INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

NEW FUSION PROTEIN SYSTEMS STATISTICALLY DESIGNED TO AVOID INCLUSION BODY FORMATION IN *ESCHERICHIA COLI*

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

By

GREGORY DEAN DAVIS

Norman, Oklahoma

1998

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

NEW FUSION PROTEIN SYSTEMS STATISTICALLY DESIGNED TO AVOID INCLUSION BODY FORMATION IN *ESCHERICHIA COLI*

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

By

GREGORY DEAN DAVIS

Norman, Oklahoma

1998

NEW FUSION PROTEIN SYSTEMS STATISTICALLY DESIGNED TO AVOID INCLUSION BODY FORMATION IN *ESCHERICHIA COLI*

A DISSERTATION APPROVED FOR THE SCHOOL

OF CHEMICAL ENGINEERING AND MATERIALS SCIENCE

By

Advisory Committee

ACKNOWLEDGEMENTS

I would like to thank my graduate committee for their support and constructive criticism. In particular, I would like to express my sincere appreciation to Dr. Roger Harrison for his guidance, encouragement, and simply for giving me the opportunity to further my education on such a good research project. I count myself to be very lucky. I would also like to thank Dr. Matthias Nollert and Dr. Edgar O'Rear for their support, encouragement, and helpful advice while I learned how to do research.

I would like to thank several others who generously lent their facilities and expertise to this project: Dr. Daniel Riggs for his advice on cloning procedures and the use of his laboratory equipment, Dr. Bruce Roe for his advice on searching and analyzing protein sequence data using the OUGCG system, Dr. David McCarthy for his advice in plasmid biology, and Dr. Robert House of the Illinois Institute of Technology-Chicago for performing human interleukin-3 cell proliferation assays.

Much credit should be given to two outstanding undergraduate students, Claude Elisee and Denton Newham, who learned recombinant DNA technology in a relatively short amount of time and were very successful at implementing it. Their skill and effort is much appreciated. I would also like to thank Elisabeth Fayard for her friendship in the laboratory and for her willingness to listen to me destroy the French language.

This work was supported by grant No. 9502235 from National Science Foundation.

DEDICATION

This dissertation is dedicated to those people who by their love, support, and presence have guided me through my life..... my late grandfather Maurice H. Davis (I wish I had the chance to know you better) and his widow, Ruth Davis (the best golfer in the world, four hole-in-ones in four decades!), my grandfather LeRoy E. Erwin (never saw a mechanical device he couldn't improve) and his wife Yvonne Erwin (I'll never forget how you played that accordion at Christmas), my late grandmother Blanch Erwin (I also wish I could have known you better), my father Garland D. Davis (I'll always remember you playing Scott Joplin Ragtime music on the piano on Sunday mornings before church), my mother Sandra D. Davis (I'll never forget going to the Ft. Worth Museum of Natural Science with you, looking at those gigantic dinosaur bones, and learning about the "meat-eaters" [i.e. meteors]), my sister Stacy A. Deaton (I loved those two years we were in the DeKalb highschool band together, thank you for inspiring me to be there and to learn music), my Aunt Sandy Davis (you always ask me such hard questions!), my cousin (or should I say brother?) and lifelong friend David M. Erwin and his family, Gene, Donna, and Holly.

Also, my school, music, and mountain hiking compadré, Gary Jacobs, just for being there when I needed him (I'll never forget that hike we made in the early morning in March of 1993, walking up through and above the clouds, to the top of Guadalupé Peak, the top of Texas), my church family at St. Stephen's United Methodist Church, specifically, Rev. Craig Stinson for giving my heart the capacity to hold both science and faith, and Dr. Mary Jo Ruggles who gave me the opportunity to continue to express myself through music in the St. Stephen's United Methodist Choir.

And, finally, my wife and best friend in the world, Shannon E. Davis (Matthiesen). You'll never know how lucky I feel that I get to share my life with you. You will never know. You are the best. Love always. -- Greg

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Figure **Page**

LIST OF TABLES

Appendix Tables

LIST OF APPENDICES

ABSTRACT

Using human interleukin-3 (hIL-3) as a model heterologous insoluble protein, three gene fusions were constructed that code for a native *E. coli* protein at the N-terminus and hIL-3 at the C-terminus. The three native *E. coli* proteins, NusA, GrpE, and bacterioferritin (BFR), were chosen based on their favorable cytoplasmic solubility characteristics as predicted by a statistical solubility model for recombinant proteins in *E. coli*. Modeling predicted the probability of soluble fusion protein expression in the following order: NusA (most soluble), GrpE, BFR, and thioredoxin (least soluble). Expression experiments showed that NusA/hIL-3 fusion protein was expressed almost completely in the soluble fraction while GrpE/hIL-3 and BFR/hIL-3 exhibited partial solubility at 37°C. Thioredoxin/hIL-3 was expressed almost completely in the insoluble fraction. hIL-3 was purified to homogeneity from the NusA/hIL-3 fusion protein using an N-terminal histidine tag, factor Xa protease cleavage, and anion exchange chromatography. These results have a broad significance to the biotechnology field since insoluble heterologous protein expression in *E. coli* continues to be a pervasive problem in both industrial and academic research. Also, these results represent a successful example of how statistical modeling can be used to design a novel protein expression system for a specific cell type.

CHAPTER I

INTRODUCTION

Goals of Research and Project Overview

A major benefit resulting from the rise of recombinant DNA technology has been the availability of relatively large amounts of pure protein for medical studies, industrial production, and fundamental academic research. While production of foreign proteins in bacterial hosts has been implemented successfully in the biotechnology industry, there have been numerous occasions where bacterial expression systems have failed to yield properly folded, biologically active proteins. For example, when high expression levels are achieved, recombinant proteins are frequently expressed in *Escherichia coli* as insoluble protein aggregates termed "inclusion bodies" which have been the subject of many protein folding studies (reviewed by Georgiou and Valax, 1996; King et al., 1996). Although initial purification of inclusion body material by cell lysis and centrifugation is relatively simple, the agglomerated protein must be subsequently refolded into active form, which is typically a cumbersome trial and error procedure yielding a process in which the costs are very significant. A detailed economic analysis by Ernst et al. (1997) concluded that soluble protein production (even at modest expression levels) is less expensive than insoluble expression since the low concentrations required for refolding can cause the cost of production to increase exponentially. Thus, considering these refolding costs and the general doubt associated with basic research conducted using refolded proteins, it is much more desirable to maximize soluble heterologous protein expression whenever possible.

One strategy to avoid inclusion body formation is to fuse the protein of interest (a.k.a. the target protein) to a protein known to possess high expression levels in *E. coli* (a.k.a. the carrier protein). The success of such a method depends largely upon the ability to predict the combined characteristics of the two proteins as a fusion. This raises a very common biochemical question which is central to the design of this study: how can protein structure and function be accurately predicted based upon amino acid content alone? The particular "function" focus of this study is simply recombinant protein solubility in the *E. coli* cytoplasm at 37°C. The general hypothesis is that if a carrier protein is highly soluble in the *E. coli* cytoplasm, it will allow a normally insoluble protein to be expressed in soluble form when both proteins are expressed together as a fusion.

Currently, the most widely implemented carrier proteins for fusion protein expression in *E. coli* are *Schistosoma japonicum* glutathione S-transferase (GST) (Smith and Johnson, 1988), *E. coli* maltose binding protein (MBP) (di Guan et al., 1988), and *E. coli* thioredoxin (LaVallie et al., 1993). The primary reason for selecting GST and MBP as fusion partners was the combined advantage of high level expression and affinity purification, while thioredoxin was chosen primarily for high level *soluble* protein expression. 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. Thus, it is the primary purpose of this study to discover new carrier proteins systematically by considering their solubility potential based on a statistical solubility model for recombinant proteins and create a

protein expression system which has been optimized for soluble protein expression *and* affinity purification. A parallel goal of this research is to characterize the ability of the solubility model, in its present form, to predict the solubility of a carrier protein fused to a target protein.

The Wilkinson-Harrison Solubility Model

With the purpose of being able to correctly predict inclusion body formation in *E. coli* based on amino acid content, a correlation has been developed called the Wilkinson-Harrison solubility model (Wilkinson and Harrison, 1991). This model predicts the probability of protein solubility in the *E. coli* cytoplasm based on the content of eight amino acids: arginine, lysine, aspartate, glutamate, asparagine, glycine, proline, and serine. The basis for this model is a discriminant statistical analysis of 81 different proteins that have been previously characterized as soluble or insoluble when overexpressed in *E. coli* at 37°C. In the development of this model, five characteristic parameters related to amino acid content were analyzed with respect to inclusion body formation. Of these five, two showed significant correlation to inclusion body formation, namely: turn forming residue content (i.e. numbers of Asn, Gly, Pro, and Ser) and approximate charge average (i.e. the numbers of Arg and Lys minus the number of Asp and Glu divided by the total number of amino acids). In short, the statistical analysis resulted in a simple algebraic equation:

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

where:

CV = canonical variable (a.k.a. composite parameter)
n = number of amino acids in protein
N,G,P,S = numbers of turn forming residues
R,K,D,E = numbers of charged residues
λ_1 , λ_2 = optimized parameters (15.43 and -29.56 respectively)

When CV was calculated and averaged for all proteins, the result was the discriminant, CV' = 1.71. Subtracting individual protein CV values from the discriminant CV' estimates the likelihood that a protein will form inclusion bodies. If CV-CV' is positive, the protein is predicted to be insoluble. If CV-CV' is negative, the protein is predicted to be soluble. The solubility probability is then calculated by using a polynomial fit to estimate the integral of the frequency distribution of CV-CV' values, namely:

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

Before this study started, the model was incorporated into a Microsoft Word Excel spreadsheet which required the manual input of amino acid data to perform solubility calculations. One accomplishment of this project is that the model has been incorporated into a custom C program which can count all of the relevant amino acids, calculate the solubility potential of several thousand proteins, and tabulate the results in an external text file. Consequently, the solubility probability of all proteins in the *E. coli* genome can be calculated and sorted in a matter of seconds.

Experimental Summary

A flow chart which covers the major steps of this project is shown in Figure 1. The various carrier and target proteins selected for this study are shown in Figure 2. The carrier proteins NusA, GrpE, BFR, and YjgD were selected by screening the approximately 4,000 known *E. coli* protein sequences registered in the SWISS-PROT databank using the two-parameter solubility model. Thioredoxin was chosen as a control carrier protein due to its widespread use as a soluble fusion partner, and hIL-3 was chosen as a model insoluble target protein since it is expressed in inclusion body form at 37°C as a fusion to thioredoxin (LaVallie et al., 1993) and when expressed alone (Donahue et al., 1988). Bovine growth hormone (bGH) and human interferon-γ (hIFN-γ) were chosen as additional target proteins for NusA to further characterize the NusA fusion system. A factor Xa protease has been included in all fusion proteins for the release of the target protein.

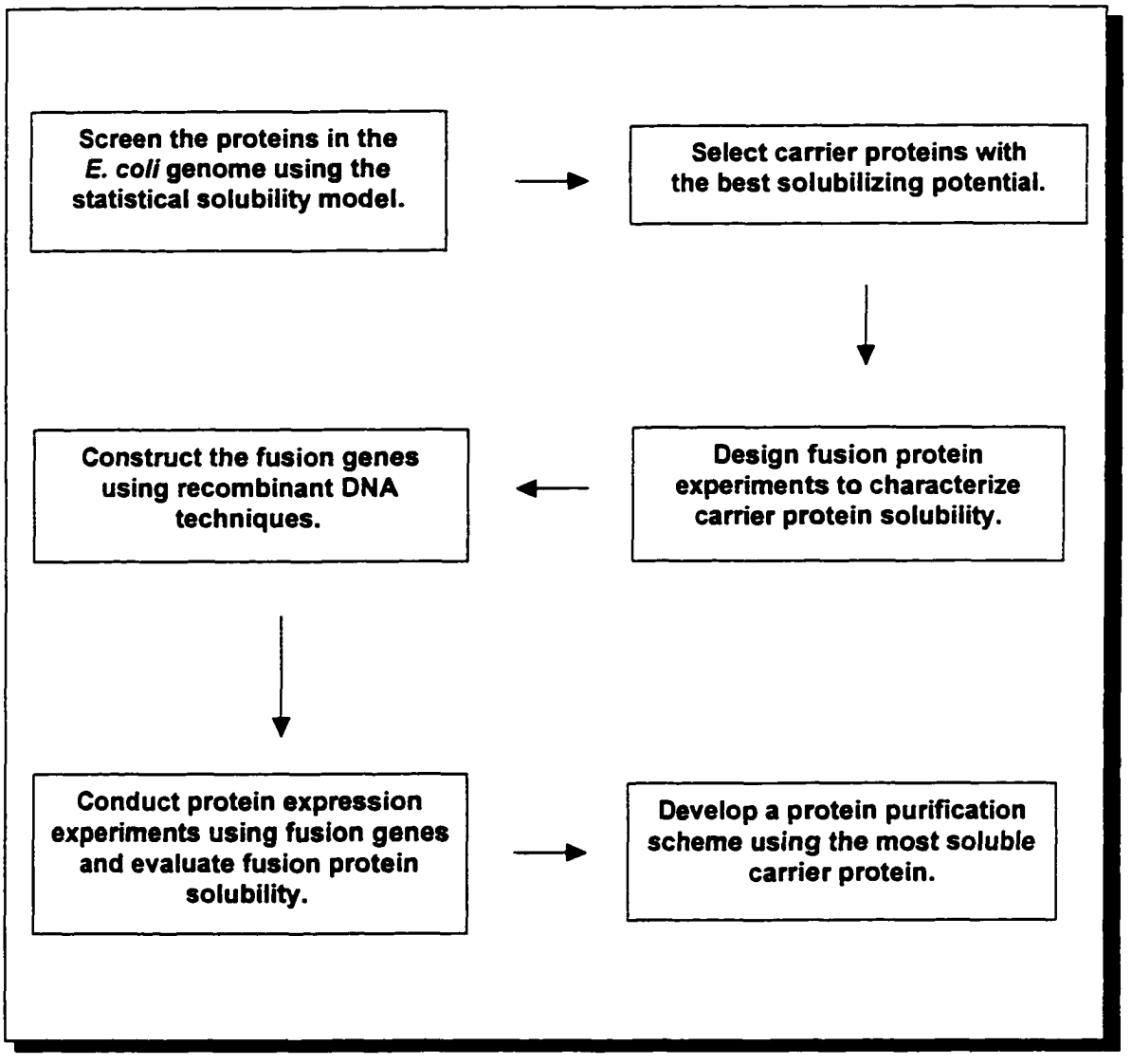


Figure 1. Flow chart showing the overall structure of this project.

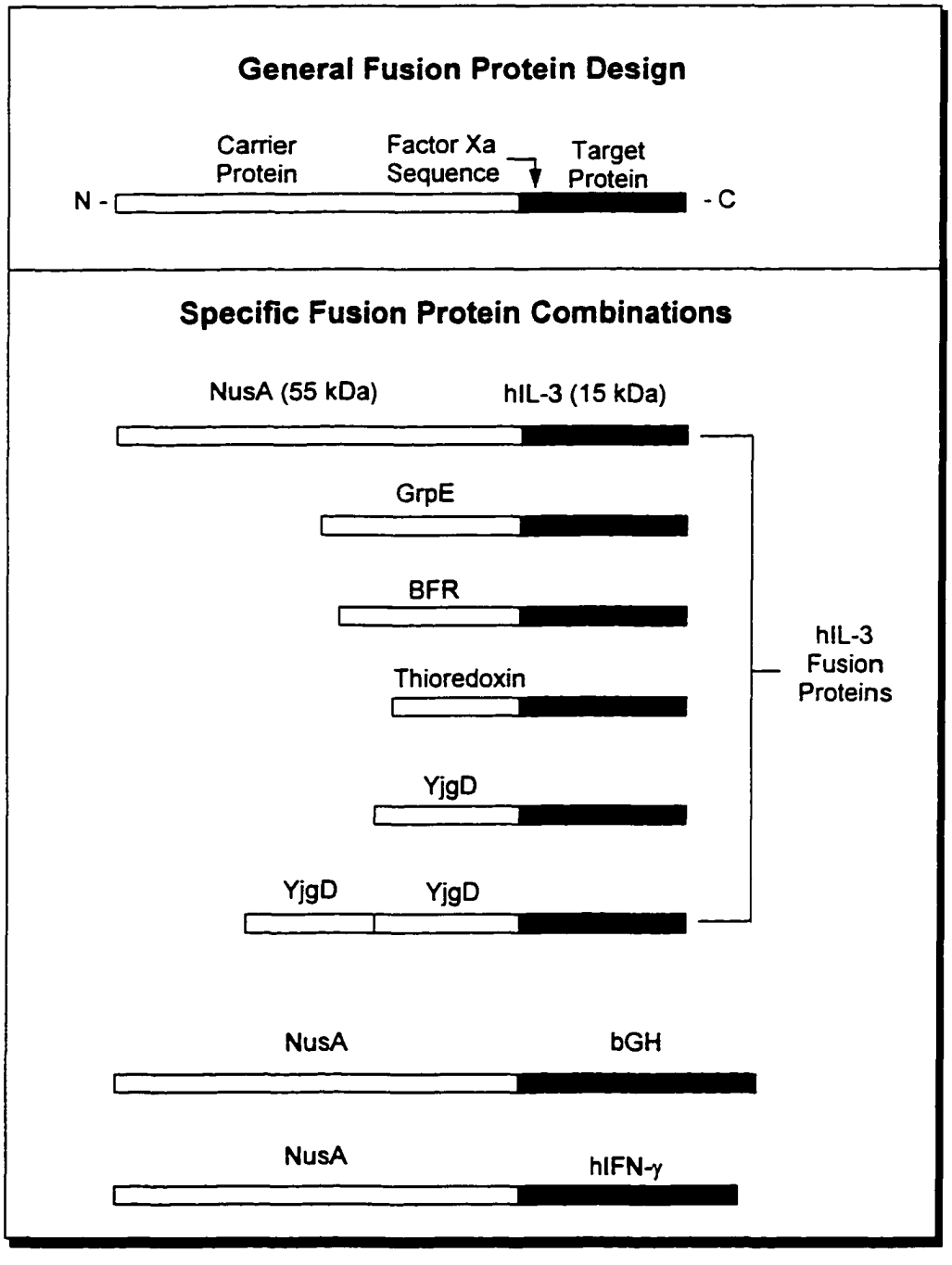


Figure 2. General structure and relative sizes (not to scale) of fusion proteins constructed for this study. A factor Xa protease site has been inserted between each carrier and target protein. Six hIL-3 fusion proteins were constructed to screen various carrier proteins for solubilizing potential. Two additional target proteins, bGH and hIFN-γ, were constructed to test the versatility of the NusA protein in solubilizing other proteins.

Current Fusion Protein Technology

Amongst the assortment of current methods for heterologous protein expression in *E. coli,* fusion protein technology has certainly proven itself as one of the most pragmatic and reliable methods. This is evident from the frequent use of fusion proteins in basic research and from the fact that several life science companies have marketed fusion protein systems. Fusion proteins have several advantages in that the carrier protein can confer upon the target protein favorable characteristics such as novel purification options (i.e. affinity "tags"), localization within (or secretion from) the cell, and best of all, high level expression. Considering the multiple domain structure of many wild type proteins resulting from natural evolution, the fusion protein approach to enhancing protein expression can be considered, in itself, a form of directed evolution: mixing and mingling together various portable protein functionalities to achieve a protein with the right personality for the task at hand. Fusion protein components can generally be divided into four categories: 1) carrier proteins which facilitate high level expression, 2) small affinity tags and signal sequences, 3) protease cleavage sequences, and 4) target proteins of interest. However, these categories are not always mutually exclusive. For example, some carrier proteins can facilitate both high level expression and affinity purification.

Carrier Proteins

The commercial advantages of fusion protein technology were first demonstrated in the production of somatostatin (Itakura et al., 1977) and the insulin A and B chains (Goeddel et al., 1979) as β-galactosidase fusion proteins. These fusions to β-galactosidase

resulted in a high level expression of the target proteins in inclusion body form, protecting them from host cell protease degradation. The localization of the fusion protein in inclusion bodies provided a method which achieved purification of the fusion protein by simple centrifugation, virtually separating the fusion protein from all soluble cell proteins in one step. The small size of somatostatin (14 amino acids) and the insulin A and B chains (21 and 30 amino acids, respectively) allowed for efficient recovery of biological activity from inclusion body material by refolding with denaturants. Over the years, the need for recombinant production of larger proteins has increased, and the refolding of larger proteins from inclusion bodies has proven to be a more arduous task (Rudolph and Lilie, 1996). Thus, soluble expression has become of major importance with current fusion protein technology. As mentioned earlier, the most popular fusion protein systems are GST, MBP, and thioredoxin.

The primary advantage of the GST and MBP systems is that they are able to allow the high level expression of heterologous proteins; and, if the resulting fusion protein is soluble, both systems offer marvelous affinity purification properties. GST fusion proteins can be purified using glutathione-Sepharose media, and MBP fusion proteins can be purified with an amylose resin. However, after several years of application, it has not been uncommon for GST fusion proteins to be found insoluble, and considerable effort has been spent attempting to develop standard refolding procedures for GST fusion proteins (Frangioni and Neel, 1993).

The improvements in solubility gained with the use of thioredoxin as a fusion partner are steadily increasing thioredoxin's popularity, and it is currently marketed by

Invitrogen (Carlsbad, CA). To date, it is the smallest carrier protein which has been found to dramatically effect fusion protein solubility. Thioredoxin has been shown to be a useful carrier protein in the soluble expression of human atrial naturetic peptide (Wilkinson et al., 1995) and several human and murine cytokines in *E. coli* (LaVallie et al., 1993). Since thioredoxin has no known natural affinity for other biomolecules strong enough to be useful in protein purification, some modifications have been made to the fusion protein expression system. In one study, spatially adjacent amino acid residues in thioredoxin were mutated to histidine to enhance binding in immobilized metal affinity chromatography (Lu et al., 1996). More recent research has focused on an *in vivo* biotinylation of thioredoxin fusion proteins (Smith et al., 1998). In this system, a short 'biotinylation peptide' is fused to the N terminus of thioredoxin which targets the fusion protein for biotinylation by the natural biotinylation system in *E. coli*. Thus, with the recent development of these affinity modifications, the thioredoxin system has easily assimilated the affinity purification advantages of the GST and MBP fusion protein systems while retaining a significant advantage in soluble expression. It should be noted, however, that it was very common that reductions in expression temperature were necessary (sometimes as low as 15°C) for the majority of thioredoxin fusion proteins expressed by LaVallie et al. (1993). While this is not a detrimental problem for lab scale protein preparation, decreases in growth rates due to temperature reductions can significantly impact large scale process economics.

With respect to overall protein yield, one of the most successful carrier proteins has been ubiquitin (Pilon et al., 1996). It has been possible to produce yields up to 709

mg/L with this system where ubiquitin fusion proteins can represent up to 90% of the total soluble cell protein. It was found that maximal specific yields were obtained when induction with IPTG and a 12°C heat shock (from 30°C to 42°C) were performed simultaneously. It is quite impressive that the fusion protein can comprise 90% of the soluble cell protein and still remain soluble at 42°C. While the small size of ubiquitin, 76 amino acids, has limited its application to small target peptides (the largest target protein expressed by Pilon et al. was 53 amino acids in length), the ubiquitin expression system has tremendous potential for the economical large scale production of peptides.

Very recently, Zhang et al. (1998) performed a series of fusion protein experiments with a mutant form of *E. coli* DsbA ($DsbA^{mut}$) as the solubilizing carrier protein. In its best case, $DsbA^{mut}$ was able to increase the percent solubility of transforming growth factor-β-2 (TGFβ-2) from 8.6 to 86%, which was only 7% more than the solubility of the thioredoxin/TGFβ-2 fusion protein. In addition to the $DsbA^{mut}$ fusion protein studies, Zhang et al. performed experiments which they based on the solubility parameters of the Wilkinson-Harrison solubility model. A particular target protein, IGF-binding protein-3 (IGFBP-3), was subjected to several amino acid substitutions, Asn to Asp, which in effect increased the approximate charge average of IGFBP-3 in the negative direction. The solubility of *each* mutant IGFBP-3 was increased over the wild type protein, providing further experimental evidence that the approximate charge average, one of the parameters in the solubility model, is a valuable parameter to consider when overexpressing proteins in *E. coli*. Furthermore, Zhang et al. coexpressed a protein prolyl isomerase, cyclophilin (L-), with a $DsbA^{mut}$ fusion protein; and while the

total amounts of expressed protein were increased, there was little effect on fusion protein solubility. Their reasoning was that cyclophilin would catalyze the folding rate of proline turns during translation of the target, thus avoiding the kinetic trap of misfolded intermediate protein aggregation. It was very interesting that the experiments of Zhang et al. indicate that the overall charge of a protein has a more dramatic effect on solubility than the use of a prolyl isomerase. These results are consistent with five parameter version of the solubility model (Wilkinson and Harrison, 1991) in the sense that the weighted λ value for the turn-forming residue content (λ_{1w}= 0.50) is almost 42% lower, in absolute value, than the approximate charge average (λ_{2w}= -0.81).

Small Affinity Tags

The first application of fusion protein technology to assist in affinity protein chromatography was developed by Uhlen et al. (1983) in which staphylococcal protein A was fused to *E. coli* β-galactosidase. Using IgG-sepharose, the fusion protein could be affinity purified with high efficiency while retaining β-galactosidase activity. Since the protein A study, several fusion protein systems which provide affinity purification have been developed, some of which are listed in Table 1. Probably the single most important development in affinity purification of recombinant proteins in recent years has been the N-terminal polyhistidine peptide (a.k.a. hexahistidine tag or His_6-tag) (Ljungquist et al., 1989). The main advantages of the polyhistidine peptide are 1) that its small size (usually 6 residues) reduces interference with native protein function, thus eliminating the need for protease cleavage prior to structure/function studies in many cases, 2) the inexpensive

Fusion partner	Size	Ligand	Relative Use in Research*	Primary Reference
Staphylococcal protein A	57 kDa	IgG-sepharose	5	Uhlen et al. (1983)
Glutathione-S-transferase	25 kDa	Glutathione-sepharose	100	Smith and Johnson (1988)
Maltose binding protein	40 kDa	Amylose resin	24	di Guan et al. (1988)
His_6	6 aa	Ni^{2+}- nitriloacetic acid	13	Ljungquist et al. (1989)
Cellulose binding domain	11 kDa	cellulose	1	Ong et al. (1995)
Calmodulin binding peptide	26 aa	Calmodulin	1	Zheng et al. (1997)
Biotin peptide tag	13 aa	avidin/ streptavidin media	<1	Smith et al. (1998)

*Based on a 1993-1998 survey of PubMed abtstracts. The survey was limited to studies classified under the substance heading, [SUBS], called "Recombinant Fusion Proteins". Numbers are relative to Glutathione-S-transferase = 100, with 1,468 citations found. It is likely that the number of citations using the His_6 tag is underestimated since several names are in use for polyhistidine peptide tags.

Table 1. Fusion partners which faciliate affinity purification of fusion proteins.

purification method of immobilized metal affinity chromatography (IMAC), 3) the fact that proteins can sometimes be purified under denaturing conditions, and 4) the easy insertion of the six histidine sequence during PCR which provides for quick and easy cloning methods. In recent years, antibodies have also been developed for polyhistidine tails allowing quick screening using ELISA and western blot procedures. One advantage that the GST system has over the polyhistidine approach is that if a GST fusion protein is expressed in inclusion body form, it is sometimes easier to refold the protein with denaturants than compared with refolding the target protein alone (Ray et al., 1993). This has also been found to be true for insoluble MBP and thioredoxin fusion proteins (Sachdev and Chirgwin, 1998).

Ong et al. (1995) were able to express and purify human interleukin-2 (hIL-2) by fusing it to the cellulose-binding domain of *Cellulomonas fimi* endo-β-1,4-glucanase A (CenA). When overexpressed in in *E. coli*, it was necessary to recover the CenA/hIL-2 fusion protein from the inclusion body fraction. Once recovered, the CenA/hIL-3 fusion protein was purified by adsorption onto cellulose and hydrolyzed *in situ* to release active hIL-2. A potential advantage of this expression system is that when cellulose particles, fabric, or membranes are used as purification media, downstream processing could become very inexpensive for larger scale processes.

Hasenwinkle et al. (1997) have developed a secretion/expression system for *E. coli* which consists of the cellulose binding domain (CBD) and the leader sequence from the exoglucanase, Cex, of *Cellulomonas fimi*. The Cex/CBD fusion protein was produced at soluble levels up to 8.2 g/L, representing 70% of the total protein in the supernatant at

the end of fermentations. However, this study did not include a target protein and only focused on high level expression of the Cex/CBD fusion protein, so the benefits of this system remain to be demonstrated.

Calmodulin-binding peptide (CBP), a 26 amino acid peptide, is derived from muscle myosin light chain kinase (MLCK) and binds to calmodulin with a high affinity. Zheng et al. (1997) at Stratagene Cloning Systems have studied proteins fused in frame with CBP and shown that they can be purified from crude *E. coli* lysates in a single step using calmodulin affinity chromatography. Since the CBP-calmodulin affinity is calcium-dependent, fusion proteins can be eluted from the resins with any buffer containing ethylene glycol-bis(β-aminoethyl ether) N,N,N',N'-tetraacetic acid (EGTA) and used directly for many applications.

As mentioned earlier, *in vivo* biotinylation using a small 13 aa peptide tag has been developed (Schatz, 1993) and has been incorporated into a thioredoxin fusion protein system (Smith et al., 1998). Also, a truncated rat neurotensin receptor (NTR) has been purified using the biotin tag and was found to give higher yields than hexahistidine tag purification of NTR (Tucker and Grisshammer, 1996).

In general, the trend in affinity purification of fusion proteins has been towards smaller and smaller fusion partners, and there is probably much truth in the statement that a six histadine sequence represents the extreme minimum of this trend. There is still room for research, however, with respect to increasing yields, since in some cases more than half of a histidine tagged fusion protein can remain refractory to binding to immobilized metal chelating media (Van Dyke et al., 1992).

Protease Cleavage

Probably the single most limiting aspect of fusion protein technology is the fact that the target protein must be released and purified from the carrier protein. In medical applications it is highly desirable, if not exclusively required, that the recombinantly produced therapeutic protein be identical to the wild type protein. Several chemicals and enzymes are available for protease cleavage, including cyanogen bromide, factor Xa, thrombin, enterokinase, renin, and collagenase. Factor Xa and enterokinase are best for wild type protein production since they cut only on the carboxy-terminal side of the cleavage sequence. Thrombin, although a widely used and efficient protease, cuts at two sequences which do not leave an authentic target protein. These sequences are 1) h1-h2-Pro-Arg/Lys-n1-n2, where h1 and h2 are hydrophobic amino acids and n1 and n2 are nonacidic amino acids, and 2) g1-Arg/Lys-g2, where g1 or g2 are Gly. Cyanogen bromide digestion has also been frequently used, but is relatively non-specific, thus limiting its application to small peptide target proteins. In terms of popularity, factor Xa and thrombin are probably the most frequently used proteases for the separation of large proteins (> 10 kDa). Since factor Xa and thrombin have yet to be recombinantly produced, they remain quite expensive to use at larger scales. Several studies have addressed this problem in a search for recombinant production of known proteases, such as enterokinase (LaVallie et al., 1993) and for new proteases which lend themselves to recombinant production, such as Kex2 protease (Ghosh and Lowenstein, 1996). Kex2 has a good potential for practical use since it is readily overproduced in yeast, is reasonably specific, and cuts at the carboxy terminus of its cleavage sequence Leu-Phe-

Lys-Arg.

Typically, protease cleavage is performed after the fusion protein has been purified to homogeneity, but there have been several attempts to move this step upstream in the purification process. One method has been developed in which the protease contains the same affinity tag as the fusion protein while still retaining specific protease activity (Walker et al., 1994). The protease and fusion protein are loaded onto the same affinity column, and the target protein can be cleaved and isolated in one purification step. Walker et al. have created a recombinant version of human rhinovirus protease 3C ($3C^{pro}$), which cleaves the amino acid sequence Glu-Thr-Leu-Phe-Gln/Gly-Pro and contains a His_6 affinity tag on the N terminus. The only drawback to this method is obtaining a non-authentic target protein with glycine and proline attached to the N terminus. However, if the glycine and proline residues do not significantly affect target protein activity, this type of purification would be very economical because the 6xHis-$3C^{pro}$ fusion protease has already been produced by recombinant methods. Walker et al. also claim that His_6-$3C^{pro}$ cleaves more specificallly than thrombin, factor X_a, or enterokinase.

An even more impressive protease cleavage system (Perez-Martin et al., 1997) has accomplished the feat of simultaneous *in vivo* fusion protein expression and protease cleavage to release a soluble His_6 tagged target protein. *E. coli* DH5α was co-transformed with two separate plasmids, one expressing an MBP-fusion protein and one expressing the NIa protease from the plum pox potyvirus. The His_6 tag was not essential for the cleavage and could have been deleted.

A new class of protein cleavage, termed "protein splicing", has recently been

implemented in fusion protein technology (Chong et al., 1997). Protein splicing is defined as the excision of an intervening protein sequence (a protein intron or "intein") from a protein precursor and the concomitant ligation of the flanking protein fragments (the "exteins") to form a mature extein protein and the free intein (Cooper and Stevens, 1995). Chong et al. have developed a fusion protein system (see Figure 3) where an intein is fused to the C terminus of the target protein, and a chitin binding domain (CBD) is placed at the C terminus of the intein. The fusion protein is purified over immobilized chitin media, and release of the authentic target protein is achieved on the column by adding dithiothreitol (DTT). The DTT causes an N-terminal cytseine in intein protein to become exposed, which then reacts with another domain of the intein that facilitates cleavage. The requirement of a reducing agent for protein cleavage has disadvantages for protein production in *E. coli* since *E. coli* has a reducing cytoplasmic environment. This system has a striking combination of advantages, including: 1) purification yielding authentic proteins with no additional amino acids remaining from fusion protein cleavage, 2) simultaneous purification and cleavage of fusion proteins, and 3) release of fusion partner without use of expensive proteases. One major limitation, however, is that the target protein is N-terminal and is transcribed and translated first. The vast majority of fusion proteins are constructed with the carrier protein at the N terminus. The primary reason for this is that high level expression requires that the mRNA encoding the protein to be expressed contain a ribosome-binding site that is not blocked by mRNA secondary

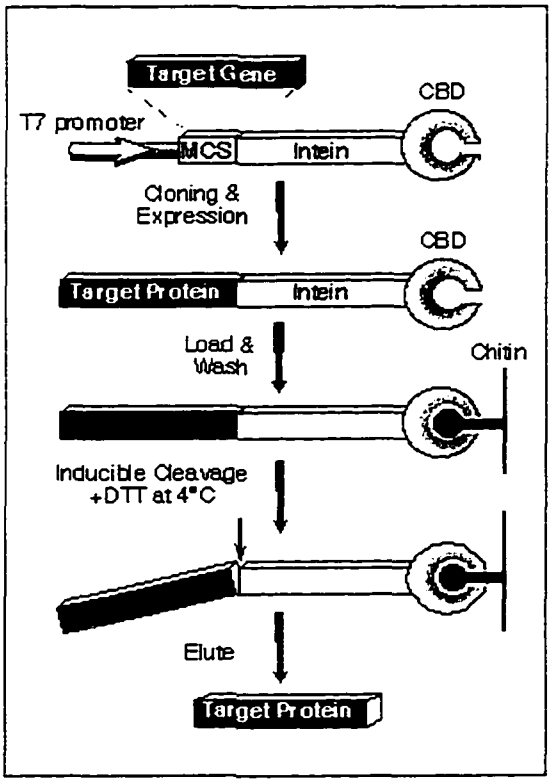


Figure 3. Illustration of the IMPACT T7 fusion protein system developed by New England Biolabs (illustration from an on-line technical bulletin at www.neb.com).

structure (Riggs and LaVallie, 1994). By encoding a protein with high level expression characteristics (e.g. GST, MBP, etc.), there is more certainty that the ribosome-binding site will remain free. A second reason is that N-terminal target proteins sometimes cause the fusion protein to misfold. This phenomena is nicely illustrated in a study of an MBP fusion protein containing aspartic proteinase (Sachdev and Chirgwin, 1998). When bacterial MBP was N-terminal, the fusions were soluble, and when the aspartic proteinase was first, the fusions formed insoluble inclusion bodies. Thus, the question becomes: who can develop an intein fusion protein where the target protein is C-terminal? Once this is accomplished, other domains can be added to the N terminus of the intein that facilitate high level, soluble protein expression.

Non-Fusion Methods for Increasing Protein Expression in *E. coli*

There are two inherent disadvantages that fusion proteins technology: one, as mentioned before, is the need for protease cleavage, and another is a reduction in target protein yield due to the metabolic requirements that the carrier protein places on the host cell for synthesis of the carrier protein itself. When an active target protein can be produced without a carrier protein attached to it, the yields can be quite impressive if purification procedures are efficient and well established. Currently, recombinant bovine growth hormone (rbGH) is produced in inclusion body form and refolded with gram per liter production levels (Monsanto, St. Louis, MO). These levels of production have brought the market value of rbGH down to about $10.00 per milligram, making this an affordable product for the dairy industry. Enhancement of protein expression is therefore

a very active area of research, and there are several reviews concerning the optimization of protein expression in *E. coli* (Hanning and Makrides, 1998; Georgiou and Valax, 1997; Makrides, 1996). Upon examining these reviews it appears that, aside from fusion protein technology, the following methods having proven most successful at producing correctly folded proteins: (1) coexpression of chaperones and/or foldases, (2) reductions in the rate of protein translation to reduce the intracellular concentration of aggregation prone intermediates, (3) localization of the recombinant protein to the periplasm and/or secretion into the growth medium, (4) simply reducing the growth temperature of the culture, (5) refolding of the inclusion body material or solubilization of the protein upon cell lysis using detergents and (6) amino acid substitutions. Several current developments in these areas are summarized below.

Chaperones

The study of molecular chaperones is currently a very active and controversial area of protein research. Chaperones are believed to assist in protein folding and to prevent the aggregation of proteins in inclusion bodies. According to Georgiou et al. (1996), there are three major chaperone systems within the *E. coli* cell: the GroEL/GroES system, the DnaK/DnaJ/GrpE system, and the ClpA/Clpx system. The most extensive research has been performed on the GroEL and DnaK systems. There have been several studies where it has been demonstrated that co-expression of chaperones leads to higher levels of soluble protein production; however, it appears that the effectiveness of chaperone activity is protein specific, thus, limiting the application of chaperones (Hannig

and Makrides, 1998). It has been speculated that fusion proteins may act as chaperones on which translating target proteins can anchor themselves to facilitate proper folding (LaVallie and McCoy, 1995). This idea is particularly interesting to consider in light of this study since GrpE, part of the DnaK system, is one of the fusion partners chosen for hIL-3. Does GrpE have a peptide binding domain on its surface which, when exposed to a misfolded target protein, will bind to it and block any self association of the misfolded target protein? Questions of this sort remain to be answered by protein-protein interaction studies.

Foldases

The term foldase used here refers to any protein which catalyzes protein folding by either disulfide bond formation or *cis/trans* isomerization about proline residues. The foldase currently receiving the most intense focus is an oxidoreductase from *E. coli* named DsbA. In one study (Joly et al., 1998) it has been shown that DsbA can double the overall accumulation of insulin-like growth factor (IGF)-I to 8.5 g/L in the periplasm of *E. coli.* Yet despite the overall increase in yield, the amount of soluble IGF-I was actually decreased when DsbA was coexpressed. This is an interesting finding in light of the results obtained by Zhang et al.(1998) since their DsbA fusion proteins were generally found to increase protein solubility.

SurA is a native *E. coli* protein which has been identified as a protein required for survival during the stationary growth phase. The role of SurA in cell survival has yet to be understood; however, Lazar and Kolter (1996) have determined that SurA participates

in *the cis-trans* isomerization of proline residues and is required for the proper *in vivo* folding of the OmpA, OmpF, and LamB outer membrane proteins.

Rat protein disulfide isomerase (rPDI) has been shown to increase the yield of bovine pancreatic trypsin inhibitor (BPTI) severalfold when coexpressed with DsbB in the *E. coli* periplasm (Ostermeier et al., 1996). The yield of a monoclonal antibody Fab' fragment has also been increased using coexpression of a human PDI in the *E. coli* cytoplasm.

A native *E. coli* prolyl isomerase, trigger factor, has received considerable attention with respect to functional studies (Kandror et al., 1997), but has yet to be tested with respect to enhancing protein over expression. Furthermore, no study has yet been attempted where a prolyl isomerase was expressed as a fusion protein with a target protein. Trigger factor would make for a very interesting fusion protein study considering its catalytic activities and its favorable solubility (86% chance of solubility) as assessed by the Wilkinson-Harrison solubility model.

As mentioned earlier, the solubility of several cytokines has been increased by fusing them to thioredoxin (LaVallie et al., 1993). In a study by Yasukawa et al. (1995), the solubilities of eight vertebrate proteins was increased *by co-expressing E. coli* thioredoxin. This suggests that the mechanism by which thioredoxin confers solubility on the target protein is not dependent on fusion of the target protein to thioredoxin. Yasakuwa et al. speculate that the increases in solubility result from the ability of thioredoxin to affect the redox state of the cell.

Secretion into the Medium

According to Hannig and Makrides (1998), *E. coli* normally secretes few proteins, and the secretion mechanisms are thought to be quite complex and poorly understood. Proteins that are destined for secretion in *E. coli* must traverse two different membranes, which is an extremely cumbersome and specific molecular process. Also, proteins which have a tendency to form insoluble aggregates intracellularly and proteins containing hydrophobic transmembrane regions have been found to be difficult to secrete in *E. coli* (Murby et al., 1996). Thus, secretion has not been proven to be a viable alternative to intracellular expression and consequently has not been widely implemented.

There is a paradox in that the bacterial secretion mechanisms commonly used for protein export have evolved to export proteases into the surrounding environment in order to breakdown extracellular proteins for metabolic use within the cell. Thus, it has been the common experience with bacteria such as *Staphylococcus aureus* and *Bacillus lichenformis* that excreted recombinant proteins undergo considerable protease degradation in the external medium (Piers et al., 1993; van Leen et al., 1991). While some cytoplasmic-protease deficient strains of *E. coli* have been developed (Strauch and Beckwith, 1988; Baneyx and Georgiou, 1991), a corresponding *Bacillus or Staphylococcus* strain has yet to be made which can prevent significant extracellular protease degradation. Considering that *E. coli* excretes few proteases, it stands to reason that if an efficient and reliable mechanism for secretion is developed, it would have tremendous benefits for recombinant protein production.

Amino Acid Substitutions

Amazingly, it has been found that single amino acid substitutions can be the determining factor for soluble versus insoluble protein expression. In a study of the expression of interleukin-1β by Wetzel and Chrunyk (1994), it was found that a Lys → Val mutation (K97V) was sufficient to induce much more inclusion body formation than the wild type protein. Such folding sensitivity is striking and adequately demonstrates the frustrating unpredictability of the protein folding problem. Another case studied by Betton and Hofnung (1996) showed that a G32D and I33P double mutation in maltose binding protein induced inclusion body formation both in the periplasm and in the cytoplasm. This case makes more sense in light of the fact that the both mutations involve turn-forming residues. Adding a proline to a protein could certainly induce a new fold and removing a glycine could interrupt a natural fold, hence possibly exposing an otherwise buried hydrophobic region of the protein. A case where mutations have resulted in higher soluble expression was illustrated by Lehrman et al. (1991). In this study, a region of bovine growth hormone (aa 109-133) was replaced by the same region of human growth hormone. In this region of amino acids the human sequence differs from the bovine sequence in only eight amino acids, and the human sequence was thus designated as 8H. The mutant bovine growth hormone, 8H-bGH, was found to aggregate significantly less than wild type bGH. While amino acid substitutions which favor higher expression of proteins are valuable and also provide information on which to base theoretical work, their general unpredictability precludes their practical use toward increasing protein production.

Refolding

Despite extraordinary new developments in fusion protein expression systems, attempts are still frequently made to refold proteins from inclusion bodies (Shi et al., 1997; Kopitar et al., 1996; Barnfield, 1996). In some cases it has been found that the exact structure of a protein can be revived using refolding procedures (Schmid, 1996). Chen and Gouaux (1997) have developed a fractional factorial screen which explores various refolding conditions using 12 parameters and 16 experiments. While this factorial screen may be useful as a last resort method at large scale, economic production of a valuable protein, the extensive laboratory work required to create the variety of refolding conditions indicate that, to date, no uniform approach has been found to be useful in refolding proteins.

Strain Development and Growth Conditions

Expression levels of recombinant protein in a strain of ppGpp-deficient *E. coli* has shown to increase overall recombinant protein production (Dedhia et al., 1997). When *E. coli* is subjected to carbon and amino acid starvation, the unusual nucleotides ppGpp and pppGpp have been shown to accumulate. The synthesis of rRNA is known to be inhibited by ppGpp; thus, it stands to reason that eliminating ppGpp should increase protein production indirectly by increasing rRNA levels.

Reductions in growth temperature have been shown to increase soluble expression levels in a number of studies (Tsai et al., 1995; LaVallie et al., 1993; Pryor and Leiting,

1997). The temperature dependence of inclusion body formation indirectly implies that protein aggregation is strongly dependent on transcription/translation kinetics and has been explored extensively by King et al. (1996). As mentioned earlier, while reductions in growth temperature are not a major obstacle to protein production at the lab scale, temperature changes can significantly effect large scale process economics.

Lower limits on temperature reduction have been shown to exist for certain promoters. In a study by Vasina and Baneyx (1997), the temperature dependence of two promoters, *cspA* and *tac*, were investigated; and it was shown that while the tac promoter became completely inhibited at 10°C, the *cspA* promoter was able to continue protein synthesis for 2 hours.

In the expression of a recombinant metalloenzyme, phosphomannose isomerase (PMI), it was found that the addition of a high level of zinc in the fermentation medium resulted in a fourfold increase in soluble protein expression (Proudfoot et al., 1996). This result is interesting in light of this study since a metalloenzyme, bacterioferritin (BFR), has been chosen as a fusion partner with hIL-3. Other small molecule additives such as sugars and alcohols have been shown to increase production of biologically active protein and are reviewed by Georgiou and Valax (1996).

CHAPTER II

MATERIALS

Plasmids and Bacterial/Viral Strains

Plasmid pKK223-3 (Amersham Pharmacia Biotech, Piscataway, NJ) was used for all fusion gene constructions. A map of this plasmid showing its general features, such as the strong *tac* promoter, is shown in Figure 4. Plasmids carrying the various carrier and target genes were kindly provided as follows: pGS281 containing the BFR gene, Dr. Simon Andrews of the University of Sheffield (UK); pWKG20 containing the GrpE gene, Dr. Debbie Ang, Université de Genéve (Swizterland); pHA-200A containing the NusA gene, Dr. Michelle Hanna at the University of Oklahoma; pLB4 containing the hIL-3 gene, Dr. Lambert Dorssers, Dr. Daniel de Hood Cancer Center (Netherlands). Plasmids carrying the thioredoxin (pTrx-2) and bGH (p8300-10A) genes were purchased from ATCC (Rockville, MD). λ phage DD962, a charon 37 based vector which harbors the YjgD gene, was provided by Dr. Frederick Blattner at the University of Wisconsin Laboratory of Genetics. *E. coli* JM105 was used as the host for both vector construction and protein expression. *E. coli* LE392 was used to propagate and isolate genes carried on λ phage vectors.

Media

For LB media, yeast extract and tryptone were purchased from Difco Laboratories (Detroit, MI). For NZC media, casamino acids were also purchased from Difco. All

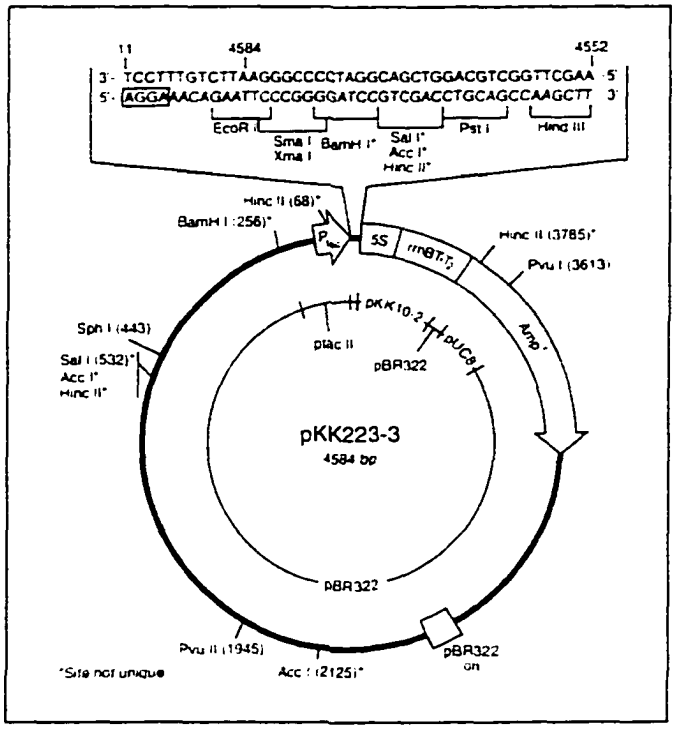


Figure 4. The pKK223-3 Expression Vector. General features of the pKK223-3 expression vector including the *tac* promoter, ampicillin resistance, and the multiple clong site (from Amersham Pharmacia Biotech catalog, 1998).

other media additives were purchased from Sigma Chemical Co. (St. Louis, MO).

Enzymes

Expand™ PCR reagents and T4 DNA ligase were purchased from Boehringer Mannheim Corp. (Indianapolis, IN). Restriction enzymes *EcoR*I and *Hind*III and factor Xa protease were purchased from New England BioLabs (Beverly, MA). *Age*I restriction enzyme was purchased from Promega Corporation (Madison, WI). Calf intestinal alkaline phosphatase (CIAP) was purchased from CLONTECH (Palo Alto, CA).

DNA Synthesis, Purification, and Analysis

Primers for the PCR amplification of gene fragments were synthesized by the Molecular Biology Resource Facility at the University of Oklahoma Health Sciences Center (Oklahoma City, OK). Gene fragments were agarose gel purified prior to ligation according to the Geneclean protocol (BIO101, Vista, CA). Double stranded plasmid DNA was purified using reagents in the Pharmacia FlexiPrep Kit. SeaKem LE agarose was purchased from FMC Bioproducts (Rockland, ME). The sequencing of all gene fusions was performed by the Recombinant DNA/Proteins Resource Facility at Oklahoma State University (Stillwater, OK) using DNA primers that we specified.

SDS-PAGE and Immunoblotting

The PROTEAN™ II large format and small format electrophoresis cells (BioRad, Hercules, CA) were used for protein analysis by SDS-PAGE. Chemiluminescent western

blot reagents and secondary antibodies were also purchased from BioRad. Murine monoclonal hIL-3 and hIFN-γ antibodies and were purchased from R&D Systems (Minneapolis, MN). Rabbit polyclonal bGH antibody was purchased from Biogenesis Ltd. (Sandown, NH). GBX developer, fixer, and BioMax MR x-ray film were purchased from Kodak (Rochester, NY) and were used to document the chemiluminescence produced during western blotting. Nitrocellulose membranes, NitroBind, were purchased from Micron Separations Inc. (Westborough, MA).

Densitometry

Coomassie blue stained SDS-PAGE gels were scanned into a UMAX Vista S-12 scanner (UMAX Technologies, Inc., Fremont, CA) using Photoshop v3.0 software (Adobe Systems, Inc., Mountain View, CA). The density of protein bands was quantified using Imagequant densitometry software (Molecular Dynamics, Inc., Sunnyvale, CA).

Protein Purification

Chelating and Q Fast Flow Sepharose® media were purchased from Amersham Pharmacia Biotech (Piscataway, NJ). Econo-column chromatography columns (1 cm and 2.5 cm diameter) and flow adapters were purchased from BioRad (Hercules, CA). In order to detect proteins at 280 nm, an ISCO UA-5 chart recorder and absorbance/fluorescence detector were used with a type 10 optical unit (Isco, Lincoln, NE). A Gilson MiniPuls3 peristaltic pump was used during protein purification (Gilson Medical Electronics Inc., Middletown, WI).

CHAPTER III

METHODS

Computer Programming

The Wilkinson-Harrison model was incorporated into two different C programs: Swisslight and Entrezlight. The C code for each program is illustrated in appendix A along with the instructions for running the program on UNIX servers. Swisslight will read SwissProt formatted files which are available at http://expasy.hcuge.ch, and Entrezlight will read Entrez formatted files available at the National Center for Biotechnology Information (NCBI) at http://www.ncbi.nlm.nih.gov. Both the SwissProt and NCBI web servers offer options for downloading a single text file containing all *E. coli* sequences (i.e. a batch file). Recently the Entez server has become redundant with respect to protein sequences and returns about 24,000 sequences when queried for *E. coli* proteins. However, the SwissProt server has taken measures to eliminate redundancy usually giving only about 4,500 sequences when queried for *E. coli* proteins. Protein structures were visualized using RASMOL, a free program available on the web at http://www.umass.edu/microbio/rasmol.

Primer Design for PCR

PCR primers were designed to have a region of complementary bases and a region containing non-complementary restriction sites for site directed ligation. The following is an example of a set of primers used to amplify the NusA gene from its cloning plasmid (a

complete list of all primer sequences is shown in Appendix C). The sequence for the 5' PCR primer is

Docking *Eco*RI Complementary to *nusA* 5' end
5' - CG / TTA / GCC / **GAA / TTC** / ATG / AAC / AAA / GAA / ATT / TTG / GC - 3'

It should be noted that the complementary region of the 5' PCR primer is complementary to the *anti-sense* strand of the double stranded template. The sequence for the 3' PCR primer is

Docking *Age*I Complementary to *nusA* 3' end
5' - CG / CGC / ATT / **ACC / GGT** / CGC / TTC / GTC / ACC / GAA / CCA / GC - 3'

The complementary region of the 3' PCR primer is complementary to the *sense* strand of the double stranded template. The template *nusA* DNA sequence was 1485 bp in length and had the following sequence at the 5' and 3' ends:

start
5' - ATG / AAC / AAA / GAA / ATT / TTG / GC .....GC / TGG / TTC / GGT / GAC / GAA / GCG - 3'

The eight base pair 5'-docking sequence in the primers is a non-complementary nonsense sequence that serves as a docking site for the restriction enzymes used to create site specific ends for directed ligation. The criteria for the length of the complementary sequence of the primers was a minimum of 20 bases or more depending on the melting temperature. The melting temperature was calculated by assigning 2°C to A and T residues, and 4°C to G and C residues (Sambrook et al., 1989). Only complementary bases were used for this calculation. It has been noted in the past that non-complementary bases can reduce the melting temperature appreciably (Sambrook et al.,

1989); however, no attempt to calculate their effect was made here. If the melting temperature was below 50°C, additional bases were added, and the complementary sequence was lengthened or shortened so that the last two bases on the 3' end were either G or C (i.e. to create a 3' GC clamp).

Upon receiving the primers from the Molecular Biology Resource Facility at the University of Oklahoma Health Sciences Center, they were diluted with dH_2O to 15 μM by the following equation:

$$\mu l \text{ of } dH_2O \text{ to add to dried primer to make 15 } \mu M \text{ stock} = 6929.1\left(\frac{OD}{N}\right)$$

where:

OD = optical density at 260 nm of the primer prior to spin drying (found in oligo synthesis report).

N = number of bases in primer.

This equation assumes a relation of 33 $\mu g/OD_{260}$ of single stranded DNA and a molecular weight average of 317.5 per base.

Construction of Gene Fusions

A general outline of the entire directional cloning scheme is shown for the NusA/hIL-3 fusion gene in Figure 5 and was used for the construction of all fusion genes, with the exception of the 2X-YjgD/hIL-3 fusion gene (described at the end of this section). Each gene was amplified from a parental cloning plasmid in 100 μl reactions using the Expand™ High Fidelity PCR System (Boehringer Mannheim). This system uses a proprietary mixture of *Taq* polymerase (for high yield) and *Pwo* polymerase (for 3'-5' exonuclease proofreading ability) which was developed by Barnes (1994). The Expand

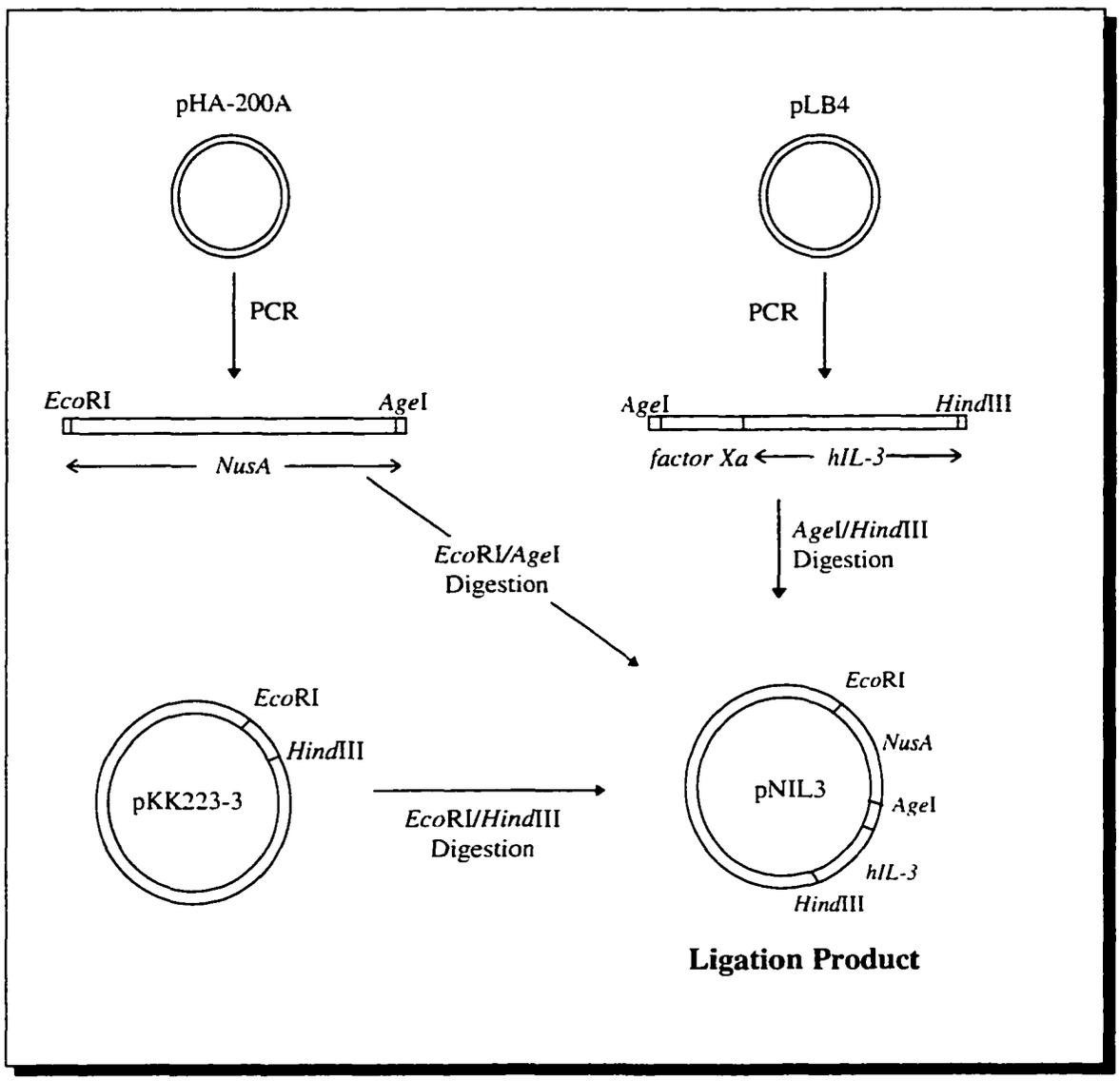


Figure 5. Construction of NusA/hIL-3 expression vector. First, the genes are amplified by PCR from their parental plasmids using primers that insert *Eco*RI, *Age*I and *Hind*III sites. Next, the genes are digested with the appropriate restriction enzymes and ligated together with plasmid pKK223-3. This scheme will be common to all fusion protein combinations using the same restriction sites, with the exception of the 2X-YjgD/hIL-3 fusion gene.

PCR system was chosen to obtain a high yield of PCR product while also minimizing the chances of random mutations. The optimal PCR reaction compositions and thermal cycling parameters are shown in Table 2. Since all genes were of reasonably similar size (320 to 1500 bp) and since the GC content was very similar (49% ± 6%), the parameters in Table 2 could be used for the amplification of all genes with minor adjustments in the volume of template added. For low template concentrations, template volumes were increased in PCR reactions (10 µl was the maximum ever used). For template preparation, double stranded plasmid DNA and λ phage DNA were purified as outlined in Appendix B.

For each gene, four identical 100 µl PCR reactions were combined and ethanol/ammonium acetate precipitated (see Appendix B). The DNA pellets were resuspended in 100 µl of dH_20, and 5 µl of the concentrated sample was run on a 1% (w/v) agarose gel to estimate the PCR product concentration. The site specific ends on the vector and PCR products were created using the restriction digestion parameters shown in Table 3. The composition of the restriction enzyme buffers is shown in Table 4. The entire volume (20 µl) of the restriction enzyme digestions was run on an agarose gel. The appropriate fragments were cut from the gel and agarose gel-purified by a modified GeneClean™ protocol (see Appendix B).

Ligation reactions were set up with the composition shown in Table 5. The amounts of the PCR products were varied according to the yield obtained from PCR reactions such that the mass of each PCR product was roughly equal to that of the linearized pKK223-3 vector. Prior to adding the last component (the ligase) to the

A. Reaction composition.

	Volume (µl)	Stock Concentration	Final Concentration
dH_2O^+	75.25		
Buffer*	10	10 X	1X buffer
$MgCl_2$	6	25 mM	1.5 mM
dNTP's	2	10 mM	200 µM (each dNTP)
5' primer	2	15 µM	300 nM
3' primer	2	15 µM	300 nM
template	2	0.1 µg/µl	0.2 µg
enzyme	0.75	3.5 U/µl	2.6 units
Total	100		

* The PCR buffer used does not contain $MgCl_2$.

+ dH_20 is sterilized, deionized tap water.

B. Cycling parameters.

Segment	No. of Cycles	Time	Temperature (°C)
1	2	1 min	94
2	29	45 sec	94
		45 sec	50
		2 min	72
3	1	2 min	72
4	1	24 hr	4

Table 2. Reaction composition and cycling parameters for the polymerase chain reaction. DNA fragments generated from these reactions were used in site directed construction of fusion genes. The template volume added to the reaction varied from 2 to 10 µl depending on the template conentration.

A. Carrier gene PCR Product.

Component	Volume (µl)	Stock Concentration	Final Concentration
Carrier gene PCR Product	16	~30.0 ng/µl	~24.0 ng/µl
NEB buffer #1	2	10 X	1X
AgeI+	1	2 U/µl	0.1 U/µl
*Eco*RI	1	20 U/µl	1 U/µl

+ *Age*I was added first and incubated for 1 h at 25°C. Then *Eco*RI enzyme was added and incubated a second hour at 37°C.

B. Target gene PCR Product.

Component	Volume (µl)	Stock Concentration	Final Concentration
Target gene PCR Product	16	~30.0 ng/µl	~24.0 ng/µl
NEB buffer #2	2	10 X	1X
AgeI+	1	2 U/µl	0.1 U/µl
*Hind*III	1	20 U/µl	1 U/µl

+ *Age*I was added first and incubated for 1 h at 25°C. Then *Hind*III enzyme was added and incubated a second hour at 37°C.

C. Expression vector pKK223-3.

Component	Volume (µl)	Stock Concentration	Final Concentration
pKK223-3	16	~30.0 ng/µl	~24.0 ng/µl
NEB buffer #2	2	10 X	1 X
EcoRI+	1	2 U/µl	0.1 U/µl
*Hind*III	1	20 U/µl	1 U/µl

+ *EcoR*I and *Hind*III enzymes were added at the same time and incubated for 2 hours at 37°C.

Table 3. Reaction compositions for restriction enzyme digestions. To create site specific ends, PCR fragments were digested with the appropriate enzymes. Certain reaction buffers (as shown above) were found to be compatible with more than one restriction enzyme so that buffer changes were unnecessary.

A. NEB buffer #1 composition (pH 7.0).

Component	Concentration (mM)
Bis Tris Propane-HCl	10
$MgCl_2$	10
Dithiotheitol	1

B. NEB buffer #2 composition (pH 7.9).

Component	Concnetration (mM)
NaCl	50
Tris-HCl	10
$MgCl_2$	10
Dithiothreitol	1

Table 4. Chemical composition of New England Biolabs' restriction enzyme buffers (from 1996/1997 Catalog). Buffer #1 is normally recommended for *Age*I and buffer #2 is recommended for *Hind*III. It was found that *Eco*RI could also cut in both buffers 1 and 2.

Component	Volume (µl)	Stock Concentration	Final Concentration
pKK223-3*	4	~18 ng/µl	3.6 ng/µl
Carrier gene PCR Product*	6	~18 ng/µl	3.6 ng/µl
Target gene PCR Product*	6	~18 ng/µl	3.6 ng/µl
Ligase Buffer	2	10 X	1 X
T4 DNA Ligase	1	1 U/µl	0.05 U/µl
dH_2O	1		
Total	20		

* These linear DNA fragments have been cut with the appropriate enzymes and agarose gel-purified.

Table 5. Ligation reaction composition for creating gene fusions. The volume of carrier gene, target gene, and vector were varied according to stock DNA concentrations such that approximately equal masses of each DNA fragment were present in the reaction. It was found that equal masses of vector and PCR fragments provided a sufficient molar excess of PCR fragment (relative to vector) for high efficiency ligation.

reaction, the buffer and DNA fragment mixture was heated at 65°C for 2 min. It has been shown that heating prior to ligation can "scramble" the DNA mixture, causing PCR fragments which might otherwise dimerize to bind to the correct vector restriction site (Ausubel et al., 1987). The ligation reaction was incubated for 16 hr at 15°C. For desalting prior to electroporation, each ligation reaction was precipitated with *n*-butanol and isopropanol (see Appendix B). Forty microliters of fresh electrocompetent cells were added, and the mixture was then and electroporated (see Appendix B for competent cell preparation and electroporation).

Following electroporation, the cells were screened by picking colonies and examining them for DNA inserts by comparing supercoiled plasmid sizes using a modified method of Unlap and Hu (1995). In some cloning experiments, clones were, instead, screened by protein expression by comparing OD_{600} readings of induced and uninduced cultures (Davis and Harrison, 1995) (both methods are described Appendix B).

For the 2X-YjgD/hIL-3 fusion gene (see Figure 2 for the general structure), it was necessary to insert an additional YjgD gene fragment at the *Age*I site in between an already existing YjgD gene and the hIL-3 gene. Since only one restriction enzyme (*Age*I) was used, the ligation was not directional and required the use of calf intestinal alkaline phosphatase (CIAP) to dephosphorylate the *Age*I cut YjgD/hIL-3 vector. One unit of CIAP was added to the *Age*I linearized YjgD/hIL-3 vector (pYIL3) immediately following AgeI cutting. When compared, the *Age*I and CIAP buffers were found to be compatible, so CIAP was added directly to the restriction digestion. The CIAP reaction

was incubated at 37°C for 30 min. The CIAP reaction was then deactivated by heating at 75°C for 10 min. For further reassurance of CIAP deactivation, the reaction was then phenol:chloroform extracted and ethanol precipitated (see Appendix B). The ethanol precipitated, linearized pYIL3 was used directly in a ligation reaction with *AgeI* cut/GeneCleaned YjgD PCR product.

Sequencing Methods

The Recombinant DNA/Protein Resource Facility used the method of cycle sequencing with fluorescent dye-terminators according to the ABI PRISM™ Ready Reaction Kit (Perkin Elmer, Foster City, CA). All sequence data were confirmed by comparison to GenBank registered sequences using the GCG multiple alignment program PileUp (version 7.0) available on the University of Oklahoma Genetic Computer Group. Alignments were conducted by inputting the data from two overlapping primers along with the entire fusion gene sequence and then performing the multiple alignment. The Multiple Sequence Format (MSF) file resulting from this alignment was then processed by the BOXSHADE computer program (version 3.21), which has the ability to convert the MSF file to a Rich Text File (RTF) which can be viewed using Microsoft Word 6.0. The BOXSHADE program can be used on-line at:

http://ulrec3.unil.ch/software/BOX_form.html

The BOXSHADE program creates a consensus sequence and colors all correct sequence data in red for rapid visual inspection.

In general, all primers were designed to begin reading sequence data 150 base

pairs apart in hopes of obtaining a double confirmation of the sequence in the forward direction (5' to 3'). In most cases primers contained a "GC clamp" on the 3' end. That is, the last two base pairs of the primer contained either a G or C. It has been found that having a GC clamp on the primer will extend the reading length of a primer. After several sequencing runs, it was found that the reading length of the primers could be further extended by choosing primers which had GC clamps longer than two base pairs. All sequencing primers along with the consensus sequences of the fusion genes are shown in Appendix C.

Solubility Tests: Protein Expression and Fractionation

For each solubility test, three 4 ml cultures of *E. coli* JM105 harboring pKK223-3 with the fusion gene of interest were grown to mid-log phase (OD_{600} = 0.4) at 37°C in shake flasks in LB medium containing 100 μg/ml ampicillin and 1% glucose. In two of the cultures, the *tac* promoter was induced by adding isopropyl-1-thio-β-D-galactopyranoside (IPTG) to a final concentration of 1 mM. When IPTG is added, it prevents the *lac* repressor protein from binding to the *tac* promoter region, thereby allowing transcription to occur. One additional feature of the *E. coli* JM105 host is that it contains a mutation (termed *lacI*^{q}) which causes it to express higher than normal levels of the *lac* repressor. The *lacI*^{q} mutation minimizes expression of the recombinant proteins, ensuring that recombinant proteins do not inhibit cell growth prior to induction with IPTG. An additional 100 μg/ml of ampicillin was also added at induction. The remaining third culture was left uninduced. The cells were grown for an additional 3 h at

37°C and harvested by centrifugation for 10 min at 1000 x *g*. One induced cell pellet was resuspended in 10 ml of sonication buffer containing 50 mM NaCl and 1 mM EDTA at pH 8.0. The cells were lysed by sonication at 4°C for 30 sec at 9 watts per ml of lysate and then allowed to cool for 30 sec on ice. This cycle was repeated four times for a total sonication time of 2 min. The inclusion body material was then removed from the cell lysate by centrifugation at 12,000 x *g* for 30 min at 4°C. The supernatant was then lyophilized. An SDS-PAGE of the uninduced, induced, soluble fraction, and insoluble fraction was then run. SDS-PAGE gel loading buffer was added to the four fractions such that the soluble and insoluble fusion protein bands were equal, when added together, to the amount of fusion protein present in the whole cell induced lysate. This allowed for direct visual evaluation of the solubilizing potential of the particular carrier protein used in the experiment.

SDS-PAGE, Immunoblotting, and Protein Assays.

Cellular proteins were separated on 12% or 8% (w/v) SDS-PAGE and detected with Coomassie brilliant blue (see Appendix B). Total protein measurements were done according to the BCA protein assay (Pierce, Rockford, IL). Western blots were performed according to the Immunstar Chemiluminescent Kit from BioRad (Hercules, CA). Optimized western blot procedures are also shown in Appendix B. hIL-3 cell proliferation assays using a TF-1 cell line were kindly performed by Dr. Robert House at the Illinois Institute of Technology Research Institute (Chicago).

Densitometry

Quantitative protein determination was made using ImageQuant densitometry software to read band intensities of Coomassie stained gels which were scanned in at a grayscale pixel density of 300 dpi on a UMAX Vista S-12 scanner. Pixel values were normalized by selecting a rectangular area corresponding to the shape of the lane, and averaging the pixel values over the width of the rectangle. The averaged pixel values were then plotted versus the length of the lane, and the resulting peaks were integrated to determine the percent of total cell protein. In conjunction with total protein measurements from the BCA total protein assay, the amount of fusion protein expressed could be quantified.

Protein Purification

All purification steps were conducted at 4°C. The entire purification scheme is outlined in Figure 6. Following expression of the His_6NusA/hIL-3 fusion protein, the cells from 300 ml of culture were harvested by centrifugation and resuspended in 300 ml of sonication/loading buffer (50 mM Tris-HCl, 300 mM NaCl, and 1 mM PMSF at pH 8.0). The cells were lysed by sonication in 20 ml aliquots at 4°C for 30 sec at 4.5 watts per ml of lysate and then allowed to cool for 30 sec on ice. This cycle was repeated four times for a total sonication time of 2 min. The inclusion body material was then removed from the cell lysate by centrifugation at 12,000 x *g* for 30 min. Following sonication and centrifugation, 5 mM imidazole was added to reduce non-specific protein binding during

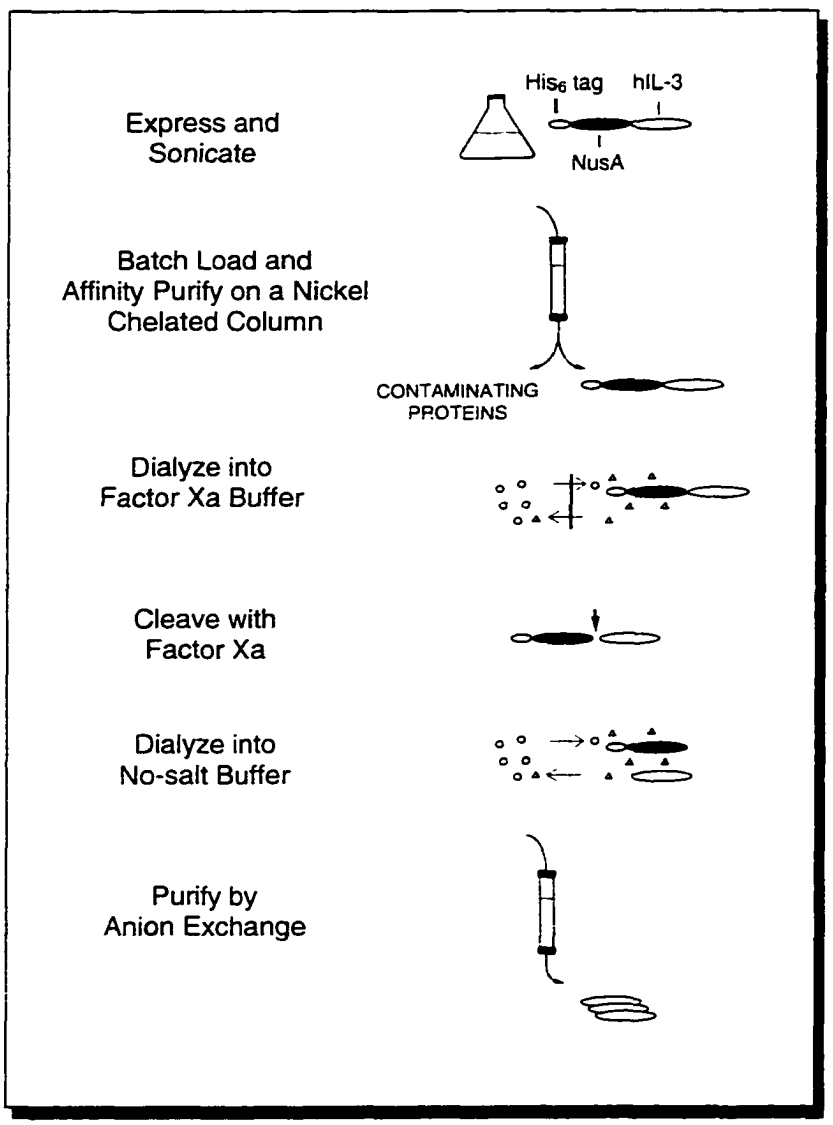


Figure 6. Scheme for the Purifcation of hIL-3 from the NusA Expression System.

loading. Fast Flow Chelating Sepharose (4.5 ml with bound Ni^{2+}) was added to the lysate and gently stirred for 1 hr at 4°C. The lysate/Sepharose mixture was then centrifuged for 10 min at 1000 x *g*. The pelleted Sepharose was then removed, and the lysate supernatant was extracted twice more with 4.5 ml of fresh chelated Sepharose. The Sepharose aliquots were combined and packed into a 2.5 cm diameter column for a final volume of 13.5 ml. The column was washed to baseline with loading buffer (50 mM Tris-HCl, 300 mM NaCl, pH 8.0) at 3.0 ml/min (37 cm/hr). The column was then washed with a buffer containing 50 mM Na_2HPO_4, 300 mM NaCl, and 20 mM imidazole at pH 6.0 to elute contaminating proteins. The fusion protein was eluted with a buffer containing 50 mM Tris-HCl, 100 mM NaCl, and 500 mM imidazole at pH 8.0. The eluted fusion (24.5 ml) protein was then dialyzed into 1.0 L of factor Xa protease buffer (20 mM Tris-HCl, 100 mM NaCl at pH 8.0) overnight. The following day, the spent buffer was discarded and another 1.0 L dialysis was performed for 2 hr. Just prior to cleavage, 2 mM $CaCl_2$ was added. Factor Xa was added at a level of 1 μg per 50 μg of fusion protein substrate, and the reaction was incubated for 3 hr at 23°C. The reaction was then dialyzed into 1.0 L of 20 mM Tris-HCl at pH 8.0 overnight. The following day, the spent buffer was discarded, and another 1.0 L dialysis was performed for 2 hr. The sample was then loaded onto a 5 ml Q Sepharose column (1 cm diameter) at 0.5 ml/min (38 cm/hr). The purified hIL-3 was found in the flowthrough during loading while all contaminating proteins remained bound to the column.

CHAPTER IV

RESULTS AND DISCUSSION

Screening the *E. coli* Genome for Soluble Proteins

The solubility model was applied to the entire set of known *E. coli* protein sequences found in the SwissProt protein database using the program SwissLight (see Appendix A for instructions on the use of SwissLight). The restrictions placed on the search were: 1) the proteins must have a greater than 90% chance of solubility, and 2) they must be larger than 100 amino acids. Proteins meeting this criteria are listed in Table 6. The reason for restricting the search to proteins containing more than 100 amino acids was to ensure that predicted solubility characteristics were not simply limited by the molecular weight of the carrier protein relative to the chosen target protein, hIL-3 (15.1 kDa). The list of possible carrier proteins was then reduced to those which were stable when expressed in *E. coli* as shown in the available literature.

A summary of all carrier and target proteins chosen for study is shown in Table 7. Reviewing the literature resulted in the selection of NusA, GrpE, and bacterioferritin, which have all been cloned and overexpressed previously in *E. coli* (Nakamura and Mizusawa, 1985; Zylicz et al., 1987; Andrews et al., 1989). Thioredoxin was chosen as a control carrier protein due to its widespread use as a soluble fusion partner. The cytokine human interleukin -3 (hIL-3) was chosen as a model insoluble target protein since it is expressed in *E. coli* in inclusion body form at 37°C as a fusion to thioredoxin (LaVallie et al., 1993) and when expressed alone (Donahue et al., 1988). Since the NusA/hIL-3 fusion

Sorted by AminoAcid Length ↓

SWISS-PROT Protein Name	Total No of Amino Acids	CV-CV'	Number of Amino Acids used for CV-CV' calculation							
			Ala	Asn	Arg	Gln	Gly	Ile	Pro	Ser
RPSD_ECOLI	613	-2.48	46	19	54	71	24	34	19	29
FTSY_ECOLI	497	-3.15	18	11	22	79	30	33	22	15
AMY2_ECOLI	495	-2.17	20	21	44	40	40	20	23	14
NUSA_ECOLI	495	-2.62	33	19	42	56	28	26	14	15
YRFI_ECOLI	294	-2.67	12	16	23	26	19	7	14	7
MAZG_ECOLI	263	-3.03	21	7	21	32	10	12	7	7
S3AD_ECOLI	263	-2.38	16	5	18	27	13	8	14	12
SSEB_ECOLI	261	-2.24	9	6	12	34	17	13	15	14
YCHA_ECOLI	252	-2.42	13	11	18	26	9	8	12	15
YAGJ_ECOLI	243	-2.26	14	5	19	26	12	15	9	10
YFBN_ECOLI	238	-2.65	15	10	19	23	9	13	4	3
NARJ_ECOLI	236	-2.44	13	4	19	19	11	8	9	11
NARW_ECOLI	231	-2.36	13	6	21	20	11	9	10	13
YECA_ECOLI	221	-2.54	8	6	13	28	12	11	14	11
CHEZ_ECOLI	214	-2.49	14	5	22	15	7	6	9	13

Table 6. *Escherichia coli* proteins of 100 amino acids or greater in length in the SWISS-PROT protein databank which have a calculated CV-CV' value of -2.10 or less, which are predicted by the two parameter solubility model of Wilkinson and Harrison to have a solubility probability of 90% or greater when expressed in the *E. coli* cytoplasm. The proteins marked were chosen for study.

SWISSPROT	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
GRPE_ECOLI	197	-2.34	11	8	13	26	8	13	9	7
SLYD_ECOLI	196	-2.98	4	6	19	17	29	6	5	5
YJAG_ECOLI	196	-3.33	11	5	12	25	9	6	5	10
YIEJ_ECOLI	195	-2.77	6	6	16	21	17	10	8	7
YGFB_ECOLI	194	-3.22	4	8	18	16	17	4	9	8
YJDC_ECOLI	191	-2.13	15	4	15	15	7	6	7	5
YCDY_ECOLI	184	-3.12	9	3	12	21	8	3	11	12
AADB_ECOLI	177	-2.65	13	2	13	20	14	4	9	5
FLAV_ECOLI	175	-4.09	4	4	20	17	15	9	5	5
FLAW_ECOLI	173	-3.32	2	5	16	16	16	8	6	7
YCED_ECOLI	173	-2.57	7	4	12	19	5	8	12	10
YFHE_ECOLI	171	-2.62	13	2	14	16	3	8	3	9
ASR_ECOLI	169	-2.62	14	8	0	5	5	18	10	9
YGGD_ECOLI	169	-2.76	4	6	17	9	5	8	5	9
YHBS_ECOLI	167	-2.39	11	3	15	13	16	2	6	6
FTN_ECOLI	165	-2.21	4	8	7	19	5	9	4	12
MENG_ECOLI	161	-2.44	8	7	15	14	21	2	3	7
YBEL_ECOLI	160	-2.13	14	4	9	22	7	7	5	8
BFR_ECOLI	158	-2.68	9	10	16	18	6	8	5	4
SMG_ECOLI	157	-5.53	7	6	12	23	5	3	4	3
HYCI_ECOLI	156	-2.55	5	7	14	13	16	4	10	2
SECB_ECOLI	155	-2.18	4	7	8	13	9	3	7	8
YBEY_ECOLI	155	-4.29	3	4	9	23	8	5	8	9

Table 6. (conitnued)

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ELAA_ECOLI	153	-2.39	6	4	12	12	10	5	6	7
YFJX_ECOLI	152	-2.37	6	5	11	11	11	1	7	8
MIOC_ECOLI	146	-2.23	2	4	10	16	15	7	7	10
YJGD_ECOLI	138	-9.37	3	3	22	25	8	8	4	3
HYFJ_ECOLI	137	-2.51	7	1	7	13	5	4	4	8
RL16_ECOLI	136	-2.45	14	2	3	7	13	16	7	2
RS6_ECOLI	135	-3.17	12	4	10	20	5	6	5	4
YHHG_ECOL	133	-2.17	10	3	14	7	7	3	1	8
GCSH_ECOL	129	-4.06	1	3	11	17	8	6	5	12
TRD5_ECOLI	129	-4.88	6	3	16	15	8	2	10	4
MSYB_ECOLI	124	-8.22	4	4	14	24	8	1	4	2
RS12_ECOLI	123	-2.23	15	5	3	4	11	13	7	6
RL7_ECOLI	120	-2.38	1	1	6	16	8	13	2	6
YACL_ECOLI	120	-3.85	6	5	7	18	9	4	0	5
YBFG_ECOLI	120	-3.11	5	3	8	12	7	3	3	6
RL20_ECOLI	117	-5.17	16	3	4	2	6	14	0	4
HYPA_ECOLI	116	-2.27	8	0	5	12	7	4	1	4
PTCA_ECOLI	116	-2.64	3	3	7	12	6	8	1	5
YZPK_ECOLI	115	-2.76	23	2	2	5	8	3	5	7
HYBF_ECOLI	113	-2.90	5	0	6	11	5	3	2	8
FER_ECOLI	110	-3.41	6	3	9	14	5	3	6	7
YR7J_ECOLI	110	-2.33	8	2	1	1	4	9	5	7
GLPE_ECOLI	108	-2.35	3	4	10	5	8	3	3	4
YGGL_ECOLI	108	-2.23	8	3	8	15	6	8	2	5

Table 6. (continued)

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
CYAY_ECOLI	106	-3.74	5	4	12	10	9	3	1	5
YEHK_ECOLI	105	-2.50	11	4	8	14	2	3	6	4
YR7G_ECOLI	105	-4.14	3	3	13	12	7	4	7	7
YQFB_ECOLI	103	-2.81	5	2	9	9	5	6	2	3
YCCD_ECOLI	101	-2.52	8	3	5	10	4	1	3	2
RS14_ECOLI	100	-2.17	14	3	5	5	5	11	4	8

Table 6. (continued)

Protein	MW (kDa)	Amino Acid Length	Probability of Solubility or Insolubility
YjgD	15.6	138	>97% soluble
NusA	55.0	495	95% soluble
BFR	18.5	158	95% soluble
GrpE	21.7	197	92% soluble
thioredoxin	11.7	109	72% soluble
bGH	21.6	189	58% insoluble
hIFN-γ	17.1	146	60% insoluble
hIL-3	15.1	133	79% insoluble
Fusion Protein			
2X-YjgD/hIL-3	47.3	417	> 97% soluble
YjgD/hIL-3	31.3	277	96% soluble
NusA/bGH	77.4	690	85% soluble
NusA/hIL-3	70.6	634	84% soluble
NusA/hIFN-γ	72.7	647	82% soluble
GrpE/hIL-3	37.3	336	67% soluble
BFR/hIL-3	34.1	297	66% soluble
thio/hIL-3	26.8	248	58% soluble

Table 7. Predicted solubilities of carrier and target proteins. The carrier proteins have relatively high solubility probabilities while the target proteins do not. Note: The Ile-Glu-Gly-Arg (0.4 kDa) sequence for factor Xa cleavage and the amino acids Thr-Gly (0.2 kDa) created by an *AgeI* restriction site are included in the MW and solubility calculations.

gave the highest soluble expression levels of hIL-3, as shown later in this chapter, two additional insoluble target proteins, bovine growth hormone (bGH) and human interferon-γ (hIFN-γ), were chosen for study with NusA.

Since roughly 38% of the proteins in the *E. coli* genome are of unknown function (Blattner et al., 1997), several proteins having extremely high solubility probability were identified which had no previous documentation of overexpression. These uncharacterized proteins, which are often referred to as "hypothetical open reading frames", are classified in the SwissProt database under four-letter names beginning with the letter Y (e.g. YRFI, YCHA, YECA etc.). One particularly interesting protein sequence was YjgD. Amazingly, 34% of the amino acids in YjgD are either aspartic or glutamic acid, making for a highly charged, extremely acidic protein with a pI of 3.5. Two fusion proteins were constructed with YjgD and hIL-3: one having a single copy of YjgD (YjgD/hIL-3), and one having two copies of (2X-YjgD/hIL-3). The 2X-YjgD/hIL-3 protein was chosen for study based on the expression results obtained from the single copy, YjgD/hIL-3 fusion protein (described later in this chapter).

Of the individual carrier proteins, YjgD, NusA, and BFR were predicted to be the most soluble (95% chance of solubility or greater), while hIL-3 was predicted to be the least soluble (79% chance of insolubility). The solubility probability for chosen carrier and target proteins *as fusion proteins* is also shown in Table 7. When the solubility prediction is performed on the fusion proteins, 2X-YjgD, YjgD and NusA possess the largest chances of solubility, while thioredoxin has the least chance. GrpE and BFR both have intermediate solubilities of 67 and 66%, respectively. Both NusA fusion proteins

with bGH and hIFN-γ were predicted to have a greater than 80% chance of solubility.

Biochemical and Structural Characteristics of the Chosen Carrier Proteins

Since proteins, as a class of macromolecules, possess such a wide array of functions, it is entirely likely that the native function of the proteins chosen for this study could affect the health of the host cell and/or the fusion protein expression level. *E. coli* BFR functions as an iron storage protein which serves a dual purpose: 1) to protect cells from the cytotoxicity of excessive iron concentrations, and 2) to provide an iron source in iron deficient environments (Andrews et al., 1989). Thus, there may be possibilities for a unique purification procedure for BFR fusion proteins which implements an immobilized iron affinity column. In cloning and expression experiments, Andrews expressed BFR to 14% of the total cell protein in soluble form using plasmid pGS281 in *E. coli* JM101.

GrpE protein belongs to the Hsp70 family of heat shock proteins which are essential components in protein folding, assembly, and protection against heat damage (Zylicz et al., 1987). Thus, if protein intermediates resembling heat denatured proteins are formed during the expression of GrpE fusions in this study, possibilities exist for stabilization of incorrectly folded intermediates by heat shock proteins.

The NusA protein of *E. coli* is a transcription elongation factor which binds directly to core RNA polymerase and is also involved in many celluar and viral termination, attenuation, and antitermiation processes (Zhang and Hanna, 1995). Zhang and Hanna have purified NusA as a glutathione S-transferase fusion protein with an ultimate yield of 5 mg protein/g cells showing that it can be expressed to high levels

without considerable interference with host cell protein expression processes. NusA has also been expressed to levels of 6.5% of the total cell protein in soluble form by Nakamura and Mizusawa (1985).

Fortunately, the tertiary structures are known for three of the carrier proteins examined in this study: thioredoxin, BFR, and GrpE. It is interesting to examine these protein structures with respect to the parameters in the Wilkinson-Harrison solubility model. Figure 7 shows the ribbon structure of thioredoxin, BFR, and GrpE. It is apparent from comparing the turn residue fraction to the overall protein structure that a large number of turns results in a tighter, more compact protein (e.g. thioredoxin). The structure of GrpE is extraordinarily different from thioredoxin, having a long, uninterrupted α-helix at its N-terminus. These differences imply that α-helical structures are much easier to form than tightly folded structures. Figure 8 shows the charge distribution of Asp, Glu, Lys, and Arg residues for these same three proteins. It can be seen that whereas the charged residues seem to be evenly distributed over the GrpE protein structure, BFR and thioredoxin both have large regions where charged residues are clustered together on one side of the protein. This charge polarization could lead to protein aggregation due to binding of non-charged protein surfaces.

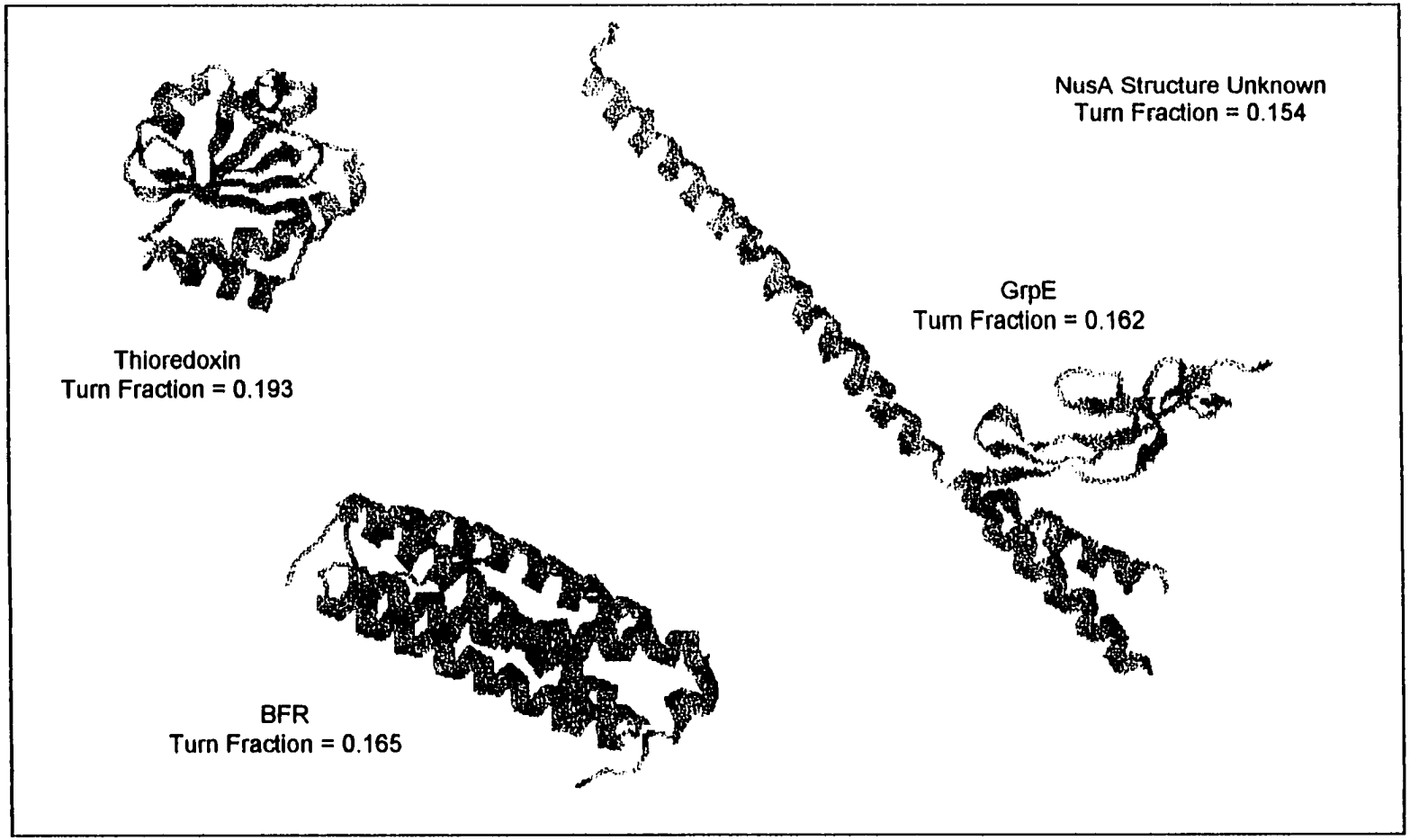


Figure 7. Comparison of the tertiary structure of thioredoxin, GrpE, and BFR. The C terminus of each protein is highlighted in red. It can be seen that as the fraction of turn residues decreases, protein structures become less compact. Structures were drawn using RASMOL from the following Brookhaven Protein Databank registered structures: 1BCF.PDB for BFR, 1DKG.PDB for GrpE, and 2TRX.PDB for thioredoxin.

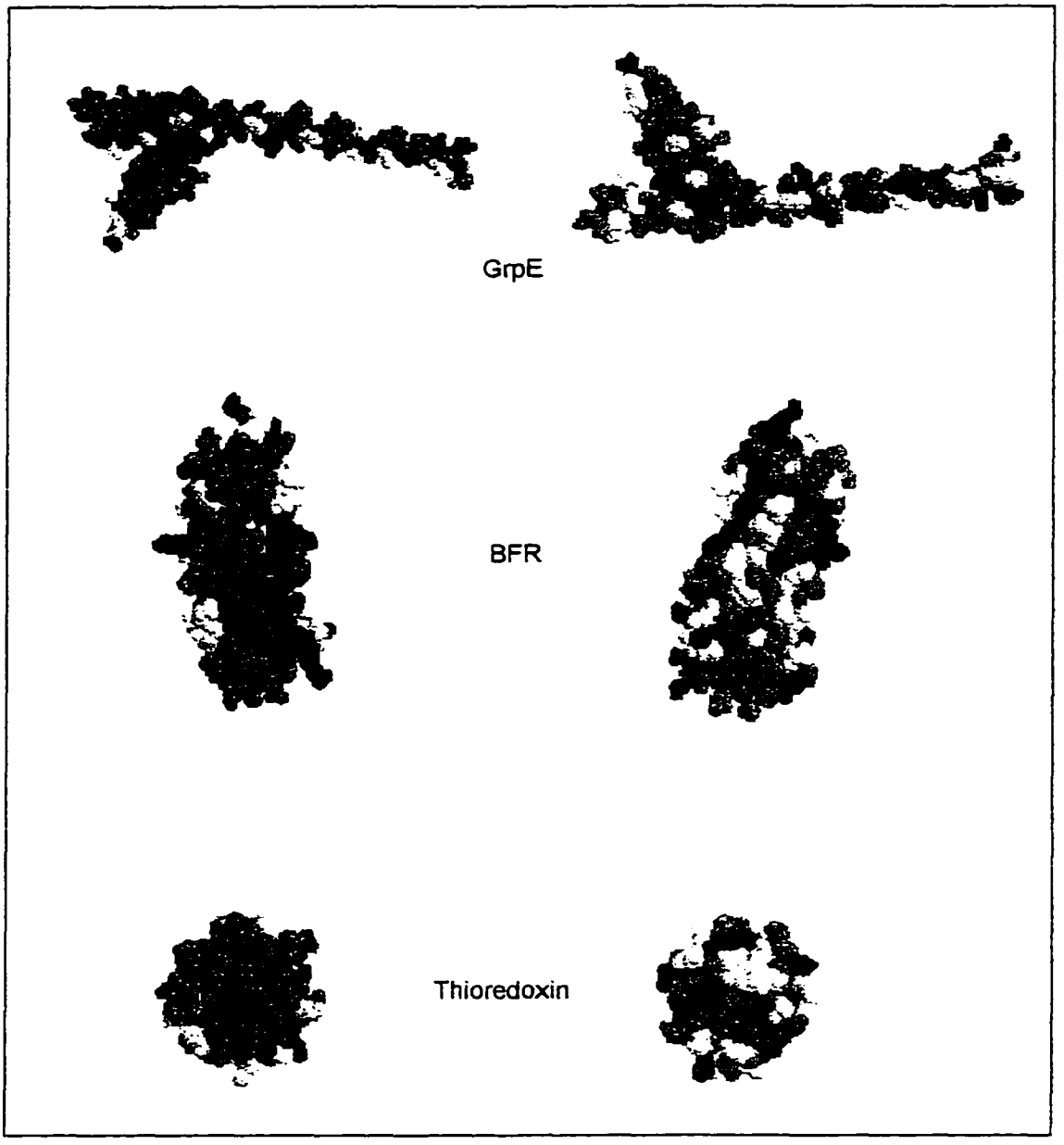


Figure 8. Locations of charged amino acids on carrier proteins GrpE, BFR, and thioredoxin. Residues Asp and Glu are yellow, residues Lys and Arg are red, the N-terminal residues are green, and the C-terminal residues are blue. All other residues are colored gray. The view on the left shows regions of low charged residue density and the view on the left shows regions high charged residue density. It is apparent that BFR and thioredoxin have a significant amount of charge polarization within their structures.

Construction of Fusion Genes

For a description of the general structure of the fusion genes described in this section, see Figure 5 (pg. 35) in the Methods section.

The NusA/hIL-3 fusion gene.

Figure 9A shows the PCR products obtained from reactions containing the templates for the NusA and hIL-3 genes. Three ligation reactions were performed under various conditions: 1) at 15°C without a phosphatase treated vector, 2) at 15°C with a phosphatase treated vector, and 3) at 23°C with a phosphatase treated vector. Upon transforming the three reactions into *E. coli* JM105, one colony was obtained from each ligation reaction. Amazingly, one of the three colonies (colony c, from the 23°C ligation with phosphatase treatment) contained the NusA/hIL-3 fusion gene within the pKK223-3 multiple cloning site (Figures 9B and C). Figure 9C shows that cutting with *EcoR*I releases a fragment of the correct length corresponding to the NusA gene plus 319 base pairs of the hIL-3 gene. This result was expected since the hIL-3 gene contains an *EcoR*I site 80 base pairs from its 3' end. This clone, with its recombinant plasmid now designated as pNIL3, was also shown to express a protein of approximately the correct molecular weight (70.6 kDa) as shown in the protein expression results discussed later in this chapter. The consensus DNA sequence resulting from the sequencing of the NusA/hIL-3 fusion gene is shown in Appendix C.

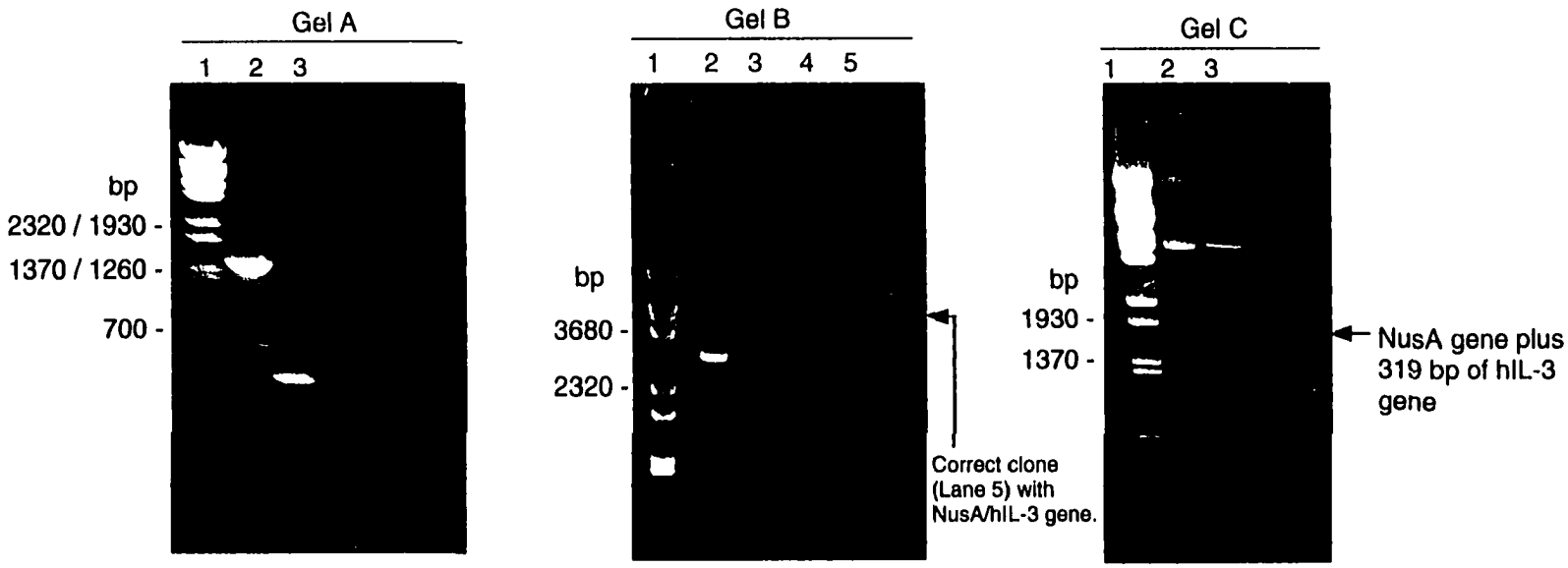


Figure 9. Cloning of the NusA/hIL-3 fusion gene. (**A**) PCR products. Lane 1, λ *Bst*EII marker. Lane 2, NusA PCR product (1485 bp); Lane 3, hIL-3 PCR product (399 bp); (**B**) Screening of colonies from ligation reaction. Lane 1, λ *Bst*EII marker; Lane 2, pKK223-3 uncut; Lane 3, colony a; Lane 4, colony b; Lane 5, colony c (a.k.a pNIL3); (**C**) Restriction digestion of a positive clone. Lane 1, λ *Bst*EII marker; Lane 2, pNIL3 uncut; Lane 3, pNIL3 cut with *Eco*RI.

The BFR/hIL-3 fusion gene.

Figure 10A shows the PCR products obtained from reactions containing the templates for the BFR and hIL-3 open reading frames. Following ligation of these two PCR fragments with linearized pKK223-3 and transformation into *E. coli* JM105, several colonies from the resulting plate were screened using a modified method of Unlap and Hu (1995). This method identifies recombinants on the basis of supercoiled plasmid sizes (see Appendix B for the detailed protocol). Three positive clones are shown in Figure 10B, and one clone (lane 13, Figure 10B) was chosen for restriction enzyme analysis (Figure 10C). It can be seen that pBIL3 cuts with all three restriction enzymes (*EcoR*I, *Age*I, and *Hind*III) and that cutting with *EcoR*I releases a fragment of the correct length corresponding to the BFR gene plus 319 base pairs of the hIL-3 gene. The clone harboring plasmid pBIL3 in Figure 10C was shown to overexpress a protein of roughly the correct molecular weight as described in the protein expression results discussed later in this chapter. The consensus DNA sequence resulting from the sequencing of the BFR/hIL-3 fusion gene is shown in Appendix C.

The GrpE/hIL-3 fusion gene.

Figure 11A shows the GrpE PCR product. Upon ligation of the GrpE PCR product, the hIL-3 PCR product, and linearized pKK223-3, several colonies were obtained. A new method of colony screening, which involves comparing the growth rate of induced and uninduced cultures, was tested during the GrpE/hIL-3 vector construction.

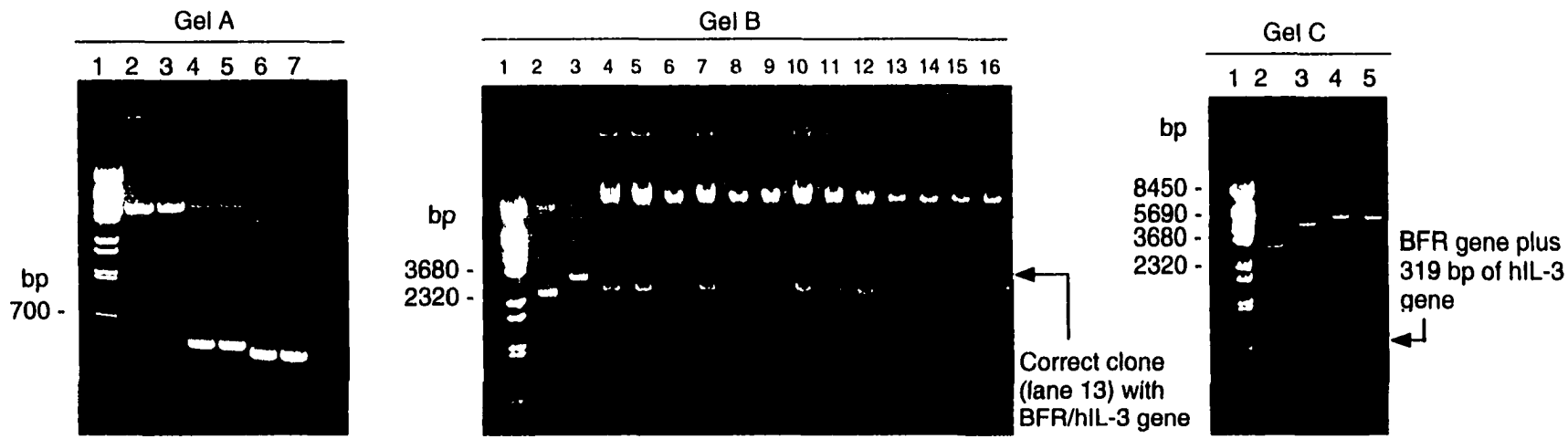


Figure 10. Cloning of the BFR/hIL-3 fusion gene. (**A**) PCR products. Lane 1, λ *Bst*EII marker; Lanes 2-3, *Eco*RI/*Hin*dIII linearized pKK223-3; Lanes 4-5, BFR PCR product (474 bp); Lane 6-7, hIL-3 PCR product (399 bp); (**B**) Screening of colonies from ligation reaction. Lane 1, λ *Bst*EII marker; Lane 2, pKK223-3 uncut; Lane 3, pBIL (harboring an unexpressible BFR/hIL-3 gene from previous cloning attempt); Lanes 4-8,10-12, 15-16, negative clones; Lanes 9, 13-14, positive clones; (**C**) Restriction digestion of a positive clone (from Gel B, lane 13, a.k.a. pBIL3). Lane 1, λ *Bst*EII marker; Lane 2, pBIL3 uncut; Lane 3, pBIL3 cut with *Eco*RI; Lane 4, pBIL3 cut with *Age*I; Lane 5, pBIL3 cut with *Hin*dIII.

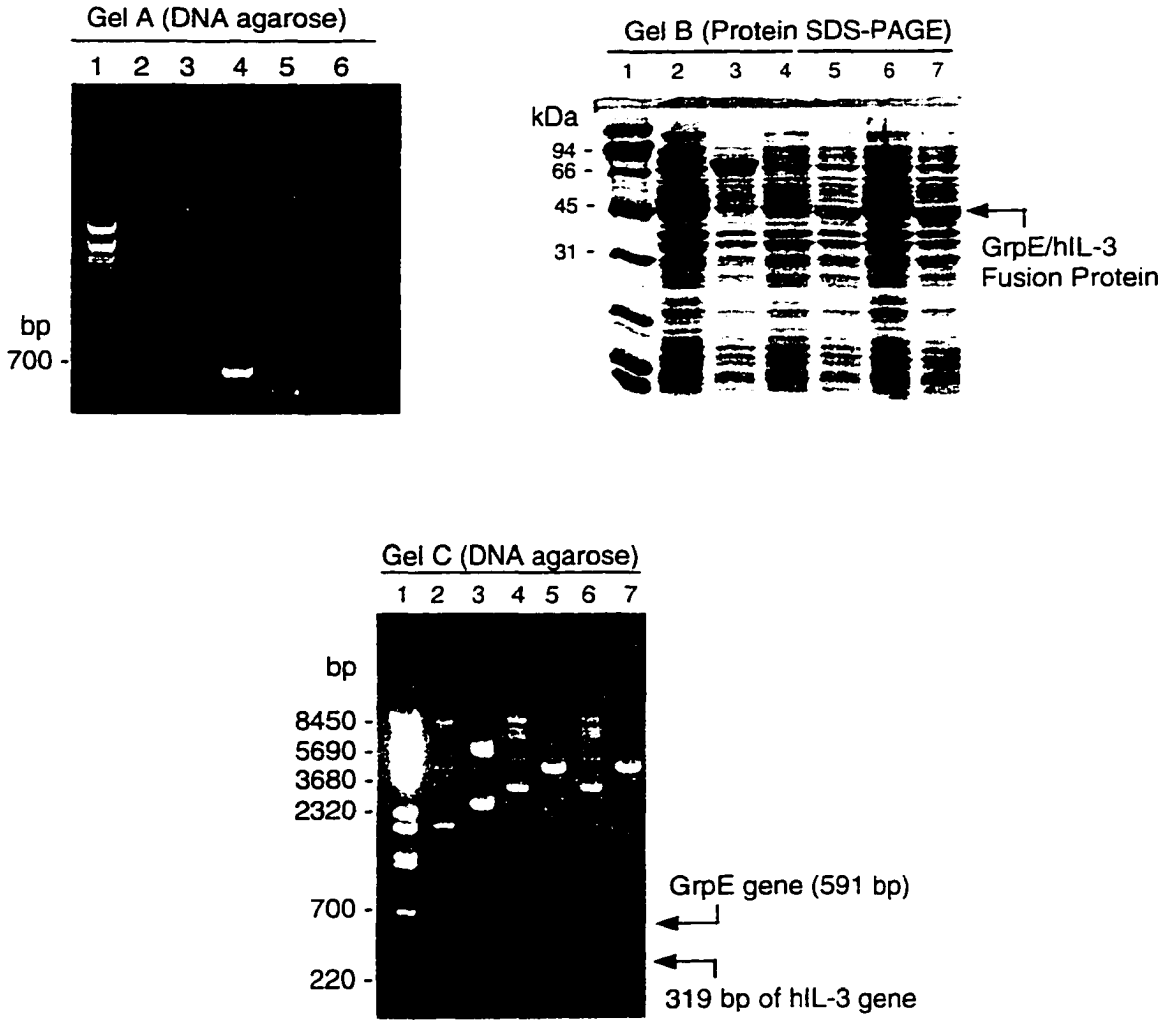


Figure 11. Cloning of the GrpE/hIL-3 fusion gene. (**A**) PCR products. Lane 1, λ *Bst*EII marker; Lanes 4, GrpE PCR product (591 bp); Lanes 6, hIL-3 PCR product (399 bp); (**B**) Screening of colonies from ligation reaction. Lane 1, protein marker; Lane 6, clone #8 uninduced; Lane 7, clone #8 induced; (**C**) Restriction digestion of clone #8 (from Gel B, lanes 6-7, a.k.a. pGIL3). Lane 1, λ *Bst*EII marker; Lane 3, pKK223-3 uncut; Lane 6, pGIL3 uncut; Lane 7, pGIL3 cut with *EcoR*I and *Age*I.

It was noticed during the growth and expression of the NusA/hIL-3 fusion protein that after three hours of induction, the induced culture appeared to have less cell density than the uninduced culture. The difference in cell density was noticed simply by holding test tubes up to the light and visually comparing the induced and uninduced cultures. This difference implied that induction was slowing down the growth rate of the cells. It was thought that measuring this difference in cell density by measuring the OD_{600} would provide a quick way to directly screen a large number of clones from the ligation reactions. The method, as developed after the construction of several fusion genes, is described in detail in Appendix B and has been published (Davis and Harrison, 1998). It is hereafter referred to as "screening by protein expression" (see Table 8 for a summary of OD_{600} data for the GrpE/hIL-3, thioredoxin/hIL-3, and YjgD/hIL-3 fusion proteins and a discussion of this data on page 72). After screening 30 colonies from the GrpE/hIL-3 ligation reaction by protein expression, several positive clones were identified, one of which is shown in Figure 11B (lanes 6 and 7). A restriction digest of the GrpE/hIL-3 fusion gene is shown in Figure 11C, and fragment sizes were consistent with the presence of both the GrpE and hIL-3 genes. The consensus DNA sequence resulting from the sequencing of the GrpE/hIL-3 fusion gene is shown in Appendix C.

The thioredoxin/hIL-3 fusion gene.

Figure 12A shows the thioredoxin PCR product. Upon ligation of the thioredoxin PCR product, the hIL-3 PCR product, and linearized pKK223-3, several colonies were obtained. After screening 30 colonies from the thioredoxin/hIL-3 ligation reaction by

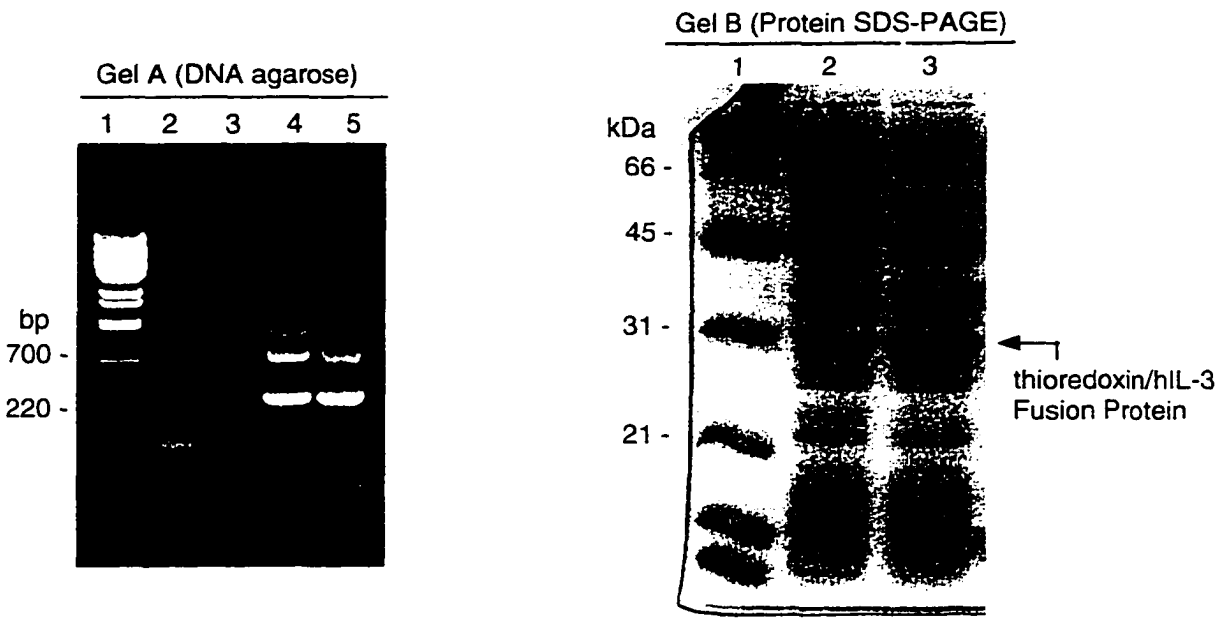


Figure 12. Cloning of the thioredoxin/hIL-3 fusion gene. (**A**) PCR products. Lane 1, λ *Bst*EII marker; Lanes 4-5, thioredoxin PCR product (329 bp); (**B**) Screening of colonies from ligation reaction by protein expression. Lane 1, protein marker; Lane 6, clone #2 (a.k.a. pTIL3) uninduced; Lane 7, pTIL3 induced.

protein expression, several positive clones were identified, one of which is shown in Figure 12B (lanes 2 and 3). No restriction analysis was performed on this clone; however, sequencing of the entire thioredoxin/hIL-3 fusion gene revealed the presence of all restriction sites (*EcoR*I, *Age*I, and *Hind*III) and both the thioredoxin and hIL-3 genes in correct positions (Appendix C).

The Yjgd/hIL-3 and 2X-YjgD/hIL-3 fusion genes.

Figure 13A shows the YjgD PCR product. Upon ligation of the YjgD PCR product, hIL-3 PCR product, and linearized pKK223-3, several colonies were obtained. After screening 30 colonies from the YjgD/hIL-3 ligation reaction by protein expression, several positive clones were identified, one of which is shown in Figure 13B (lanes 2 and 3). No restriction analysis was performed on this clone; however, sequencing of the entire YjgD/hIL-3 fusion gene in plasmid pYIL3 revealed the presence of all restriction sites (*EcoR*I, *Age*I, and *Hind*III) and both the YjgD and hIL-3 genes in correct positions. A one base pair deletion was discovered in the hIL-3 gene which is shown in Appendix C.

Figure 14A shows a YjgD PCR product (lane 5) which, instead of having *EcoR*I and *Age*I sites at the 5' and 3' ends, respectively, has an *Age*I site at both ends. This PCR fragment was ligated into the *Age*I site of the YjgD/hIL-3 expression plasmid (pYIL3, described in the preceding paragraph) to give a fusion gene with two copies of the YjgD gene upstream of the hIL-3 gene. After screening several colonies from the ligation reaction by protein expression, several positive clones were identified (see Figure 14B).

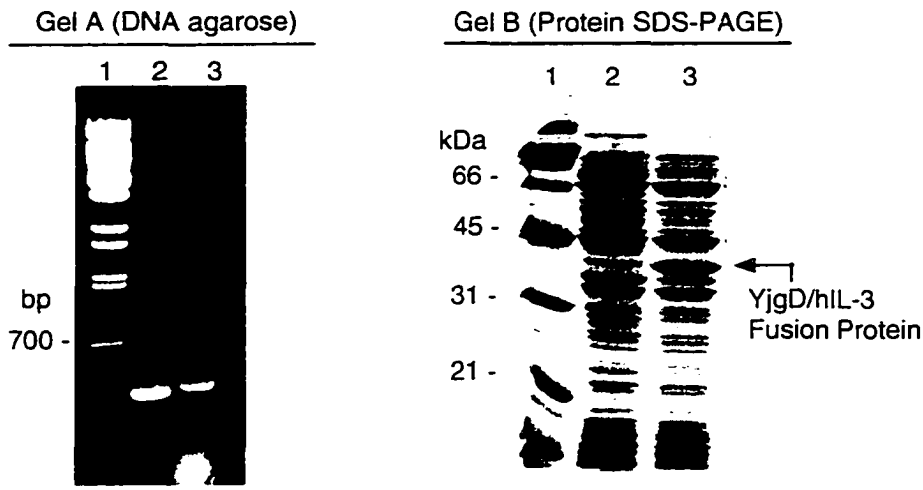


Figure 13. Cloning of the YjgD/hIL-3 fusion gene. (**A**) PCR product. Lane 1, λ *Bst*EII marker; Lane 3, YjgD PCR product (417 bp); (**B**) Screening of colonies from ligation reaction by protein expression. Lane 1, protein marker; Lane 2, clone 6 (a.k.a. pYIL3) uninduced; Lane 7, pYIL3 induced.

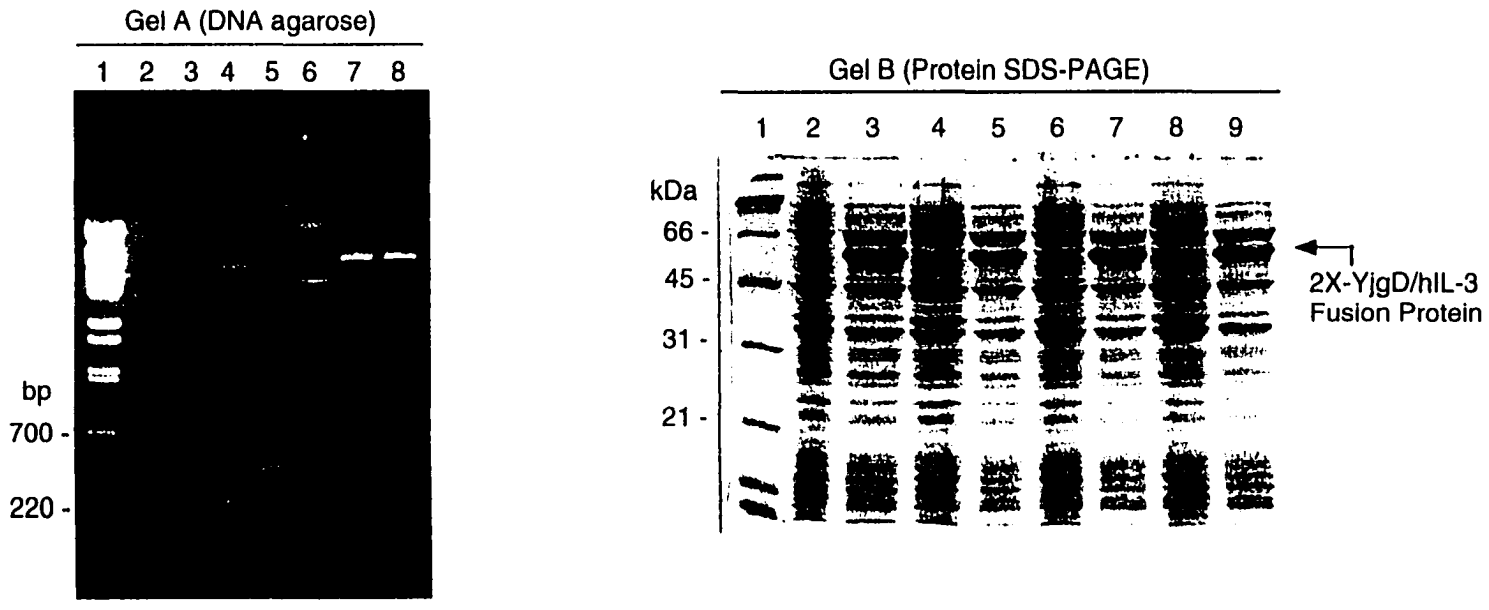


Figure 14. Cloning of the 2X-YjgD/hIL-3 fusion gene. (**A**) PCR products. Lane 1, λ *Bst*EII marker; Lane 5, YjgD PCR product with *Age*I sites on both ends (417 bp); (**B**) Screening of colonies from ligation reaction by protein expression. Lane 1, protein marker; Lanes 2-3, clone 1 induced and uninduced, respectively; Lanes 4-5,6-7,8-9, clones 3, 5, and 6 induced and uninduced, respectively.

The 2X-YjgD/hIL-3 fusion gene has not been sequenced. The reason for not sequencing the 2X-YjgD-hIL-3 clone is its rather low soluble expression level, which is characteristic of both pYIL3 and p2X-YIL3 as shown later in this chapter by western blotting. Since plasmid pYIL3 was used for insertion of the extra YjgD gene, it is likely that the hIL-3 gene in plasmid p2X-YIL3 also contains a deletion in the hIL-3 gene.

The NusA/bGH and NusA/hIFN-γ fusion genes.

Figure 15A shows the bGH and hIFN-γ PCR products. Plasmid pNIL3, which harbors the NusA/hIL-3 fusion gene, was used as the host vector in these cloning experiments. The hIL-3 gene was cut out of pNIL3 using *Age*I and *Hind*III, and the bGH gene was then ligated into the plasmid at the *Age*I and *Hind*III sites. In a separate experiment, the hIL-3 gene in pNIL3 was replaced with the hIFN-γ gene. Upon ligation and transformation, several colonies were obtained for both the bGH and hIFN-γ experiments. After screening several colonies from the NusA/bGH and NusA/ hIFN-γ ligation reactions by protein expression, several positive clones were identified (Figure 15B and C). No restriction analysis was performed on these clones; however, sequencing of the bGH portion of the NusA/bGH fusion gene revealed the presence of both restriction sites (*Age*I and *Hind*III) and the presence of the entire correct bGH gene (Appendix C). The clones used for later expression experiments were bGH clone #5 and hIFN-γ clone #22. It was not possible to efficiently identify hIFN-γ clone #22 by SDS-PAGE alone. After the sequencing of two incorrect hIFN-γ clones, a dot-immunoblot

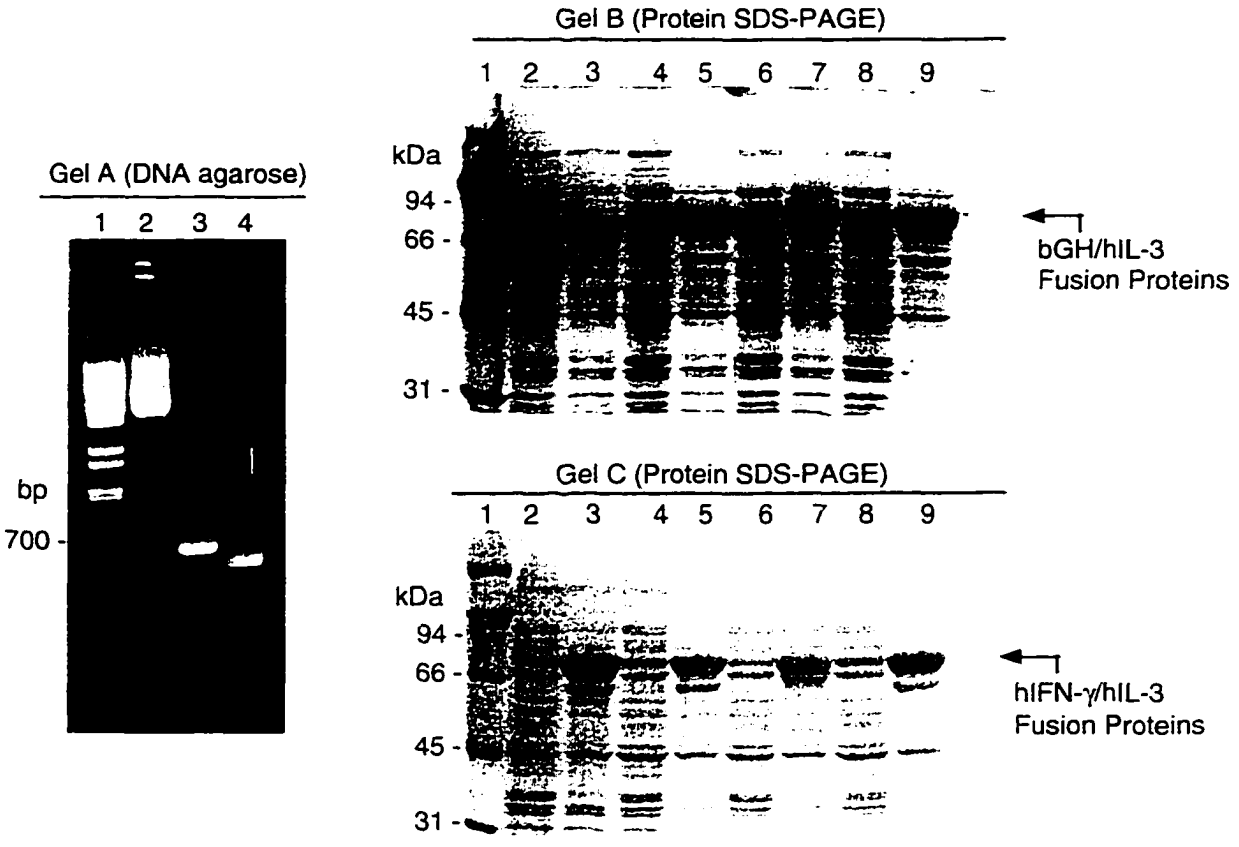


Figure 15. Cloning of the bGH/hIL-3 and hIFN-γ fusion genes. (**A**) PCR products. Lane 1, λ *Bst*EII marker; Lane 3 bGH PCR product (567 bp); Lane 4, hIFN-γ PCR product (438 bp); (**B**) Screening of colonies from ligation reaction (NusA/bGH). Lane 1, protein marker; Lanes 2-3, 4-5, 6-7, 8-9, represent clones 4, 5, 8, and 14 induced and uninduced, respectively; (**C**) Screening of colonies from ligation reaction (NusA/hIFN-γ). Lane 1, protein marker; Lanes 2-3, 4-5, 6-7, 8-9, represent clones 10, 16, 18, and 22 induced and uninduced, respectively.

method was developed which uses a hIFN-γ monoclonal antibody. This method identified a clone (#22) expressing a fusion protein of the right theoretical size for the NusA/hIFN-γ fusion protein which was also reactive with the antibody. This dot blot method is described in Appendix B. No restriction analysis was performed on clone #22; however, sequencing of the hIFN-γ portion of the NusA/hIFN-γ fusion gene revealed the presence of both restriction sites (*Age*I and *Hind*III) and the presence of the entire correct hIFN-γ gene (Appendix C).

The His_6-NusA/hIL-3 fusion gene

A PCR primer was designed which inserted a His_6 sequence between the start codon (ATG) and the first codon of the NusA gene (AAC) and is shown in Appendix C. The remaining primers were the same as used in the construction of the NusA/hIL-3 gene. Following PCR, the fragments were cut by restriction digestion (e.g. with the same restriction sites in the same positions as in the NusA/hIL-3 construction), ligated together and transformed into *E. coli* JM105. Upon screening the colonies by protein expression, several positive clones were identified. One clone was shown to overexpress a protein of the correct size (as shown later in this chapter in the expression results), and the fusion gene of this clone was sequenced (see Appendix C). The His_6 sequence was confirmed, and the sequence of the hIL-3 gene was shown to be completely correct. The NusA gene was not completely sequenced; however, the presence of the NusA gene was confirmed to be immediately downstream of the His_6 sequence and immediately upstream of the factor

Xa/hIL-3 sequence. This fusion gene was used to express the His_6-NusA/hIL-3 fusion protein which was used in developing a purification protocol for the NusA expression system.

Summarized Data from Screening by Protein Expression

A summary of the data obtained when screening by protein expression is shown in Table 8 for the thioredoxin, YjgD, and GrpE fusion proteins containing hIL-3. In typical screening experiment, seven colonies with the highest OD_{600} differences (e.g. OD_{600} of the uninduced culture minus the OD_{600} of the induced culture at 3 hours post-induction) were analyzed by SDS-PAGE: some colonies were found to overexpress a protein and some were not. The seven colonies with the lowest OD_{600} differences were also analyzed by SDS-PAGE and invariably showed no protein expression. The highest and lowest OD_{600} differences averaged 0.12 ± 0.04 and -0.02 ± 0.01, respectively. It is interesting that the highest OD_{600} differences for the YjgD/hIL-3 fusion protein are roughly twice that of the thioredoxin/hIL-3 and GrpE/hIL-3 fusion proteins. Perhaps the acidity of the YjgD protein has a negative effect on cell growth. It is also useful to know that the OD_{600} differences are detectable at lower levels of expression (e.g. 7% of the total cell protein for the GrpE/hIL-3 fusion protein).

Expression of Fusion Proteins

Figure 16 shows an SDS-PAGE and western blot of various hIL-3 fusion proteins. All fusion proteins were expressed at a high level with respect to the percent of the total

	Thioredoxin/hIL-3 OD_{600} Differences				YjgD/hIL-3 OD_{600} Differences				GrpE/hIL-3 OD_{600} Differences			
Clone #	**Top 7 Expressed**	**Top 7 No Protein**	**Bottom 7**	**Clone #**	**Top 7 Expressed**	**Top 7 No Protein**	**Bottom 7**	**Clone #**	**Top 7 Expressed**	**Top 7 No Protein**	**Bottom 7**	
1	0.12			6	0.14			8	0.1			
2	0.12			7	0.14			15	0.09			
5	0.13			13	0.2			16	0.09			
9	0.1			16	0.2			24	0.09			
12	0.11			29	0.25			10		0.06		
39	0.09			15		0.21		12		0.05		
30		0.08		21		0.19		30		0.05		
4			0	4			-0.02	2			-0.03	
10			0	10			-0.01	4			-0.03	
11			0	11			0	9			-0.04	
16			-0.03	25			-0.05	11			0	
27			-0.02	32			-0.03	17			-0.03	
35			0	33			0	18			-0.03	
37			0	37			0	29			-0.04	
Average	**0.11**	**0.08**	**-0.01**		**0.19**	**0.20**	**-0.02**		**0.09**	**0.05**	**-0.03**	

- "Top 7 Expressed" indicates the subset of the seven clones with the highest OD_{600} differences which were shown to overexpress the correct size protein.
- "Top 7 no protein" indicates the subset of the seven clones with the highest OD_{600} differences which expressed no protein.
- "Bottom 7" indicates the seven clones with the lowest OD_{600} differences. These clones invariably expressed no protein.

Table 8. Summarized data for the screening of colonies by protein expression as outlined in Appendix B. OD_{600} differences are defined as the OD_{600} of the uninduced culture minus the OD_{600} of the induced culture at 3 hours post induction.

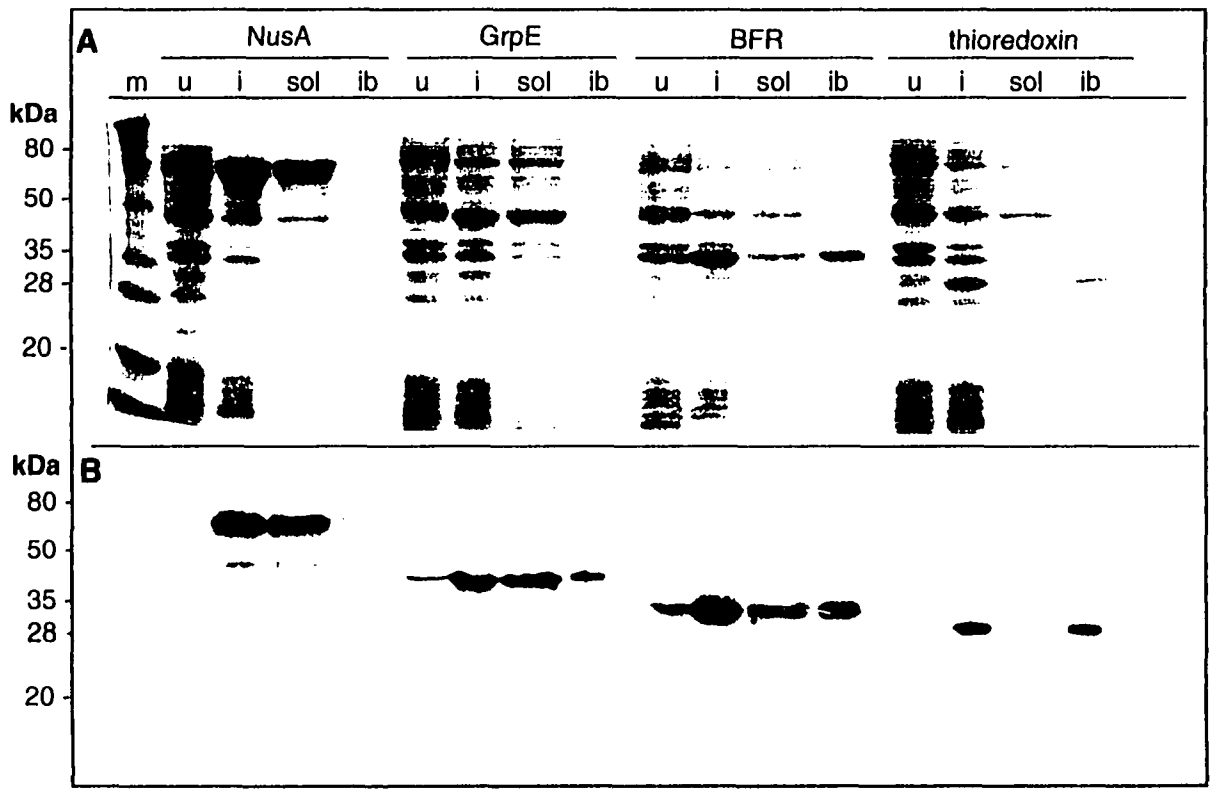


Figure 16. SDS-PAGE (A) and Western Blot (B) of NusA, GrpE, BFR, and thioredoxin fusion proteins containing hIL-3. Equal portions of cell lysate, soluble fraction, and insoluble fraction were loaded. **(m)** markers, **(u)** uninduced whole cell lysate, **(i)** induced whole cell lysate, **(sol)** soluble fraction, **(ib)** inclusion body fraction. 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 western 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.

cell protein, namely: NusA, 18.4%; GrpE, 7.4%; BFR, 13.7%; and thioredoxin, 8.5%. 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). The insolubility of the thioredoxin/hIL-3 fusion protein at 37°C is consistent with the results of LaVallie et al. (1993). The western blot (Figure 16B) shows, with more clarity, the distribution of hIL-3 among the soluble and insoluble fractions. BFR/hIL-3 and GrpE/hIL-3 are partially 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 (Andrews et al., 1989) and serves as a convenient marker for confirmation of protein expression.

One striking finding from the western blot data (Figure 16B) is that the level of expression of hIL-3 in the soluble fraction was significantly higher in the NusA fusion than in the other fusions. Thus, the large size of NusA (55.0 kDa) was not at all limiting in being able to express high levels of hIL-3. Upon investigating the known biological function of NusA, it was learned that it is a transcriptional elongation factor which binds directly to the core RNA polymerase. One interesting phenomenon associated with NusA is transcriptional pausing. When NusA is present during transcription, RNA polymerase will pause for longer periods of time at certain pausing sites along the template DNA (Zhang and Hanna, 1995). Since transcription and translation are coupled in prokaryotes, it is possible that NusA may be able to increase soluble levels of proteins by slowing

down translation at the transcriptional pauses and allowing critical folding events to occur. In one paper by Crombie et al. (1992), it was shown that the removal of a translational pause caused a decrease in the expression of indoleglycerol phosphate synthetase activity. While this paper is not directly related to NusA, it may apply to this idea in a general sense.

When comparing the predicted solubilities of fusion proteins in Table 7 and the amounts of each fusion protein expressed in the soluble and insoluble fractions in Figure 16, it is apparent that the solubility model correctly predicted the *trend* of the fusion protein solubilities. The model was correct in its prediction of protein solubility when the predicted probability was high (84% for the NusA/hIL-3 fusion) but was not correct when the predicted solubility was low (58% for the thioredoxin/hIL-3 fusion). For two cases of intermediate values of predicted solubility (67% and 66% for GrpE/hIL-3 and BFR/hIL-3 fusions, respectively), the model was at least partially correct since part of the fusion protein expressed for each case appeared in the soluble fraction.

Another conclusion that one might draw from the data and predictions in Table 7 and the experimental results in Figure 16 is that the size of the carrier protein is important. For all the fusion proteins shown in Figure 16, the solubility of the fusion protein increases as the size of the carrier increases. This leads to the hypothesis that solubility could be improved by linking two of the same carrier proteins together. For example, since the predicted solubility of GrpE alone is similar to that of NusA alone, the solubility of a fusion protein with two GrpE's linked together should be better than with only one GrpE. It seems unlikely, however, that this strategy would give a higher overall

level of expression of the target protein, since a larger fusion protein would have to be made for each target protein made. Thus, it is doubtful that the excellent results obtained with NusA could be improved upon.

Figure 17A shows an SDS-PAGE of the overexpression of the YjgD/hIL-3 fusion protein, and it can be seen by comparing the soluble and insoluble fractions that the fusion protein is expressed primarily in the insoluble fraction. This was a very surprising result considering that YjgD has the highest CV-CV' value (-9.37) of any of the carrier proteins chosen. This highly negative CV-CV' value is comprised of the largest charge average (-0.327) and the lowest turn fraction (0.130) of any of the carrier proteins chosen. YjgD represents a very interesting datapoint since it is the smallest carrier protein chosen, and also has the highest predicted solubility. Thus, the experimental insolubility of the YjgD/hIL-3 fusion protein suggests that the turn forming residue content and approximate charge average have little solubilizing influence when the molecular weight of the carrier protein (15.6 kDa for YjgD) is similar to that of the target protein (15.1 kDa for hIL-3).

As mentioned previously, a one base pair deletion was found at the 5' end of the hIL-3 gene in the YjgD/hIL-3 fusion gene. This deletion has the effect of changing the last codon of the hIL-3 gene (TTT) to TT. This results in a frame shift that adds nine additional amino acids to the C-terminal end of the hIL-3 protein. Considering that this mutation is at the C-terminal end of hIL-3 (and therefore translated last), and that the YjgD/hIL-3 fusion protein was found to have 57% native hIL-3 activity, it is likely that the mutation has little effect on the solubility of the fusion protein.

Since the YjgD/hIL-3 fusion protein was found to be insoluble, an attempt was

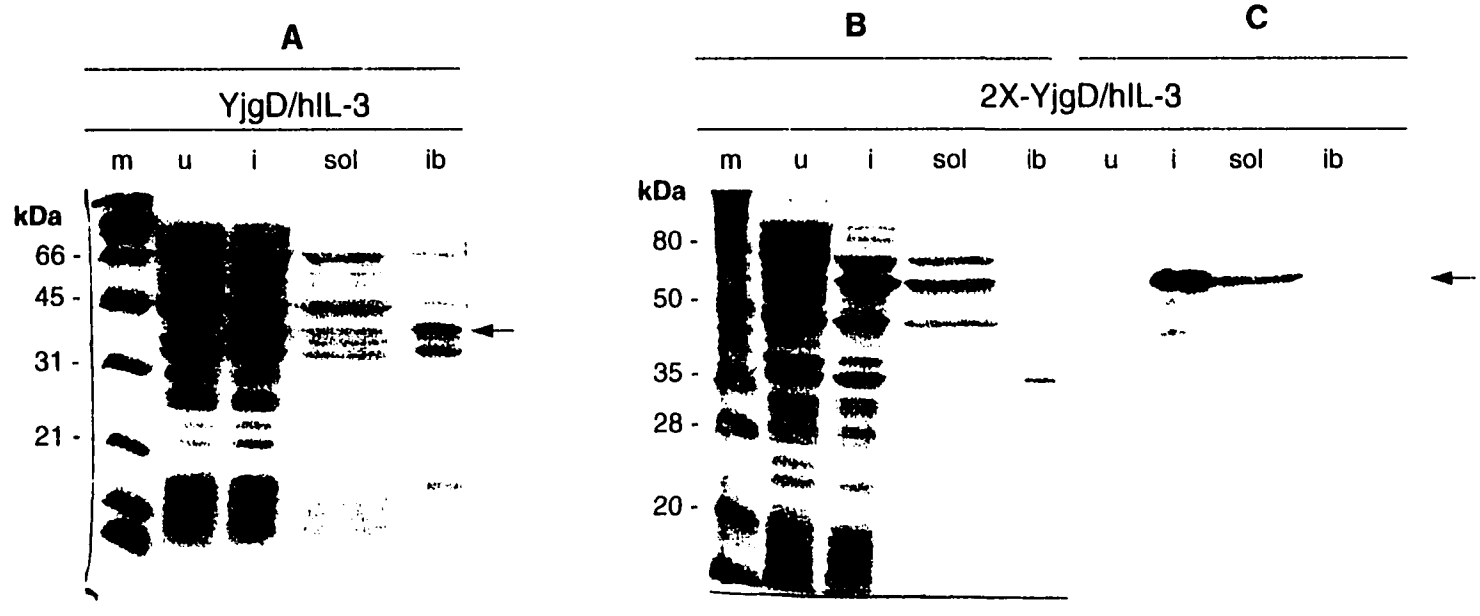


Figure 17. SDS-PAGE (A) of YjgD/hIL-3 fusion protein, and SDS-PAGE (B) and Western Blot (C) of 2X-YjgD/hIL-3 fusion protein. Equal portions of cell lysate, soluble fraction, and insoluble fraction were loaded. **(m)** markers, **(u)** uninduced whole cell lysate, **(i)** induced whole cell lysate, **(sol)** soluble fraction, **(ib)** inclusion body fraction. 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.

made to place another copy of the YjgD protein upstream of the hIL-3 protein to see if two copies of the YjgD gene could solubilize hIL-3. As shown in Figures 17B and C, the 2X-YjgD/hIL-3 fusion protein appeared to be entirely soluble. Thus, an important qualitative conclusion drawn from the results shown in Figures 16 and 17 is that if a researcher wished to have a high consistency in soluble expression with a fusion protein expression system, he/she should choose a system in which the carrier protein is somewhat larger than the target protein. It is apparent from the western blot in Figure 17C that the 2X-YjgD/hIL-3 fusion protein undergoes considerable protease degradation upon lysis of the cells. A possible remedy to this problem is the use of protease deficient strains of *E. coli* (Murby et al., 1996) or an optimized combination of protease inhibitors.

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 (Table 9). hIL-3 activity was found to be present in all fusion proteins, with the highest amount of native activity present in the NusA/hIL-3 fusion protein. It should be noted that reductions in native activity are most likely due to the presence of the carrier protein, which may interfere with the receptor binding properties of hIL-3.

The effect of temperature on the expression of the BFR/hIL-3 and YjgD/hIL-3 was assessed since these two fusion proteins produced the least amount of soluble hIL-3 (Figure 18). As indicated in the western blot, reducing the expression temperature to 23°C did not result in a significant increase in solubility for the BFR/hIL-3 fusion protein; however, the YjgD/hIL-3 fusion protein appears to become entirely soluble.

hIL-3 Fusion Protein	hIL-3 (µg/ml) Est. by SDS-PAGE and BCA Total Protein Assay	hIL-3 (µg/ml) Cell Proliferation Assay+	Percent Native Activity
NusA/hIL-3	7.5	5.0	67
GrpE/hIL-3	5.2	0.3	6
BFR/hIL-3*	8.7	1.0	12
YjgD	5.3	3.0	57
2X-YjgD	4.7	1.2	26
Thioredoxin/hIL-3*	7.5	3.6	48

* Expressed at 23°C to ensure adequate levels of hIL-3 for the cell proliferation assay. All other fusion proteins were grown at 37°C.

+ Erythroleukemic cell line TF-1 (Kitamura et al., 1989) was used to measure the cell proliferation activity of hIL-3.

Table 9. Activity of Recombinant Human Interleukin-3. The activity of hIL-3 was determined in the soluble crude cell lysates of various fusion protein constructs. Reductions in native activity are most likely due to the presence of the carrier protein.

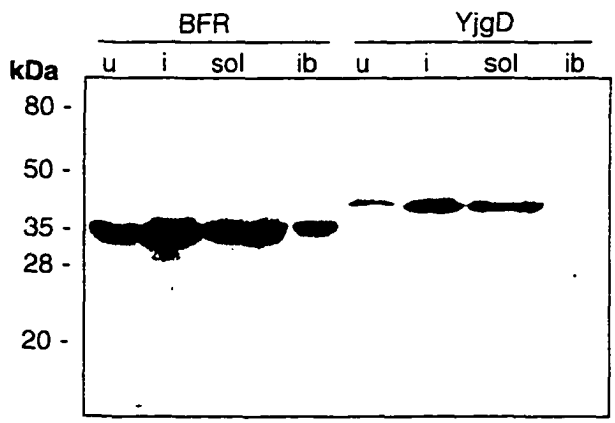


Figure 18. Western Blot of BFR and YjgD fusion proteins containing hIL-3 expressed at 23°C. Equal portions of cell lysate, soluble fraction, and insoluble fraction were loaded. **(u)** uninduced whole cell lysate, **(i)** induced whole cell lysate, **(sol)** soluble fraction, **(ib)** inclusion body fraction. Fusion proteins were expressed from plasmid pKK223-3 under control of the *tac* promoter in *E. coli* JM105 at 23°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.

Since adding Zn^{2+} to cultures expressing the metalloenzyme phosphomannose isomerase (PMI) was found to result in increased levels of soluble PMI expression (Proudfoot et al., 1996), and since BFR is known to bind iron and zinc (Le Brun et al., 1995), the effect of Fe^{2+} and Zn^{2+} on the soluble expression of the BFR/hIL-3 fusion protein was investigated (Figure 19). The addition of 10 mM $FeCl_2$ was found to severely inhibit cell growth to the point that no overexpressed protein could be detected by western blotting. The same result was also obtained for 10 mM $ZnCl_2$ and 1 mM $ZnCl_2$. The addition of 1 mM $FeCl_2$ was not lethal to the cells, but does not appear to significantly increase soluble levels of BFR/hIL-3 when compared to the results presented in Figure 16. It should be noted that the metals were not added to the culture until the time of induction (OD_{600} = 0.4), and were dispensed into the culture from pre-dissolved, concentrated solutions.

Since the overall expression level of the GrpE/hIL-3 fusion protein was the lowest of all the fusion proteins shown in Figure 16 (7.4% of the total cell protein), induction periods greater than three hours were attempted. The results of an 8 hour induction period are shown in Figure 20. It is evident that after 2 hours of induction, the level of GrpE/hIL-3 in the culture remains constant for 6 subsequent hours of incubation.

Since the NusA/hIL-3 fusion protein appears to give the highest level of soluble expression (Figure 16), it was decided to test its versatility by solubilizing two other proteins known to be insoluble: bovine growth hormone (bGH) and human interferon-γ (hIFN-γ). Previous data attest to the insolubility of bGH (Nikolaeva et al., 1993; Langley et al., 1987) and the folding properties of bGH have also received considerable attention

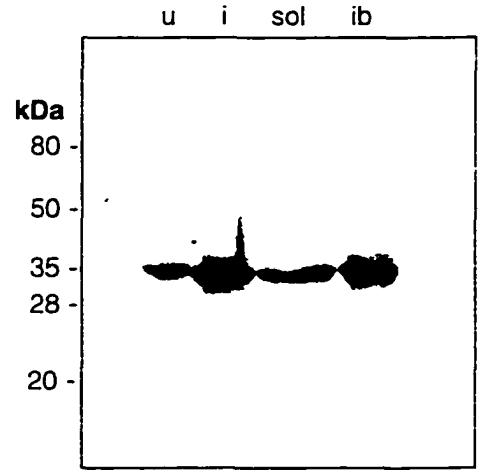


Figure 19. Western Blot of BFR/hIL-3 Fusion Protein with 1 mM Fe^{2+} in Growth Media. Equal portions of cell lysate, soluble fraction , and insoluble fraction were loaded. **(u),** uninduced. **(i),** induced. **(sol),** soluble fraction. **(ib),** insoluble fraction. The BFR/hIL-3 fusion protein was 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. The western blot was probed with mouse anti-hIL-3 monoclonal antibody and visualized using chemiluminescence.

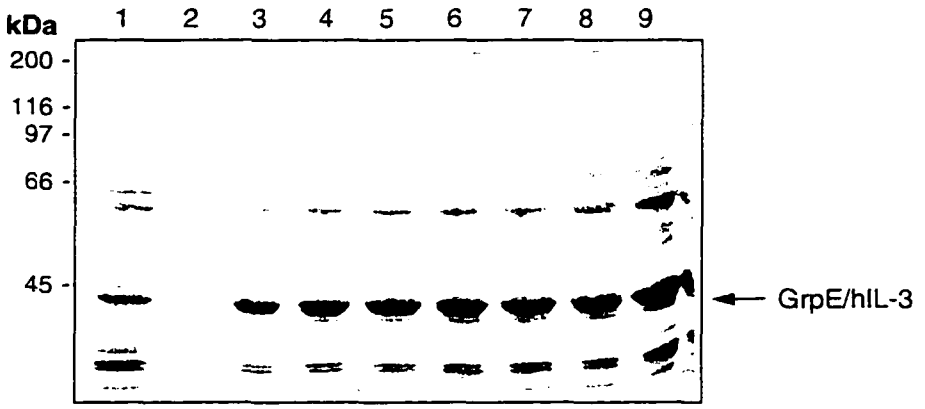


Figure 20. SDS-PAGE of 8-Hour Induction Period for the GrpE/hIL-3 Fusion Protein. Lane 1, uninduced whole cell lysate; Lanes 2-9, whole cell induced lysate at 1, 2, 3, 4, 5, 6, 7 and 8 hours post-induction, respectively. The GrpE/hIL-3 fusion protein was expressed from plasmid pKK223-3 under control of the *tac* promoter in *E. coli* JM105 at 37°C. Cells were induced with 1 mM IPTG. Note: The protein markers were not visible on this gel. The markers in Figure 21 were used to estimate protein size on this gel since the SDS-PAGE gels in Figures 20 and 21 were cast from the same batch of acrylamide and run at the same time, under the same conditions.

(Lehrman et al., 1991; Brems and Havel, 1989). Figure 21 shows an SDS-PAGE and western blot of the expression of bovine growth hormone as a fusion to NusA. It is clear that the vast majority of the fusion protein is expressed in the soluble fraction (estimated as 89%). To our knowledge, this is the first evidence of soluble expression of bGH in *E. coli*. Previous literature has also established the insolubility of hIFN-γ (Arora and Khanna, 1996; Haelewyn and De Ley, 1995). Figure 22 shows an SDS-PAGE and western blot of the expression of hIFN-γ as a fusion to NusA. It is clear that the fusion protein is expressed mostly in the soluble fraction (estimated as 87%). To our knowledge, this is also the first evidence of soluble expression of hIFN-γ in *E. coli*. These results presented in Figures 21, 22 and 16 suggest that NusA is an excellent carrier protein for soluble protein expression. To further develop the NusA expression system, a protocol was developed for the purification of hIL-3 from the NusA/hIL-3 fusion protein.

Purification of hIL-3 from the NusA System

A His_6 tag was added to the N-terminus of NusA for purification by immobilized metal affinity chromatography (IMAC) using an Ni^{2+}-IDA affinity column. The six histidine sequence was added using the same cloning procedures as outlined in the methods section. The PCR primer which guides the insertion of the six histidine sequence is shown in Appendix C.

The chromatogram of the IMAC column is shown in Figure 23, and an SDS-PAGE of the protein eluted with 500 mM imidazole is shown in Figure 24 (lane 7). A concentration of 20 mM imidazole was chosen to elute contaminating proteins since a

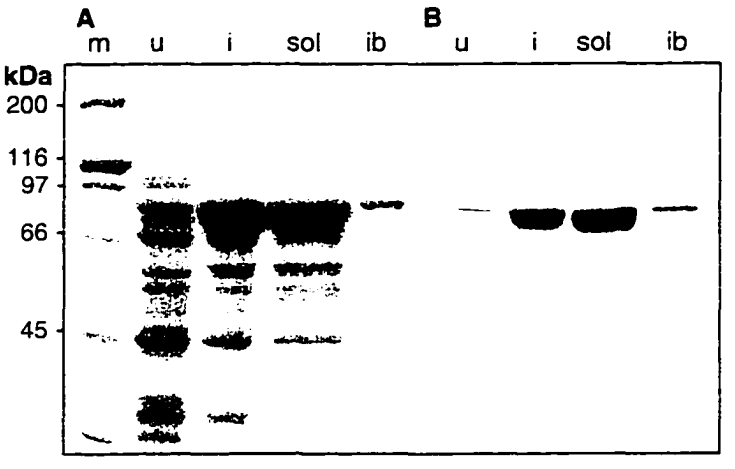


Figure 21. SDS-PAGE (A) and Western Blot (B) of Bovine Growth Hormone (bGH) Expressed as a Fusion to NusA. Equal portions of cell lysate, soluble fraction , and insoluble fraction were loaded. **(m)**, molecular weight markers. **(u)**, uninduced. **(i)**, induced. **(sol)**, soluble fraction. **(ib)**, insoluble fraction. The NusA/bGH 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. The western blot was probed with rabbit anti-bGH polyclonal antibody and visualized using chemiluminescence. The percentage solubility based on the western blot is 89%.

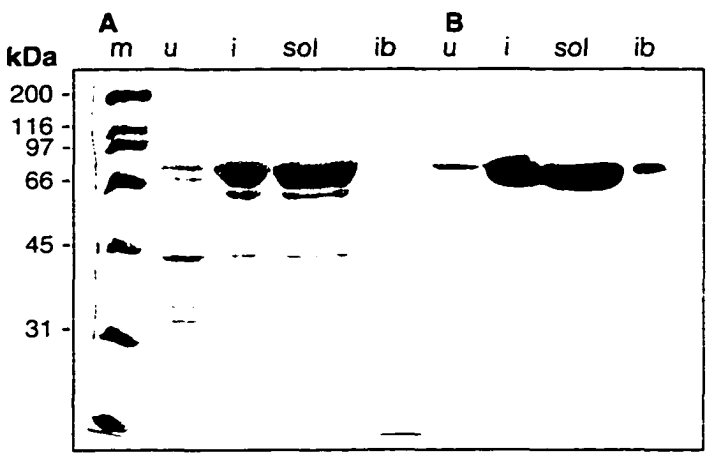


Figure 22. SDS-PAGE (A) and Western Blot (B) of Human Interferon-γ (hIFN-γ) Expressed as a Fusion to NusA. Equal portions of cell lysate, soluble fraction , and insoluble fraction were loaded. **(m)**, molecular weight markers. **(u)**, uninduced. **(i)**, induced. **(sol)**, soluble fraction. **(ib)**, insoluble fraction. The NusA/hIFN-γ 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. The western blot was probed with rabbit anti-hIFN-γ monoclonal antibody and visualized using chemiluminescence. The percentage solubility based on the western blot is 87%.

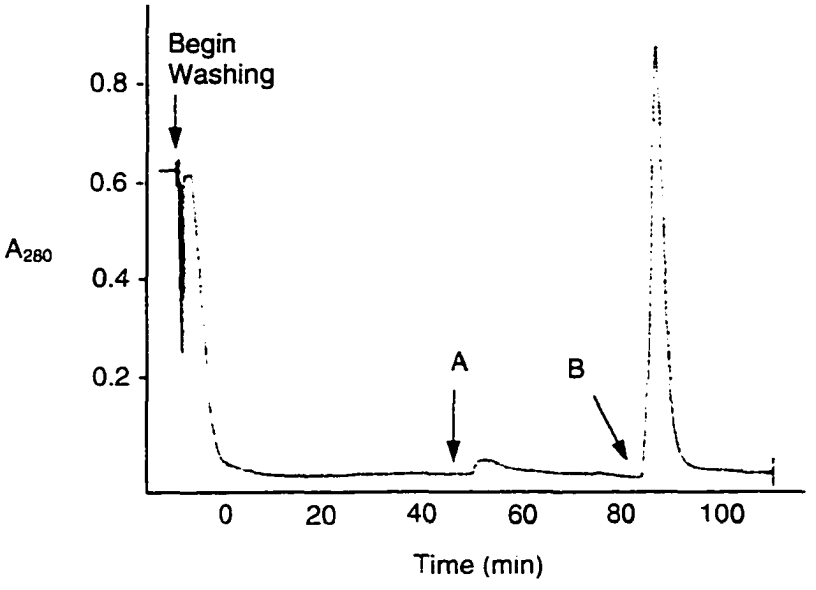


Figure 23. Chromatograph of metal affinity chromatography using a step imidazole gradient of *E. coli* culture expressing the His6-NusA/hIL-3 fusion protein. Wash buffer was 50 mM Tris-HCl and 300 mM NaCl at pH 8.0. (A) Elution buffer containing containing 50 mM Na_2HPO_4, 300 mM NaCl, and 20 mM imidazole at pH 6.0. (B) Elution buffer containing containing 50 mM Tris-HCl, 100 mM NaCl, and 500 mM imidazole at pH 8.0.

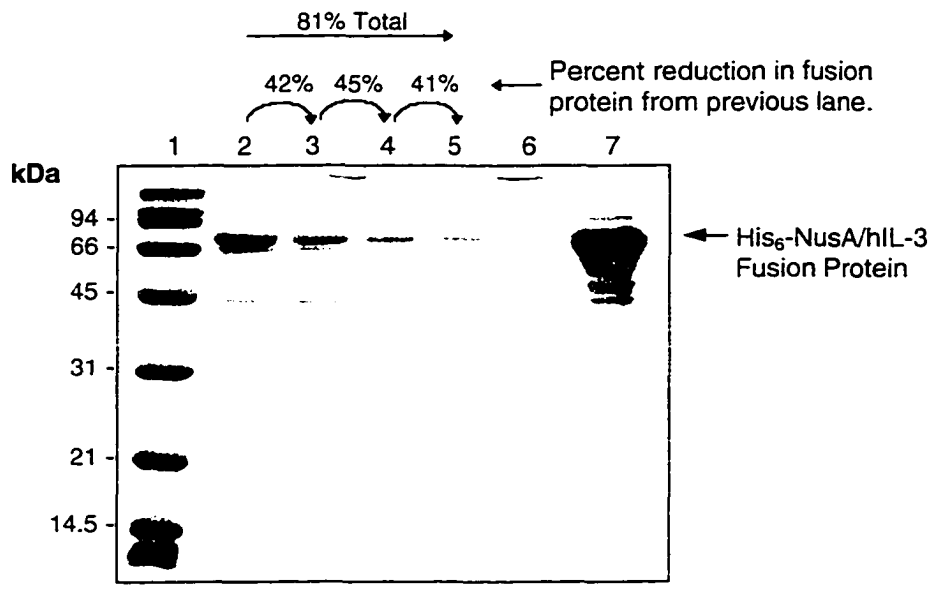


Figure 24. SDS-PAGE of Batch Incubations of His_6-NusA/hIL-3 Fusion Protein with Nickel Chelated Sepharose. Lane 1, markers; Lane 2, soluble lysate; Lane 3, supernatant after 1 hour incubation of 300 ml of soluble lysate with 4.5 ml of nickel chelated Sepharose media; Lane 4, supernatant after a second hour of incubation with 4.5 ml of fresh nickel chelated media; Lane 5, supernatant after a third hour of incubation with fresh 4.5 ml of nickel chelated media; Lane 6, 20 mM imidazole elution from column; Lane 7, 500 mM imidazole elution from column. Note: the volume of nickel chelated Sepharose media is measured as loaded (e.g. gravity sedimented) in a chromatography column. The three 4.5 ml aliquots of media were combined and loaded onto a chromatography column for elution of the fusion protein. The percent reductions in fusion protein were determined by densitometry.

100 mM imidazole elution buffer resulted in significant loss of the fusion protein. Also, it should be noted that when sodium phosphate buffers (Na_2HPO_4) were used to elute the fusion protein at 500 mM imidazole, precipitates were consistently found after dialysis into the factor Xa buffer. This precipitate was eventually identified as an interaction of the phosphate in the elution buffer and the calcium chloride in the factor Xa buffer. When stock solutions of Na_2HPO_4 and $CaCl_2$ were mixed together, large precipitates formed immediately. The insolubility of calcium phosphate has previously been documented in CRC Handbook of Chemistry and Physics. Thus, the 500 mM imidazole buffer was changed to a Tris-HCl buffer and precipitation ceased to be a problem.

An SDS-PAGE of the supernatant of the soluble lysate following batch incubations with 4.5 ml aliquots of nickel chelated Sepharose is shown in Figure 24. A reduction in the amount of fusion protein in the supernatant was found after each incubation. This reduction was quantified by densitometry and, as shown in Figure 24, ranged from 41 to 45% for each incubation. The percent reduction from the soluble lysate (lane 2) to after the third batch incubation (lane 5) was 81%. Thus, the loss after three batch incubations in the soluble lysate was 19%.

Following IMAC purification and dialysis into factor Xa digestion buffer, the His_6NusA/hIL-3 fusion protein was digested with factor Xa protease. As shown in Figure 25 (lanes 2-4), there were several unexpected protein bands which resulted from the factor Xa cleavage. While factor Xa has proven to be generally useful with other fusion protein systems, non-specific cleavage is not unprecendented (Doskeland et al., 1996). The non-specific cleavage of the His_6-NusA/hIL-3 was only observed upon addition of

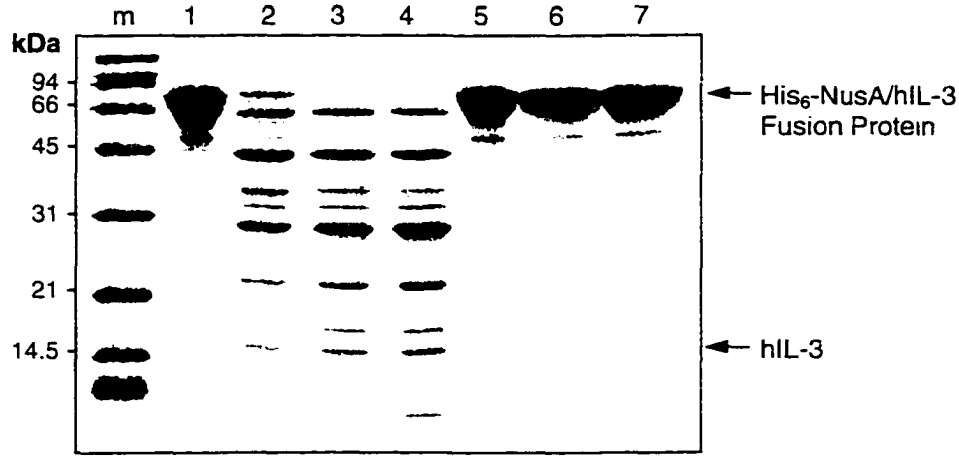


Figure 25. SDS-PAGE of Factor Xa Digestions of Purified His_6-NusA/hIL-3 Fusion Protein. Lane m, markers: Lane 1, IMAC purified His_6-NusA/hIL-3 fusion protein; Lanes 2-4, 1, 2, and 3 hr (23°C) factor Xa digestions respectively; Lanes 5-7, 1, 2, and 3 hr (23°C) digestions without factor Xa to determine protein stability.

the factor Xa enzyme to the reaction. Without factor Xa, the fusion protein appears to be quite stable for three hours at room temperature (lanes 5-6, Figure 25). Attempts were then made to decrease the amount of non-specific cleavage by decreasing the amount of factor Xa enzyme in the reaction. The results for various dilutions of factor Xa are shown in Figure 26. Over the three hour incubation period, even a 1:5 dilution of the factor Xa enzyme failed to yield significant cleavage of the fusion protein. It seems unlikely that factor Xa concentrations between 1:1 and 1:5 would reduce non-specific cleavage since a 1:5 dilution of factor Xa also resulted in a small, but visible, amount of non-specific cleavage (lane 3, Figure 26). In the New England Biolabs catalog (1998/99) there is a description of factor Xa activity which identifies Gly-Arg as a secondary recognition site for factor Xa cleavage. Upon inspection of the His_6-NusA/hIL-3 fusion protein for Gly-Arg sites, only one was found within the NusA protein (residues 319 and 320). If factor Xa cleaved at this site, the two resulting proteins would be 35.8 kDa (N-terminal) and 19.1 kDa (C-terminal). Upon inspection of Figure 25 (Lanes 2, 3, and 4), there appears to be one cleavage product at 23 kDa and three cleavage products between 29 and 35 kDa. Thus, it is very likely that these cleavage products could have been created by cleavage at the Gly-Arg site of NusA. It is shown later by western blotting (Figure 28B, lane FXa) that one of the non-specific cleavage products contains the hIL-3 protein. The apparent molecular weight of this protein band is 34 kDa, which roughly corresponds to the hIL-3 protein (15.1 kDa) plus the 19.1 kDa NusA fragment created by Gly-Arg. An explanation consistent with this result is that cleavage at the Gly-Arg site renders the Ile-Glu-Gly-Arg site inaccessible to factor Xa, thus inhibiting further cleavage. Another fact

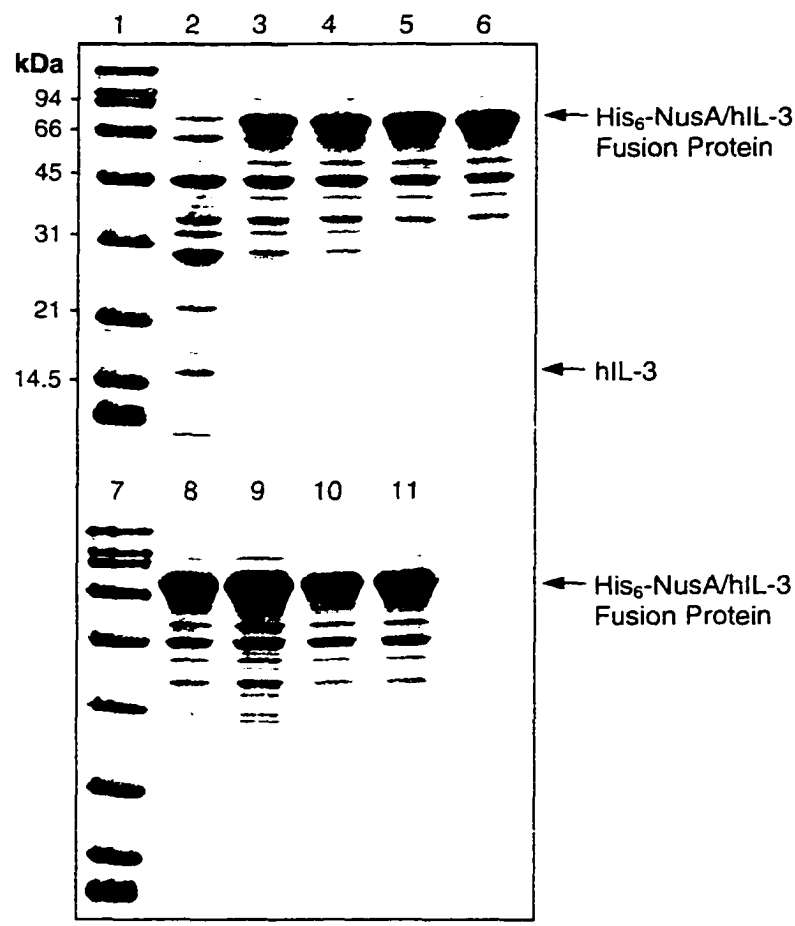


Figure 26. SDS-PAGE of Factor Xa Digestions of Purified His_6-NusA/hIL-3 Fusion Protein. Lane 1, markers; Lane 2, 3 hr (23°C) factor Xa digestion of IMAC purified His_6-NusA/hIL-3 fusion protein at 1 μg factor Xa per 50 μg fusion protein; Lanes 3-6 and 8-10 are factor Xa digestions (3 hr at 23°C) at 1:5, 1:10, 1:50, 1:100, 1:500, 1:1,000 and 1:10,000 dilutions, respectively, of factor Xa level in Lane 2; Lane 11, uncut starting fusion protein.

worthy of mentioning is that the Gly-Arg site in NusA is preceded by an Ile residue, thus creating an Ile-Gly-Arg site which might constitute enough specificity to mimic Ile-Glu-Gly-Arg recognition.

Factor Xa protease was chosen for cleavage here because of previous good experience in our laboratory with this protease cleaving other fusion proteins. For fusion proteins with NusA, however, it appears that cleavage with thrombin would be a better choice, since Zhang and Hanna (1995) did not obtain non-specific cleavage of NusA when they used thrombin to cleave NusA from a fusion protein with glutathione S-transferase. Thrombin, however, is not the ideal candidate for protease cleavage since it leaves N-terminal Lys-Gly residues on the target protein; however, it could be useful in lab scale protein characterizations.

Following factor Xa digestion of the fusion protein at 1 μg of factor Xa per 50 μg of fusion protein (23°C for three hours) and dialysis into a buffer containing no salts, the protease digest was passed over a Q Sepharose anion exchange chromatography column to remove hIL-3 from the digest mixture. A chromatogram and SDS-PAGE of the column is shown in Figure 27. Pure hIL-3 was found to elute in the flowthrough during loading of the Q Sepharose column.

Results for SDS-PAGE and western blotting for the expression and purification of hIL-3 are shown in Figure 28, and yield and percent recovery results are shown in Table 10. hIL-3 was purified to homogeneity after the anion exchange chromatography at a yield of 4.3 mg/L of expression culture and a 13% recovery. The largest loss of hIL-3 occurred in the IMAC step. The 32% yield in the IMAC step is consistent with the

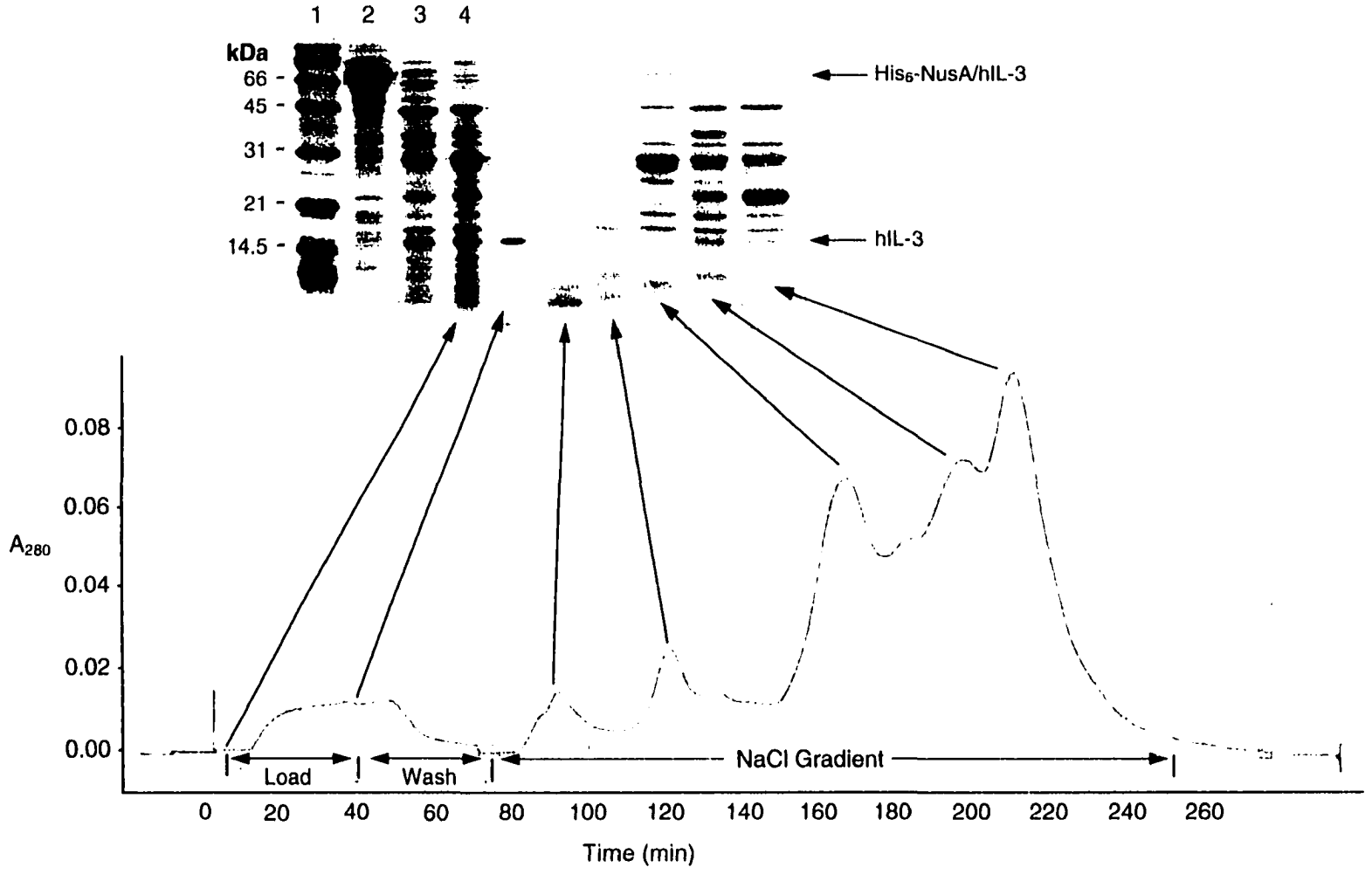


Figure 27. Chromatograph and SDS-PAGE of anion exchange chromatography using an NaCl gradient from 0 to 360 mM over 175 minutes at 0.5 ml/min (38 cm/hr). Pure hIL-3 was found to elute during the loading of the column. For visualization by SDS-PAGE, 20 µl of the above peak samples were combined with 5 µl of gel loading buffer, and loaded onto an SDS-PAGE mini-gel. Lane 1, marker; Lane 2, IMAC purified His_6-NusA/hIL-3; Lane 3, factor Xa cut His_6-NusA/hIL-3, Lane 4, dialyzed factor Xa digest (ready for column loading).

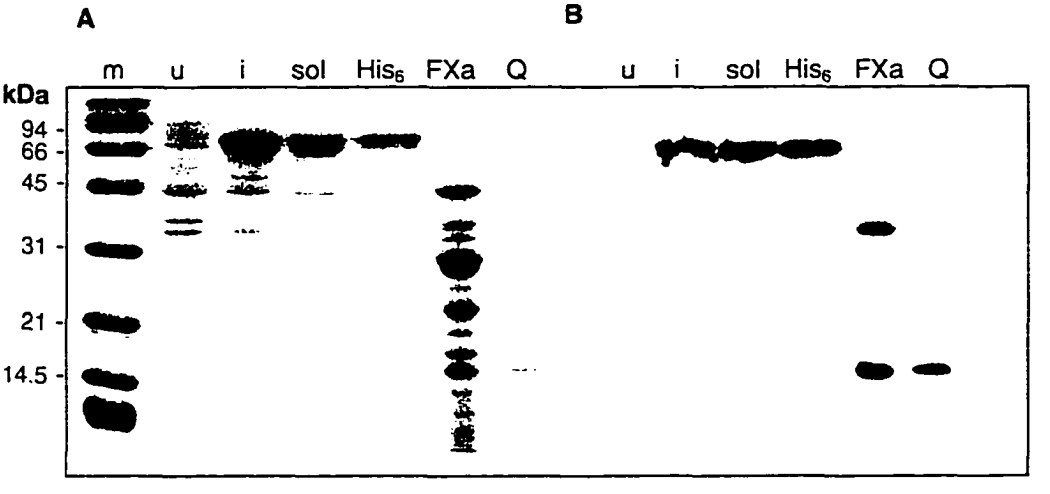


Figure 28. SDS-PAGE (A) and Western Blot (B) of Expression and Purification of hIL-3 using the NusA Expression System. A His_6 tag was placed at the N-terminus of NusA for affinity purification. **(m)**. molecular weight markers. **(u)**, uninduced whole cell lysate of cells harboring the His_6NusA/hIL-3 gene. **(i)**, induced whole cell lysate of cells with His_6NusA/hIL-3 gene. **(sol)**, soluble lysate. **(His_6)**, His_6NusA/hIL-3 protein eluted from IMAC column. **(FXa)**, Factor Xa digest of IMAC purified His_6NusA/hIL-3 protein. **(Q)**, hIL-3 purified from the Factor Xa digest using a Q-Sepharose column. The His_6NusA/hIL-3 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. The western blot was probed with mouse anti-hIL-3 monoclonal antibody and visualized using chemiluminescence.

Purification Step	hIL-3
Soluble lysate	9.9 mg
IMAC purification	3.2 mg
Factor Xa Digest	2.7 mg
Anion Exchange Chromatography	1.3 mg
Recovery	13.1 %
Yield	4.3 mg/L

Table 10. Yield and Percent Recovery of hIL-3. Production of human interleukin-3 from expression as a His_6-NusA fusion protein through IMAC and anion exchange chromatography. Quantities of hIL-3 and calculations are based on a 300 ml expresion culture. hIL-3 levels were quantified by densitometry of Coomassie Blue stained SDS-PAGE gels and the BCA total protein assay.

recovery results obtained by Van Dyke et al. (1992) for the purification by IMAC of an upstream stimulatory factor (USF) with a His_6 at the N terminus. The percent recovery for the wild type and mutant forms of USF were reported to be 17 and 28%, respectively, for the IMAC step.

Various Qualitative Laboratory Observations

During the overexpression of proteins, it was noticed on several occasions that there is a difference in the color of the cell pellets. At the end of the 3 hour post-induction growth period, 4 ml cultures were typically centrifuged for 10 min at 1000 x *g* (as described in Appendix B) and then stored at -20°C until further analysis. The pellet of the uninduced culture was a light-brown to yellowish color, while the induced cell pellet was considerably lighter, almost entirely white. While not rigorously recorded and analyzed, this phenomena was observed for several different fusion proteins and is therefore not protein specific.

At one point, it was attempted to shorten the centrifugation time of the 4 ml expression cultures at the end of the 3 hour post-induction period from 10 min to 2 min. While a 2 min centrifugation at 1000 x g resulted in no significant sedimentation, steadily increasing the centrifugation time in increments from 2 min to 10 min revealed a partial sedimentation of cells. The partial sedimentation of induced and uninduced cultures was not the same. The induced cells sedimented faster than the uniniduced cells.

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

This study has resulted in the creation of a soluble fusion protein expression system based on the *E. coli* NusA protein. NusA was shown to solubilize three target proteins (hIL-3, bGH, and hIFN-γ) which have been shown in previous research to be expressed as inclusion bodies in *E. coli*. In addition, a method of purification was developed in which a target protein, hIL-3, was purified to homogeneity from a histidine-tagged version of the NusA expression system. These results indicate that the NusA fusion protein expression system has a legitimate potential for broad implementation in the biotechnology field.

Several other carrier proteins were identified which were also shown to increase the solubility of hIL-3, namely bacterioferritin, GrpE, and the 2X-YjgD protein. While the increases in solubility of these carrier proteins was not as dramatic as the case for NusA, each resulted in an increase in solubility compared to the use of thioredoxin. This is a very significant result considering that the thioredoxin fusion protein expression system is a marketed life science product (Invitrogen, Carlsbad, CA) and has found wide use in biotechnology research.

The two-parameter version of the Wilkinson-Harrison solubility model was shown to be useful in correctly identifying soluble carrier proteins. This is apparent since every carrier protein chosen, with the exception of YjgD, was shown to produce soluble hIL-3 at the optimal growth temperature of 37°C for *E. coli*. The agreement of

experimental results and modeling predictions attest to the solubility model's potential for use in other applications. It is very possible that the same statistical approach to modeling protein solubility in *E. coli* could applied to other cell types such as yeast, mammalian, or insect cells. At the moment however, eukaryotic cell types, aside from yeast, possess rather low protein expression levels compared to those found in prokaryotes. So, it is likely that inclusion body formation, as a result of high level expression, will not be a major problem for eukaryotic systems in the near future. However, the solubility model could find some limited application with respect to inherently insoluble proteins which are shown to aggregate even at low expression levels. Another complication for mammalian systems is a lack of expression data. *E. coli* has been studied and utilized so intensely in the biotechnology field that a plethora of data is available on which to base modeling. The most important aspect of this data is that it provides a large enough sample size for a valid statistical analysis.

YjgD had the lowest CV-CV' value (-9.37) of any *E.coli* protein larger than 100 amino acids. The CV-CV' value of -9.37 is lower than any of the CV-CV' values for the proteins on which the solubility model is based (the lowest CV-CV' value for the proteins used for the solubility model was -4.3). Thus, YjgD represents a data point outside the range of the solubility model. As shown with the other carrier proteins (e.g. NusA, GrpE, and BFR), significant increases in solubility were obtained at CV-CV' values less negative than -9.37. So, is it necessary to have a protein with CV-CV' values lower than that of NusA, BFR, and GrpE? The experimental results associated with YjgD suggest that an *extremely* high approximate charge average (-0.327 charge average for YjgD

versus an average charge average of -0.112 for NusA, GrpE, and BFR) can negatively affect protein expression. For example, in Table 8, it was shown that induction of YjgD results in the largest decrease in growth rate of *E. coli* relative to thioredoxin and GrpE fusion proteins. Also, Figure 17C shows a significant amount of protease degradation of the 2X-YjgD/hIL-3 fusion protein upon sonication. Thus, the numerous proteins listed in Table 6 with CV-CV' values less negative than that of YjgD probably have a more realistic chance than YjgD of becoming useful carrier proteins which can express at high levels.

The only limitation in implementing any of the remaining carrier proteins is committing the time required to evaluate them by subcloning and overexpression. Towards this end, much thought would have to be put into developing a rapid, high-throughput method of screening for soluble carrier proteins. Could an efficient reporter molecule like green fluorescent protein (GFP) be used? It has been shown that wild-type GFP partitions between soluble and insoluble fractions in *E. coli* (Patterson et al., 1997), and therefore might actually be practical as a quick screening method. Solubility would simply be correlated with light emitting intensity. There would also be a bottle-neck with respect to DNA template acquisition and preparation for each of the potential carrier proteins. Contacting 40 to 50 different researchers and preparing the DNA in sufficient quantities from different cloning vectors (e.g. plasmid DNA, viral DNA, etc.) would prove to be arduous. A more efficient method would be to PCR amplify the carrier genes of interest directly from the *E. coli* genome using preparations of genomic DNA.

One theme that seems to be consistent in the results obtained in this study is that

the carrier protein must be of a sufficient molecular weight in order to solubilize a target protein. However, it is not believed that a large molecular weight carrier protein is all that is required. All the proteins studied here had solubility probabilities above 90% as individual proteins. So, most likely, for the fusion proteins selected to study the odds are stacked in favor of solubility beyond a certain molecular weight limitation. It would be nice if the development of soluble carrier proteins could follow the path that affinity tags have taken; that is, starting out with a huge affinity carrier proteins such as Staphylococcus protein A and eventually honing the size down to a six-histidine peptide sequence which is equally, if not more effective. This study, however, indicates that for the solubility parameters to have a significant effect, a carrier protein with a sufficient molecular weight is required for soluble expression.

Despite its predicted solubility, the YjgD/hIL-3 protein was found to be insoluble at 37°C. Upon adding another YjgD protein to the YjgD/hIL-3 fusion protein, the resulting 2X-YjgD/hIL-3 fusion protein was found to be completely soluble at 37°C. This result suggests that the same "multiple carrier protein" approach could be applied to other carrier proteins in this study. For instance, two BFR proteins could be linked together as was done with the YjgD protein. Or a combination of BFR and YjgD could be constructed which might offer multiple avenues for protein purification in one fusion protein. BFR would be placed at the N terminus since it seems to have a higher expression level than YjgD. The acidity of YjgD could be useful in ion exchange chromatography, and it would also be interesting to know if BFR offers any novel purification strategies with immobilized metal affinity chromatography due to its iron and

zinc binding properties.

Despite considerable non-specific cleavage of the NusA/hIL-3 protein by factor Xa, it was possible to purify hIL-3 to homogeneity by anion exchange chromatography. Reducing the concentration of factor Xa in the reaction did not improve protease specificity. Thus, an efficient target protein release method has yet to be developed for the NusA expression system. There are several approaches which can be taken to increase cleavage specificity, namely: 1) the use of different proteases, 2) using factor Xa purified from a source other than New England Biolabs, 3) mutating NusA to remove the Ile-Gly-Arg sequence, 4) extending cleavage times beyond 3 hours at diluted levels of factor Xa, or 5) inserting a linker sequence between the carboxy terminus of the NusA protein and the amino terminus of the factor Xa sequence.

Two proteases, bovine enterokinase light chain (EK_L) (LaVallie et al., 1993) and *Saccharomyces cerevisiae* Kex2 (Ghosh and Lowenstein, 1996), both cut at the carboxy terminus of their cleavage sequences (Asp_4-Lys and Leu-Phe-Lys-Arg, respectively) and are not present in the NusA sequence. It has been claimed that EK_L has absolute specificity for the Asp_4-Lys sequence (LaVallie et al., 1993), however, it has also been suggested that EK_L will cleave at other basic residues depending on protein conformation (New England Biolabs catalog, 1998/99). One advantage of EK_L is that it is recombinantly produced, thus avoiding the previous problems of contamination from other proteases during the purification of enterokinase from bovine intestines. It is possible that the specificity of factor Xa may be somewhat limited since it is currently not produced recombinantly. The specificity of Kex2 was extensively investigated by

Brenner and Fuller (1992). Several alternative cleavage sites were characterized including Tyr-Lys-Arg, Leu-Lys-Arg, and Leu-Arg-Arg with Tyr-Lys-Arg having the most favorable kinetics. It is possible that either EK_L or Kex2 might prove to be more compatible with NusA than factor Xa.

Factor Xa has been criticized for cleaving slowly and inefficiently (Pryor and Leiting, 1997), so it would be worth investigating the effect of a 24 hour incubation at the 1:10 factor Xa dilution shown in Figure 4 (lane 4). As for linker sequences, it has been shown that a glycine-rich linker (Pro-Gly-Ile-Ser-Gly_5) placed immediately downstream of a thrombin cleavage site can greatly increase the cleavage efficiency of several GST fusion proteins (Guan and Dixon, 1991). Another study using the H64A subtilisin protease found that increasing the length of the linker resulted in an increase in cleavage efficiency (Polyak et al., 1997). Increasing the accessibility of the factor Xa sequence by placing a glycine-rich linker upstream of the factor Xa recognition sequence might improve factor Xa specificity.

An aspect of this project which has not been explored extensively is the kinetics of the expression levels of the various fusion proteins. An 8 hour expression of the GrpE/hIL-3 fusion protein was investigated, and it was found that a steady state level of fusion protein expression was reached after 1 hour with no further accumulation of fusion protein. Assuming that the GrpE/hIL-3 fusion protein is being expressed continually over the 8 hour period and considering the swift rise within 1 hour to a steady expression level, it stands to reason that there must be some *in vivo* proteolytic degradation of the GrpE/hIL-3 fusion protein. Various upstream strategies for decreasing proteolytic

degradation of proteins have recently been reviewed and suggest several alternatives such as secretion and the use of protease deficient strains (Murby et al., 1996). Despite the constant level of GrpE/hIL-3 fusion protein expression, the kinetics of the BFR, NusA, and YjgD fusion proteins might prove to be altogether different over an 8 hour induction period. It is certainly worth investigating the kinetics of the NusA system to maximize its yield potential since it has been shown in this study that NusA can solubilize three insoluble target proteins.

Non-fusion chaperone expression systems have an advantage over fusion protein technology in that they possess catalytic activity; one chaperone molecule can catalyze the proper folding of several thousand target proteins. Obviously, less chaperone protein is required than in the case of fusion protein expression where equimolar quantities of the target and carrier protein are synthesized. This equimolar requirement is bound to place some metabolic burden on the cell. A major drawback to the practical implementation of chaperone systems is their specificity for folding only certain proteins (Makrides, 1996). The relatively non-specific nature of fusion protein technology is readily apparent and, indeed, is the reason for its broad practical implementation. How could the problem of specificity be overcome in chaperone systems? Could a universal peptide be attached to the target protein which marked it for refolding? Could that peptide, when attached to an insoluble target protein, direct the target to the chaperone machinery? If a system like this could be developed, which supplied control over the specificity of the chaperone, the system would take on the broad application advantage of fusion protein technology and not place such a high metabolic burden on the cell.

As mentioned earlier in the expression results for the NusA/hIL-3 fusion protein, it was learned that NusA is a transcriptional elongation factor which binds directly to the core RNA polymerase. One interesting phenomenon associated with NusA is transcriptional pausing. When NusA is present during transcription, RNA polymerase will pause for longer periods of time at certain pausing sites along the template DNA. Since transcription and translation are coupled in prokaryotes, it is possible that NusA may be able to increase soluble levels of proteins by slowing down translation at the transcriptional pauses and allowing critical folding events to occur. This is a very interesting prospect since the kinetics of transcription, translation, and protein folding have received much attention in recent years with respect to production of active proteins in *E. coli* (Georgiou and Valax, 1996; Makrides, 1996; King et al., 1996). Therefore, it would be interesting to investigate the solubilizing potential of NusA not as a fusion protein, but when co-overexpressed with the target protein. As mentioned before, this kind of work has been done for thioredoxin with rather impressive results (Yasakuwa et al., 1995). This would determine if the activity of NusA has any solubilizing affect and could potentially allow higher levels of target protein expression; that is, since NusA and the target protein would not necessarily have to be produced in equimolar amounts. Another possibility is to truncate various portions of the NusA protein to reduce its size and allow higher levels of expression for the target protein. However, it is difficult to plan this kind of research since the crystal structure of NusA is not known.

A word on laboratory protocol is merited. One of the major bottle-necks of this project was having an efficient method for screening a large number of clones. The

method of screening by protein expression, which was developed during this study, proved to be very reliable in identifying correct clones. However, it is quite laborious and requires a non-stop, 12-hour day of laboratory work to screen 30 colonies using four milliliter test tube cultures. In comparison, the dot-immunoblot method used to identify the NusA/hIFN-γ clone was less laborious and gives more information as to the identity of the fusion protein being overexpressed. While both methods have their own merits, it is believed that screening using antibodies should be done whenever an antibody is available.

BIBLIOGRAPHY

Andrews, S.C., Harrison, P.M. and J.R. Guest. 1989. Cloning, sequencing, and mapping of the bacterioferritin gene (*bfr*) of *Escherichia coli* K-12. *Journal of Bacteriology*. **171**(7): 3940-3947.

Ausubel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith and Struhl, K. 1987. *Current Protocols in Molecular Biology*. Wiley Interscience, Boston, MA.

Baneyx, F. and Georgiou, G.J. 1991. Construction and characterization of *Escherichia coli* strains deficient in multiple secreted proteases: protease III degrades high-molecular-weight substrates *in vivo*. *J. Bacteriol.* **173**(8):2696-2703.

Barnes, W.M. 1994. PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc. Natl. Acad. Sci. U.S.A.* **91**(6):2216-2220.

Barnfield-Frej, A.K. 1996. Expanded bed adsorption for recovery of renatured human recombinant interleukin-8 from *Escherichia coli* inclusion bodies. *Bioseparation.* **6**(5):265-271.

Betton J. and Hofnung, M. J. 1996. Folding of a mutant maltose-binding protein of *Escherichia coli* which forms inclusion bodies. *J. Biol. Chem.* **271**(14):8046-8052.

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W. et al. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science.* 277(5331):1453-1474.

Brenner, C. and Fuller, R.S. 1992. Structural and enzymatic characterization of a purified prohormone-processing enzyme: secreted, soluble Kex2 protease. 1992. *Proc. Natl. Acad. Sci. U.S.A.* **89**(3):922-926.

Chen, G.Q. and Gouaux, E. 1997. Overexpression of a glutamate receptor (GluR2) ligand binding domain in *Escherichia coli*: application of a novel protein folding screen. *Proc. Natl. Acad. Sci. U.S.A.* **94**(25):13431-13436.

Chong, S., Mersha, F.B., Comb, D.G., Scott, M.E., Landry, D., Vence, L.M., Perler, F.B., Benner J., Kucera, R.B., Hirvonen, C.A., Pelletier, J.J., Paulus, H. and Xu, M.Q. 1997. Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene.* **192**(2):271-281.

Cooper, A.A., Stevens, T.H. 1995. Protein splicing: self-splicing of genetically mobile elements at the protein level. *Trends Biochem. Sci.* **20**(9):351-356.

Davis, G.D. and R.G. Harrison. 1998. Rapid screening of fusion protein recombinants by measuring effects of protein overexpression on cell growth. *BioTechniques.* **24**(3): 360-362.

Dedhia, N., Richins, R., Mesina, A. and Chen, W. 1997. Improvement in recombinant protein production in ppGpp-deficient *Escherichia coli. Biotechnol. Bioeng.* **53**(4): 379-386.

di Guan, C., Li, P,, Riggs, P.D. and Inouye, H. 1988. Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein.
Gene. **67**(1):21-30.

Donahue, R.E., Seehra, J., Metzger, M., Lefebvre, D., Rock, B., Carbone, S., Nathan, D.G., Garnick, M., Sehgal, P.K., Laston, D., et al. 1988. Human IL-3 and GM-CSF act synergistically in stimulating hematopoiesis in primates. *Science.* **241**(4874):1820-1823.

Doskeland, A.P., Martinez, A., Knappskog, P.M. and Flatmark, T. 1996. Phosphorylation of recombinant human phenylalanine hydroxylase: effect on catalytic activity, substrate activation and protection against non-specific cleavage of the fusion protein by restriction protease. *Biochem. J.* **313**(Pt 2):409-414.

Ernst, S., Garro, O. A., Winkler, S., Venkataraman, G., Langer, R., Cooney, C.L. and Sasisekharan, R. 1997. Process simulation for recombinant protein production: cost estimation and sensitivity analysis for heparinase I expressed in *Escherichia coli. Biotechnol. Bioeng.* **53**(6): 575-582.

Frangioni, J.V. and Neel, B.G. 1993. Solubilization and purification of enzymatically active glutathione S-transferase (pGEX) fusion proteins. *Anal. Biochem.* **210**(1):179-187.

Georgiou, G. and Valax, P. 1996. Expression of correctly folded proteins in *Escherichia coli. Curr. Opin. Biotechnol.* 7(2):190-197.

Ghosh, S. and Lowenstein, J.M. 1996. A multifunctional vector system for heterologous expression of proteins in *Escherichia coli.* Expression of native and hexahistidyl fusion proteins, rapid purification of the fusion proteins, and removal of fusion peptide by Kex2 protease. *Gene.* **176**(1-2):249-255.

Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K. and Riggs, A.D. 1979. Expression in *Escherichia coli* of chemically synthesized genes for human insulin. *Proc. Natl. Acad. Sci. U.S.A.* **76**(1):106-110.

Guan, K.L., Dixon, J.E. 1991. Eukaryotic proteins expressed in *Escherichia coli*: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione-S-transferase. *Anal. Biochem.* **192**(2):262-267.

Itakura, K., Hirose, T., Crea, R., Riggs, A.D., Heyneker, H.L., Bolivar, F. and Boyer, H.W. 1977. Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. *Science.* **198**(4321):1056-1063.

Hannig, G. and Makrides, S.C. 1998. Strategies for optimizing heterologous protein expression in *Escherichia coli. Trends Biotechnol.* **16**(2):54-60.

Hasenwinkle, D., Jervis, E., Kops, O., Liu, C., Lesnicki, G., Haynes, C.A. and Kilburn, D.G. 1997. Very high-level production and export in *Escherichia coli* of a cellulose binding domain for use in a generic secretion-affinity fusion system. *Biotechnol. Bioeng.* **55**(6):854-863.

Joly, J.C., Leung, W.S., Swartz, J.R. 1998. Overexpression of *Escherichia coli* oxidoreductases increases recombinant insulin-like growth factor-I accumulation. *Proc. Natl. Acad. Sci. U.S.A.* **95**(6):2773-2777.

Kandror, O., Sherman, M., Moerschell, R., Goldberg, A.L. 1997. Trigger factor associates with GroEL *in vivo* and promotes its binding to certain polypeptides. *J. Biol. Chem.* **272**(3):1730-1734.

King, J., Haase-Pettingell, C., Robinson, A.S., Speed, M. and Mitraki, A. 1996. Thermolabile folding intermediates: inclusion body precursors and chaperonin substrates. *FASEB J.* **10**(1):57-66.

Kitamura, T., Tange, T., Terasawa, T., Chiba, S., Kuwaki, T., Miyagawa, K., Piao, Y.F., Miyazono, K., Urabe, A. and Takaku, F. 1989. Establishment and characterization of a unique human cell line that proliferates dependently on GM-CSF, IL-3, or erythropoietin. *J. Cell. Physiol.* **140**(2):323-334.

Kopitar, G., Dolinar, M., Strukelj, B., Pungercar, J. and Turk, V. 1996. Folding and activation of human procathepsin S from inclusion bodies produced in *Escherichia coli. Eur. J. Biochem.* **236**(2):558-562.

LaVallie, E.R. and McCoy, J.M. 1995. Gene fusion expression systems in *Escherichia coli. Curr. Opin. Biotechnol.* **6**(5):501-506.

LaVallie, E.R., Rehemtulla, A., Racie, L.A., DiBlasio, E.A., Ferenz, C., Grant, K.L., Light, A. and McCoy, J.M. 1993. Cloning and functional expression of a cDNA encoding the catalytic subunit of bovine enterokinase. *J. Biol. Chem.* **268**(31):23311-23317

LaVallie, E.R., DiBlasio, E. A., Kovacic, S., Grant, K.L., Schendel, P.F. and J.M. McCoy. 1993. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Bio/Technology.* **11**:187-193.

Lazar, S.W. and Kolter, R. 1996. SurA assists the folding of *Escherichia coli* outer membrane proteins. *J. Bacteriol.* **178**(6):1770-1773.

Lehrman, S.R., Tuls, J.L., Havel, H.A., Haskell, R.J., Putnam, S.D., Tomich, C.S. 1991. Site-directed mutagenesis to probe protein folding: evidence that the formation and aggregation of a bovine growth hormone folding intermediate are dissociable processes. *Biochemistry.* **30**(23):5777-5784.

Ljungquist, C., Breitholtz, A., Brink-Nilsson, H., Moks, T., Uhlen, M. and Nilsson, B. 1989. Immobilization and affinity purification of recombinant proteins using histidine peptide fusions. *Eur. J. Biochem.* **186**(3):563-569.

Lu, Z., DiBlasio-Smith, E.A., Grant, K.L., Warne, N.W., LaVallie, E.R., Collins-Racie, L.A., Follettie, M.T., Williamson, M.J. and McCoy, J.M. 1996. Histidine patch thioredoxins. Mutant forms of thioredoxin with metal chelating affinity that provide for convenient purifications of thioredoxin fusion proteins. *J. Biol. Chem.* **271**(9):5059-5065.

Makrides, S.C. 1996. Strategies for achieving high-level expression of genes in *Escherichia coli. Microbiol. Rev.* **60**(3):512-538.

Murby, M., Uhlen, M. and Stahl, S. 1996. Upstream strategies to minimize proteolytic degradation upon recombinant production in *Escherichia coli. Protein Expr. Purif.* 7(2):129-136.

Nakamura, Y. and S. Mizusawa. 1985. *In vivo* evidence that the *nus*A and *inf*B genes of *E. coli* are part of the same multi-gene operon which encodes at least four proteins. *The EMBO Journal.* **4**(2): 527-532.

Ong. E., Alimonti. J.B., Greenwood. J.M., Miller, R.C. Jr., Warren, R.A. and Kilburn, D.G. 1995. Purification of human interleukin-2 using the cellulose-binding domain of a prokaryotic cellulase. *Bioseparation.* **5**(2):95-104.

Ostermeier, M., De Sutter, K. and Georgiou, G. 1996. Eukaryotic protein disulfide isomerase complements *Escherichia coli* dsbA mutants and increases the yield of a heterologous secreted protein with disulfide bonds. *J. Biol. Chem.* **271**(18):10616-10622.

Patterson, G.H., Knobel, S.M., Sharif, W.D., Kain, S.R. and Piston, D.W. 1997. Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophys. J.* 73(5):2782-2790.

Perez-Martin J., Cases, I. and de Lorenzo V. 1997. Design of a solubilization pathway for recombinant polypeptides *in vivo* through processing of a bi-protein with a viral protease. *Protein Eng.* **10**(6):725-730.

Piers, K.L., Brown, M.H. and Hancock, R.E. 1993. Recombinant DNA procedures for producing small antimicrobial cationic peptides in bacteria. *Gene.* **134**(1):7-13.

Pilon, A.L., Yost, P., Chase, T.E., Lohnas, G.L. and Bentley, W.E. 1996. High-level expression and efficient recovery of ubiquitin fusion proteins from *Escherichia coli. Biotechnol. Prog.* **12**(3):331-337.

Polyak, S.W., Forsberg, G., Forbes, B.E., McNeil, K.A., Aplin, S.E. and Wallace, J.C. 1997. Introduction of spacer peptides N-terminal to a cleavage recognition motif in recombinant fusion proteins can improve site-specific cleavage. *Protein Eng.* **10**(6):615-619.

Proudfoot, A.E., Goffin, L., Payton, M.A., Wells, T.N. and Bernard, A.R. 1996. *In vivo* and *in vitro* folding of a recombinant metalloenzyme, phosphomannose isomerase. *Biochem. J.* **318**(Pt 2):437-442.

Pryor, K.D. and Leiting, B. 1997. High-level expression of soluble protein in *Escherichia coli* using a His_6-tag and maltose-binding-protein double-affinity fusion system. *Protein Expr. Purif.* **10**(3):309-319.

Ray, M.V., Van Duyne, P., Bertelsen, A.H., Jackson-Matthews, D.E., Sturmer, A.M., Merkler, D.J., Consalvo, A.P., Young, S.D., Gilligan, J.P. and Shields, P.P. 1993. Production of recombinant salmon calcitonin by *in vitro* amidation of an *Escherichia coli* produced precursor peptide. *Bio/Technology.* **11**(1):64-70.

Riggs, P. and LaVallie, E.R. 1994. Introduction to expression by fusion protein vectors. *in Current Protocols in Molecular Biology* Supplement 28. Wiley Interscience, Boston, MA.

Rudolph, R. and Lilie, H. 1996. *In vitro* folding of inclusion body proteins. *FASEB J.* **10**(1):49-56.

Sachdev, D. and Chirgwin, J.M. 1998. Order of fusions between bacterial and mammalian proteins can determine solubility in *Escherichia coli. Biochem. Biophys. Res. Commun.* **244**(3):933-937.

Sambrook, J., E.F. Fritsch and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schatz, P.J. 1993. Use of peptide libraries to map the substrate specificity of a peptide-modifying enzyme: a 13 residue consensus peptide specifies biotinylation in *Escherichia coli*. *Bio/Technology*. **11**(10):1138-1143.

Schmid, B., Kromer, M. and Schulz, G.E. 1996. Expression of porin from Rhodopseudomonas blastica in *Escherichia coli* inclusion bodies and folding into exact native structure. *FEBS Lett.* **381**(1-2):111-114.

Shi, P.Y., Maizels, N. and Weiner, A.M. 1997. Recovery of soluble, active recombinant protein from inclusion bodies. *Biotechniques*. 23(6):1036-1038.

Smith, D.B. and K.S. Johnson. 1988. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene*. **67**: 31-40.

Smith, P.A., Tripp, B.C., DiBlasio-Smith, E.A., Lu, Z., LaVallie, E.R. and McCoy, J.M. 1998. A plasmid expression system for quantitative *in vivo* biotinylation of thioredoxin fusion proteins in *Escherichia coli*. *Nucleic Acids Res*. **26**(6):1414-1420.

Strauch, K.L. and Beckwith, J. 1988. An *Escherichia coli* mutation preventing degradation of abnormal periplasmic proteins. *Proc. Natl. Acad. Sci. U.S.A.* **85**(5):1576-1580.

Thomas, M. 1994. Simple, effective cleanup of DNA ligation reactions prior to electro-transformation of *E. coli*. *Biotechniques*. **16**(6): 988.

Tsai, A.M., Betebaugh, M.J. and J. Shiloach. 1995. The Kinetics of RCC1 inclusion body formation in *Escherichia coli*. *Biotechnology and Bioengineering*. **48**:715-718.

Tucker, J. and Grisshammer, R. 1996. Purification of a rat neurotensin receptor expressed in *Escherichia coli*. *Biochem. J.* **317**(Pt 3):891-899.

Uhlen, M., Nilsson, B., Guss, B., Lindberg, M., Gatenbeck, S. and Philipson, L. 1983. Gene fusion vectors based on the gene for staphylococcal protein A. *Gene*. **23**(3):369-378.

Unlap, T and X. Hu. 1995. Direct identification of recombinant clones in *Escherichia coli* utilizing supercoiled DNA. *Biotechniques*. **18**(4): 590-592.

Van Dyke, M.W., Sirito, M. and Sawadogo, M. 1992. Single-step purification of bacterially expressed polypeptides containing an oligo-histidine domain. *Gene*. **111**(1):99-104.

van Leen, R.W., Bakhuis, J.G., van Beckhoven, R.F.W.C., Burger, H., Dorssers, L.C.J., Hommes, R.W.J., Lemson, P.J., Noordam, B., Persoon, L.M. and G. Wagemaker. 1991. Production of Human Interleukin-3 using industrial microoorganisms. *Bio/Technology*. **9**: 47-52.

Vasina, J.A. and Baneyx, F. 1997. Expression of aggregation-prone recombinant proteins at low temperatures: a comparative study of the *Escherichia coli* cspA and tac promoter systems. *Protein Expr. Purif.* **9**(2):211-218.

Wetzel, R. and B.A. Chrunyk. 1994. Inclusion body formation by interleukin-1β depends on the thermal sensitivity of a folding intermediate. *FEBS Letters*. **350**: 245-248.

Walker, P.A., Leong, L.E.C., Ng, P.W.P, Tan, S.H., Waller, S., Murphy, D. and A.G. Porter. 1994. Efficient and rapid affinity purification of proteins using recombinant fusion proteases. *Bio/Technology*. **12**:601-605.

Wilkinson, D.L. and R.G. Harrison. 1991. Predicting the solubility of recombinant proteins in *Escherichia coli*. *Bio/Technology*. **9**: 443-448.

Wilkinson, D.L., Ma, N.T., Haught, C. and R.G. Harrison. 1995. Purification by immobilized metal affinity chromatography of human atrial natriuretic peptide expressed in a novel thioredoxin fusion protein. *Biotechnology Progress*. **11**: 265-269.

Zhang, Y., Olsen, D.R., Nguyen, K.B., Olson, P.S., Rhodes, E.T. and Mascarenchas, D. 1998. Expression of eukaryotic proteins in soluble form in *Escherichia coli*. *Protein Expr. Purif.* **12**(2):159-165.

Zhang, Y. and M.M. Hanna. 1995. Expression and functional characterization of *Escherichia coli* NusA and lamda Q as glutathione S-transferase fusion proteins. *Protein Expr. Purif.* **6**: 625-631.

Zheng, C.F., Simcox, T., Xu, L. and Vaillancourt, P. 1997. A new expression vector for high level protein production, one step purification and direct isotopic labeling of calmodulin-binding peptide fusion proteins. *Gene*. **186**(1):55-60.

Zylicz, M., Ang, D. and C. Georgopoulos. 1987. The grpE protein of *Escherichia coli*. *Journal of Biological Chemistry*. **262**(36): 17437-17442.

APPENDIX A
Computer Programming

SwissLight

The SwissLight program will calculate the solubility probability of recombinant proteins in the *E. coli* cytoplasm based on the Wilkinson-Harrison statistical model. The C program can be compiled and run on the UNIX servers at the University of Oklahoma Engineering Computer Network (ECN). A text file must be provided in the same directory as this program which contains the SwissProt formatted sequences of all the proteins to be evaluated. Each protein sequence in the text file should be separated by two forward slash characters (i.e. //). A list of all *E. coli* sequences can be created by going the SwissProt site on the World Wide Web at http://expasy.hcuge.ch, searching for the proteins you are interested in, and then having the server condense your search results into one text file which may then be downloaded through Netscape. To compile and execute this program, at the UNIX prompt type:

cc -o SwissLight SwissLight.c -lm

Then execute the program by typing:

SwissLight

The program will prompt you for the name of the sequence input file. To see the results of the calculation, use a text editor to open the file 'solout' created by the program.

The C code for the SwissLight program is:

```
#include <stdio.h>
#include <dirent.h>
#include <math.h>
#include <sys/types.h>
#include <sys/stat.h>
#include <string.h>
```

```
#define MAXBUFLEN 1024
#define MAXLINELEN 128
#define MAXSEQLEN 1024*128

main()
{

  char buf[MAXLINELEN];
  char seq[MAXSEQLEN];
  char fname[MAXSEQLEN];
  int mode, abode, nseq, c;
  FILE *in, *out;
  char *z;
  char id[11];
  float r,n,d,e,g,k,p,s,cv;

  printf("Enter the name of a GCG formatted file to read....\n\n");
  scanf("%s", fname);
  printf("\n");

  if ((in = fopen (fname, "r")) == NULL)
   {
      printf ("\nCannot find the file %s.\n\n", fname);
      printf("%s must be in this directory for program execution.\n\n",
fname);
      return 0;
   }

  out = fopen("solout", "w");
  fprintf (out, "Protein Name      aa       CV-CV'\n");
  fprintf (out, "---------------------------------\n");

  while (!feof(in))
  {
    fgets(buf, sizeof(buf), in);
    if (strncmp(buf, "ID", 2)==0)
      {
        for (c=5; c < 15; c++)
            id[c-5] = buf[c];
            id[10] = 0;
      }
    mode=1;
    abode=1;
    nseq=0;
    while(!feof(in) && mode)
    {
      fgets(buf, sizeof(buf), in);
      if (strncmp(buf, "SQ", 2)==0)
          mode=0;
    }

    while (!feof(in) && abode)
    {
```

```
      z = fgets(buf, sizeof(buf), in);
      if (strncmp(buf, "//", 2)!=0)
         {
          for (z=buf;*z;z++)
        if (isalpha(*z)) seq[nseq++] = *z;
          }
      if (strncmp(buf, "//", 2)==0)
          abode=0;
    }

/* ------------- COUNT AMINO ACIDS ------------------------- */

   r=0;
   n=0;
   d=0;
   e=0;
   g=0;
   k=0;
   p=0;
   s=0;
   cv=0;

   for (c=0; c < nseq; c++)
     {
            if (seq[c]=='R') r=r++;          /* Arginine */
            if (seq[c]=='N') n=n++;        /* Asparagine */
            if (seq[c]=='D') d=d++;        /* Aspartic Acid */
            if (seq[c]=='E') e=e++;        /* Glutamic Acid */
            if (seq[c]=='G') g=g++;        /* Glycine */
            if (seq[c]=='K') k=k++;        /* Lysine*/
            if (seq[c]=='P') p=p++;        /* Proline */
            if (seq[c]=='S') s=s++;        /* Serine */
     }

   if (nseq !=0)
     cv = 15.43*((n+g+p+s)/(float)nseq)-29.56*fabs(((r+k-d-
e)/(float)nseq-0.03))-1.71;
    if (nseq !=0 && cv > -100.00)
      fprintf(out, "%s       %d       %4.2f    %2.0f   %2.0f   %2.0f
%2.0f   %2.0f   %2.0f   %2.0f   %2.0f   %2.0f\n",
          id, nseq, cv, r, n, d, e, g, k, p, s);
      printf(".");
  }
  printf("\nThe results can be viewed in the file named 'solout'.\n");
  fclose(out);
  fclose(in);
}
```

EntrezLight

The EntrezLight program is the same as SwissLight shown above except that it reads Entrez formatted files from the NCBI Entrez server at http://www.ncbi.nlm.nih.gov. To compile and execute this program, at the UNIX prompt type:

cc -o EntrezLight EntrezLight.c -lm

Then execute the program by typing:

EntrezLight

The program will prompt you for the name of the sequence input file. The C code for the EntrezLight program is:

```
#include <stdio.h>
#include <dirent.h>
#include <math.h>
#include <sys/types.h>
#include <sys/stat.h>
#include <string.h>

#define MAXBUFLEN 1024
#define MAXLINELEN 128
#define MAXSEQLEN 1024*128

main()
{

  char buf[MAXLINELEN];
  char seq[MAXSEQLEN];
  char fname[MAXSEQLEN];
  int mode, abode, nseq, c;
  FILE *in, *out;
  char *z;
  char id[7];
  float r,n,d,e,g,k,p,s,cv;

  printf("Enter the name of a GCG formatted file to read....\n\n");
  scanf("%s", fname);
  printf("\n");

  if ((in = fopen (fname, "r")) == NULL)
   {
      printf ("\nCannot find the file %s.\n\n", fname);
      printf("%s must be in this directory for program execution.\n\n",
fname);
```

```
    return 0;
  }

 out = fopen("solout", "w");
 fprintf (out, "Protein Name      aa       CV-CV'\n");
 fprintf (out, "---------------------------------\n");

 while (!feof(in))
 {
   fgets(buf, sizeof(buf), in);
   if (strncmp(buf, "LOCUS", 5)==0)
     {
       for (c=12; c < 19; c++)
           id[c-12] = buf[c];
           id[7] = 0;
     }
   mode=1;
   abode=1;
   nseq=0;
   while(!feof(in) && mode)
   {
     fgets(buf, sizeof(buf), in);
     if (strncmp(buf, "ORIGIN", 6)==0)
         mode=0;
   }

   while (!feof(in) && abode)
   {
     z = fgets(buf, sizeof(buf), in);
     if (strncmp(buf, "ORIGIN", 6)!=0)
        {
         for (z=buf;*z;z++)
       if (isalpha(*z)) seq[nseq++] = *z;
        }
     if (strncmp(buf, "//", 2)==0)
         abode=0;
   }

/* ------------- COUNT AMINO ACIDS ------------------------- */

   r=0;
   n=0;
   d=0;
   e=0;
   g=0;
   k=0;
   p=0;
   s=0;
   cv=0;

   for (c=0; c < nseq; c++)
     {
           if (seq[c]=='r') r=r++;          /* Arginine */
           if (seq[c]=='n') n=n++;        /* Asparagine */
```

```
            if (seq[c]=='d') d=d++;        /* Aspartic Acid */
            if (seq[c]=='e') e=e++;        /* Glutamic Acid */
            if (seq[c]=='g') g=g++;        /* Glycine */
            if (seq[c]=='k') k=k++;        /* Lysine*/
            if (seq[c]=='p') p=p++;        /* Proline */
            if (seq[c]=='s') s=s++;        /* Serine */
     }

   if (nseq !=0)
     cv = 15.43*((n+g+p+s)/(float)nseq)-29.56*fabs(((r+k-d-
e)/(float)nseq-0.03))-1.71;
    if (nseq !=0 && cv < 10)
      fprintf(out, "%s       %d       %4.2f    %2.0f   %2.0f   %2.0f
%2.0f   %2.0f   %2.0f   %2.0f   %2.0f\n",
          id, nseq, cv, r, n, d, e, g, k, p, s);
      printf(".");
  }
  printf("\nThe results can be viewed in the file named 'solout'.\n");
  fclose(out);
  fclose(in);
}
```

APPENDIX B

Laboratory Protocols

Storage of Unstable Reagents and Bacteria

Plates of bacteria can typically be stored at 4°C for up to 2 months before the cells die. For bacteria with specialized (e.g. very important) expression plasmids, always store liquid culture samples in 50% glycerol at -80°C, since there is a possibility of mutation, cell death, and plasmid loss when stored at 4°C. Ampicillin can be stored as a 100 mg/ml solution in dH_20 for up to two months at 4°C.

Plasmid Isolation from *E. coli* Hosts

The following protocol was taken from the Pharmacia Biotech FlexiPrep™ Kit.

Solution I: 100 mM Tris-HCI (pH 7.5), 10 mM EDTA, 400 μg/ml RNase I.

Solution II: 1 M NaOH, 5.3% (w/v) SDS (30 ml of a 5.3X concentrated solution).

Solution III: 3 M potassium, 5 M acetate solution.

Sephaglas™ FP: Sephaglas FP suspended in a buffered solution of guanidine-HCl [7 M guanidine-HCI, 50 mM Tris-HCl (pH 7.5), 10 mM EDTA].

Wash Buffer: 20 mM Tris-HCI (pH 7.5), 2 mM EDTA, 200 mM NaCl, 60 % (v/v) ethanol.

Cell Lysis

1. Transfer 1.5 ml of an overnight culture of *E. coli* to a microcentrifuge tube and centrifuge at full speed for 30 seconds to pellet the cells.

2. Remove the supernatant by aspiration without disturbing the cell pellet, but leaving it as dry as possible. If the cell pellet is small, steps 1 and 2 maybe repeated to increase the size of the cell pellet, but generally this is not required.
3. Resuspend the pellet in 200 µl of Solution I by vigorous vortexing.
4. Add 200 µl of Solution II and mix by inverting the tube several times. Note: The bacterial suspension should clear as cell lysis occurs.
5. Add 200 µl of Solution III and mix by inverting the tube several times. Centrifuge at full speed for 5 minutes at room temperature. Transfer the supernatant to a clean centrifuge tube and try to avoid transferring any precipitated cell debris (it contains enzymes that could possibly degrade the plasmid DNA during the prep).
6. Add 420 µl (0.7 volume) of ambient-temperature isopropanol to the supernatant and vortex to mix. Incubate for 10 minutes at room temperature. Two volumes of 100% ethanol can be substituted for isopropanol, but ethanol will precipitate more salt which may or may not effect subsequent binding of plasmid DNA to the Sephaglas™ glass beads.
7. Centrifuge at full speed for 10 minutes to pellet the plasmid DNA. Remove the supernatant by aspiration and invert the tube to drain.

DNA Purification

1. Suspend the Sephaglas FP slurry by shaking the bottle. Add 150 µl of the suspension to the DNA pellet and vortex gently for 1 minute to dissolve the pellet. (No further incubation is required.)

2. Centrifuge at full speed for 15 seconds in a microcentrifuge. Carefully remove the supernatant without disturbing the Sephaglas pellet.
3. Wash the Sephaglas pellet by adding 200 µl of Wash Buffer and vortexing gently to resuspend it. Centrifuge at full speed for 15 seconds in a microcentrifuge. Remove the supernatant without disturbing the pellet.
4. Add 300 µl of 70% ethanol to the Sephaglas pellet and vortex gently to resuspend it. Centrifuge at full speed for 15 seconds in a microfuge. Remove the supernatant without disturbing the pellet.
5. Invert the tube to dry. Allow the pellet to air dry for 10 minutes. Note: Do not dry under vacuum. Extensive drying will make it difficult to elute the DNA from the Sephaglas.
6. To elute the bound DNA from the Sephaglas, add 50 µl of TE buffer and vortex to resuspend the pellet. Incubate for 5 minutes at room temperature, vortexing occasionally to keep the Sephaglas in suspension. Sterile dH_20 (e.g. deionized water available from the faucet in the lab) can be used instead of TE, but TE buffer has been shown to give higher yields.
7. Centrifuge at full speed for 1 minute in a microcentrifuge. Transfer the supernatant to a clean microcentrifuge tube. If any Sephaglas is in the eluted sample, centrifuge the eluant and carefully transfer the supernatant to a clean tube, taking care to avoid the Sephaglas pellet at the bottom of the tube.

- **Note**: Do not store the eluted DNA in the presence of Sephaglas. Depending on the application requirements, the DNA may be further concentrated by ethanol precipitation.

Isolation of Lambda Page Genomic DNA

The following protocol was receieved from Frederick Blattner at the University of Wisconsin Laboratory of Genetics. Note: phage lysates containing the phage clones harboring the λ vector were provided by the Blattner laboratory. The phage lysate is a sample of *E. coli* LE392 culture which has been grown in the presence of the λ phage and has undergone lysis by the phage. It is used as the starting material for purifying λ vector DNA.

1. Plate out *E. coli* LE392 from -80°C storage on an NZC plate (Table B.1.A) and incubate overnight at 37°C.
2. Using a sterile pipet tip, pick a single colony into 4 ml of NZC broth (Table B.1.A). Grow for 4 to 6 hr at 37°C at 250 rpm. Make five serial dilutions (100 µl each) of the phage lysate using SM buffer (Table B.1.B). Mix the diluted phage with 200 µl of the LE392 culture. Make 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}, and 10^{-10} dilutions. Shake at 250 rpm and 37°C for 15 min to allow for the phage and cells to mix well. Mix the dilutions with 2.0 ml of NZC top agar (pre-cooled to 37°C) (Table B.1.A) and pour onto an NZC plate. Incubate overnight at 37°C. Start 4 ml of an overnight culture of LE392 in NZC broth at 37°C.
3. Using a sterile pipet tip, pick a single phage plaque into 300 µl of LE392 culture. Incubate for 15 min at 37°C at 250 rpm. Empty the contents into 50 ml of NZC broth and shake at 250 rpm overnight at 37°C.

A. NZC broth and plates.

Component	
NZ amine	10.47 g
NaCl	5.23 g
$MgSO_4$	1.26 g
Casamino acids	1.05 g
Agar (top agar only)	5.5 g
Agar (plates only)	11 g
Fill to 1.0 L w/ dH_20 and autoclave.	

B. SM buffer.

Component	
NaCl	4.64 g
2 M Tris-HCl pH 7.4	20 ml
$MgSO_4$	0.78 g
Gelatin 2% (w/v)	4 ml
Fill to 800 ml with dH_20 and autoclave	

C. 20%PEG/2MNaCl

Component	
SM Buffer	100 ml
PEG 8000	20 g
NaCl	11.7 g
Autclave to dissolve PEG.	

Table B.1. Buffer components for λ phage DNA isolation.

4. The following morning, check for lysis. If lysis has occurred, the culture will appear less turbid and cell debris will have agglomerated into large, stringy masses. If no lysis has occurred, transfer 10 ml of the culture to 40 ml of NZC broth and continue shaking until lysis occurrs. Once the culture has lysed, start to process the phage. It should not be left in the incubator or the refrigerator at this stage. When good lysis is achieved, add 1 ml of chloroform and continue shaking for another 10 minutes.
5. Decant into a centrifuge tube, leaving the chloroform behind, and spin at 5000 rpm at 4°C for 30 min.
6. Split the supernatant into two 25 ml aliquots and add 3.5 µl each of 10 mg/ml RNAseA and 10 mg/ml DNAseI for a final concentration of 1 µg/ml. Incubate shaking for 30 min.
7. Add an equal volume of 20% PEG/2MNaCl in SM buffer (Table B.1.C), and incubate on ice overnight in a 4°C refigerator.
8. The next day centrifuge at 5000 rpm at 4°C for 30 min.
9. Decant and discard the supernatant. Dry the pellet by inverting the tube on a paper towel and allow it to drain. Resuspend the pellet in 500 µl of SM buffer and transfer into a 1.5 ml Eppendorf tube.
10. Add 7 µl 10% SDS and 8.75 µl 0.4 M EDTA (pH 8.0) and incubate for 15 min at 68°C.
11. Add an equal volume (500 µl) of phenol (preheated to 55°C), mix the layers well and let sit for 5 minutes. Re-mix and spin for 5 min in the microfuge. Remove the

aqueous (top) phase into a clean Eppendorf tube.

12. Add an equal volume (500 µl) of chloroform, mix, and spin for 2.5 min in microfuge. Transfer the aqeuous (top) phase into a clean Eppendorf tube.
13. Add an equal volume of isopropanol, mix and incubate 5 min at room temperature. Centrifuge for 5 minutes and discard the supernatant.
14. Wash with 500 µl of cold 70% EtOH. Incubate 5 min at room temperature. Centrifuge for 5 min and discard the supernatant.
15. Invert the tube to dry over the air stream in a sterile hood for 10 min.
16. Resuspend in 50 µl of dH_20 or TE buffer. Run a gel to test to quantify the yield of DNA.

Agarose Gel Electrophoresis

The following protocol is for making a 1 % (w/v) agarose gel using SeaKem® LE agarose (FMC Bioproducts) using a BRL Life Technologies Horizon 58 electrophoretic cell with a total gel volume of 35 ml.

1. Assemble the electrohporetic cell.
2. Weigh out 0.35 g of agarose in a beaker. Add 35 ml of 1X TAE buffer (40 mM Tris-acetate, 1 mM EDTA). Microwave on high for 90 sec. Let the solution cool to 55°C and add 35 µl of 0.5 mg/ml ethidium bromide stock (mix the solution well to evenly distribute the ethidium bromide). Pour the gel into the cell and wait 20-30 min for solidification.
3. Add 5 ml of 50X TAE buffer and 250 µl of 0.5 mg/ml ethidium bromide to 245 ml of dH_20. Once the gel has solidified pour the TAE/EtBr solution over the gel until it

covers the gel by about 1 mm. The gel is now ready to be loaded.

4. To each DNA sample to be loaded, add 5 μl of gel loading buffer (GLB). The composition of GLB is 0.25% (w/v) bromophenol blue and 30% (v/v) glycerol in water. The maximum volume of DNA sample is 20 μl for each well of the 8 comb gel which has a total volume of 25 μl.
5. For the marker lane add 5 μl of λ BstEII digest to 5 μl of GLB and run in lane 1.

Ethanol/Ammonium Acetate Precipitation of DNA

(taken from laboratory course manual for MBIO 5822, Applications in Molecular Biology, Fall 1995)

1. Add 3 volumes of 1 M ammonium acetate (in 100% ethanol) to the DNA sample.
2. Incubate in the -80°C freezer for 10 min then pellet the DNA by spinning the tube for 15 min in the microfuge at maximum speed. Remove the supernatant with a micropipet.
3. Rinse the DNA in 200 μl of 80% ethanol (do not resuspend the pellet, just add the 80% ethanol and vortex briefly for 2 sec). Centrifuge at maximum speed for 2 min. Remove as much of the supernatant as possible with a micropipet.
4. Invert the tube and air dry for 10 min. Be sure that no ethanol droplets remain before proceeding. Incubate for a longer period or under an air flow if needed.
5. Resuspend the DNA in the desired volume of dH_20 water or TE buffer.

GeneClean™ Procedure (modified from BIO101)

1. Using a scalpel, cut out the desired DNA band from the agarose gel.
2. Centrifuge the gel piece in a 1.5 ml Eppendorf tube and estimate its volume.
3. Add 3 volumes (~ 700 µl) of 6 M sodium iodide and incubate for 5 min at 45 to 50°C. Make sure the gel piece dissolves. If not, add more sodium iodide solution and re-heat the solution until the gel dissolves.
4. Add 5 µl of Glassmilk, mix thoroughly by pipette, and incubate at room temperature for 5 min.
5. Centrifuge at high speed for 15 sec, discard the supernatant, and add 200 µl of New Wash (stored at -20°C). Briefly vortex the sample.
6. Repeat step 5.
7. Centrifuge the sample, discard the supernatant, and invert to dry for 10 min over the air intake grill in a sterile hood (this allows for faster drying). If no air stream is available, let the sample stay inverted for 20 min.
8. Elute the bound DNA with 30 µl of dH_20. Mix the suspension with a pipette to ensure that there is a maximum contact area between the glassbeads and the water. Incubate for 5 min at room temperature.
9. Centrifuge the sample for 1 minute and remove the supernatant to a new tube, being careful not to disturb the glass pellet.
10. More DNA may be recovered by performing a second 30 µl extraction of the glassbeads.

Phenol Extraction

1. Add one volume of 1:1 phenol:chloroform to the restriction digest and vortex for 10 sec. See Sambrook et al. (1989) for preparation and storage of phenol:chloroform mixtures.
2. Centrifuge at maximum speed for 2 min and remove the top layer to a new tube.
3. Ethanol precipitate the top layer and resuspend in the original volume of water.

***n*-Butanol/Isopropanol Precipitation of Ligation Reaction**

(modified from Thomas, 1994)

This protocol removes salts from the ligation reaction which will cause arcing during electroporation. It has been modified from Thomas et al. by the addition of a subsequent isopropanol precipitation.

1. Dilute the DNA sample (usually a 20 µl ligation reaction) to 50 µl with dH_20, and then add 500 µl of *n*-butanol and vortex for 5 sec.
2. Centrifuge at maximum speed (~10,000x*g*) for 10 min.
3. Pipet off the supernatant and dry the pellet (dry under vacuum or invert the tube under an air flow for 10 min).
4. Add 50 µl of dH_20 and vortex for 30 sec to redissolve the pellet (note: it is likely that the pellet is invisible at this point). Add 50 µl of isopropanol and vortex to mix. Incubate for 15 min at room temperature and then centrifuge for 15 min.

6. Air dry the pellet and resuspend in electrocompetent cells (described below) by pipet mixing thoroughly.

Preparation of Electrocompetent Cells.

This modified procedure is based on the protocol in the instruction manual for the Bio-Rad *E. coli* GenePulser™ (1995).

1. Inoculate a 100 ml LB-Broth which has been equilibrated to 37°C with a 1/100 volume (1 ml) of a fresh overnight culture.
2. Grow cells at 37°C with shaking at 250 rpm to an OD_{600} of 0.5.
3. To harvest, centrifuge cells in cold centrifuge bottles in a cold rotor at 4000 X g for 15 minutes at 4 °C.
4. From this point on, **keep the cells within 0 to 4 °C all the time**. Remove as much of the LB-supernatant as possible. Gently resuspend the cells in 100 ml of ice-cold 10% glycerol by slow pipetting with a 1 ml pipet volume. Take care not to lyse the cells with vigorous shaking or pipetting. Centrifuge as in step 3.
5. Resuspend in 50 ml of ice-cold 10% glycerol. Centrifuge as in step 3.
6. Resuspend to a final volume of 400 µl of ice-cold 10% glycerol. For highest transformation efficiencies, use these cells for transformation within an hour.
7. This suspension may be frozen in aliquots on dry ice, and stored at -70°C. The cells are good for at least 6 months under these conditions for transforming concentrated plasmid samples, but **should not** be used for transforming small quantities of DNA (e.g. recombinant DNA from ligation reactions).

Cell Transformation

Note: In addition to the samples to be transformed, perform positive and negative controls. Use ~0.05 μg of pKK223-3 (or your preferred vector) for the positive control and no DNA for the negative control.

1. Cool an electroporation cuvette on ice.
2. Add 40 μl of electrocompetent *E. coli* cells to the DNA sample. Mix well and transfer to the electroporation cuvette.
3. Shock the cells at 1.8 kV in a 0.1 cm gap cuvette.
4. **Immediately** add 1.0 ml of SOCg media (see Table B.2) preheated to 37°C. Use a pipette to mix the added SOCg media with the cells. Record the time constant. The 2% glucose ensures that the *tac* promoter in pKK223-3 is inhibited as much as possible.
5. Place the cells in a 1.5 ml Eppendorf tube, and then place the tube in a 1.0 L Erlenmyer flask. Shake the flask at 37°C for 1 hr. The Eppendorf tube will shake around haphazardly in the bottom of the flask, so shut the Eppendorf lid tight to prevent the sample from leaking.
6. Spin down the cells by centrfuging at maximum speed for 30 s. Resuspend the cells in 100 μl of SOCg media preheated to 37°C.
7. Mix the SOCg media and cell pellet by pipetting and spread onto a *tac* agar plate (LB-ampicillin (100 μg /ml) with 2 % (w/v) D-glucose).
8. Incubate the plate for 16 hr at 37°C and then screen the colonies.

Component	Amount
Tryptone	20 g
Yeast Extract	5 g
NaCl	0.5 g
KCl	0.19 g
Add dH_2O to 1.0 L and autoclave.	
Glucose	20 g
$MgCl_2 \cdot 6H_2O$	2 g

Note: glucose and $MgCl_2$ are added after sterilization.

Table B.2. Compsition of SOCg media for cell transformation.

Screening of Clones by Supercoiled Plasmid DNA Size

(modified from Unlap and Hu, 1995)

1. Pick out several colonies from a plate and grow each in 3 ml of pre-heated LB-broth with ampicillin for 4 h at 37°C.
2. Perform the cell lysis portion FlexiPrep procedure on 1.5 ml of each culture (outlined above) with the following modifications: a) incubate soluble lysate with two volumes of ethanol (instead of isopropanol) for 1 minute, b) centrifuge DNA pellet for only 5 min.
3. Discard the supernatant and invert the tube for 10 min of drying. Be sure that the pellet is completely dry (if the pellet is not dry at this point, the ethanol will cause the DNA sample to flow upwards, out of the well during loading of the agarose gel, destroying the sample). Resuspend the ethanol precipitated DNA pellet in 15 µl of dH_2O, add 5 µl of gel loading buffer, and run on an agarose gel along with the original uncut vector (pKK223-3) to check for inserts.

Screening Clones by Comparing OD_{600} Measurements of Induced and Uninduced Cultures (Davis and Harrison, 1998)

1. Pick 30 colonies into 1 ml of LB-Amp (100 µg/ml) with 1% glucose in a 1.5 ml Eppendorf tube. Place the tubes in a 500 ml Erlenmyer flask and shake at 250 rpm at 37°C for 2 hours.
2. Dilute each culture with 4 ml of additional LB-Amp-glucose media (pre-dispensed into a 13 X 100 mm glass test tube) and continue incubation **at 280 rpm** to an OD_{600}

of about 0.5. **Note:** different recombinants will reach an OD_{600} of 0.5 at different times.

3. Add an additional 4 ml of LB-Amp-glucose and mix well by pouring the mixture between two test tubes approximtely 7 times. It is **very important** that the culture be well mixed at this point. Once the diluted culture is well-mixed take a 100 µl sample and combine it with 100 µl of glycerol in a 1.5 ml Eppendorf tube: mix it well by pipet and store it at -80°C.
4. Split the remaining culture into two separate 13 X 100 mm test tubes (about 4.5 ml each) and induce one culture with 1mM IPTG by adding 90 µl of a 50 mM IPTG solution which also contains 4.5 µg/ml of ampicillin (stored in 1 ml aliquots at -20°C). Note: The additional ampicillin is included on the advice of Dr. Thomas Baldwin at the Texas A&M Dept. of Biochemistry and Biophysics. Baldwin has found that when the $OD_{600} = 0.1$, the ampicillin concentration is usually zero.
5. Continue the incubation for 3 hours.
6. At the end of the three hour incubation, measure the OD_{600} of each culture (induced and uninduced) and then centrifuge the samples at 1000 x *g* for 10 min. Discard the supernatant and store the bacterial pellets at -20°C until used for analysis by SDS-PAGE.

Rapid Screening of Clones by Dot-Immunoblotting

Note: This method requires an antibody to the target protein.

1. Pick 30 colonies into 1 ml of media (LB with 1% glucose and 100 μg/ml ampicllin) and grow them for 2 hours in a 1.5 ml Eppendorf tube at 37°C with shaking at 250 rpm.
2. For each colony, mix 50 μl of cells with 50 μl of glycerol and store the mixture at -80°C for future use. In the remaining culture (~950 μl) induce the cells by adding IPTG to 1 mM (add an additional 100 μg/ml ampicillin with the IPTG). Continue incubation for 1 hr.
3. Centrifuge the cells in a microcentrifuge for 1 min at maximum speed.
4. Cell lysis: resuspend the cells in 100 μl of 10% SDS with 50 mM Tris at pH 7.0 and heat at 100°C for 2 min.
5. Allow the mixture to cool. The liquid should be clarified at this point.
6. Cut out a piece of nitrocellulose paper and mark the edges to make a grid. Place 2 μl of lysate from each colony in the grid and allow the membrane to air dry for 30 min.
7. Follow the western blot procedure outlined below beginning with step 6. This step immediately follows the electroblot transfer and begins with a TBS wash of the membrane.
8. Run an SDS-PAGE and western blot of positive colonies to confirm the size of the expressed protein.

Fractionation of Cells (Soluble/Insoluble Proteins)

Note: this protocol is optimized for expression cultures done in 4 ml of LB-Media with 100 μg/ml ampicillin and 1% (w/v) glucose. Colonies from a plate should be grown in 1 ml of media for 2 hours (37°C and 250 rpm) and then treated as follows: for a typical solubility analysis, inoculate three 13 X 100 mm test tubes containing 4 ml LB-amp-glucose media each with 300 μl of the 2 hr culture. At an OD_{600} of 0.4 induce two of the tubes (one tube is used to observe the whole cell induced protein and the other is used for cell fractionation). Do not induce the third tube. At three hours post-induction, pellet the cells at 1000 x *g* for 10 min and freeze the pellets at -20°C.

1. Resuspend a frozen cell pellet from an induced culture in 10 ml of sonication buffer (50 mM NaCl and 1 mM EDTA pH 8.0) in a small 20 ml beaker, measure the OD_{600}, and place it on an ice water bath.
2. Place the sonication horn into the solution as far down as possible without touching the bottom of the beaker.
3. Sonicate 30 sec at 90 W (e.g. 30% power, the power setting dial should be at level 4) and then allow the solution to cool for 30 sec on ice. Repeat 3 more times for a total sonication time of 2 min (2 min sonication + 2 min total cooling time = 4 total minutes for the procedure). Measure the OD_{600} of the lysate: it should be less than the original OD_{600}.
4. Centrifuge the lysate at 12,000 x *g* for 30 min at 4°C.
5. At this point, take care **not to disturb** the inclusion body pellet. Carefully take the

top 5 ml of the supernatant by drawing it into the pipet just under the water surface as far away from the insoluble pellet as possible and place in a labeled freeze drying flask. **Immediately** freeze the supernatant in the -80°C refrigerator. Discard the remaining supernatant and freeze the inclusion body pellet at -20°C.

6. Turn on the cold trap and the freeze dryer and place about 300 ml of 95% ethanol in the cold trap chamber. Make sure the freeze trap valve is open so that the vacuum pump can draw the air from the freeze dryer chamber. Allow about 2 hours for the cold trap the cool. The vacuum pump will automatically turn on when the trap is ready.
7. Place the frozen lysate supernatant samples on the freeze dryer overnight for sublimation.
8. Store the freeze dried solid at -20°C until SDS-PAGE analysis.
9. For SDS-PAGE analysis resuspend the freeze dried supernatant in about 400 µl of SLB (see SDS-PAGE Analysis of Proteins for composition) and the inclusion body pellet, the uninduced pellet, and the induced pellet in 800 µl of SLB.

SDS-PAGE Analysis of Proteins

Using the Bio-Rad Mini-PROTEAN®II Cell

1. Assemble the glass plates on the gel casting stand as described in the instructions for the Bio-Rad Mini-PROTEAN II Cell.
2. Mix the gel components together in the amounts and order described in Table B.3.A for either the 12% or 8% separating gels. Use a 12% gel for proteins ranging in size

A.

Mini-Gels

[glass plates (8.0 cm W) X (7.3 cm L) X (0.75 mm spacer)]

Component	Stacking Gel (4%)	Separating Gel (12%)	Separating Gel (8%)
dH_2O	7.35 ml	4.0 ml	5.6 ml
1.5 M Tris-HCl pH 8.8	-	3.0 ml	3.0 ml
1 M Tris-HCl pH 6.8	1.25 ml	-	-
10 % (w/v) SDS	100 µl	120 µl	120 µl
Acrylamide (29%) Bis (1%)	1.33 ml	4.8 ml	3.2 ml
Ammonium Persulfate (10%)	50 µl	60 µl	60 µl
TEMED	10 µl	12 µl	12 µl
Total	10 ml	12 ml	12 ml

B.

Large Format Gels

[glass plates (16 cm W) X (16 cm L) X (1.5 mm spacer)]

Component	Stacking Gel (5%)	Separating Gel (12%)	Separating Gel (10%)	Separating Gel (8%)
dH_2O	13.6 ml	9.9 ml	11.9 ml	13.9 ml
1.5 M Tris-HCl pH 8.8	-	7.5 ml	7.5 ml	7.5 ml
1 M Tris-HCl pH 6.8	2.5 ml	-	-	-
10 % (w/v) SDS	200 µl	300 µl	300 µl	300 µl
Acrylamide/Bis (30 %)	3.4 ml	12.0 ml	10.0 ml	8.0 ml
Ammonium Persulfate (10%)	200 µl	300 µl	300 µl	300 µl
TEMED	20 µl	12 µl	12 µl	18 µl
Total	20 ml	30 ml	30 ml	30 ml

Table B.3. Reagents required for the casting of mini-SDS-PAGE gels and large format SDS-PAGE gels.

from 16 to 68 kDa and an 8% gel for proteins ranging from 36 to 94 kDa.

3. After adding the TEMED to the separating gel solution, mix thoroughly by pipet and immediately fill the glass plates to 5 cm from the bottom. Then immediately add 1 ml of isopropanol on top of the gel to prevent oxygen from inhibiting the polymerization. Wait 15 to 20 min for solidification. Note: 10% ammonium persulfate can be stored in aliquots at -20°C, the acrylamide solution can be stored at 4°C, and all other components at room temperature.
4. Pour off the isopropanol and rinse three times with dH_2O to remove any residual isopropanol. Mix the components for the 4% stacking gel and pour on top of the separating gel to within 1 mm of the top of the lowest glass plate. Insert the well-comb and wait 15 to 20 min for solidification.
5. Prepare the sample loading buffer (SLB) by making a 1:20 dilution of β-mercaptoethanol (BME) into SDS-PAGE loading buffer (SDS-LB). Typically, add 9.5 ml of SDS-LB to 500µl of BME. The composition of SDS-LB is shown inTable B.4.A.
6. For a 4 ml culture with an approximate OD_{600} of 0.8, add 800 µl of SLB to the centrifuged cell pellet. This value may vary (600 to 1000 µl) for different expression cultures. Note: SDS-LB should not be stored with BME since BME will lose reducing potential over time.
7. Vortex the cell pellet so that it becomes evenly distributed in the SLB. Heat the sample at 100°C for 2 min. Allow the sample to cool for at least one minute before loading.

A. SDS-PAGE loading buffer (SDS-LB)

Component	Volume (ml)
dH_2O	59
1 M Tris-HCl pH 6.8	6
10 % (w/v) SDS	20
0.1 % (w/v) Bromophenol blue	5
Glycerol	10
Total	100

B. 5X Running buffer (pH 8.3)

Component	Mass (g)
Tris base	15
Glycine	72
SDS	5
Add dH_2O to 1.0 L	

Table B.4. SDS-PAGE loading buffer and 5X running buffer compostions.

8. Assemble the solidified gels onto the upper buffer chamber (see Bio-Rad instructions). Dilute 100 ml of 5X running buffer (see Table B.4.B) with 400 ml of dH2O and fill the lower chamber with ~500ml of 1X running buffer.
9. Fill the upper chamber with 1X running buffer until the top of the wells are covered by about 1 cm. Using the upper chamber running buffer, use a 100 µl pipet to flush out any loose acrylamide particles from the wells.
10. Using gel loading tips (Perfector Scientific Cat. No. 1590), load 10 µl (15 µl for large format gels) of each protein sample into the wells. For mini gels, run the gel at a constant voltage of 150 V for about 1 hr or until the dye front exits the bottom of the gels. The running conditions for the large format gels are 25 mA (constant amperage) for the stacking gel and 35 mA for the separating gel. Total run time was 4.5 to 5 hr for the large format gels.
11. Remove the upper buffer chamber and disassemble the glass plates so that the gel remains attached to only one of the glass plates in which it was cast. Cut off the stacking gel with a scalpel and discard it. Mark the gel by cutting off a piece of a selected corner so that it can be identified later. Float the glass plate gel-side down into a tray containing ~200 ml of stain [45% (v/v) dH_20, 45% (v/v) methanol, 10% (v/v) acetic acid, 0.25% (w/v) Coomassie Brilliant Blue R250]. Pick up the glass plate, and the surface tension of the stain will remove the gel from the plate into the stain. Repeat for any additional gels and stain for 1 hr with gentle shaking (~2 rpm is sufficient). If shaking is not possible, an overnight incubation will be sufficient to stain the gel.

12. While holding down the gels in the tray, pour the stain back into its container (it can be re-used several times) and rinse the tray 3 times with tap water. Add ~200 ml of destain (the same as stain without Coomassie Blue). Protein bands are then visible immediately on a light table. Incubate with destain until the desired resolution of protein bands is reached. It maybe necessary to wash twice in 200 ml of destain for about 1 hr each. Pour any used destain into a waste container and save it. It can be regenerated by vacuum filtration through activated carbon.

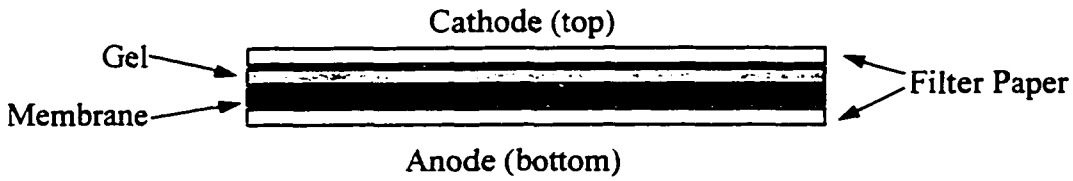
Western Blotting

Note: The following procedure is optimized for SDS-PAGE mini gels (gel thickness of 0.75 mm). It was developed using the BioRad ImmunStar Chemiluminescent Kit. The following reagents were supplied with the kit: non-fat dry milk (powdered), mouse or rabbit secondary anti-body conjugated to alkaline phosphatase, the substrate (a phosphorylated molecule which releases light upon de-phosphorylation), and enhancer solution (unknown composition which speeds up the chemiluminescent reaction).

1. Run an SDS-PAGE gel of the protein sample to be analyzed. Use pre-stained standards (BioRad) for molecular weight markers.
2. Cut one piece of nitrocellulose membrane (NitroBind, Micron Separations Inc.) and two pieces of filter paper (Trans-Blot Paper, BioRad) to the size of the SDS-PAGE gel (about 5 X 8 cm)
3. Equilibrate the gel and the nitrocellulose membrane separately in 30 ml of Towbin transfer buffer (25 mM Tris, 192 mM glycine, 20% (v/v) methanol, pH 8.3) for 15

min with gentle shaking.

4. Briefly soak the two pieces of filter paper in Towbin buffer and assemble a sandwich with the gel and membrane as shown below:



5. Place the gel sandwich in the Trans-Blot Electrophoretic Transfer Cell (BioRad). Using a 10 ml pipet, roll out any air bubbles in the sandwich. Wipe up any excess liquid on the transfer cell surface and close the cell. Run the transfer at 10 V for 25 min. Typically the amperage will start at 150 mA and end at about 50 mA. Do not exceed 200 mA per mini-gel. The pre-stained markers will indicate the efficiency of the transfer of the proteins from the SDS-PAGE gel to the nitrocellulose membrane.
6. Wash the membrane in 30 ml of TBS (20 mM Tris, 500 mM NaCl, pH 7.5) for 5 min with gentle shaking. Repeat wash once.
7. Pour off wash solution and immerse the membrane in 30 ml of blocking solution (TBS with 0.2% (w/v) non-fat dry milk). Incubate for 30 min with gentle shaking.
8. Pour off the blocking solution and add 30 ml of TTBS (TBS with 0.1% (v/v) Tween 20). Incubate for 5 min with gentle shaking.
9. Decant the TTBS and add 30 ml of antibody buffer (TTBS with 0.2% (w/v) non-fat dry milk). Add the primary antibody at the appropriate dilution (usually 1:3000) and incubate for one hour with gentle shaking.
10. Decant the antibody buffer and wash twice for 5 min with 30 ml of TTBS.

11. Decant the TTBS and add 30 ml of antibody buffer. Add 2.5 μl (1:12,000 dilution) of the secondary antibody and incubate for 30 min with gentle shaking.
12. Decant the antibody buffer and wash three times for 5 min each with TTBS.
13. Mix 500 μl of enhancer with 10 ml of substrate solution. Decant the TTBS and place the membrane on a piece of Saran wrap. Pour just enough substrate/enhancer solution on the membrane to cover the surface. Incubate for 5 min. Pour off the substrate solution and enclose the membrane entirely in Saran wrap. Make sure the Saran wrap is flat against the surface to be developed.
14. In a dark room, develop the blot using Kodak BioMax MR x-ray film by placing the x-ray film directly on the membrane enclosed in Saran wrap. Exposure time may vary from 1 to 10 min. Develop the film by placing in developer for 30 seconds with gentle agitation. Fix the image by placing the film in fixer for 30 seconds with gentle agitation. Rinse the film with water and hang to dry.

APPENDIX C

Primers and Consensus Sequences

Primers for PCR of Carrier and Target Genes

The general design and use of these primers is described in the methods section. The particular restriction enzyme sequence (or other relevant sequence) within each primer is also indicated.

NusA 5' Primer (*EcoR*I - Start):
5' - Cg / TTA / gCC / gAA / TTC / ATg / AAC / AAA / gAA / ATT / TTg / gC - 3'

NusA 5' Primer (*Eco*RI - Start - His_6):
5' - Cg / TTA / gCC / gAA / TTC / ATg / CAT / CAC / CAT / CAC / CAT / CAC / AAC / AAA / gAA / ATT / TTg / gCT / gTA / gTT / gAA / gC - 3'

NusA 3' Primer (*Age*I):
5' - Cg / CgC / ATT / ACC / ggT / CgC / TTC / gTC / ACC / gAA / CCA / gC - 3'

GrpE 5' Primer (*EcoR*I - Start):
5' - gC / ATT / gAC / gAA / TTC / ATg / AgT / AgT / AAA / gAA / CAg / AAA / ACg - 3'

GrpE 3' Primer (*Age*I):
5' - Cg / CTA / ACg / ACC / ggT / AgC / TTT / TgC / TTT / CgC / TAC / AgT / AAC - 3'

BFR 5' Primer (*EcoR*I - Start):
5' - Cg / TTA / gCC / gAA / TTC / ATg / AAA / ggT / gAT / ACT / AAA / gTT / ATA / AAT / TAT / C - 3'

BFR 3' Primer (*Age*I):
5' - Cg / CgC / ATT / ACC / ggT / ACC / TTC / TTC / gCg / gAT / CTg / Tg - 3'

YjgD 5' Primer (*EcoR*I - Start):
5' - Cg / TTA / gCC / gAA / TTC / ATg / gCA / AAC / CCg / gAA / CAA / CTg / gAA / gAA / CAg / C - 3'

YjgD 5' Primer (*Age*I):
5' - Cg / TTA / gCC / ACC / ggT / ATg / gCA / AAC / CCg / gAA / CAA / CTg / gAA / gAA / CAg / C - 3'

YjgD 3' Primer (*Age*I):
5' - Cg / CgC / ATT / ACC / ggT / gTg / gCg / AAC / TCC / gTC / ATC / gTC / TTC / ATC / g - 3'

Thioredoxin 5' Primer (*Eco*RI - Start):
5' - Cg / TTA / gCC / gAA / TTC / ATg / AgC / gAT / AAA / ATT / ATT / CAC / CTg / ACT / gAC / g - 3'

Thioredoxin 3' Primer (*Age*I):
5' - Cg / CgC / ATT / ACC / ggT / CgC / CAg / gTT / AgC / gTC / gAg / gAA / CTC - 3'

hIL-3 5' Primer (*Age*I - Factor Xa):
5' - Cg / CgC / ATT / ACC / ggT /ATC / gAA / ggT / CgA / gCT / CCC / ATg / ACC / CAg / ACA / ACg - 3'

hIL-3 3' Primer (*Hind*III - Stop):
5' - Cg / ATT / CgC / AAg / CTT / TCA / AAA / gAT / CgC / gAg / gCT / CAA / Ag - 3'

bgH 5' Primer (*Age*I - Factor Xa):
5' - Cg / CgC / ATT / ACC / ggT / ATC / gAA / ggT / CgA / TTC / CCA / gCC / ATg / TCC / TTg / TCC - 3'

bgH 3' Primer (*Hind*III - Stop):
5' - Cg / ATT / CgC / AAg / CTT / TCA / gAA / ggC / gCA / gCT / ggC / CTC / CCC - 3'

hIFN-γ 5' Primer (*Age*I - Factor Xa):
5' - Cg / CgC / ATT / ACC / ggT / ATC / gAA / ggT / CgA / TgT / TAC / TgC / CAg / gAC / CCA / TAT / g - 3'

hIFN-γ 3' Primer (*Hind*III - Stop):
5' - Cg / ATT / CgC / AAg / CTT / TTA / CTg / ggA / TgC / TCT / TCg / ACC / TCg - 3'

Primers for the Sequencing of Fusion genes

Below, in Tables C.1 through C.8, are the primers used for the sequencing of the various fusion genes. The primers are shown as postitioned in the fusion genes to illustrate the over all sequencing approach. All primer sequences are underlined, and all restriction sites are in uppercase bold letters. The forward and reverse directions of the primers are indicated by arrows. The factor Xa sequence following the *Age*I restriction sites is also shown in uppercase bold letters.

Table C.1. Sequencing Primers for the NusA/hIL-3 Fusion Gene.

```
                                        PKPROM-->
   1  ccaatgcttc tggcgtcagg cagccatcgg aagctgtggt atggctgtgc
  51  aggtcgtaaa tcactgcata attcgtgtcg ctcaaggcgc actcccgttc
 101  tggataatgt tttttgcgcc gacatcataa cggttctggc aaatattctg
                                        pKKECO-->
 151  aaatgagctg ttgacaatta atcatcggct cgtataatgt gtggaattgt
 201  gagcggataa caatttcaca caggaaacag AATTCatgaa caaagaaatt
 251  ttggctgtag ttgaagccgt atccaatgaa aaggcgctac ctcgcgagaa
                                        NUS1-->
 301  gattttcgaa gcattggaaa gcgcgctggc gacagcaaca aagaaaaaat
 351  atgaacaaga gatcgacgtc cgcgtacaga tcgatcgcaa aagcggtgat
 401  tttgacactt tccgtcgctg gttagttgtt gatgaagtca cccagccgac
                                        NUS2-->
 451  caaggaaatc acccttgaag ccgcacgtta tgaagatgaa agcctgaacc
 501  tgggcgatta cgttgaagat cagattgagt ctgttacctt tgaccgtatc
 551  actacccaga cggcaaaaca ggttatcgtg cagaaagtgc gtgaagccga
                                        NUS3-->
 601  acgtgcgatg gtggttgatc agttccgtga acacgaaggt gaaatcatca
 651  ccggcgtggt gaaaaaagta aaccgcgaca acatctctct ggatctgggc
 701  aacaacgctg aagccgtgat cctgcgcgaa gatagctgcc gcgtgaaaac
                                        NUS4-->
 751  tttcgccctg gcgaccgcgt tcgtggcgtg ctctattccg ttcgcccgga
 801  acggtggcgc gcaactgttc gtcactcgtt ccaagccgga aatgctgatc
 851  gaactgttcc gtattgaagt gccagaaatc ggcgaagaag tgattgaaat
                                        NUS5-->
 901  taaagcagcg gctcgcgatc cgggttctcg tgcgaaaatc gcggtgaaaa
 951  ccaacgataa acgtatcgat ccggtaggtg cttgcgtagg tatgcgtggc
1001  gcgcgtgttc aggcggtgtc tactgaactg ggtggcgagc gtatcgatat
                                        NUS6-->
1051  cgtcctgtgg gatgataacc cggcgcagtt cgtgattaac gcaatggcac
1101  cggcagacgt tgcttctatc gtggtggatg aagataaaca caccatggat
1151  atcgccgttg aagccggtaa cttggcgcag gcgattggcc gtaacggtca
                                        NUS7-->
1201  gaacgtgcgt ctggcttcgc agctgagcgg ttgggaactc aacgtgatga
1251  ccgttgacga cctgcaggct aagcatcagg cggaagcgca cgcagcgatc
1301  gacaccttca ccaaatatct cgacatcgac gaagacttcg cgactgttct
                                        NUS8-->
1351  ggtagaagaa ggcttctcga cgctggaaga attggcctat gtgccgatga
1401  aagagctgtt ggaaatcgaa ggccttgatg agccgaccgt tgaagcactg
1451  cgcgagcgtg ctaaaaatgc actggccacc attgcacagg cccaggaaga
                                        NUS9-->
1501  aagcctcggt gataacaaac cggctgacga tctgctgaac cttgaagggg
1551  tagatcgtga tttggcattc aaactggccg cccgtggcgt ttgtacgctg
1601  gaagatctcg ccgaacaggg cattgatgat ctggctgata tcgaagggtt
                                        NUS10-->
1651  gaccgacgaa aaagccggag cactgattat ggctgcccgt aatatttgct
1701  ggttcggtga cgaagcgACC GGTATCGAAG GTCGAgctcc catgacccag
1751  acaacgccct tgaagacaag ctgggttaac tgctctaaca tgatcgatga
                                        IL1--->
1801  aattataaca cacttaaagc agccaccttt gcctttgctg gacttcaaca
```

Table C.1. *(continued).*

```
1851  acctcaatgg ggaagaccaa gacattctga tggaaaataa ccttcgaagg
1901  ccaaacctgg aggcattcaa cagggctgtc aagagtttac agaacgcatc
                                      IL2--->
1951  agcaattgag agcattctta aaaatctcct gccatgtctg cccctggcca
2001  cggccgcacc cacgcgacat ccaatccata tcaaggacgg tgactggaat
2051  gaattccgga ggaaactgac gttctatctg aaaacccttg agaatgcgca
2101  ggctcaacag acgactttga gcctcgcgat cttttgaAAG CTTctgtttt
2151  ggcggatgag agaagatttt cagcctgata cagattaaat cagaacgcag
```

Table C.2. Sequencing Primers for the BFR/hIL-3 Fusion Gene.

```
                                          PKPROM-->
   1  ccaatgcttc tggcgtcagg cagccatcgg aagctgtggt atggctgtgc
  51  aggtcgtaaa tcactgcata attcgtgtcg ctcaaggcgc actcccgttc
 101  tggataatgt tttttgcgcc gacatcataa cggttctggc aaatattctg
                        PKPROM2-->
 151  aaatgagctg ttgacaatta atcatcggct cgtataatgt gtggaattgt
 201  gagcggataa caatttcaca caggaaacaG AATTCatgaa aggtgatact
 251  aaagttataa attatctcaa caaactgttg ggaaatgagc ttgtcgcaat
                   BFR1-->
 301  caatcagtac tttctccatg cccgaatgtt taaaaactgg ggtctcaaac
 351  gtctcaatga tgtggagtat catgaatcca ttgatgagat gaaacacgcc
 401  gatcgttata ttgagcgcat tctttttctg gaaggtcttc caaacttaca
                   BFR2-->
 451  ggacctgggc aaactgaaca ttggtgaaga tgttgaggaa atgctgcgtt
 501  ctgatctggc acttgagctg gatggcgcga agaatttgcg tgaggcaatt
 551  ggttatgccg atagcgttca tgattacgtc agccgcgata tgatgataga
                  BFR3-->
 601  aattttgcgt gatgaagaag gccatatcga ctggctggaa acggaacttg
 651  atctgattca gaagatgggc ctgcaaaatt atctgcaagc acagatccgc
 701  gaagaaggtA CCGGTATCGA AGGTCGAgct cccatgaccc agacaacgcc
 751  cttgaagaca agctgggtta actgctctaa catgatcgat gaaattataa
                              IL1-->
 801  cacacttaaa gcagccacct ttgcctttgc tggacttcaa caacctcaat
 851  ggggaagacc aagacattct gatggaaaat aaccttcgaa ggccaaacct
 901  ggaggcattc aacagggctg tcaagagttt acagaacgca tcagcaattg
                               IL2-->
 951  agagcattct taaaaatctc ctgccatgtc tgcccctggc cacggccgca
1001  cccacgcgac atccaatcca tatcaaggac ggtgactgga atgaattccg
1051  gaggaaactg acgttctatc tgaaaaccct tgagaatgcg caggctcaac
1101  agacgacttt gagcctcgcg atcttttgaA AGCTTctgtt ttggcggatg
                              <--PKK2
1151  agagaagatt ttcagcctga tacagattaa atcagaacgc agaagcggtc
```

Table C.3. Sequencing Primers for the GrpE/hIL-3 Fusion Gene.

```
                                          PKPROM-->
   1  ccaatgcttc tggcgtcagg cagccatcgg aagctgtggt atggctgtgc
  51  aggtcgtaaa tcactgcata attcgtgtcg ctcaaggcgc actcccgttc
 101  tggataatgt tttttgcgcc gacatcataa cggttctggc aaatattctg
                                          PKKECO-->
 151  aaatgagctg ttgacaatta atcatcggct cgtataatgt gtggaattgt
 201  gagcggataa caatttcaca caggaaacaG AATTCatgag tagtaaagaa
 251  cagaaaacgc ctgaggggca agccccggaa gaaattatca tggatcagca
                                            ga-->
 301  cgaagagatt gaggcagttg agccagaagc ttctgctgag caggtggatc
 351  cgcgcgatga aaaagttgcg aatctcgaag ctcagctggc tgaagcccag
 401  acccgtgaac gtgacggcat tttgcgtgta aaagccgaaa tggaaaacct
                                          gb-->
 451  gcgtcgtcgt actgaactgg atattgaaaa agcccacaaa ttcgcgctgg
 501  agaaattcat caacgaattg ctgccggtga ttgatagcct ggatcgtgcg
 551  ctggaagtgg ctgataaagc taacccggat atgtctgcga tggttgaagg
                                                grpe2-->
 601  cattgagctg acgctgaagt cgatgctgga tgttgtgcgt aagtttggcg
 651  ttgaagtgat cgccgaaact aacgtcccac tggacccgaa tgtgcatcag
 701  gccatcgcaa tggtggaatc tgatgacgtt gcgccaggta acgtactggg
                                            gc-->
 751  cattatgcag aagggttata cgctgaatgg tcgtacgatt cgtgcggcga
 801  tggttactgt agcgaaagca aaagctACCG GTATCGAAGG TCGAgctccc
 851  atgacccaga caacgccctt gaagacaagc tgggttaact gctctaacat
                                              IL1-->
 901  gatcgatgaa attataacac acttaaagca gccacctttg cctttgctgg
 951  acttcaacaa cctcaatggg gaagaccaag acattctgat ggaaaataac
1001  cttcgaaggc caaacctgga ggcattcaac agggctgtca agagtttaca
                                              IL2-->
1051  gaacgcatca gcaattgaga gcattcttaa aaatctcctg ccatgtctgc
1101  ccctggccac ggccgcaccc acgcgacatc caatccatat caaggacggt
1151  gactggaatg aattccggag gaaactgacg ttctatctga aaacccttga
1201  gaatgcgcag gctcaacaga cgactttgag cctcgcgatc ttttgaAAGC
                                                <--PKK2
1251  TTctgttttg gcggatgaga gaagattttc agcctgatac agattaaatc
1301  agaacgcaga agcggtctga taaaacagaa tttgcctggc ggcagtagcg
1351  cggtggtccc acctgacccc atgccgaact cagaagtgaa acgccgtagc
```

Table C.4. Sequencing Primers for the Thioredoxin/hIL-3 Fusion Gene.

```
                                        PKPROM-->
   1  ccaatgcttc tggcgtcagg cagccatcgg aagctgtggt atggctgtgc
  51  aggtcgtaaa tcactgcata attcgtgtcg ctcaaggcgc actcccgttc
 101  tggataatgt tttttgcgcc gacatcataa cggttctggc aaatattctg
                                        PKKECO-->
 151  aaatgagctg ttgacaatta atcatcggct cgtataatgt gtggaattgt
 201  gagcggataa caatttcaca caggaaacaG AATTCatgag cgataaaatt
 251  attcacctga ctgacgacag ttttgacacg gatgtactca aagcggacgg
                                        THIO1-->
 301  ggcgatcctc gtcgatttct gggcagagtg gtgcggtccg tgcaaaatga
 351  tcgccccgat tctggatgaa atcgctgacg aatatcaggg caaactgacc
 401  gttgcaaaac tgaacatcga tcaaaaccct ggcactgcgc cgaaatatgg
 451  catccgtggt atcccgactc tgctgctgtt caaaaacggt gaagtggcgg
        THIO2-->
 501  caaccaaagt gggtgcactg tctaaaggtc agttgaaaga gttcctcgac
 551  gctaacctgg cgACCGGTAT CGAAGGTCGA gctcccatga cccagacaac
 601  gcccttgaag acaagctggg ttaactgctc taacatgatc gatgaaatta
                                     IL1-->
 651  taacacactt aaagcagcca cctttgcctt tgctggactt caacaacctc
 701  aatggggaag accaagacat tctgatggaa ataaccttc gaaggccaaa
 751  cctggaggca ttcaacaggg ctgtcaagag tttacagaac gcatcagcaa
                                     IL2-->
 801  ttgagagcat tcttaaaaat ctcctgccat gtctgcccct ggccacggcc
 851  gcacccacgc gacatccaat ccatatcaag gacggtgact ggaatgaatt
 901  ccggaggaaa ctgacgttct atctgaaaac ccttgagaat gcgcaggctc
 951  aacagacgac tttgagcctc gcgatctttt gaAAGCTTct gttttggcgg
                                  <--PKK2
1001  atgagagaag attttcagcc tgatacagat taaatcagaa cgcagaagcg
1051  gtctgataaa acagaatttg cctggcggca gtagcgcggt ggtcccacct
1101  gaccccatgc cgaactcaga agtgaaacgc cgtagcgccg atggtagtgt
```

Table C.5. Sequencing Primers for the YjgD/hIL-3 Fusion Gene.

```
                                          PKPROM-->
   1  ccaatgcttc tggcgtcagg cagccatcgg aagctgtggt atggctgtgc
  51  aggtcgtaaa tcactgcata attcgtgtcg ctcaaggcgc actcccgttc
 101  tggataatgt tttttgcgcc gacatcataa cggttctggc aaatattctg
                          PKPROM2-->
 151  aaatgagctg ttgacaatta atcatcggct cgtataatgt gtggaattgt
 201  gagcggataa caatttcaca caggaaacaG AATTCatggc aaacccggaa
 251  caactggaag aacagcgtga agaaacacgt ttgattattg aagaattact
                     YJGD1-->
 301  ggaagatggc agcgatccgg acgcgctgta caccatcgaa caccatcttt
 351  ccgcagacga tctggaaacc ctggaaaaag cagcagttga agcgtttaaa
 401  ctcggttacg aagtgaccga tccagaagag ctggaagttg aagatggtga
                           YJGD2-->
 451  tatcgtgatt tgctgcgaca tcctcagcga gtgcgcgttg aatgccgatc
 501  tgatcgatgc ccaggttgaa caactgatga cgctggcaga gaaatttgac
 551  gttgagtacg acggatgggg cacttacttt gaagatccta acggcgaaga
                      YJGD3-->
 601  tggcgacgat gaagattttg tcgatgaaga cgatgacgga gttcgccact
 651  ACCGGTATCG AAGGTCGAgc tcccatgacc cagacaacgc ccttgaagac
 701  aagctgggtt aactgctcta acatgatcga tgaaattata acacacttaa
                       IL1-->
 751  agcagccacc tttgcctttg ctggacttca acaacctcaa tggggaagac
 801  caagacattc tgatggaaaa taaccttcga aggccaaacc tggaggcatt
 851  caacagggct gtcaagagtt tacagaacgc atcagcaatt gagagcattc
                      IL2-->
 901  ttaaaaatct cctgccatgt ctgcccctgg ccacggccgc acccacgcga
 951  catccaatcc atatcaagga cggtgactgg aatgaattcc ggaggaaact
1001  gacgttctat ctgaaaaccc ttgagaatgc gcaggctcaa cagacgactt
1051  tgagcctcgc gatcttttga AAGCTTctgt tttggcggat gagagaagat
                      <--PKK2
1101  tttcagcctg atacagatta aatcagaacg cagaagcggt ctgataaaac
```

Table C.6. Sequencing Primers for the NusA/bGH Fusion Gene.

```
   1  atgcactggc caccattgca caggcccagg aagaaagcct cggtgataac
                 NUS9-->
  51  aaaccggctg acgatctgct gaaccttgaa ggggtagatc gtgatttggc
 101  attcaaactg gccgcccgtg gcgtttgtac gctggaagat ctcgccgaac
 151  agggcattga tgatctggct gatatcgaag ggttgaccga cgaaaaagcc
                 NUS10-->
 201  ggagcactga ttatggctgc ccgtaatatt tgctggttcg gtgacgaagc
 251  gACCGGTATC GAAGGTCGAt tcccagccat gtccttgtcc ggcctgtttg
 301  ccaacgctgt gctccgggct cagcacctgc atcagctggc tgctgacacc
                        BGH1-->
 351  ttcaaagagt ttgagcgcac ctacatcccg gagggacaga gatactccat
 401  ccagaacacc caggttgcct tctgcttctc tgaaaccatc ccggccccca
 451  cgggcaagaa tgaggcccag cagaaatcag acttggagct gcttcgcatc
                         BGH2-->
 501  tcactgctcc tcatccagtc gtggctcggg ccctgcagt tcctcagcag
 551  agtcttcacc aacagcttgg tgtttggcac ctcggaccgt gtctatgaga
 601  agctgaagga cctggaggaa ggcatcctgg ccctgatgcg ggagctggaa
                   BGH3-->
 651  gatggcaccc cccgggctgg gcagatcctc aagcagacct atgacaaatt
 701  tgacacaaac atgcgcagtg acgacgcgct gctcaagaac tacggtctgc
 751  tctcctgctt ccggaaggac ctgcataaga cggagacgta cctgagggtc
 801  atgaagtgcc gccgcttcgg ggaggccagc tgcgccttct gaAAGCTTct
                                            <--PKK2
 851  gttttggcgg atgagagaag attttcagcc tgatacagat taaatcagaa
```

Table C.7. Sequencing Primers for the NusA/hIFN-γ Fusion Gene.

```
   1  atgcactggc caccattgca caggcccagg aagaaagcct cggtgataac
                NUS9-->
  51  aaaccggctg acgatctgct gaaccttgaa ggggtagatc gtgatttggc
 101  attcaaactg gccgcccgtg gcgtttgtac gctggaagat ctcgccgaac
 151  agggcattga tgatctggct gatatcgaag ggttgaccga cgaaaaagcc
                NUS10-->
 201  ggagcactga ttatggctgc ccgtaatatt tgctggttcg gtgacgaagc
 251  gaccggtatc gaaggtcgat gttactgcca ggacccatat gtacaagaag
 301  cagaaaacct taagaaatat tttaatgcag gtcattcaga tgtagcggat
            IFN1-->
 351  aatggaactc ttttcttagg cattttgaag aattggaaag aggagagtga
 401  cagaaaaata atgcagagcc aaattgtctc cttttacttc aaacttttta
 451  aaaactttaa agatgaccag agcatccaaa agagtgtgga gaccatcaag
                        IFN2-->
 501  gaagacatga atgtcaagtt tttcaatagc aacaaaaaga aacgagatga
 551  cttcgaaaag ctgactaatt attcggtaac tgacttgaat gtccaacgca
 601  aagcaataca tgaactcatc caagtgatgg ctgaactgtc gccagcagct
                        IFN3-->
 651  aaaacaggga agcgaaaaag gagtcagatg ctgtttcgag gtcgaagagc
 701  atcccagtaa aagcttctgt tttggcggat gagagaagat ttcagcctg
          <--PKK2
 751  atacagatta aatcagaacg cagaagcggt ctgataaaac agaatttgcc
```

Table C.8. Sequencing Primers for the His_6-NusA/hIL-3 Fusion Gene.

```
                                        PKPROM-->
   1  ccaatgcttc tggcgtcagg cagccatcgg aagctgtggt atggctgtgc
  51  aggtcgtaaa tcactgcata attcgtgtcg ctcaaggcgc actcccgttc
 101  tggataatgt tttttgcgcc gacatcataa cggttctggc aaatattctg
                                        PKKECO-->
 151  aaatgagctg ttgacaatta atcatcggct cgtataatgt gtggaattgt
 201  gagcggataa caatttcaca caggaaacag AATTCatgCA TCACCATCAC
 251  CATCACaaca aagaaatttt ggctgtagtt gaagccgtat ccaatgaaaa
 301  ggcgctacct cgcgagaaga ttttcgaagc attggaaagc gcgctggcga
                                      .
                                      .
                                      .
                                                    NUS9-->
1501  tgcacaggcc caggaagaaa gcctcggtga taacaaaccg gctgacgatc
1551  tgctgaacct tgaaggggta gatcgtgatt tggcattcaa actggccgcc
1601  cgtggcgttt gtacgctgga agatctcgcc gaacagggca ttgatgatct
                                                   NUS10-->
1651  ggctgatatc gaagggttga ccgacgaaaa agccggagca ctgattatgg
1701  ctgcccgtaa tatttgctgg ttcggtgacg aagcgACCGG TATCGAAGGT
1751  CGAgctccca tgacccagac aacgcccttg aagacaagct gggttaactg
                                                     IL1-->
1801  ctctaacatg atcgatgaaa ttataacaca cttaaagcag ccacctttgc
1851  ctttgctgga cttcaacaac ctcaatgggg aagaccaaga cattctgatg
1901  gaaaataacc ttcgaaggcc aaacctggag gcattcaaca gggctgtcaa
                                                     IL2-->
1951  gagtttacag aacgcatcag caattgagag cattcttaaa aatctcctgc
2001  catgtctgcc cctggccacg gccgcaccca cgcgacatcc aatccatatc
2051  aaggacggtg actggaatga attccggagg aaactgacgt tctatctgaa
2101  aacccttgag aatgcgcagg ctcaacagac gactttgagc ctcgcgatct
                                                  <--PKK2
2151  tttgaAAGCT Tctgttttgg cggatgagag aagattttca gcctgataca
2201  gattaaatca gaacgcagaa gcggtctgat aaaacagaat ttgcctggcg
2251  gcagtagcgc ggtggtccca cctgacccca tgccgaactc agaagtgaaa
```

Consensus Sequences from Sequencing of Fusion genes

Below, in Tables C.9 through C.16, are the consensus sequences obtained from the sequencing of the fusion genes using the primers illustrated in the previous section. Unless otherwise indicated at the end of the sequence, no mutations were found. The Genbank sequence number of sequence to which the consensus sequence was compared is also given after each sequence listing.

Table C.9. Consensus Sequence for the NusA/hIL-3 Fusion Gene.

```
            NusA -->
    1  GAATTCatga acaaagaaat tttggctgta gttgaagccg tatccaatga
   51  aaaggcgcta cctcgcgaga agattttcga agcattggaa agcgcgctgg
  101  cgacagcaac aaagaaaaaa tatgaacaag agatcgacgt ccgcgtacag
  151  atcgatcgca aaagcggtga ttttgacact ttccgtcgct ggttagttgt
  201  tgatgaagtc acccagccga ccaaggaaat caccccttgaa gccgcacgtt
  251  atgaagatga aagcctgaac ctgggcgatt acgttgaaga tcagattgag
  301  tctgttacct ttgaccgtat cactacccag acggcaaaac aggttatcgt
  351  gcagaaagtg cgtgaagccg aacgtgcgat ggtggttgat cagttccgtg
  401  aacacgaagg tgaaatcatc accggcgtgg tgaaaaaagt aaaccgcgac
  451  aacatctctc tggatctggg caacaacgct gaagccgtga tcctgcgcga
  501  agatatgctg ccgcgtgaaa acttccgccc tggcgaccgc gttcgtggcg
  551  tgctctattc cgttcgcccg gaagcgcgtg gcgcgcaact gttcgtcact
  601  cgttccaagc cggaaatgct gatcgaactg ttccgtattg aagtgccaga
  651  aatcggcgaa gaagtgattg aaattaaagc agcggctcgc gatccgggtt
  701  ctcgtgcgaa aatcgcggtg aaaaccaacg ataaacgtat cgatccggta
  751  ggtgcttgcg taggtatgcg tggcgcgcgt gttcaggcgg tgtctactga
  801  actgggtggc gagcgtatcg atatcgtcct gtgggatgat aacccggcgc
  851  agttcgtgat taacgcaatg gcaccggcag acgttgcttc tatcgtggtg
  901  gatgaagata aacacaccat ggatatcgcc gttgaagccg gtaacctggc
  951  gcaggcgatt ggccgtaacg gtcagaacgt gcgtctggct tcgcagctga
 1001  gcggttggga actcaacgtg atgaccgttg acgacctgca ggctaagcat
 1051  caggcggaag cgcacgcagc gatcgacacc ttcaccaaat atctcgacat
 1101  cgacgaagac ttcgcgactg ttctggtaga agaaggcttc tcgacgctgg
 1151  aagaattggc ctatgtgccg atgaaagagc tgttggaaat cgaaggcctt
 1201  gatgagccga ccgttgaagc actgcgcgag cgtgctaaaa atgcactggc
 1251  caccattgca caggcccagg aagaaagcct cggtgataac aaaccggctg
 1301  acgatctgct gaaccttgaa ggggtagatc gtgatttggc attcaaactg
 1351  gccgcccgtg gcgtttgtac gctggaagat ctcgccgaac agggcattga
 1401  tgatctggct gatatcgaag ggttgaccga cgaaaaagcc ggagcactga
 1451  ttatggctgc ccgtaatatt tgctggttcg gtgacgaagc gACCGGTATC
                hIL-3-->
 1501  GAAGGTCGAg ctcccatgac ccagacaacg cccttgaaga caagctgggt
 1551  taactgctct aacatgatcg atgaaattat aacacactta aagcagccac
 1601  ctttgccttt gctggacttc aacaacctca atggggaaga ccaagacatt
 1651  ctgatggaaa ataaccttcg aaggccaaac ctggaggcat tcaacagggc
 1701  tgtcaagagt ttacagaacg catcagcaat tgagagcatt cttaaaaatc
 1751  tcctgccatg tctgcccctg gccacggccg cacccacgcg acatccaatc
 1801  catatcaagg acggtgactg gaatgaattc cggaggaaac tgacgttcta
 1851  tctgaaaacc cttgagaatg cgcaggctca acagacgact ttgagcctcg
 1901  cgatcttttg aAAGCTT
```

GenBank sequences used for comparison:

NusA: ECAE000397
hIL-3: HUMCSFM

Table C.10. Consensus Sequence for the BFR/hIL-3 Fusion Gene.

```
          BFR-->
  1  GAATTCatga aaggtgatac taaagttata aattatctca acaaactgtt
 51  gggaaatgag cttgtcgcaa tcaatcagta ctttctccat gcccgaatgt
101  ttaaaaactg gggtctcaaa cgtctcaatg atgtggagta tcatgaatcc
151  attgatgaga tgaaacacgc cgatcgttat attgagcgca ttctttttct
201  ggaaggtctt ccaaacttac aggacctggg caaactgaac attggtgaag
251  atgttgagga aatgctgcgt tctgatctgg cacttgagct ggatggcgcg
301  aagaatttgc gtgaggcaat tggttatgcc gatagcgttc atgattacgt
351  cagccgcgat atgatgatag aaattttgcg tgatgaagaa ggccatatcg
401  actggctgga aacggaactt gatctgattc agaagatggg cctgcaaaat
                                                    hIL-3-->
451  tatctgcaag cacagatccg cgaagaaggt ACCGGTATCG AAGGTCGAgc
501  tcccatgacc cagacaacgc ccttgaagac aagctgggtt aactgctcta
551  acatgatcga tgaaattata acacacttaa agcagccacc tttgcctttg
601  ctggacttca acaacctcaa tggggaagac caagacattc tgatggaaaa
651  taaccttcga aggccaaacc tggaggcatt caacagggct gtcaagagtt
701  tacagaacgc atcagcaatt gagagcattc ttaaaaatct cctgccatgt
751  ctgcccctgg ccacggccgc acccacgcga catccaatcc atatcaagga
801  cggtgactgg aatgaattcc ggaggaaact gacgttctat ctgaaaaccc
851  ttgagaatgc gcaggctcaa cagacgactt tgagcctcgc gatcttttga
901  AAGCTT
```

GenBank sequences used for comparison:

BFR:	ECOBFR
hIL-3:	HUMCSFM

Table C.11. Consensus Sequence for the GrpE/hIL-3 Fusion Gene.

```
          GrpE-->
   1  GAATTCatga gtagtaaaga acagaaaacg cctgaggggc aagccccgga
  51  agaaattatc atggatcagc acgaagagat tgaggcagtt gagccagaag
 101  cttctgctga gcaggtggat ccgcgcgatg aaaaagttgc gaatctcgaa
 151  gctcagctgg ctgaagccca gacccgtgaa cgtgacggca ttttgcgtgt
 201  aaaagccgaa atggaaaacc tgcgtcgtcg tactgaactg gatattgaaa
 251  aagcccacaa attcgcgctg gagaaattca tcaacgaatt gctgccggtg
 301  attgatagcc tggatcgtgc gctggaagtg gctgataaag ctaacccgga
 351  tatgtctgcg atggttgaag gcattgagct gacgctgaag tcgatgctgg
 401  atgttgtgcg taagtttggc gttgaagtga tcgccgaaac taacgtccca
 451  ctggacccga atgtgcatca ggccatcgca atggtggaat ctgatgacgt
 501  tgcgccaggt aacgtactgg gcattatgca gaagggttat acgctgaatg
 551  gtcgtacgat tcgtgcggcg atggttactg tagcgaaagc aaaagctACC
                    hIL-3-->
 601  GGTATCGAAG GTCGAgctcc catgacccag acaacgccct tgaagacaag
 651  ctgggttaac tgctctaaca tgatcgatga aattataaca cacttaaagc
 701  agccaccttt gcctttgctg gacttcaaca acctcaatgg ggaagaccaa
 751  gacattctga tggaaaataa ccttcgaagg ccaaacctgg aggcattcaa
 801  cagggctgtc aagagtttac agaacgcatc agcaattgag agcattctta
 851  aaaatctcct gccatgtctg cccctggcca cggccgcacc cacgcgacat
 901  ccaatccata tcaaggacgg tgactggaat gaattccgga ggaaactgac
 951  gttctatctg aaaaccttg agaatgcgca ggctcaacag acgactttga
1001  gcctcgcgat cttttgaAAG CTT
```

GenBank sequences used for comparison:

GrpE:	ECGRPE
hIL-3:	HUMCSFM

Table C.12. Consensus Sequence for the Thioredoxin/hIL-3 Fusion Gene.

```
          thioredoxin-->
   1  GAATTCatga gcgataaaat tattcacctg actgacgaca gttttgacac
  51  ggatgtactc aaagcggacg gggcgatcct cgtcgatttc tgggcagagt
 101  ggtgcggtcc gtgcaaaatg atcgccccga ttctggatga aatcgctgac
 151  gaatatcagg gcaaactgac cgttgcaaaa ctgaacatcg atcaaaaccc
 201  tggcactgcg ccgaaatatg gcatccgtgg tatcccgact ctgctgctgt
 251  tcaaaaacgg tgaagtggcg gcaaccaaag tgggtgcact gtctaaaggt
 301  cagttgaaag agttcctcga cgctaacctg gcgACCGGTA TCGAAGGTCG
       hIL-3-->
 351  Agctcccatg acccagacaa cgcccttgaa gacaagctgg gttaactgct
 401  ctaacatgat cgatgaaatt ataacacact taaagcagcc acctttgcct
 451  ttgctggact tcaacaacct caatggggaa gaccaagaca ttctgatgga
 501  aaataacctt cgaaggccaa acctggaggc attcaacagg gctgtcaaga
 551  gtttacagaa cgcatcagca attgagagca ttcttaaaaa tctcctgcca
 601  tgtctgcccc tggccacggc cgcacccacg cgacatccaa tccatatcaa
 651  ggacggtgac tggaatgaat tccggaggaa actgacgttc tatctgaaaa
 701  cccttgagaa tgcgcaggct caacagacga ctttgagcct cgcgatcttt
 751  tgaAAGCTT
```

GenBank sequences used for comparison:

Thioredoxin:	ECORHOA
hIL-3:	HUMCSFM

Table C.13. Consensus Sequence for the YjgD/hIL-3 Fusion Gene.

```
          YjgD-->
  1  GAATTCatgg caaacccgga acaactggaa gaacagcgtg aagaaacacg
 51  tttgattatt gaagaattac tggaagatgg cagcgatccg gacgcgctgt
101  acaccatcga acaccatctt tccgcagacg atctggaaac cctggaaaaa
151  gcagcagttg aagcgtttaa actcggttac gaagtgaccg atccagaaga
201  gctggaagtt gaagatggtg atatcgtgat ttgctgcgac atcctcagcg
251  agtgcgcgtt gaatgccgat ctgatcgatg cccaggttga acaactgatg
301  acgctggcag agaaatttga cgttgagtac gacggatggg gcacttactt
351  tgaagatcct aacggcgaag atggcgacga tgaagatttt gtcgatgaag
                                              hIL-3-->
401  acgatgacgg agttcgccac tACCGGTATC GAAGGTCGAg ctcccatgac
451  ccagacaacg cccttgaaga caagctgggt taactgctct aacatgatcg
501  atgaaattat aacacactta aagcagccac ctttgccttt gctggacttc
551  aacaacctca atggggaaga ccaagacatt ctgatggaaa ataaccttcg
601  aaggccaaac ctggaggcat tcaacagggc tgtcaagagt ttacagaacg
651  catcagcaat tgagagcatt cttaaaaatc tcctgccatg tctgcccctg
701  gccacggccg cacccacgcg acatccaatc catatcaagg acggtgactg
751  gaatgaattc cggaggaaac tgacgttcta tctgaaaacc cttgagaatg
801  cgcaggctca acagacgact ttgagcctcg cgatctttga AAGCTT
                                          position of
                                           deleted t
```

The t in position 838 was found to be deleted. The last codon of the hIL-3 gene is TTT and was found to be TT.

GenBank sequences used for comparison:

YjgD:	ECOUW93
hIL-3:	HUMCSFM

Table C.14. Consensus Sequence for the NusA/bGH Fusion Gene. Since this particular NusA gene was sequenced as a part of the NusA/hIL-3 gene, only the immediate 3' end of the NusA gene was sequenced . The bGH gene was sequenced in its entirety.

```
     NusA (3' end) -->                                              bGH-->
  1  ccgtaatatt tgctggttcg gtgacgaagc gACCGGTATC GAAGGTCGAt
 51  tcccagccat gtccttgtcc ggcctgtttg ccaacgctgt gctccgggct
101  cagcacctgc aCcagctggc tgctgacacc ttcaaagagt ttgagcgcac
151  ctacatcccg gagggacaga gatactccat ccagaacacc caggttgcct
201  tctgcttctc tgaaaccatc ccggccccca cgggcaagaa tgaggcccag
251  cagaaatcag acttggagct gcttcgcatc tcactgctcc tcatccagtc
301  gtggctTggg cccctgcagt tcctcagcag agtcttcacc aacagcttgg
351  tgtttggcac ctcggaccgt gtctatgaga agctgaagga cctggaggaa
401  ggcatcctgg ccctgatgcg ggagctggaa gatggcaccc cccgggctgg
451  gcagatcctc aagcagacct atgacaaatt tgacacaaac atgcgcagtg
501  acgacgcgct gctcaagaac tacggtctgc tctcctgctt ccggaaggac
551  ctgcataaga cggagacgta cctgagggtc atgaagtgcc gccgcttcgg
601  ggaggccagc tgcgccttct gaAAGCTT
```

Two substitutions were found: one at position 112 (T to C) and one at position 307 (C to T). The susbsitution at position 112 results in the same His residue. The substitution at position 307 results in the same Leu residue.

GenBank sequences used for comparison:

NusA:	ECAE000397
bGH:	BOVGHC

Table C.15. Consensus Sequence for the NusA/hIFN-γ Fusion Gene. Since this particular NusA gene was sequenced as a part of the NusA/hIL-3 gene, only the immediate 3' end of the NusA gene was sequenced . The hIFN-γ gene was sequenced in its entirety.

```
      NusA (3' end) -->                                          hIFN-γ -->
   1  ccgtaatatt tgctggttcg gtgacgaagc gACCGGTATC GAAGGTCGAt
  51  gttactgcca ggacccatat gtaaaagaag cagaaaacct taagaaatat
 101  tttaatgcag gtcattcaga tgtagcggat aatggaactc ttttcttagg
 151  cattttgaag aattggaaag aggagagtga cagaaaaata atgcagagcc
 201  aaattgtctc cttttacttc aaacttttta aaaactttaa agatgaccag
 251  agcatccaaa agagtgtgga gaccatcaag gaagacatga atgtcaagtt
 301  tttcaatagc aacaaaaaga aacgagatga cttcgaaaag ctgactaatt
 351  attcggtaac tgacttgaat gtccaacgca aagcaataca tgaactcatc
 401  caagtgatgg ctgaactgtc gccagcagct aaaacaggga agcgaaaaag
 451  gagtcagatg ctgtttcgag gtcgaagagc atcccagtaa AAGCTT
```

GenBank sequences used for comparison:

NusA:	ECAE000397
hIFN-γ:	HSIFR15

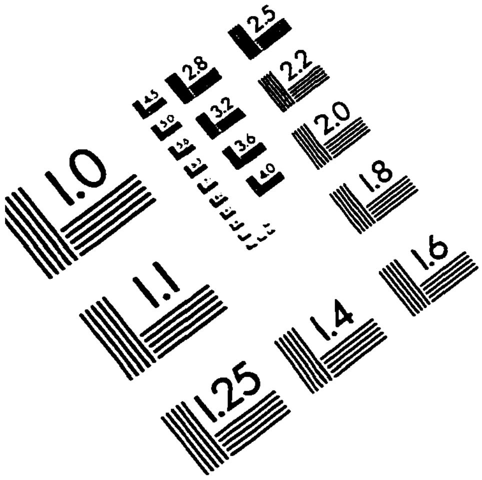
Table C.16. Consensus Sequence for the His_6-NusA/hIL-3 Fusion Gene. Only the immediate 5' and 3' ends of the NusA gene were sequenced . The hIL-3 gene was sequenced in its entirety.

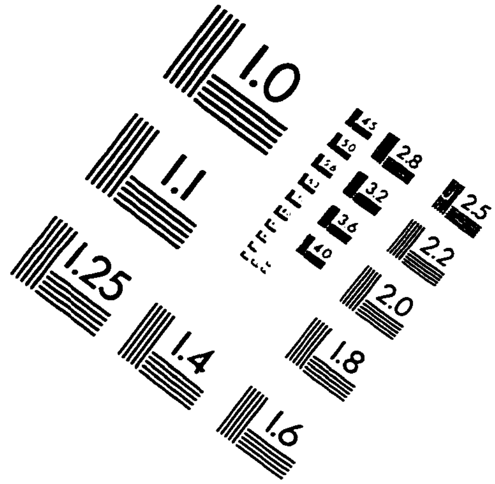
```
                |<-----His6----->| NusA (5' end)-->
  1  GAATTCatgC ATCACCATCA CCATCACaac aaagaaattt tggctgtagt
 51  tgaagccgta tccaatgaaa aggcgctacc tcgcgagaag at


      NusA (3' end) -->
  1  atcgtgattt ggcattcaaa ctggccgccc gtggcgtttg taACCGGTAT
                hIL-3-->
 51  CGAAGGTCGA gctcccatga cccagacaac gcccttgaag acaagctggg
101  ttaactgctc taacatgatc gatgaaatta taacacactt aaagcagcca
151  cctttgcctt tgctggactt caacaacctc aatggggaag accaagacat
201  tctgatggaa aataaccttc gaaggccaaa cctggaggca ttcaacaggg
251  ctgtcaagag tttacagaac gcatcagcaa ttgagagcat tcttaaaaat
301  ctcctgccat gtctgccccc ggccacggcc gcacccacgc gacatccaat
351  ccatatcaag gacggtgact ggaatgaatt ccggaggaaa ctgacgttct
401  atctgaaaac ccttgagaat gcgcaggctc aacagacgac tttgagcctc
451  gcgatctttt gaAAGCTT
```

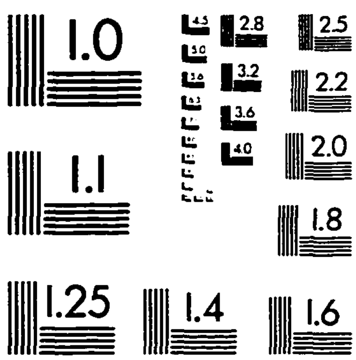
GenBank sequences used for comparison:

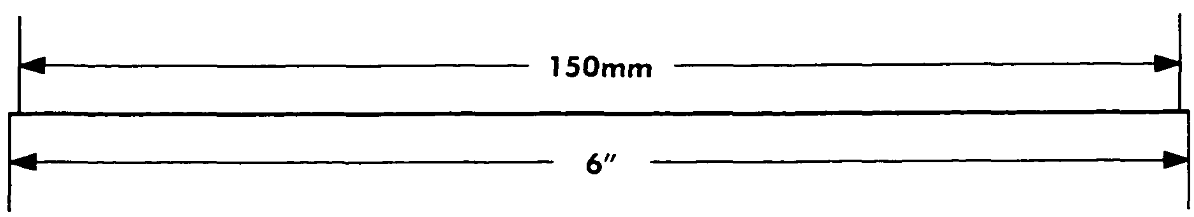
NusA: ECAE000397
hIL-3: HUMCSFM

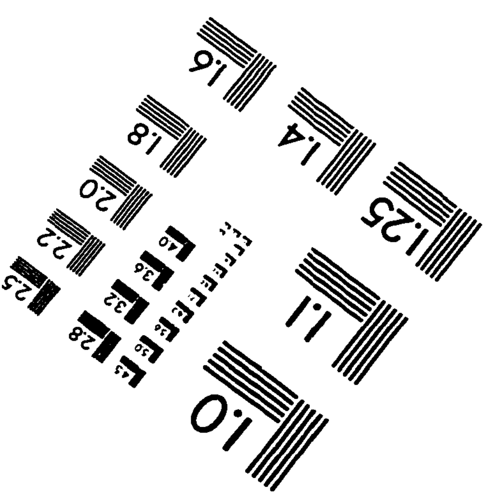
IMAGE EVALUATION TEST TARGET (QA-3)











APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989