Available online at www.sciencedirect.com
SCIENCE DIRECT
Protein Expression and Purification 46 (2006) 23–32
Protein Expression & Purification
www.elsevier.com/locate/yprep

An optimized system for expression and purification of secreted bacterial proteins

Brian V. Geisbrecht [a,*], Samuel Bouyain [b], Mihai Pop [c]

[a] *Division of Cell Biology and Biophysics, School of Biological Sciences, University of Missouri-Kansas City, 5100 Rockhill Road, Kansas City, MO 64110, USA*
[b] *Department of Biophysics and Biophysical Chemistry, The Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, MD 21205, USA*
[c] *The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA*

Received 22 May 2005, and in revised form 27 August 2005
Available online 4 October 2005

Abstract

In this report, we describe an optimized system for the efficient overexpression, purification, and refolding of secreted bacterial proteins. Candidate secreted proteins were produced recombinantly in *Escherichia coli* as Tobacco Etch Virus protease-cleavable hexahistidine-c-myc eptiope fusion proteins. Without regard to their initial solubility, recombinant fusion proteins were extracted from whole cells with guanidium chloride, purified under denaturing conditions by immobilized metal affinity chromatography, and refolded by rapid dilution into a solution containing only Tris buffer and sodium chloride. Following concentration on the same resin under native conditions, each protein was eluted for further purification and/or characterization. Preliminary studies on a test set of 12 secreted proteins ranging in size from 13 to 130 kDa yielded between 10 and 50 mg of fusion protein per liter of induced culture at greater than 90% purity, as judged by Coomassie-stained SDS–PAGE. Of the nine proteins further purified, analytical gel filtration chromatography indicated that each was a monomer in solution and circular dichroism spectroscopy revealed that each had adopted a well-defined secondary structure. While there are many potential applications for this system, the results presented here suggest that it will be particularly useful for investigators employing structural approaches to understand protein function, as attested to by the crystal structures of three proteins purified using this methodology (B.V. Geisbrecht, B.Y. Hamaoka, B. Perman, A. Zemla, D.J. Leahy, J. Biol. Chem. 280 (2005) 17243–17250).

Keywords: Pathogens; Secreted proteins; Virulence factors; Toxins

Introduction

The continually expanding database of genomic sequence, coupled with efficient informatics tools for analyzing this data have increased greatly the numbers of proteins available for study in many aspects of modern biology. This is particularly true in the field of bacterial pathogenesis, as newly completed genome sequences of medically important prokaryotes are made available nearly monthly. Proteins secreted by pathogenic bacteria participate in a variety of interactions with their host, many of which promote infection by allowing these bacteria to enter host cells and tissue layers or by interfering with host defense mechanisms (e.g., blood clotting or recognition by the immune system) [2]. As a consequence, detailed study of secreted bacterial proteins should reveal fundamental mechanisms of host-pathogen interactions and may identify potential vaccine/drug targets and/or diagnostic agents for widespread microbial diseases.

Unfortunately, the isolation of secreted proteins produced in situ is often a difficult and labor-intensive process. First, many species of pathogenic bacteria are difficult and dangerous to culture, requiring specialized containment facilities that can be cost-prohibitive to many investigators.

* Corresponding author. Fax: +1 816 235 1593.
E-mail address: GeisbrechtB@umkc.edu (B.V. Geisbrecht).

1046-5928/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2005.09.003

Second, many secreted proteins are present in only trace-quantities or remain cell-associated making them difficult to detect in dilute complex mixtures such as conditioned culture medium. Finally, secreted proteins can be subject to extensive post-translational modifications (including glycosylation, lipidation, and proteolytic degradation), which may be undesirable for biochemical and structural studies (e.g., macromolecular crystallography or NMR). Thus, the development of a recombinant system suitable for expressing a wide-range of proteins is essential for a thorough and systematic study of secreted bacterial proteins.

Because of the large number of potentially important proteins involved, we sought to develop and optimize a system for highly parallel expression and purification of sufficient quantities of secreted bacterial proteins for biochemical and structural characterization. To begin, DNAs encoding the mature forms of target proteins are subcloned into the vector pT7HMT, which allows for expression of hexahistidine-myc-tagged, Tobacco Etch Virus (TEV)[1] Virus, protease-cleavable fusion proteins in *Escherichia coli*. Following protein expression, the induced cells are lysed by incubation in a denaturing buffer, and the fusion proteins are purified by denaturing chelating chromatography. Purified proteins are then refolded by rapid dilution into a native buffer and subsequently concentrated by native chelating chromatography. Proteins eluted from this column are suitable for biochemical and functional characterization and/or digestion with TEV protease for further purification toward structural studies. While this system takes advantage of a number of existing features in commercially available expression and purification schemes, no single integrated expression and purification system incorporating each of these features has been reported until now.

Materials and methods

Expression and purification of recombinant TEV protease

A DNA fragment encoding an octahistidine-tagged variant of TEV protease was amplified by standard techniques to include *Sal*I and *Not*I restriction sites at the 5′ and 3′ ends, respectively. Following digestion, this fragment was subcloned into the corresponding sites of the vector pMBPT. This vector is identical to the maltose-binding protein (MBP) fusion vector pMBP [3], with the exception that it contains a TEV protease recognition site (-Glu-Asn-Leu-Tyr-Phe-Gln-^-Gly-) immediately upstream of and in frame with its *Sal*I site. The resulting plasmid pMBPT-His_8TEVp was transformed into *E. coli* strain BL21(DE3) and cells from the transformed strain were grown, induced at 18 °C, and harvested as previously described [3,4]. Induced cells from 1 L were resuspended 100 ml of Ni–NTA binding buffer (20 mM Tris (pH 8.0), 500 mM NaCl, 10 mM imidazole, 10% glycerol), lysed by microfluidization according to manufacturer's instructions (Microfluidics; Newton, MA), and a clarified cell extract was prepared by centrifugation as previously described [3]. The extract was applied to a 7.5 ml column of Ni–NTA Sepharose at 4 °C, the column was washed with 10 column volumes (CV) of Ni–NTA wash buffer (20 mM sodium phosphate (pH 6.0), 500 mM NaCl, 20 mM imidazole, 10% glycerol), and bound proteins were eluted with 2.5 CV of elution buffer (wash buffer with 500 mM imidazole). This eluate was buffer exchanged into cation-exchange binding buffer (20 mM sodium phosphate (pH 6.0), 10 mM DTT, 10% glycerol), immediately applied to a 10 ml SP-Sepharose fast-flow cation exchange column, and eluted with a gradient to 500 mM NaCl in binding buffer over 7.5 CV. Fractions containing recombinant protease were pooled, quantitated by UV-absorption spectrophotometry, and glycerol was added to final concentration of 50% for storage at −70 °C for at least 1 year. The yield of purified, recombinant TEVp was approximately 40 mg/L of original culture [5].

Expression vector construction

The vector pT7HMT is a variant of the commercially available prokaryotic expression vector pET28 (Novagen) and was assembled to contain the desired modifications in two steps. First, the region encoding both the c-myc epitope (recognized by the monoclonal antibody 9E10 [6]) and the TEV protease cleavage site [7] was subcloned into *Nde*I–*Xho*I digested pET28 using a pair of complementary, 5′-phosphorylated oligonucleotides with the appropriate overhangs. The resulting plasmid was digested with *Sal*I and *Xho*I and treated with calf intestinal phosphatase and the region comprising the polycloning site (5′-*Sal*I–*Bam*HI–*Nhe*I–*Sac*II–*Not*I–*Xho*I-3′) was subcloned into this vector using a second pair of complementary, 5′-phosphorylated oligonucleotides with the appropriate overhangs. The sequence of modified regions was confirmed using standard methods following each step.

Subcloning of DNAs encoding secreted bacterial proteins

Candidate proteins predicted to undergo Type-I signal peptidase-dependent secretion were identified from bacterial genomic sequences through a computer-based approach that incorporates the composition, location, and sequence context of likely signal peptides (M. Pop and B.V. Geisbrecht, Unpublished Results). The highest probability signal peptidase cleavage site was determined using the SignalP server (www.cbs.dtu.dk/services/SignalP/). DNA fragments encoding either mature polypeptides or known stable domains of previously characterized proteins were amplified from the appropriate template by PCR and contained *Sal*I and *Not*I sites at the 5′ and 3′ ends, respectively. Each PCR product was digested with *Sal*I and *Not*I, subcloned into the same sites of pT7HMT, and sequenced in its

[1] *Abbreviations used:* TEV, Tobacco Etch; MBP, maltose-binding protein; CV, column volumes; TB, Terrific Broth; TCEP, trichloroethylphosphine.

Table 1
Summary and yield of refolded secreted bacterial proteins prepared using the pT7HMT system

Protein name	Accession number and published reference	First-step Ni–NTA yield (mg purified/liter culture)	MW obs (kDa); MW expected (kDa)
Ehp	NP_371679	10.7	11 ; 9.5
Efb	P68799 [16]	49.6	19 ; 16
$ClfA_t$	NP_373997 [17]	23.8	39 ; 37
$ClfB_t$	NP_373154 [17]	36.6	N/D ; 39
EapH1	NP_372729[1]	21.6	12 ; 12
EapH2	NP_371505[1]	21.8	14 ; 14
MPT64	NP_216496[18]	36.1	22 ; 23
PNGase-F	PNFMGF [19]	35.3	N/A ; 35
$SdrC_t$	NP_371085[20]	44.9	33 ; 25
SdrC	NP_371085[20]	12.6	97 ; 74
SdrD	NP_371086[20]	11.5	N/A ; 117
SdrE	NP_371087[20]	11.0	N/A; 97

DNAs corresponding to the mature extracellular region or independent domains of several secreted bacterial proteins were subcloned into pT7HMT, and the corresponding proteins were expressed in *E. coli* and purified by denaturing chelating chromatography. Following refolding, the proteins were concentrated by native chelating chromatography, and several of these were digested by TEV protease and purified to homogeneity. In addition, the dispersity and apparent molecular mass of each purified protein was assessed by analytical gel filtration chromatography. The following table presents the Accession Numbers, yield of refolded, tagged target-protein per liter of culture, and the approximate experimental molecular mass of each purified protein.

entirety. A list of GenBank Accession Numbers and references to stable domains can be found in Table 1.

Bacterial culture and induction of protein expression

All proteins were expressed from the *E. coli* strain B834(DE3) and routine manipulations of bacterial cultures were carried out as previously described [3]. During the course of these studies, expression cultures were grown at 37 °C in 250 ml Terrific Broth (TB) and induced at an OD_{600} of 2.0 by adding IPTG to 1 mM final concentration. The cultures were incubated with vigorous shaking for an additional 18 h to achieve maximal cell density and the induced cells were harvested by centrifugation [3].

Analysis of total bacterial protein and Western blotting

Following protein expression the OD_{600} of each bacterial culture was measured and the cells from 1 ml of each culture were harvested by centrifugation. The cells were resuspended in a volume of water equal to 100× the OD_{600} in microliters, 6× Laemmli sample buffer was added to working concentration, and each sample was heated to 100 °C for 5 min. For analysis of protein induction, 5 μl of each suspension were separated by SDS–PAGE and the protein bands were visualized by Coomassie staining. For Western blot analysis, 0.5 μl of each suspension was separated as above and blotting was performed as previously described [8] using the 9E10 anti-c-myc monoclonal antibody at 1 μg/ml concentration (Sigma–Aldrich; [6]).

Bacterial cell lysis, chelating affinity chromatography, and protein refolding

Pellets of induced cells were concomitantly resuspended and lysed by stirring in 5% of the original culture volume of denaturing lysis buffer (0.1 M Tris (pH 8.0), 6 M guanidine–HCl). The cell pellet was stirred at 300–500 rpm for 30 min at room temperature and the solubilized, denatured proteins were separated from the cell debris by centrifugation (30 min at 25,000*g*). Following centrifugation, the clarified cell extract was decanted and applied by gravity flow to a 2 ml column of Ni^{2+}–NTA Sepharose (Qiagen) that had previously been equilibrated to room temperature in denaturing wash buffer (20 mM sodium phosphate (pH 6.0), 0.5 M NaCl, 10 mM imidazole, 8 M urea). Unbound proteins and contaminants were removed from the column by applying 5 CV of denaturing wash buffer. Tagged proteins were eluted from the resin with 2.5 CV of denaturing elution buffer (20 mM sodium phosphate (pH 6.0), 0.5 M NaCl, 0.2 M imidazole, 8 M urea), though the first 0.5 CV of eluate was discarded since it contained negligible protein.

Following purification under denaturing conditions, recombinant proteins were refolded by a "rapid dilution" technique, and concentrated by native chelating chromatography. First, TCEP–HCl (trichloroethylphosphine) was added to a final concentration of 1 mM and the denatured sample was incubated at 37 °C for 1 h; it should be noted that this step was included only if the target protein contained cysteine residues. Next, the denatured sample was drawn into an appropriate syringe with needle and refolded by quickly injecting the entire amount into a 10-fold volume excess of rapidly stirring, room temperature native buffer (20 mM Tris (pH 8.0), 0.5 M NaCl). The diluted sample was allowed to stir for 5–10 min longer under these conditions, at which time it was reapplied to the previous 2 ml column of Ni^{2+}–NTA Sepharose that had since been regenerated according to manufacturer's suggestions and equilibrated in native wash buffer (20 mM Tris (pH 8.0), 0.5 M NaCl, 10 mM imidazole). Excess urea was removed by washing with 5 CV of native wash buffer, and the bound,

refolded proteins were eluted as before with 2.5 CV of native elution buffer (20 mM Tris (pH 8.0), 0.5 M NaCl, 500 mM imidazole).

TEV protease digestion and ion-exchange chromatography

Recombinant TEV protease (see above) was used to digest the fusion tag away from the target protein as previously described [9]. All reactions were performed according to manufacturer's suggestions with the following modifications: digests were carried out at 4 °C to avoid protein precipitation and reducing agents were not added in cases where the formation of disulfide bonds was suspected in the target protein. Following digestion, each of the target proteins was purified further by ion exchange chromatography. In each case, the proteolytic digest was desalted into a suitable 20 mM buffer without salt, applied to a 6 ml Resource S or Q ion-exchange column (Amersham-Pharmacia Biotechnology), and eluted with a gradient to 1 M NaCl over 10 CV. The following section describes the buffer, pH, and column used for each protein: Ehp, ethanolamine, 9.0, S; Efb, ethanolamine, 9.0, S; $ClfA_t$, bis-Tris, 6.0, Q; $ClfB_t$, bis-Tris, 6.0, Q; EapH1, formate, 3.5, S; EapH2, formate, 3.5, S; MPT64, bis-Tris, 6.0, Q; $SdrC_t$, bis-Tris, 6.0, Q; and SdrC, bis-Tris, 6.0, Q. The concentration of each purified protein was determined by ultraviolet absorbance using approximate extinction coefficients derived from the protein primary sequence.

Analytical gel filtration chromatography

The dispersity of each purified protein was assessed by analytical gel filtration chromatography. In each case, 100 μg of purified protein was separated at 0.75 ml/min on a Superdex 75 10/300 column (Amersham-Pharmacia Biotechnology) and the approximate molecular mass of the eluted peak was estimated by linear regression against a standard curve of globular proteins (Amersham-Pharmacia Biotechnology).

Circular dichroism spectroscopy

The folding state of each purified protein was assessed by circular dichroism spectroscopy using a Jasco J-700 spectropolarimeter. Purified protein samples were dissolved in ddH_2O to a concentration where the OD_{215nm} was equal to 15.0 in a 1 cm pathlength UV spectrophotometry quartz cuvette; the samples were then immediately placed into a 1.0 mm pathlength quartz spectropolarimetry cell. CD spectra were collected in scan mode at 10 nm/min from 260 to 190 nm using a 0.5 nm resolution increment, 1.0 nm bandwidth, 1 s response time and a sensitivity of 20 mdeg. Five individual spectra were averaged and the noise from a water blank spectrum was subtracted prior to smoothing the composite spectrum according to manufacturer's suggestions (Jasco).

Miscellaneous

Matrix-assisted laser desorption ionization time-of-flight mass spectra were collected on a Perkin-Elmer Biosystems Voyager system that was previously calibrated with Equine apomyoglobin (16952.27 Da $(M+H)^+$ average; Sigma–Aldrich) according to manufacturer's suggestions. The plasmid encoding the Tobacco-Etch Virus protease was a gift from Dr. Cynthia Wolberger. Genomic DNA from *Staphylococcus aureus* strain Mu50 was prepared from 1 L of stationary-phase bacteria with a described modification of the Marmur protocol [1,10]. Genomic DNA from *Mycobacterium tuberculosis* strain H37Rv was a gift of Dr. Katherine Sacksteder and the plasmid encoding PNGase-F was a gift from Dr. Christopher Garcia.

Results

Design of the pT7HMT vector and expression of secreted bacterial proteins

pT7HMT is a variant of the pET28 family of plasmids available commercially from Novagen. This vector provides for IPTG-inducible, T7 polymerase-driven expression of hexahistidine-tagged fusion proteins and can be maintained by growth in the presence of kanamycin. While pET28 has many useful features, such as directing high-level protein expression and allowing for the use of chelating chromatography in purification schemes, it has two major drawbacks. First, the vector encodes a "T7 tag" that allows for detection by a monoclonal antibody. While this is a very useful feature, it is not cost efficient for routine use since the antibody is only available from the commercial supplier and the hybridoma is not available for purchase. Second, the hexahistidine–epitope tag is cleavable from the target protein by thrombin digestion. Although, thrombin has proven useful for digesting many fusion proteins, it is also expensive to use and can, in many cases, result in unwanted proteolysis of the target protein due to its inherently lower specificity.

In pT7HMT, we have replaced these elements of the parent plasmid with more cost-effective and useful features (Fig. 1). To begin, a TEV protease recognition sequence was inserted immediately upstream of a modified multicloning site. TEV protease offers several advantages over other proteases for digestion of fusion proteins, including increased specificity, and efficient activity at 4 °C (which may promote stability of the target protein) [7,11]. Furthermore, TEV protease can be efficiently expressed and purified from recombinant strains of *E. coli* (See Materials and methods), making it cost-efficient to use routinely in large-scale purification protocols [5,11,12]. Next, the T7 tag was replaced with the more commonly used c-myc epitope tag [6]. This offers two distinct advantages. First, the hybridoma (9E10) which secretes the anti-myc monoclonal antibody is readily available from several suppliers (e.g., ATCC); this allows investigators to easily generate large quantities of the antibody for routine use in

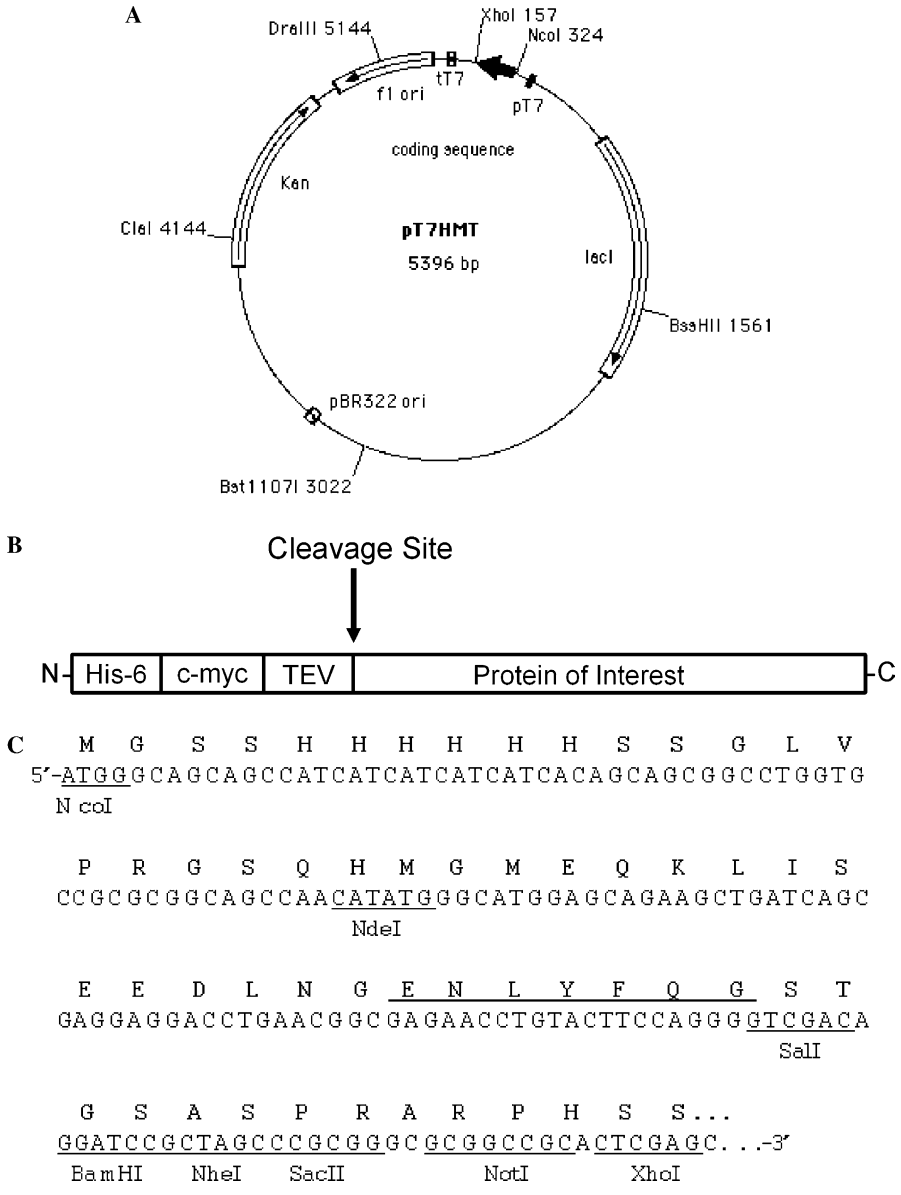

Fig. 1. pT7HMT, a vector optimized for the expression of secreted bacterial proteins. (A) A map of the pT7HMT bacterial expression vector is shown along with the location of selected unique restriction sites. (B) A diagram of a fusion protein expressed from pT7HMT is shown. (C) The DNA and protein sequences of the pT7HMT translated region and MCS are depicted. The TEV protease recognition site is shown in bold underline; hydrolysis of the peptide bond occurs between the Glutamine (Q) and Glycine (G) residues. Unique restriction endonuclease sites are shown with underlines. For clarity, only the final four residues of the 5′-most *Nco*I site are depicted.

detection systems. Second, both tagged and untagged versions of the same protein can be easily generated from a single expression vector, since the c-myc tag is amino-terminal to the TEV protease site. This feature is particularly useful when characterizing novel proteins or protein–ligand interactions because it facilitates competition experiments, immunoprecipitations and pulldowns, immunofluorescence assays, and ELISAs.

To test the suitability of this system for the expression of secreted bacterial proteins, we subcloned DNAs encoding the putative mature forms of 12 proteins either known or predicted to undergo Type-I signal-peptidase-dependent secretion into pT7HMT and carried out small-scale expression trials. As shown in Fig. 2A, all 12 proteins were expressed to high levels, typically between 5 and 10% of total cellular protein, as judged by SDS–PAGE and Coomassie staining. There was little qualitative difference in the expression levels of these proteins, despite the fact that they ranged in size from 13 to 130 kDa. In addition, we observed similar relative protein expression levels when the host cells were grown in minimal media suitable for stable ^{15}N and/or ^{13}C or seleno-L-methionine labeling applications (data not shown). The expression levels of each protein were also monitored by Western blotting against the vector-encoded c-myc epitope; this revealed that the hexahistidine and epitope tag was intact in every case (Fig. 2B).

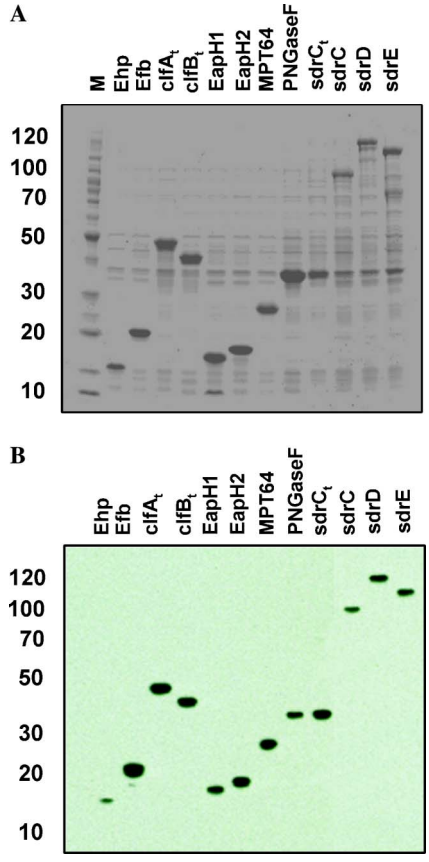

Fig. 2. High-level overexpression and detection of secreted bacterial proteins using the pT7HMT system. (A) Equal portions of whole-cell extracts prepared from overexpressing *E. coli* strains were separated by SDS–PAGE and visualized by Coomassie staining. (B) Equal portions of the same whole-cell extracts were separated by SDS–PAGE and analyzed by Western blotting against the c-myc epitope tag. A concise description of each protein can be found in Table 1.

Denaturing purification and efficient refolding of secreted bacterial proteins

Of the numerous fusion protein tags available, we chose the hexahistidine tag for our studies because it has the unique advantage of facilitating purification for both native and denatured proteins. In this respect, it can be advantageous to purify a protein under denaturing conditions for several reasons. First, purification under denaturing conditions can aid the stability of labile proteins by inactivating the proteolytic enzymes present in crude samples. Second, far fewer contaminating proteins are capable of binding to chelating resins in the presence of denaturing buffers. Finally, the higher resolution of chelating columns under denaturing conditions can greatly simplify cell fractionation protocols, thereby decreasing the total time spent in purifying individual protein samples. These benefits, particularly the last two, are of primary concern when processing many samples in parallel (e.g., high throughput overexpression and purification for antigen characterization or structural genomics projects).

In light of these considerations, we purified each target fusion protein by denaturing chelating chromatography without regard to their solubility states following expression (See Materials and methods). The induced cell pellet for each protein was lysed by stirring in a buffer containing 6 M guanidine hydrochloride for 30 min and the solubilized, denatured proteins were separated from insoluble components by centrifugation. The supernatants from each cell pellet were applied to separate chelating columns and the denatured, hexahistidine-tagged proteins were eluted from the columns following a single, stringent washing step. In every case, the column eluate was highly purified, containing greater than 90% hexahistidine-tagged protein as judged by SDS–PAGE and Coomassie staining (data not shown).

Following the initial purification, we tested whether we could efficiently refold these secreted, bacterial proteins with a simple buffer system that permitted immediate concentration of the renatured, hexahistidine-tagged proteins by native chelating chromatography. Using a simple, but previously unreported modification of the "rapid-dilution" method [13], the entire denatured protein sample was drawn into a syringe and immediately diluted into a rapidly stirring, 10-fold excess volume of a solution consisting of only Tris-buffer (pH 8.0) and sodium chloride (0.5 M). At this point, the imidazole concentration in the samples was sufficiently low (approximately 20 mM) to allow efficient binding of the diluted, renatured hexahistidine-tagged proteins to the chelating resin. Each sample was subjected to a second round of chelating chromatography and the soluble, refolded proteins were subsequently eluted in a highly purified, concentrated form (Fig. 3A). It is worth noting that the eluates from each column remained clear, with no evidence of precipitation or aggregation. Furthermore, each sample was quantitatively recovered during this procedure, indicating that the refolding and concentration steps were highly efficient (Table 1). Nine of these proteins were digested with recombinant TEV protease and purified to greater than 95% purity (as judged by Coomassie-strained SDS–PAGE) for further characterization (Fig. 3B).

Assessment of the dispersity and folding state of recombinant secreted proteins

While solubility at high concentration is often a reasonable indicator of both physical dispersity and proper folding of recombinant proteins, it does not rule out the possibility that these refolded proteins were present in solution as large molecular weight aggregates. Although, the fact that the hexahistidine-c-myc fusion tag could be successfully proteolyzed from each protein suggested that this was unlikely (Fig. 3B), we tested this possibility directly by subjecting each purified protein to analytical gel filtration chromatography. In each case, the observed molecular weight of the purified recombinant protein was consistent with a soluble, monomeric species (Table 1). To control for the possibility that these proteins were monomeric but

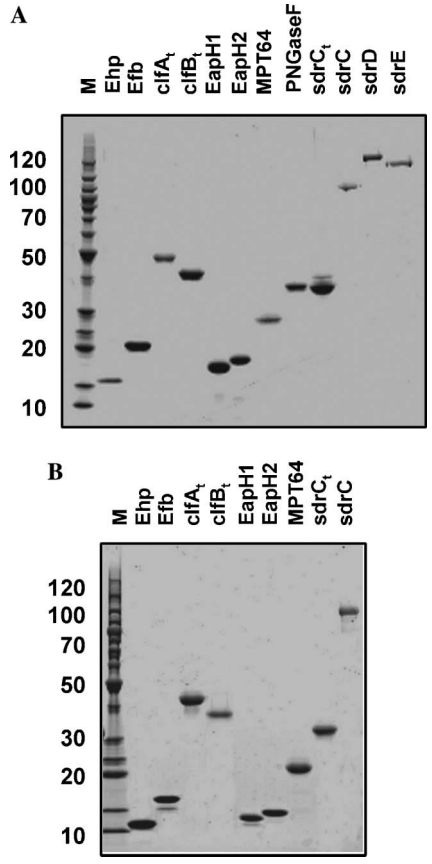

Fig. 3. Efficient purification and refolding of secreted bacterial proteins expressed from pT7HMT. Cells of overexpressing *E. coli* strains were lysed under denaturing conditions and the tagged, target proteins were purified by chelating chromatography. Each protein was refolded (please see Materials and methods) and concentrated by native chelating chromatography. (A) Equal portions (2 μl of 4 ml) of soluble, refolded protein eluted from the second chelating column were separated by SDS–PAGE and detected by Coomassie blue staining. Nine of the proteins presented here were digested with recombinant TEV protease, and subsequently purified to homogeneity (please see Materials and methods). (B) 2 μg of each purified protein were separated by SDS–PAGE and detected by Coomassie blue staining. Note the change in electrophoretic mobility for each protein resulting from proteolytic digestion of the 4.8 kDa his-myc fusion tag.

unfolded, we also collected circular dichroism spectra for each purified protein. As is shown in Fig. 4, each purified protein exhibited a unique CD spectrum. Furthermore, signals were observed in these CD spectra for each type of basic structural element (α-helix, β-sheet, and not defined), implying that the ability to undergo successful refolding using this technique is not limited to proteins of particular structure subtypes.

Discussion

The reemergence and spread of antibiotic-resistant strains of pathogenic bacteria has renewed interest in the study of these organisms. In the present work, we have described an integrated *E. coli* overexpression and purification system that provides a rapid, time- and cost-efficient purification method for generating multi-milligram quantities of secreted bacterial proteins for functional, biochemical, and structural analysis. While, we have presented results on a select group of proteins, we have currently used this system to successfully overexpress and purify more than 30 secreted proteins from the widespread pathogens *Staphylococcus aureus* and *M. tuberculosis*. These efforts have resulted in the solution of three crystal structures [1], sufficient resonance assignments to determine the solution structure of a protein by NMR [21], and diffracting crystals of two additional proteins (B.V. Geisbrecht, Manuscripts in Preparation). Although, the current study focused exclusively on proteins known or predicted to be secreted through a Type-I signal peptidase-dependent mechanism, this merely reflects the specific interests of our laboratory; there is no reason to suspect that these methods could not be successfully applied to those bacterial proteins that require either Type-II or Type-III secretion pathways. Furthermore, although we have pursued mainly structural studies following protein purification, the approach described here should prove very useful to any researcher needing to purify multi-milligram quantities of similar proteins for other purposes (such as antigen or vaccine discovery, etc.).

The main advantage of the approach described here is that it streamlines the inherently variable process of protein purification toward one capable of higher throughput. By eliminating variability at the level of choosing the appropriate chromatography scheme for any given protein of interest, we were able to process samples for up to 12 different proteins simultaneously (Figs. 2 and 3). There are several implicit features of our approach that allow for this improvement. First, since the initial separation was carried out in the same denaturing buffer, we are able to lyse the induced cells with a single chemical incubation. This procedure easily accommodates several samples in parallel, in contrast to mechanical lysis, which must be performed serially. Second, the superior resolution of immobilized affinity chromatography, especially when performed in the presence of denaturants, allows for the isolation of highly purified samples in one step. Not only does this serve to eliminate excessive chromatographies to achieve final purity, it is also easily performed in parallel and therefore increases sample throughput. Third, this system gives investigators the flexibility to generate both an epitope-tagged and untagged version of the same protein from one vector and cell preparation. This can partially eliminate the need to generate antibodies against individual proteins and may simplify certain types of functional and biochemical assays. Finally, it is worth mentioning that work done at the scale reported here did not require any automation or unusual modifications of standard equipment. This suggests that it may be easily adapted for the purposes of any single or group of laboratories working collaboratively to generate a large cohort of purified proteins.

During the course of these studies, we chose to employ ion-exchange chromatography instead of a third round of

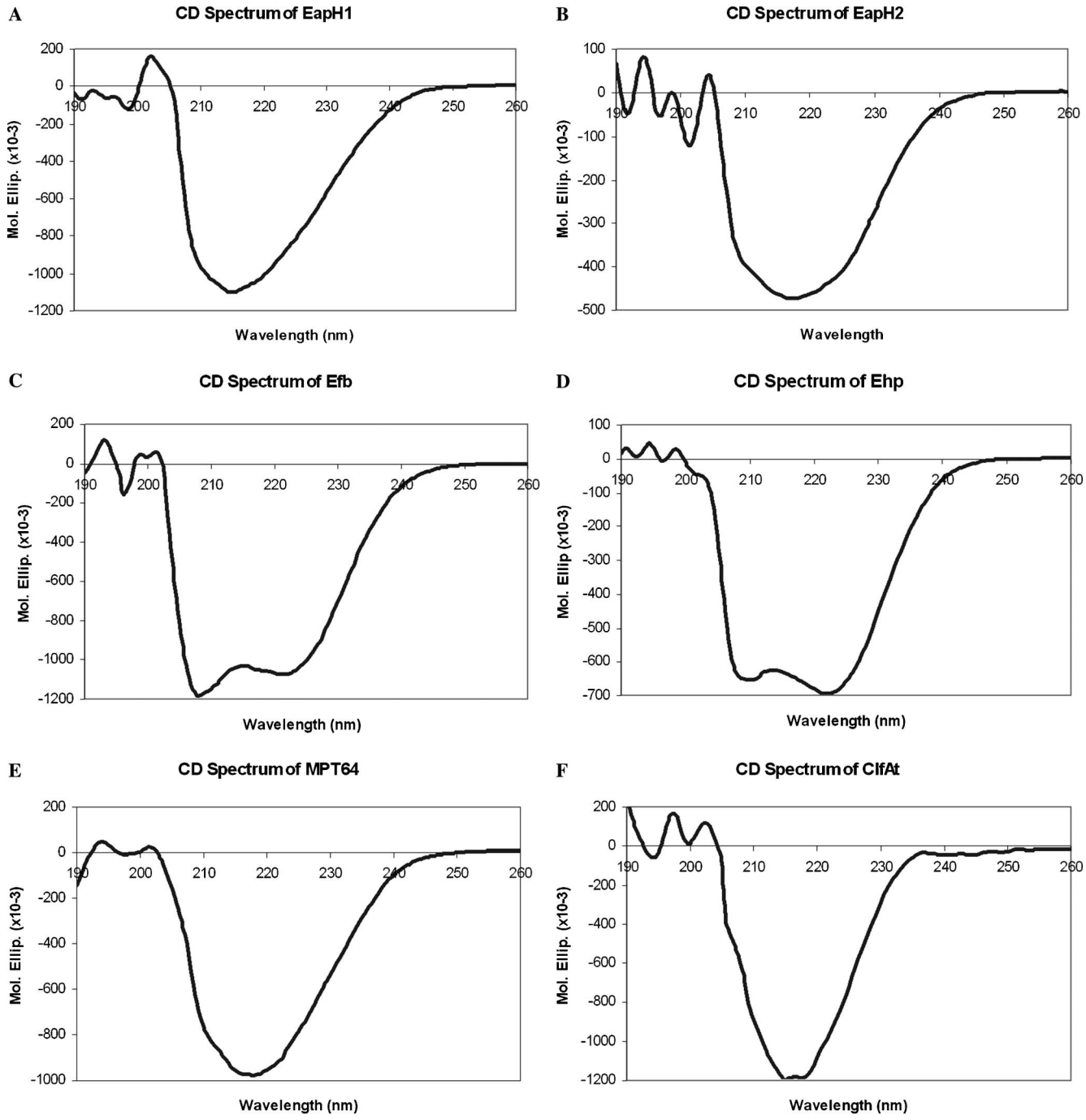

Fig. 4. Circular dichroism spectra of purified secreted bacterial proteins. Circular dichroism spectra were collected and processed within the range of 190–260 nm as described in Materials and methods. Note the distinct character of the spectrum for each different protein consistent with a unique secondary structure. (A) EapH1, (B) EapH2, (C) Efb, (D) Ehp, (E) MPT64, and (F) $ClfA_t$.

immobilized metal affinity chromatography as a final purification step following proteolytic digestion of the fusion tag. This decision was based primarily on our need for a high resolution separation versus a higher throughput "reverse" affinity chromatography (i.e., where the target protein does not bind to the resin). To illustrate, we observed that the main contaminants of refolded proteins containing intramolecular disulfide bonds (e.g., MPT64, [21]) were soluble, disulfide-linked multimers. While the buffer conditions for ion-exchange could be tailored to efficiently separate the desired protein from its misfolded counterparts, such resolution was simply not achievable when using reverse affinity chromatography. Furthermore, the use of ion-exchange as a final purification always resulted in a more concentrated protein sample for downstream applications (e.g., crystallization). This simplification could help reduce the use of expensive centrifugal or pressure-driven concentrating devices, a concern which

may be particularly relevant when planning to purify many proteins. On the other hand, we observed several instances when reverse affinity chromatography provided adequate separation of the target protein from its contaminants. This was particularly true for proteins that are highly soluble and lack cysteine residues (e.g., EapH1 and EapH2 [1]). Thus, it appears that this choice represents a compromise between resolution and throughput and ultimately depends on the downstream application(s) and number of samples being processed by an investigator.

The purification scheme presented here is predicated on the ability to efficiently refold target proteins from highly denaturing conditions into minimal aqueous buffers, and we chose to employ a simple modification of the rapid dilution refolding technique instead of using equilibrium dialysis methods. As originally described by Garboczi et al. for the refolding of MHC-I complexes, the rapid dilution methodology involves injecting a denatured sample into a large volume excess of a renaturing buffer [13,14]. An examination of several published protocols revealed that these buffers typically contain high concentrations of cosolvents (e.g., arginine, detergents, etc.) that promote protein folding, presumably by limiting aggregation of exposed hydrophobic surfaces. In our study, we employed a refolding buffer that used only sodium chloride as a cosolvent and found that this facilitated the in vitro refolding of each protein tested. Conveniently, this also allowed dilution of residual imidazole to levels low enough to permit efficient concentration of refolded proteins by affinity chromatography performed under native conditions.

A similar dilution-based approach to refolding proteins extracted from inclusion bodies was recently published by Vincentelli et al. [15]. In this work, the authors developed an innovative 96-well screen to monitor the likelihood that particular cosolvents and additives could assist in the in vitro refolding of denatured proteins, and found that approximately 60% of their target proteins from *M. tuberculosis* could be refolded with sufficient yield to be entered into crystallization trials. One interesting observation from this study is that proteins that refolded successfully often did so in multiple conditions, suggesting that refolding is a protein-specific. This is particularly noteworthy in the context of the current work, which revealed that many secreted bacterial proteins can be successfully refolded even without the use of extensive cosolvents. While it is tempting to speculate that particular biochemical properties of many of secreted bacterial proteins make them more amenable to such in vitro manipulations (e.g., extreme isoelectric point values, low cysteine content, etc.), it is clear that a much more thorough analysis is warranted before it is possible to efficiently predict refolding behavior from primary sequence characteristics.

Acknowledgments

We thank Dr. Natasha Shepulina for expert assistance with circular dichroism spectroscopy. This work was funded by the Howard Hughes Medical Institute (performed in the laboratory of Daniel J. Leahy, Ph.D., The Johns Hopkins University School of Medicine) and the Cell Biology and Biophysics Research Incentive Fund of the University of Missouri-Kansas City. S.B. is a postdoctoral fellow of the Wellcome Trust.

References

[1] B.V. Geisbrecht, B.Y. Hamaoka, B. Perman, A. Zemla, D.J. Leahy, The Crystal Structures of EAP Domains from *Staphylococcus aureus* Reveal an Unexpected Homology to Bacterial Superantigens, J. Biol. Chem. 280 (2005) 17243–17250.

[2] J.M. Patti, B.L. Allen, M.J. McGavin, M. Hook, MSCRAMM mediated adherence of microorganisms to host tissues, Annu. Rev. Microbiol. 48 (1994) 585–617.

[3] B.V. Geisbrecht, D. Zhu, K. Schulz, K. Nau, J.C. Morrell, M. Geraghty, H. Schulz, R. Erdmann, S.J. Gould, Molecular characterization of *Saccharomyces cerevisiae* Delta(3), Delta(2)-enoyl-CoA isomerase, J. Biol. Chem. 273 (1998) 33184–33191.

[4] B.V. Geisbrecht, D. Zhang, H. Schulz, S.J. Gould, Characterization of PECI, a novel monofunctional Delta(3),Delta(2)-enoyl-CoA isomerase of mammalian peroxisomes, J. Biol. Chem. 274 (1999) 21797–21803.

[5] R.B. Kapust, J. Tozser, J.D. Fox, D.E. Anderson, S. Cherry, T.D. Copeland, D.S. Waugh, Tobacco Etch Virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency, Prot. Eng. 14 (2001) 993–1000.

[6] G.I. Evan, G.K. Lewis, G. Ramsay, J.M. Bishop, Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product, Mol. Cell. Biol. 5 (1985) 3610–3616.

[7] W.G. Dougherty, J.C. Carrington, S.M. Cary, T.D. Parks, Biochemical and mutational analysis of a plant virus polyprotein cleavage site, EMBO J. 7 (1988) 1281–1287.

[8] D.I. Crane, J.E. Kalish, S.J. Gould, The *Pichia pastoris* PAS4 gene encodes a ubiquitin-conjugating enzyme required for peroxisome assembly, J. Biol. Chem. 269 (1994) 21835–21844.

[9] D.J. Leahy, C.E. Dann 3rd, P. Longo, B. Perman, K.X. Ramyar, A mammalian expression vector for expression and purification of secreted proteins for structural studies, Protein Expr. Purif. 20 (2000) 500–506.

[10] J. Marmur, A procedure for the isolation of deoxyribonucleic acid from micro-organisms, J. Mol. Biol. 3 (1961) 208–218.

[11] T.D. Parks, E.D. Howard, T.J. Wolpert, D.J. Arp, W.G. Dougherty, Expression and purification of a recombinant Tobacco Etch Virus NIa proteinase: biochemical analyses of the full-length and a naturally occurring truncated proteinase form, Virology 210 (1995) 194–201.

[12] L.J. Lucast, R.T. Batey, J.A. Doudna, Large-scale purification of a stable form of recombinant Tobacco Etch Virus protease, Biotechniques 30 (2001) 544–554.

[13] D.N. Garboczi, D.T. Hung, D.C. Wiley, HLA-A2-peptide complexes: refolding and crystallization of molecules expressed in *Escherichia coli* and complexed with single antigenic peptides, Proc. Natl. Acad. Sci. USA 89 (1992) 3429–3433.

[14] D.J. Stauber, A.D. DiGabriele, W.A. Hendrickson, Structural interactions of fibroblast growth factor receptor with its ligands, Proc. Natl. Acad. Sci. USA 97 (2000) 49–54.

[15] R. Vincentelli, S. Canaan, V. Campanacci, C. Valencia, D. Maurin, F. Frassinetti, L. Scappucini-Calvo, Y. Bourne, C. Cambillau, C. Bignon, An automated method for high throughput refolding of inclusion bodies, Protein Sci. 13 (2004) 2782–2792.

[16] M.K. Boden, J.-I. Flock, Cloning and characterization of a gene for a 19 kDa fibrinogen-binding protein from *Staphylococcus aureus*, Mol. Microbiol. 12 (1994) 599–606.

[17] C.C.S. Deivanayagam, S. Perkins, S. Danthuluri, R.T. Owens, T. Bice, T. Nanavathy, T.J. Foster, M. Hook, S.V.L. Narayana, Crystallization

of ClfA and ClfB fragments: the fibrinogen-binding surface proteins of *Staphylococcus aureus*, Acta. Cryst. D. 55 (2001) 554–556.

[18] M. Harboe, S. Nagai, M.E. Patarroyo, M.L. Torres, C. Ramirez, N. Cruz, Properties of proteins MPB64, MPB70, and MPB80 of *Mycobacterium bovis* BCG, Infect. Immun. 52 (1986) 293–302.

[19] A.L. Tarentino, G. Quinones, A. Trumble, L.-M. Changchien, B. Duceman, F. Maley, T.H. Plummer Jr., Molecular cloning and amino acid sequence of peptide-N4-(*N*-acteyl-β-D-glucosaminyl)asparagine amidase from *Flavobacterium meningosepticum*, J. Biol. Chem. 265 (1990) 6961–6966.

[20] K.W. McCrea, O. Hartford, S. Davis, D. NiEidhin, G. Lina, P. Speziale, T.J. Foster, M. Hook, The Serine-aspartate repeat (Sdr) protein family in *Staphylococcus aureus*, Microbiology 146 (2000) 1535–1546.

[21] J.M. Danahy, B.M. Potter, B.V. Geisbrecht, J.H. Laity, NMR assignment of protein Rv1980c from *Mycobacterium tuberculosis*, J. Biomol. NMR 33 (2005) 73.