (Ubp2); lane 6, purified ubiquitin-GSTPl fusion protein from lane 4; lane 7, purified GSTPl from lane 5; lane 8, TG1 containing pRB315 (Ub-GSTPl-W39C) and pRB173; lane 9, TG1 containing pRB316 (Ub-GSTPl-Q52E) and pRB173. Molecular sizes in kDa are indicated at the left. The positions of Ub-GSTPl, GSTPl, lysozyme, and ubiquitin are indicated on the right. Approximately equal amounts of protein were loaded/lane. The sample in lane 3 is overloaded so that the weak GSTPl band (arrow) can be seen.
expected for the GSTP1 protein (23.3 kDa). The ~30-kDa ubiquitin-GSTP1 fusion protein was absent from these lysates, indicating that ubiquitin cleavage was complete, and a prominent band of approximately 8 kDa was present, which is of the correct size to be the ubiquitin moiety cleaved from the fusion proteins (Fig. 3, lane 5; see also lanes 8 and 9). As expected, these crude lysates also possessed abundant GST activity. In Fig. 3, the GSTP1 protein appears to be more abundant relative to E. coli proteins than is the ubiquitin-GSTP1 protein (compare lanes 4 and 5). However, in lysates prepared on other occasions this difference was not apparent and presumably represents day-to-day fluctuations.

Similarly, when E. coli was transformed with the plasmids pRB315 and pRB316, which encode ubiquitin-GSTP1-W39C and ubiquitin-GSTP1-Q52E, respectively, an abundant protein of ~30 kDa was detected in crude lysates (data not shown). When the ubiquitin-specific protease Ubp2 was also present in these cells (expressed from pRB173), the ~30-kDa protein was cleaved to produce proteins of ~24 kDa (GSTP1 mutants) and ~8 kDa (ubiquitin) (Fig. 3, lanes 8 and 9). Again, the ~30-kDa ubiquitin-GSTP1 fusion protein was absent from these lysates, indicating that ubiquitin cleavage was complete. Despite the abundant levels of a GSTP1-sized protein in crude lysates of cells expressing mutant GSTP1 enzymes, both lysates had markedly reduced GST activity, indicating that both mutations had deleterious effects on enzyme activity (see below).

**Purification of Wild-type and Mutant GSTP1**—The ubiquitin-GSTP1 fusion protein expressed from pRB307 could be readily purified by affinity chromatography on glutathione-agarose (GSH-agarose) as described by Simons and Vander Jagt (1977) (Fig. 3, lane 6). Thus the presence of ubiquitin at the NH2 terminus of GSTP1 did not significantly interfere with GSH binding, as expected from the high GST activity of the fusion protein. The yield of the ubiquitin-GSTP1 fusion protein was 60 mg/liter. As expected, the GSTP1 protein expressed from pRB307 in conjunction with pRB173 was also readily purified by GSH-agarose affinity chromatography, with a yield of 42 mg/liter. This is a substantial (~40-fold) improvement on the yield obtained when GSTP1 was expressed unfused to ubiquitin under otherwise identical conditions (~1 mg/liter; Board et al., 1993). No uncleaved ubiquitin-GSTP1 fusion protein was detected in this preparation (Fig. 3, lane 7), indicating that ubiquitin cleavage was complete. The purified GSTP1-1 from Fig. 3, lane 7, was subjected to 10 cycles of automated NH2-terminal sequencing, which resulted in no ubiquitin protein sequence being detected, indicating that any uncleaved fusion protein was present at less than 1% of the sample. The NH2-terminal sequence data also revealed that in approximately 10% of the GSTP1-1 protein, the methionine residue had been cleaved to expose the mature GSTP1-1 NH2-terminal residue proline (Ålin et al., 1985; Suzuki et al., 1987), whereas the remaining 90% retained the NH2-terminal Met residue (see Fig. 2C). This Met cleavage is presumably the consequence of the action of the E. coli methionine amino-terminal peptidase following cleavage of ubiquitin, rather than infidelity in the site of ubiquitin-specific protease cleavage. The latter enzymes have been demonstrated to have a precise specificity for cleavage immediately after the COOH-terminal glycine of ubiquitin (Bachmair et al., 1986; Butt et al., 1989; Gonda et al., 1989). This Met residue was required in the ubiquitin-GSTP1 fusion construct because the presence of Pro immediately following ubiquitin would be expected to inhibit the action of the ubiquitin-specific protease (Bachmair et al., 1986).

In contrast to the ease of GSTP1-1 and ubiquitin-GSTP1-1 purification by GSH-agarose affinity chromatography, the bulk of each mutant protein was present in the column flow-through. The GSTP1-W39C protein failed to bind to GSH-agarose and could not be purified. The GSTP1-Q52E isoenzyme bound weakly to GSH-agarose and sufficient enzyme could be purified to permit some kinetic measurements (see below). Neither protein bound significantly better to a pentylglutathione agarose affinity matrix.

**Enzymatic Activity of Wild-type and Mutant GSTP1**—The specific activities of the wild-type and mutant proteins were determined with GST1 and 1-chloro-2,4-dinitrobenzene as substrates. The specific activity of the wild-type GSTP1-1 (95.7 μmol/min/mg; Table I) falls within the range of published values for purified recombinant GSTP1-1 (64–128 μmol/min/mg; Kong et al., 1991; Kolm et al., 1992; Widersten et al., 1992; Board et al., 1993). This indicates that the expression of GSTP1 as a ubiquitin fusion followed by ubiquitin cleavage does not interfere with its enzymatic activity, and by inference does not interfere with its folding. The W39C mutant was devoid of activity in the crude E. coli lysate, and the protein could not be purified and concentrated by GSH-affinity chromatography for further study. The activity of the Q52E mutant was only 6% of that of the wild-type GSTP1-1 protein (Table I). The kinetics of GST reactions are complex and require rate equations with many parameters for a complete description of their behavior (Mannervik, 1985). As a result of this complexity, true Kcat values are difficult to accurately determine (Stenberg et al., 1991). Despite these difficulties Kcat/Km values are given in Table 1 (cf. Danielson and Mannervik, 1985) and indicate that the Q52E mutant has a lower catalytic efficiency and has a lower apparent substrate affinity for GSH.

**Table I**

<table>
<thead>
<tr>
<th>Isoenzyme</th>
<th>Specific activity (μmol/min/mg)</th>
<th>Km (GSH) (μM)</th>
<th>Vmax (μmol/min/mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>95.7</td>
<td>126.6</td>
<td>100</td>
</tr>
<tr>
<td>Q52E</td>
<td>6.1</td>
<td>4.9</td>
<td>0.2</td>
</tr>
<tr>
<td>W39C</td>
<td>ND*</td>
<td>4.0</td>
<td>0.5</td>
</tr>
</tbody>
</table>

*ND, not detectable.

**Discussion**

We have constructed an E. coli expression vector that directs the synthesis of proteins as fusions to yeast ubiquitin. Recognition sites for several restriction enzymes at or near the 3’ end of the ubiquitin coding region allow the precise fusion of any reading frame to ubiquitin. Protein expression is under control of the inducible trp/lac (“tac”) promoter. The technique of expressing proteins and peptides as ubiquitin fusions has been reported to enhance their yield in either E. coli (Butt et al., 1989; Yoo et al., 1989) or S. cerevisiae (Ecker et al., 1989; Sabin et al., 1989). When ubiquitin-fusion proteins are expressed in a eukaryotic host such as S. cerevisiae, the ubiquitin is cleaved by endogenous ubiquitin-specific proteases. This cleavage is extremely rapid and probably occurs co-translationally, as no uncleaved fusion protein was detected. However, because bacteria such as E. coli lack ubiquitin and enzymes of the ubiquitin system, ubiquitin-fusion proteins expressed in bacteria remain intact (Butt et al., 1989; Yoo et al., 1989; Monia et al., 1989; Tobias and Varshavsky, 1991; Baker et al., 1992). Depending on the in-
tended use of the expressed protein, the benefit gained from increased yield may be negated by the presence of covalently bound ubiquitin.

The availability of cloned ubiquitin-specific proteases allows circumvention of this problem by their co-expression in E. coli with the ubiquitin-fusion protein. As we have demonstrated, the benefit of increased yield is still obtained under these conditions, in spite of the rapid and complete cleavage of ubiquitin. Thus, as is the case in eukaryotic cells, the yield-enhancing effect of ubiquitin must be exerted in the short time that the fusion protein is intact. Proposed mechanisms for this effect of ubiquitin include a chaperonin-type role and the protection of the amino-terminal portion of the nascent protein from proteolytic attack. Our results indicate that the same protective mechanism(s) operate in both prokaryotic and eukaryotic cells. Thus, this system results in high yields of protein free of ubiquitin and theoretically could be used to enhance the expression of a range of proteins and peptides. While it may be argued that the co-translational cleavage may be accomplished by expression in a eukaryotic host such as S. cerevisiae, in most cases bacterial hosts are preferred for protein expression due to ease of growth, cell lysis, and handling.

This system was tested using the expression of a human Pi class GST enzyme GSTP1-1. Our previous attempts at expression of the GSTP1 cDNA from the tac promoter resulted in low yields of approximately 1 mg/liter of E. coli culture (Board et al., 1993), whereas merely by placing a ubiquitin coding unit between the promoter and the GSTP1 initiation codon, the yield increased dramatically. Low yields of human or rat Pi class GST enzymes (3-3 mg/liter) have been observed by others (Kolm et al., 1992; Kong et al., 1992), although it is interesting to note that Nishihira et al. (1992) obtained yields of 20 mg/liter when rat Pi class GST was expressed in the pET3a system with an 18-residue amino-terminal extension. This observation suggests that the nascent Pi class GSTs may be proteolytically sensitive and/or slow to fold into a stable conformation and that their synthesis as a fusion protein to either ubiquitin or to another peptide may offer protection until a more stable conformation can be achieved. This is not true of other classes of GSTs, which may be expressed at high levels in E. coli (Board and Pierce, 1987; Smith et al., 1988; Pennington and Rule, 1992). Manoharan et al. (1992) expressed a variety of human Pi class GST mutants in E. coli and observed a yield of 3.7% of soluble protein for the wild-type enzyme, and variable yields ranging from 2 to 26% of soluble protein for different mutants. Residues Trp59 and Gin52 were not mutated in this study. The differences in yield were attributed to the translation of the unstable protein, and the observation that some mutations substantially increased yield is consistent with an intrinsic instability of Pi class GSTs when expressed in E. coli. A direct comparison of our data and that of Manoharan et al. is not possible because the latter's "wild-type" and mutant enzymes actually contained a Pro26Ala mutation erroneously introduced while engineering the cDNA for expression. This change may have altered the stability and/or other properties of the resulting proteins.

Two mutant GSTP1 cDNAs were constructed by site-directed mutagenesis and expressed using the ubiquitin-fusion technique. Neither mutation appears to affect the expression level or stability of the resulting proteins, which were both obtained in high yields (Fig. 3). Trp59 and Gin52 were chosen for mutation because they are predicted to be involved in glutathione binding from analysis of the three-dimensional structure of the pig and human Pi class GSTs (Reinemer et al., 1991, 1992) and are also 2 of the 4 residues that differ between GSTP1 and the human fatty acid ethyl ester synthetase-III (FAEES-III) enzyme deduced from a cDNA reported by Bora et al. (1991). Board et al. (1993) have constructed a cDNA that encodes the FAEES-III protein and have shown that the recombinant FAEES-III protein expressed in E. coli from this cDNA completely lacks both GST activity and FAEES activity. Our results demonstrate that the Trp59→Cys mutation alone is sufficient to eliminate glutathione binding by GSTP1-1 and thus its GST activity, while the Gin52→Glu mutation substantially reduces both glutathione binding and GST activity. Consistent with these findings, Nishihira et al. (1992) have demonstrated that Trp59 in the rat Pi class GST (equivalent to Trp59 in GSTP1) is essential for glutathione binding and GST activity. In addition, Widersten et al. (1992) constructed a Gin52→Ala mutation in GSTP1 and observed a 10-fold decrease in affinity for glutathione and a 79% reduction in GST activity, similar to the 94% reduction in the GST activity of the Gin52→Glu mutant that we observed. Thus, the lack of GST activity of the recombinant FAEES-III protein expressed by Board et al. (1993) can be explained by the single Trp59→Cys mutation. The data of Board et al. (1993) also demonstrates that the other 3 residues that differ between GSTP1 and FAEES-III cannot suppress or compensate for the Trp59→Cys mutations. These results support the conclusion of Board et al. (1993) that the FAEES-III cDNA reported by Bora et al. (1991) may have arisen from a cloning artifact.