Method for enhancing solubility of the expressed recombinant proteins in *Escherichia coli*

Sudip Ghosh[1,2], Sheeba Rasheedi[2], Sheikh Showkat Rahim[2], Sharmistha Banerjee[2], Rakesh Kumar Choudhary[2], Prachee Chakhaiyar[2], Nasreen Z. Ehtesham[1], Sangita Mukhopadhyay[2] and Seyed E. Hasnain[2,3]

[1]National Institute of Nutrition (ICMR), Hyderabad, [2]CDFD, Nacharam, Hyderabad, and [3]Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur, Bangalore, India

BioTechniques 37:418-423 (September 2004)

The production of correctly folded protein in Escherichia coli *is often challenging because of aggregation of the overexpressed protein into inclusion bodies. Although a number of general and protein-specific techniques are available, their effectiveness varies widely. We report a novel method for enhancing the solubility of overexpressed proteins. Presence of a dipeptide, glycylglycine, in the range of 100 mM to 1 M in the medium was found to significantly enhance the solubility (up to 170-fold) of the expressed proteins. The method has been validated using mycobacterial proteins, resulting in improved solubilization, which were otherwise difficult to express as soluble proteins in* E. coli. *This method can also be used to enhance the solubility of other heterologous recombinant proteins expressed in a bacterial system.*

INTRODUCTION

Escherichia coli is the most widely used system for the rapid and economical production of recombinant proteins because of its very well-characterized genetics and rapid growth rate in inexpensive culture media. One major disadvantage of *E. coli* is that heterologous proteins are often expressed as insoluble aggregates of folding intermediates known as inclusion bodies. Expression in soluble fraction is paramount for the expressed protein to be biologically active. In order to recover soluble proteins from the inclusion bodies, the inclusion bodies are solubilized in the presence of strong denaturants such as urea or guanidinium hydrochloride, followed by the removal of the denaturants under optimal conditions that favor refolding. Although considerable progress has been made for efficient refolding of proteins (1), specific folding conditions differ greatly from protein to protein. Even under optimal conditions of refolding, quite a large number of proteins are found to be recalcitrant to refolding, and the yield of renatured protein is relatively low.

Several general and protein-specific methods are available for increased solubility of expressed proteins in *E. coli*. One approach is the coexpression of molecular chaperones, which assists in the correct folding of the heterologous protein (2,3). Similarly, concomitant overexpression of thioredoxin (TrxA) is known to improve the solubility of the expressed proteins (4). Another approach that has gained considerable success in recent years is the use of gene fusion (5). Fusion partners such as glutathione-S-transferase (GST) and maltose binding protein (MBP) are known to impart solubility of many heterologous proteins in addition to serving as a tag for affinity purification. Sometimes, soluble expression can also be enhanced by supplying essential cofactors necessary for the activity of the protein in question. For example, soluble expression of human cystathionine β-synthase, a heme-containing protein, could be increased over 8-fold by the addition of the heme precursor δ-aminolevulinic acid (δ-ALA) to the culture medium (6). In some instances, coexpression of nuclear receptor partners is also found to increase the solubility of nuclear receptors expressed in *E. coli* (7). In yet other cases, specific substitution of some amino acid residues was found to enhance the solubility of the expressed proteins (8). In this report, we describe a novel method of enhancing the solubility of expressed proteins by inducing recombinant protein expression in the presence of the dipeptide glycylglycine. We deliberately chose as an example mycobacterial proteins that are known to be difficult to express as soluble proteins in *E. coli*. The solubilization of these proteins was enhanced up to 170-fold.

MATERIALS AND METHODS

Preparation of Recombinant Constructs

The open reading frames (ORFs) Rv0256c, Rv2430c (both are the members of the PPE gene family), Rv3339c (isocitrate dehydrogenase-1), and Rv1609 (anthranilate synthase) of *Mycobacterium tuberculosis* were amplified from the genomic DNA of the strain H37Rv by PCR. Oligodeoxyribonucleotide primers were chemically synthesized (Microsynth GmbH, Balgach, Switzerland), with appropriate restriction sites suitable for in-frame cloning into expression vectors, with N-terminal 6x-histidine tags. The primers used for amplification are shown in Table 1. The recombinant proteins were expressed as N-terminal His-tagged fusions. The positive clones were confirmed by DNA sequencing.

Cell Culture

Competent BL21(DE3)pLysS cells (Novagen, Abingdon, UK) were transformed with pRSET256, pRSET1609, and pRSET3339 plasmid DNA, and the colonies were grown overnight on Luria Bertani (LB) plates (9) containing 100 µg/mL ampicillin. For pQE2430, competent M15(pREP4) cells (Qiagen GmbH, Hilden, Germany) were used and grown in LB plates containing ampicillin (100 µg/mL) and kanamycin (50 µg/mL). Fresh colonies were first inoculated into 5 mL LB media containing appropriate antibiotics and grown overnight at 37°C with shaking. These overnight cultures were diluted 10-fold into 10 mL Terrific Broth (TB) medium (9), containing different concentrations (0, 50, 200, 500, and 1000 mM) of glycylglycine (Amersham Biosciences, Buckinghamshire, UK) and further grown at 37°C in an orbitory shaker till the absorbance $(A)_{600}$ = 1. The cultures were cooled to room temperature, and protein expression was induced with 0.5 mM isopropyl-β-D-thiogalactoside (IPTG) for 14–16 h at 27°C.

Table 1. Sequence of the Primers, Restriction Sites, and Vectors Used for Expressing Different Mycobacterial Proteins

Gene	Primers with Restriction Enzyme Sites	Restriction Enzyme Sites	Cloning Vector
Rv0256c	5′-CGAGATCTATGACCGCCCCGATCTGGAT-3′	*Bgl*II	pRSET B
	5′-GCGAATTCTCACTCCACCCGGGTCG CTGA-3′	*Eco*RI	
Rv2430c	5′-GGATCCATGCATTTCGAAGCGTAC-3′	*Bam*HI	pQE30
	5′-AAGCTTCTAAGTGTCTGTACGCGATGA-3′	*Hin*dIII	
Rv1609	5′-AATCTCGAGGTGTCCGAGCTCAGCGT-3′	*Xho*I	pRSET A
	5′-AATCCATGGCTGGCGTGCAACCAGATAA-3′	*Nco*I	
Rv3339c	5′-AGGATCCATGTC CAACGCACCCAAGATA-3′	*Bam*HI	pRSET A
	5′-TAAGCTTCTAATTGGCCAGCTCCTTTTC-3′	*Hin*dIII	

Restriction enzyme sites have been underlined.

Preparation of Cell Lysate and Protein Solubility Analysis

The induced cells were centrifuged at 10,000× *g*, and the cell pellet was resuspended in extraction buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl) with lysozyme to a final concentration of 1 mg/mL and incubated on ice for 30 min, followed by sonication 5 times with a burst duration of 15 s each. The sonicated lysates were centrifuged at 18,000× *g* for 10 min, and the supernatants containing the soluble proteins were collected into fresh tubes. The concentration of total protein in supernatant was estimated using a Bio-Rad DC™ Protein Assay Kit (Bio-Rad Laboratories, Hertfordshire, UK). About 60 µg of protein from each supernatant were electrophoresed in a 12% sodium dodecyl sulfate (SDS) polyacrylamide gel (10) and transblotted onto a nitrocellulose membrane (Hybond™-ECL™; Amersham Biosciences) at 300 mA for 2 h at 4°C in a transfer buffer (25 mM Tris-HCl, 190 mM glycine, 20% methanol). The membrane was blocked for 1 h in phosphate-buffered saline (PBS; Reference 9) containing 4% non-fat dry milk and then incubated with 1:200 diluted monoclonal anti-his tag antibodies (Santa Cruz Biotechnology, Santa Cruz, CA, USA) in PBS. The membrane was washed thrice with excess PBST (PBS containing 0.1% Tween® 20) for 15 min each. Goat anti-mouse immunoglobulin G (IgG)-horseradish peroxidase (HRP) conjugate (Amersham Biosciences) was used at 1:10,000 dilution as the second antibody. The membrane was again washed thrice for 5 min each with PBST. The reactive bands were developed by chemiluminescence with luminol reagents (Santa Cruz Biotechnology). All the experiments were performed at least three times, and the representative blots are presented (Figures 1–3).

Densitometric Analyses

Densitometric analyses were performed using National Institutes of Health (NIH)-Image software, available in the public domain (http://rsb.info.nih.gov/nih-image/Default.html), developed by Wayne Rasband for Macintosh® computers.

Biochemical Activity Assays

The isocitrate dehydrogenase-1 activity of the purified Rv3339c solubilized in the presence of 500 mM glycylglycine in the medium was measured spectrophotometrically by monitoring the time-dependent reduction of $NADP^+$ to NADPH at 25°C at 340 nm. The standard assay solution contained 20 mM triethanolamine chloride buffer, pH 7.5, 2 mM $NADP^+$, 0.03 mM DL-isocitrate, 10 mM $MgCl_2$, 100 mM NaCl, and 100 pmol of the enzyme in a total reaction volume of 400 µL.

RESULTS AND DISCUSSION

All the ORFs, Rv0256c, and Rv2430c belonging to the PPE family of proteins and Rv1609 (anthranilate synthase) and Rv3339c (isocitrate dehydogenase-1) of *M. tuberculosis* were found to have enhanced solubility in the presence of the dipeptide glycylglycine in TB media. Because these proteins were expressed with an N-terminal histidine tag, the relative amount of protein going into the soluble fractions was determined by subjecting equal amounts of soluble proteins to SDS polyacrylamide gel electrophoresis (PAGE) and subsequent detection by Western blot analysis using monoclonal anti-his tag antibodies. In the case of Rv0256c, the amount of soluble protein

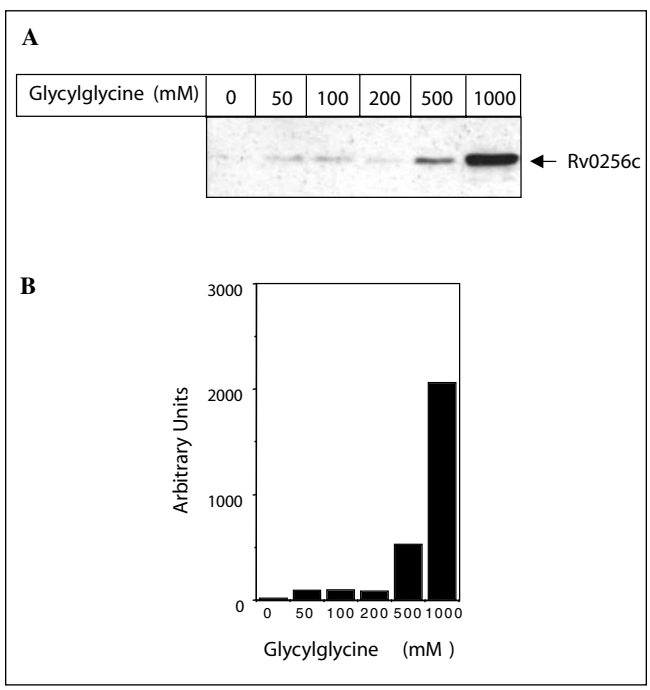

Figure 1. (A) Glycylglycine enhances the solubility of Rv0256c expressed in *Escherichia coli*. About 60 µg of soluble protein from the sonicated extract were loaded for each lane and transferred onto the nitrocellulose membrane. Upon Western transfer, the blot was probed with monoclonal anti-6x-histidine antibodies and developed by luminol reagents. The concentrations of glycylglycine used are indicated at the top of each lane. (B) Densitometric analyses of the same Western blot, in which the density of the individual bands was plotted against the glycylglycine concentrations used in the culture medium.

Table 2. Relative Levels of Soluble Proteins Expressed in the Presence of Glycylglycine for Different Mycobacterial Proteins

Protein	Solubility in Glycylglycine (mM) 0	50	100	200	500	1000	*Escherichia coli* Strain Used
Rv0256c	-/+	+	+	+	+++	++++++++	Bl21(DE3)pLysS
Rv2430c	-/+	-/+	-/+	++	++	-/+	M15(pREP4)
Rv1609	-	-	-	-	+	++	Bl21(DE3)pLysS
Rv3339c	-	-	-	+	++	+	Bl21(DE3)pLysS

- = not detectable; + = detectable

was very low when grown in the absence of glycylglycine (Figure 1A). However, with the increasing concentrations of glycylglycine, the amount of soluble protein dramatically increased, with 1 M glycylglycine being the most effective. In the presence of 1 M glycylglycine, there was a more than 170-fold increase in solubility. Use of glycylglycine higher than 1 M was found to affect the growth of the bacteria (data not shown); therefore, concentrations greater than 1 M were not used.

For the other proteins, Rv2430c, Rv1609 (anthranilate synthase), Rv3339c (isocitrate dehydrogenase-1), our initial attempts to express these proteins in soluble form employing various other strategies such as low temperature induction, induction at low IPTG concentration, etc., proved to be futile. However, when we induced these proteins in the presence of glycylglycine, soluble proteins were readily detected as compared to the undetectable level in experiments without glycylglycine (Figure 2, A–C). For Rv3339c, the maximum soluble yield was in the presence of 500 mM glycylglycine (Figure 2A), whereas for Rv1609, the maximum soluble expression was achieved in presence of 1 M glycylglycine (Figure 2B). Similarly, in the case of Rv2430c, the soluble expression was maximum in the presence of 200 mM glycylglycine (Figure 2C). The results are summarized in Table 2.

We next tested whether or not the recombinant proteins solubilized in the presence of glycylyglycine possessed biological activity. For this, we purified the soluble protein coded by the Rv3339c ORF, encoding isocitrate dehydrogenase-1, solubilized in the presence of 500 mM glycylglycine in the medium, and assayed its ability to reduce $NADP^+$ to NADPH. The purified solubilized protein was found to possess enzymatic activity following a typical Michaelis-Menten reaction kinetics (Figure 3) and was stable for over 1 week at 4°C.

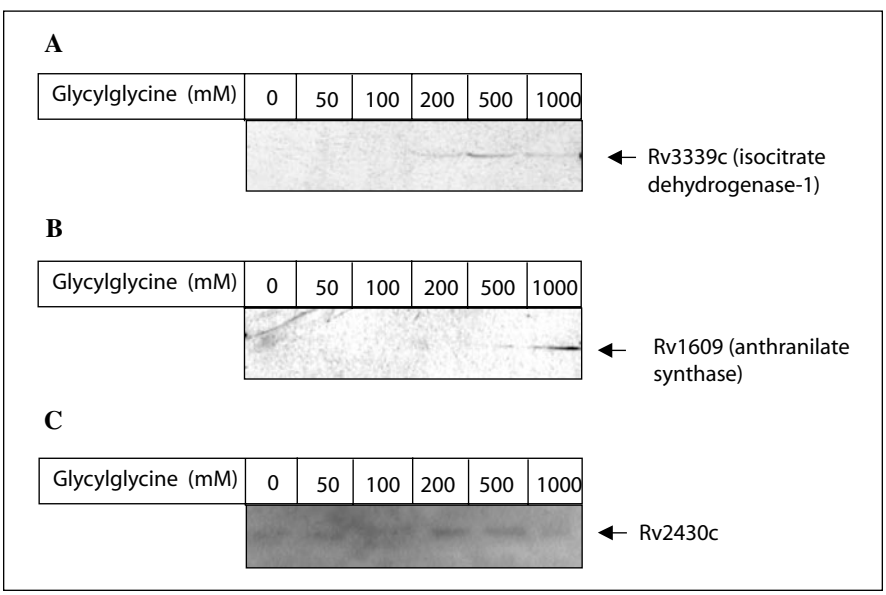

Figure 2. Glycylglycine enhances the solubility of other mycobacterial proteins expressed in *Escherichia coli*. About 60 µg of soluble protein from the sonicated extract were loaded in each lane and transferred onto the nitrocellulose membrane. The membrane was probed with monoclonal anti-6x-histidine antibodies and developed by luminol reagents. The concentrations of glycylglycine used are indicated at the top of each lane. (A) Western blot of Rv3339c. (B). Western blot of Rv1609. (C). Western blot of Rv2430c.

We provide a simple and novel way of enhancing solubility of difficult proteins that are otherwise expressed in a nonproductive fashion into inclusion bodies. Inclusion body formation may be a consequence of the rate of protein translation exceeding the capacity of the cell to fold the newly synthesized protein correctly (11). This phenomenon has been defined as secretory load for the baculovirus insect cell system (12) and has been addressed by reducing the transcription level using a weaker promoter or by allowing more time for the insect cells to process the recombinant protein (13). Therefore, decreasing the rate of protein production is one of the major strategies to overcome this problem. Some general approaches to achieve this have been to induce the cells at lower temperature (14), use low IPTG concentration (15), use weak promoters (16), etc. Yet other methods utilize various "compatible solutes" to induce osmotic stress (17,18). Improved solubility has also been reported by the use of a specific host strain in which heterologous proteins can be expressed upon osmotic induction with high salt (19). Although the method was initially developed to enhance the soluble expression of some of the mycobacterial proteins, this method is also applicable to other proteins refractory to soluble expression in *E. coli* systems.

The mechanism of glycylglycine-mediated enhanced solubilization remains to be understood. One of the possibilities could be the increased osmolarity of the medium by the dipeptide. Osmotic stress induces the expression of heat shock proteins with chaperone-like activity to assist correct folding. Another possibility is the direct interaction of glycylglycine with the expressed protein by acting as a chemical chaperone (20,21). *E. coli* is known to possess specific transporters for dipetides and oligopeptides. These in turn are of particular advantage to the bacteria, which thrive in the peptide-rich gut

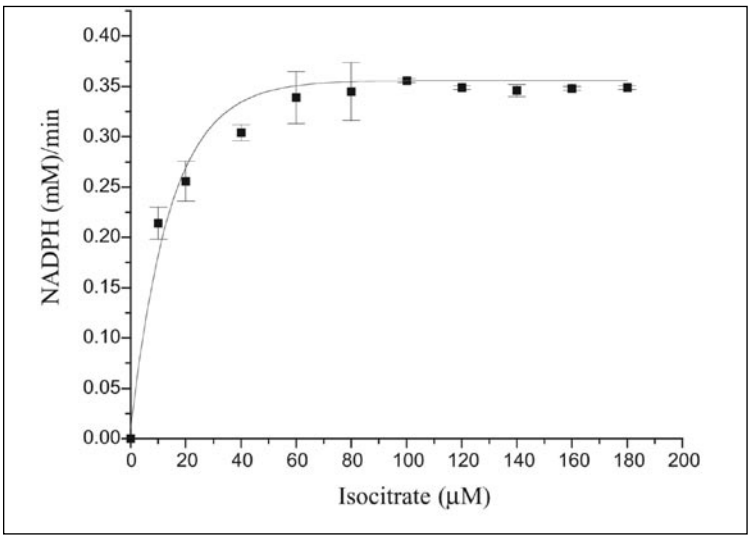

Figure 3. The recombinant protein Rv3339c purified using media containing 500 mM glycylglycine is biochemically active. *Mycobacterium tuberculosis* isocitrate dehydrogenase-1 activity was measured spectrophotometrically by monitoring the time-dependent reduction of $NADP^+$ to NADPH at 25°C at 340 nm. The standard assay solution per 400 µL contained 20 mM triethanolamine chloride buffer, pH 7.5, 2 mM $NADP^+$, 0.03 mM DL-isocitrate, 10 mM $MgCl_2$, 100 mM NaCl, and 100 pmol of the enzyme.

lumen environment (22). Glycylglycine transport behaves similar to other shock-sensitive transport systems requiring ATP for its transport (23). In the presence of higher concentrations of glycylglycine in the media, the bacteria probably ends up spending considerable energy in active glycylglycine transport, thus slowing down the overall metabolic rate including the rate of translation. This probably allows more time for the expressed proteins to be folded correctly. However, it will be interesting to study how a dipeptide can actually help proteins to be folded in its native condition.

ACKNOWLEDGMENTS

This work was supported by core support to CDFD from the Department of Biotechnology, Government of India. S.G. was supported by a Post-Doctoral Fellowship from the Department of Biotechnology, while S.R., S.B., R.K.C., P.C., and S.S.R. were the recipients of Research Fellowships from the Council of Scientific and Industrial Research, New Delhi, India.

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

REFERENCES

1. **Lilie, H., E. Schwarz, and R. Rudolph.** 1998. Advances in refolding of proteins produced in *E. coli*. Curr. Opin. Biotechnol. *9*:497-501.
2. **Amrein, K.E., B. Takacs, M. Stieger, J. Molnos, N.A. Flint, and P. Burn.** 1995. Purification and characterization of recombinant human p50csk protein-tyrosine kinase from an Escherichia coli expression system overproducing the bacterial chaperones GroES and GroEL. Proc. Natl. Acad. Sci. USA *92*:1048-1052.
3. **Goenka, S. and C.M. Rao.** 2001. Expression of recombinant zeta-crystallin in Escherichia coli with the help of GroEL/ES and its purification. Protein Expr. Purif. *21*:260-267.
4. **Yasukawa, T., C. Kanei-Ishii, T. Maekawa, J. Fujimoto, T. Yamamoto, and S. Ishii.** 1995. Increase of solubility of foreign proteins in *Escherichia coli* by coproduction of the bacterial thioredoxin. J. Biol. Chem. *270*:25328-25331.
5. **LaVallie, E.R. and J.M. McCoy.** 1995. Gene fusion expression systems in *Escherichia coli*. Curr. Opin. Biotechnol. *6*:501-506.
6. **Kery, V., D. Elleder, and J.P. Kraus.** 1995. Delta-aminolevulinate increases heme saturation and yield of human cystathionine beta-synthase expressed in Escherichia coli. Arch. Biochem. Biophys. *316*:24-29.
7. **Li, C., J.W. Schwabe, E. Banayo, and R.M. Evans.** 1997. Coexpression of nuclear receptor partners increases their solubility and biological activities. Proc. Natl. Acad. Sci. USA *94*:2278-2283.
8. **Du Bois, G.C., S.P. Song, I. Kulikovskaya, J.L. Rothstein, M.W. Germann, and C.M. Croce.** 2000. Purification and characterization of recombinant forms of murine Tcl1 proteins. Protein Expr. Purif. *18*:277-285.
9. **Sambrook, J., E.F. Fritsch, and T. Maniatis.** 1989. Molecular Cloning: A laboratory Manual. CSH Laboratory Press, New York.
10. **Laemmli, U.K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*:680-685.
11. **Kiefhaber, T., R. Rudolph, H.H. Kohler, and J. Buchner.** 1991. Protein aggregation *in vitro* and *in vivo*: a quantitative model of the kinetic competition between folding and aggregation. Biotechnology (NY) *9*:825-829.
12. **Sridhar, P. and S.E. Hasnain.** 1993. Differential secretion and glycosylation of recombinant human chorionic gonadotropin (beta hCG) synthesized using different promoters in the baculovirus expression vector system. Gene *131*:261-264.
13. **Sridhar, P., A.K. Panda, R. Pal, G.P. Talwar, and S.E. Hasnain.** 1993. Temporal nature of the promoter and not relative strength determines the expression of an extensively processed protein in a baculovirus system. FEBS Lett. *315*:282-286.
14. **Schein, C.H. and M.H.M. Noteborn.** 1988. Formation of soluble recombinant proteins in *Escherichia coli* is favored by low growth temperature. BioTechnology *6*:291-294.
15. **Sawyer, J.R., J. Schlom, and S.V. Kashmiri.** 1994. The effects of induction conditions on production of a soluble anti-tumor sFv in *Escherichia coli*. Protein Eng. *7*:1401-1406.
16. **Swartz, J.R.** 2001. Advances in *Escerichia coli* production of therapeutic proteins. Curr. Opin. Biotechnol. *12*:195-201.
17. **Barth, S., Huhn, B. Matthey, A. Klimka, E.A. Galinski, and A. Engert.** 2000. Compatible-solute-supported periplasmic expression of functional recombinant proteins under stress conditions. Appl. Environ. Microbiol. *66*:1572-1579.
18. **Blackwell, J.R. and R.A. Horgan.** 1991. Novel strategy for production of a highly expressed recombinant protein in an active form. FEBS Lett. *295*:10-12.
19. **Bhandari, P. and J. Gowrishankar.** 1997. An *Escherichia coli* host strain useful for efficient overproduction of cloned gene products with NaCl as the inducer. J. Bacteriol. *179*:4403-4406.
20. **Gekko, K. and S. Koga.** 1983. Increased thermal stability of collagen in the presence of sugars and polyols. J. Biochem. (Tokyo) *94*:199-205.
21. **Ou, W.B., Y.D. Park, and H.M. Zhou.** 2002. Effect of osmolytes as folding aids on creatine kinase refolding pathway. Int. J. Biochem. Cell Biol. *34*:136-147.
22. **Sussman, A.J. and C. Gilvarg.** 1971. Peptide transport and metabolism in bacteria. Annu. Rev. Biochem. *40*:397-408.
23. **Cowell, J.L.** 1974. Energetics of glycylglycine transport in *Escherichia coli*. J. Bacteriol. *120*:139-146.

Received 23 October 2003; accepted 8 June 2004.

Address correspondence to Seyed E. Hasnain, Centre for DNA Fingerprinting and Diagnostics, Nacharam, Hyderabad, 5000076, India. e-mail: ehtesham@www.cdfd.org.in