Carine Robichon, Jianying Luo, Thomas B. Causey, Jack S. Benner, and James C. Samuelson*

New England BioLabs, Inc., Gene Expression Division, 240 County Road, Ipswich, Massachusetts 01938

Received 19 January 2011/Accepted 4 May 2011

Recombinant His-tagged proteins expressed in *Escherichia coli* and purified by immobilized metal affinity chromatography (IMAC) are commonly coeluted with native *E. coli* proteins, especially if the recombinant protein is expressed at a low level. The *E. coli* contaminants display high affinity to divalent nickel or cobalt ions, mainly due to the presence of clustered histidine residues or biologically relevant metal binding sites. To improve the final purity of expressed His-tagged protein, we engineered *E. coli* BL21(DE3) expression strains in which the most recurring contaminants are either expressed with an alternative tag or mutated to decrease their affinity to divalent cations. The current study presents the design, engineering, and characterization of two *E. coli* BL21(DE3) derivatives, NiCo21(DE3) and NiCo22(DE3), which express the endogenous proteins SlyD, Can, ArnA, and (optionally) AceE fused at their C terminus to a chitin binding domain (CBD) and the protein GlmS, with six surface histidines replaced by alanines. We show that each *E. coli* CBD-tagged protein remains active and can be efficiently eliminated from an IMAC elution fraction using a chitin column flowthrough step, while the modification of GlmS results in loss of affinity for nickel-containing resin. The "NiCo" strains uniquely complement existing methods for improving the purity of recombinant His-tagged protein.

Over the past 25 years, several techniques and tools have been developed to express and purify recombinant proteins for protein structure-function studies, for the development of new drugs, or simply for the manufacture of enzymes. The most frequently used method for isolating recombinant protein from a cell lysate in a single purification step is immobilized metal ion affinity chromatography (IMAC). In the simplest application of this method, the target protein is tagged with a polyhistidine sequence (typically $6 \times \text{His}$), which mediates chelation to immobilized divalent metal ions such as nickel or cobalt. Other studies have demonstrated that peptides with nonconsecutive histidines are also capable of chelation to immobilized divalent metal ions (5) (U.S. patent 7,176,298 [41] and U.S. patent application 2006/0030007 A1).

Escherichia coli is the most commonly used host for highyield expression of recombinant protein, usually by exploiting the high promoter specificity and transcriptional activity of bacteriophage T7 RNA polymerase. However, several *E. coli* host proteins also contain nonconsecutive histidine residues exposed to the surface of their ternary structure. In addition, metal binding motifs often mediate binding to nickel- and/or cobalt-containing purification resins. Such host proteins are routinely copurified during IMAC procedures and are therefore referred to as "contaminants." Several metal binding proteins that behave as IMAC contaminants have been identified in recent years. For example, Bolanos-Garcia et al. reviewed this issue in detail by classifying the E. coli metal binding proteins according to their affinity for Ni-nitrilotriacetic acid (NTA) resin by determining the imidazole concentration required for elution (5). Among 17 E. coli IMAC contaminants described, 15 were reported to elute from Ni-NTA at an imidazole concentration of >55 mM, a concentration which is higher than advised for most IMAC column washing procedures. Thus, most of the cited contaminants are eluted only when the imidazole concentration is increased to a level that elutes the histidine-tagged protein of interest. Also, variable amounts of host protein contaminants are detected depending on the expression system used (genetic background of the strain and plasmid) and the culture conditions employed (medium, carbon source, oxygen, temperature, and cell density at the induction time and at the harvest time).

Various techniques for improving the purity of a His-tagged protein of interest have been described in the literature. First, an alternative to imidazole washing and elution is used to elute the protein of interest with an acidic buffer. At pH <6, the histidine side chain in most contexts becomes protonated and loses its affinity for divalent metals. However, the same is true for contaminant proteins enriched in histidine residues. Therefore, low-pH elution may not result in discrimination between elution of contaminants and elution of the protein of interest. Some impurities can also be reduced by adjusting growth conditions (culture conditions, medium composition, and the genetic background of the strain), but this method of addressing the problem is empirical. Secondary chromatographic steps, of course, may be carried out, e.g., size exclusion chromatography

^{*} Corresponding author. Mailing address: New England BioLabs, Inc., Gene Expression Division, 240 County Road, Ipswich, MA 01938. Phone: (978) 380-7288. Fax: (978) 921-1350. E-mail: samuelson@neb .com.

[†] Supplemental material for this article may be found at http://aem .asm.org/.

⁷ Published ahead of print on 20 May 2011.

[‡] The authors have paid a fee to allow immediate free access to this article.

Strain	Genotype or description	Source
BL21(DE3)	fhuA2 [lon] ompT gal (λ DE3) [dcm] Δ hsdS λ DE3 = λ sBamHIo Δ EcoRI-B int::(lacI::PlacUV5::T7 gene 1) i21 Δ nin5 T1 phage-resistant version of BL21(DE3)	NEB strain catalog no. C2527
BL21(DE3) <i>slyD-CBD</i> BL21(DE3) <i>glmS-CBD</i> CGSC5477 CGSC4823 NiCo21(DE3) NiCo22(DE3)	<i>thr-26 aceE64</i> e14 <i>trp-66 relA1 spoT1 thi-1</i> <i>aceE2 tyrT58</i> (AS) <i>trp-26 mel-1</i> BL21(DE3) glmS _{6Ala} slyD-CBD can-CBD arnA-CBD BL21(DE3) glmS _{6Ala} slyD-CBD can-CBD arnA-CBD aceE-CBD	This study This study <i>E. coli</i> Genetic Stock Center <i>E. coli</i> Genetic Stock Center This study This study

TABLE 1. Strains used in this study

(31), protein specific chromatography (Heparin affinity chromatography described by Finzi et al. [10] or immunoaffinity chromatography described by Muller et al. [29]). But these approaches usually require time-consuming optimization procedures that are dependent on the properties of the target protein. Dual affinity tags, which imply the use of a second affinity column, also provide for improved target protein purity. For example polyhistidine has been combined with glutathione S-transferase (GST) or maltose binding protein (MBP) (36, 45). However, important concerns raised by employing double tags are risks of proteolysis or aggregation of the protein after cleavage of the first tag.

Deleting the genes of the most abundant contaminants also may not be a viable solution. For example, SlyD, a peptidyl prolyl cis-trans isomerase, is cited as the most frequent IMAC copurified protein, but E. coli B and C strains lacking the expression of SlyD suffer from a significant growth defect (38, 44). Furthermore, the majority of E. coli contaminants are critical for cell viability, especially under the stressed conditions caused by high-yield protein expression (5). Therefore, E. coli knockout strains for particular contaminants (such as chaperones or other stress response factors) have not been seriously considered as solutions for avoiding contamination of His-tagged target protein. As a new strategy, we engineered BL21(DE3), the most widely used E. coli strain for high-yield expression of recombinant protein, to express the major Ni-NTA binding proteins with an alternative tag. As a result, the tagged contaminants can be rapidly removed either before or after IMAC capture of the target protein.

MATERIALS AND METHODS

Media and cultures. All strains were routinely grown using Luria-Bertani (LB) liquid broth or agar (27) at 20°C, 30°C, or 37°C. All media were supplemented with appropriate antibiotics as follows: ampicillin at 100 µg/ml and chloramphenicol 4 μ g/ml for maintaining the mini F' plasmid pFOS1-lacIq or 10 μ g/ml for maintaining pMAK705 constructs. Tests with BL21(DE3) glmS-CBD were performed on tryptone broth or agar (LT) containing 10 g tryptone, 5 g NaCl, and 15 g agar (22, 42), supplemented with appropriate antibiotics and 200 µg/liter N-acetylglucosamine, when necessary, and incubated at 30°C. The AceE-CBD in vivo activity analysis was carried out using M9 minimal medium agar supplemented with 0.2% Casamino Acids, 1 mM MgSO4, 0.3 mg/ml thiamine hydrochloride, 20 µg/ml tryptophan, 0.2% glucose, and, when indicated, 2 mM potassium acetate. The medium used in fermentations was as follows (per liter): 20 g animal-free soy peptone, 10 g yeast extract, 10 g NaCl, 30 g glycerol, 0.75 g KH₂PO₄, 0.75 g Na₂HPO₄, 0.70 g NH₄Cl, 0.085 g K₂SO₄, 0.62 g MgSO₄ · 7H₂O, 0.0125 g Fe(III) citrate, $0.015 \text{ g MnCl}_2 \cdot 7\text{H}_2\text{O}$, $0.0013 \text{ g Zn(CH}_3\text{COO})_2 \cdot 2\text{H}_2\text{O}$, 0.0025 g H₃BO₃, 0.0025 g Na₂MoO₄ \cdot 2H₂O, 0.0025 g CoCl₂ \cdot 6H₂O, 0.0015 g CuCl₂ · 2H₂O, 0.0014 g Na₂EDTA, and 0.05 g PPG Lumulese antifoam.

The fermentation conditions [for the expression of GluRS in BL21(DE3), NiCo21(DE3), and NiCo22(DE3) strains carrying pET21a-gluRS] were as fol-

lows. Ten-liter batch fermentations were performed using Bioflo 3000 fermentors from New Brunswick Scientific. The pH was controlled at 7.00 with automatic addition of 28% NH₄OH and 46 N H₃PO₄. The dissolved oxygen was kept above 20% of air saturation using proportional, integral, and differential control (PID) control of agitation (400 rpm to 600 rpm) and mixing pure oxygen with air (the gas flow rate was 0.5 vessel volume per minute [vvm]). The temperature was controlled at 30°C for both growth and induction. Once the culture reached an optical density at 600 nm (OD₆₀₀) of 8 to 10, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to give a final concentration of 20 μ M and the culture was incubated for an additional 3 h.

For shake flask expression of AlaRS in BL21(DE3) *slyD-CBD*, cells carrying pQE30-*alaRS* and pFOS1-*lacIq* (which provides a source of Lac repressor to control expression from the T5-*lacO* promoter) were grown in LB medium supplemented with 100 µg/ml of ampicillin. After outgrowth at 37°C to an OD₆₀₀ of ~0.8, 500 µM IPTG was added and the cultures were incubated for 20 h at 20°C (OD₆₀₀, ~3.2 to 3.6; cell pellet, ~5.5 to 6 g from 1 liter culture).

For shake flask expression of AlaRS in BL21(DE3), NiCo21(DE3), and NiCo22(DE3), cells were grown in 500 ml of LB medium supplemented with 100 μ g/ml of ampicillin. After incubation at 37°C to an OD₆₀₀ of ~0.8, 200 μ M IPTG was added and the cultures were incubated for 4 h at 30°C (final OD₆₀₀, ~2.2 to 2.6; cell pellet, ~2 g).

For shake flask expression of GluRS in BL21(DE3), NiCo21(DE3), and NiCo22(DE3), cells carrying pET21a-gluRS were grown in 500 ml of LB medium supplemented with 100 μ g/ml of ampicillin. After growth at 37°C to an OD₆₀₀ of \approx 0.5, 20 μ M IPTG was added and the cultures were incubated for 4 h at 30°C (final OD₆₀₀, \approx 2.4 to 2.6). Cells containing pET21a were grown using the same procedure in order to prepare mock lysates for the lysate-mixing experiment (see Fig. 7).

Enzymes and reagents. Restriction enzymes and DNA-modifying enzymes were provided by New England BioLabs. Mutagenesis was carried out using a Phusion site-directed mutagenesis kit (New England BioLabs). DNA amplification procedures utilized either Phusion or *Taq* DNA polymerases.

Strains and plasmids. Bacterial strains and genotypes are listed in Table 1. Oligonucleotides and plasmids are described in Tables S1 and S2, respectively, in the supplemental material. The pMAK-*CBD* vector for 3' end gene tagging was created by inserting the *CBD* open reading frame (ORF) from vector pTYB1 (New England BioLabs) into the polylinker region of pMAK705 (14). An amino acid linker region which encodes LQASSS(N)₁₀LQS, where the first LQ codons correspond to a PstI restriction site and the last LQS codons contain a SalI restriction site, was inserted. A unique AsiSI restriction site was inserted after the *CBD* ORF. All allele exchange procedures utilized derivatives of pMAK-*CBD* containing DNA fragments amplified from the BL21(DE3) chromosome. Genes of interest were cloned into HindIII, SphI, PstI, and/or SacI sites. Downstream DNA (3' flanking sequence) was cloned into AsiSI, Acc65I, SacI, and/or EagI sites.

To create the strain BL21(DE3) *slyD-CBD*, pMAK*slyD-CBD* was transformed into BL21(DE3) and individual clones were grown in LB liquid medium with 4 μ g/ml of chloramphenicol (Cm). The allele exchange method of Hamilton et al. (14) was followed to replace the wild-type (wt) *slyD* gene with the *slyD-CBD* allele. The strains positive for allele exchange were cured of the pMAK vector carrying the wild-type *slyD* gene using a coumermycin treatment (8). PCR amplification with primers 4059For and 4060Rev confirmed that the *slyD-CBD* allele was present at the correct locus within the chromosome of BL21(DE3) *slyD-CBD*.

To create NiCo21(DE3), the allele exchange procedure was carried out in the same manner as that described for *slyD-CBD* except (i) *can* locus analysis was accomplished using primers 4841For and 2187Rev, (ii) analysis of the *arnA-arnD* locus to confirm the *arnA-CBD-arnD* allele was accomplished using primers

5032For and 5031Rev, and (iii) analyses to confirm replacement of the wild type glmS gene by $glmS_{6Ala}$ (glms allele with six histidine-to-alanine substitutions) were accomplished using primers glmSdown-Rev, glmS466check-Rev, Hind-glmS-For, and glmS-Sac-Rev.

To create NiCo22(DE3) from NiCo21(DE3), the pMAKaceE-CBD allele exchange construct was utilized. Primers 4845For and 0076Rev were used to PCR amplify the *aceE-CBD-aceF* locus for sequence characterization.

Mutant glmS constructs were created from pMAKglmS using a Phusion sitedirected-mutagenesis kit (New England BioLabs). The glmS_{24la} gene has four mutations resulting in alanine codons at positions 62 and 65 (<u>GCTCCTCTG</u> <u>GCT</u>, modified from the wt sequence CATCCTCTGCAT). The pMAKglmS6Ala construct was generated in 3 steps. First, the plasmid pMAKglmS2Ala was amplified with primers His432Ala-Rev and His436Ala-For, followed by ligation to circularize the linear PCR product. The resulting glmS_{4Ala} mutant has the mutated sequence <u>GCTGACATTGTGGC</u>, which results in additional alanine codons at positions 432 and 436. Second, the pMAKglmS4Ala mutant template was amplified with primers His466Ala-Rev and glmS467-For, followed by the ligation of the linear PCR product, to generate the plasmid pMAKglmS(His62-65-432-436-466Ala). Third, the pMAKglmS(His62-65-432-436-466Ala) template was amplified with the primers glmS466-Rev and His467Ala-For, followed by ligation of the linear PCR product, to generate the plasmid pMAKglmS6Ala. The glmS_{6Ala} allele has additional alanine codons at positions 466 and 467 (<u>GCTGCCCGCG</u>).

pMS119-*pE* was generated by PCR amplification of the gene *E* from bacteriophage φ x174 genomic DNA (laboratory stock) with the primers Hind-*pE*-For and *pE*-Xba-Rev and then cloned into the HindIII and XbaI sites of pMS119 (11).

All the plasmids were confirmed by DNA sequencing using gene-specific primer s1260 (Ptac forward), s1233 (pMAK705 polylinker forward), or s1224 (pMAK705 polylinker reverse).

Ni²⁺ affinity chromatography using an AKTA fast-performance liquid chromatography (FPLC) system. For the purification of AlaRS from BL21(DE3) *slyD-CBD*, the cell pellet was resuspended in 30 ml of buffer A1 (50 mM HEPES, pH 7.6, 1 M NH₄Cl, 10 mM MgCl₂, 7 mM β-mercaptoethanol, 5% glycerol, and 25 mM imidazole) supplemented with 3 ml BugBuster (10×; Novagen) and sonicated. The clarified lysate was loaded onto a 5-ml HisTrap HP column (GE Healthcare Bio-Sciences), equilibrated with 5 column volumes (CVs) of buffer A1. The flowthrough was collected, and the column was washed with 15 CVs of buffer A1. Elution was completed using a gradient from 25 mM to 250 mM imidazole in buffer B1 (50 mM HEPES, pH 7.6, 0.1 M KCl, 10 mM MgCl₂, 7 mM β-mercaptoethanol, and 250 mM imidazole). Eleven CVs were collected in 28 tubes of 2 ml each.

For the Ni-NTA affinity test with GlmS, $GlmS_{2His-Ala}$, and $GlmS_{6His-Ala}$, cell pellets were resuspended in 6 ml of buffer A2 (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, and 20 mM imidazole) supplemented with 600 μ l BugBuster (10×; Novagen) and sonicated. Clarified lysates were loaded onto a 1-ml His-Trap HP column, equilibrated with 5 CVs of buffer A2. After collecting the flowthrough, the column was washed with 15 CVs of buffer A2 and then eluted using a gradient from 20 mM to 400 mM imidazole with buffer B2 (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, and 400 mM imidazole). A pool of eluted fractions was concentrated using a Vivaspin column (Vivascience) and analyzed on SDS-PAGE gel stained with Coomassie blue R-250.

For the purification of AlaRS from BL21(DE3), NiCo21(DE3), and NiCo22(DE3) strains, the cell pellets were resuspended in 20 ml of buffer A2 (20 mM sodium phosphate, pH 7.4, 500 mM NaCl, and 20 mM imidazole) supplemented with 2 ml BugBuster ($10 \times$; Novagen) and sonicated. Clarified lysates were loaded onto a 5-ml HisTrap HP column equilibrated with 5 CVs of buffer A2. The flowthrough was collected, and the column was washed with 15 CVs of buffer A2. Elution was completed using a gradient from 20 mM to 400 mM imidazole (10 CVs collected in 25 tubes of 2 ml each) with buffer B2.

 Ni^{2+} affinity chromatography-batch process. For the purification of GluRS from BL21(DE3), NiCo21(DE3), and NiCo22(DE3) strains, a cell pellet of 1 g was resuspended in 10 ml of buffer A3 (50 mM sodium phosphate, pH 8, 300 mM NaCl, and 10 mM imidazole) supplemented with 1 ml BugBuster (10×; Novagen) and sonicated. The clarified lysates were loaded onto a 1-ml Ni-NTA column (Qiagen superflow resin), previously equilibrated with 10 CVs of buffer A3. After batch incubation for 1 h at 4°C with gentle shaking, the flowthrough was collected after a short spin and the resin was washed with 10 CVs of buffer A3. Elution was completed with 5 ml buffer B3 (50 mM sodium phosphate, pH 8, 300 mM NaCl, and 250 mM imidazole).

Chitin affinity chromatography. The elution fractions obtained after Ni²⁺ affinity chromatography were pooled, when necessary, and incubated with 1, 2, or 5 ml of chitin beads (New England BioLabs), previously equilibrated with 5 CVs of buffer C (50 mM sodium phosphate, pH 7.4, 500 mM NaCl) at room temperature. After gentle shaking for 30 min at 4°C, the target protein was eluted from

the chitin bead slurry by gravity flow. The chitin beads were then washed with 5 CVs of buffer C, and a chitin bead sample was removed for analysis of bound contaminating proteins.

SDS-PAGE and Western blot analysis. Whole cells or purified protein fractions were prepared for SDS-PAGE analysis by mixing 2 parts sample with 1 part 3× sample buffer (New England BioLabs). Samples were analyzed by 4 to 20% SDS-PAGE. Proteins were visualized by Coomassie blue R-250 staining, or proteins were transferred to nitrocellulose membrane (Millipore) for immunoblotting. CBD-tagged proteins were detected by anti-CBD monoclonal antibody (New England BioLabs), and His-tagged proteins were detected by anti-His monoclonal antibody (EMD Biosciences). Horseradish peroxidase (HRP)-linked secondary antibody, and enhanced chemiluminescence reagents were supplied by Cell Signaling Technology.

MS. Protein samples of approximately 1 mg were added to 20 μ l of trypsin reaction buffer (50 mM Tris-HCl, 20 mM CaCl2, pH 8) and digested overnight at 37°C with trypsin (New England BioLabs) at a protein-to-protease ratio of 20:1. Online liquid chromatography coupled with tandem mass spectrometry (LC-MS-MS) analyses of digested fractions using an Agilent 6330 Ion Trap mass spectrometer with an integrated C18 chromatin immunoprecipitation-nanoelectrospray ionization (C18 ChIP/nano-ESI) interface were performed as described previously (40). Protein separation, digestion, and peptide analysis were repeated in triplicate for each sample. The MS-MS data were analyzed by the Spectrum Mill (Rev A.03.03.084 SR4; Agilent Technologies) search engine using parameters described previously (40), with minor modifications. Data were searched against an E. coli BL21(DE3) database that was supplemented with the three mutant GlmS protein sequences (GlmS_{2Ala}, GlmS_{4Ala}, and GlmS_{6Ala}). The search criteria were set to allow two missed cleavages by a tryptic digest with no other protein modifications. Peptides were validated by using a reverse database search and needed to have a score 2.0 or higher than any reverse score to be valid. Proteins built from these validated peptides scoring 20 or better were considered valid identifications for Spectrum Mill. Proteins identified from one detected peptide using a single spectrum were excluded. The falsepositive rate for all Spectrum Mill analyses was less than 1%.

Protein purity analysis. Target protein purity was analyzed using the Caliper LabChip GXII protein assay. Five microliters of protein solution (1:15 lysate mixture after Ni-NTA or chitin columns) was denatured and processed according to the protocol supplied by Caliper LifeSciences. Protein signals between 14 and 200 kDa were analyzed using HT Protein Express Chip version 2 in combination with a Protein Express reagent kit.

RESULTS

In order to confirm the major E. coli metal binding proteins cited in the literature, three different recombinant His-tagged target proteins were overexpressed from a vector in BL21(DE3). An automated AKTA FPLC system was used in combination with a 1-ml HisTrap column to perform a standard fractionation of the cell lysates. Mass spectrometry (MS) analysis of the target protein elution fractions revealed that the following E. coli host proteins also coeluted in significant amounts: DnaK, GlmS, AceE, EF-Tu, ArnA, RnaseE, AtpF, the Rho transcription terminator, CRP, and SlyD (data not shown). Several independent studies, performed using different conditions or different E. coli strains, also reported many of the same contaminants after Ni⁺² affinity chromatography (5, 13, 18). The most common E. coli proteins listed in these previous reports were CRP, Fur, ArgE, DnaK, SlyD, GlmS, GlgA, ODO1, ODO2, Can (YadF), ArnA (YfbG), AceE, GroES, and GroEL. Based on these previous studies and our analysis, we chose the following consistent contaminants to tag with the chitin binding domain (CBD) sequence: SlyD, GlmS, Can, ArnA, and AceE. The selection criteria also included the possibility of checking the activity of each tagged protein when expressed from the chromosome.

To create the desired expression strains, we first generated a plasmid encoding each candidate gene fused with the *CBD* ORF. The respective *E. coli* gene-*CBD* constructs were ex-

pressed in BL21(DE3) to examine the stability of the respective fusion protein. Anti-CBD immunoblots indicated that each fusion protein was not subject to in vivo proteolysis (data not shown), an outcome that would prevent the removal of the contaminant protein by tag-mediated chromatography. The plasmids encoding CBD fusion proteins were subsequently employed to modify the BL21(DE3) chromosome by homologous recombination at the native E. coli gene locus. The replacement of the native gene with the CBD fusion allele was performed using the allele exchange method described by Hamilton et al. (14). Efficient allele exchange occurs when the exchange vector contains at least 300 bp of homology to both 5' and 3' regions flanking the target site on the bacterial host chromosome. Thus, for each allele replacement step, homologous DNA sequences of at least 300 bp were cloned at the 5' and the 3' ends of the CBD sequence in the pMAK-CBD vector (see Materials and Methods). E. coli BL21(DE3) derivatives were generated by replacing each candidate allele one by one. The phenotype of each derivative strain was analyzed after each replacement step, and then each strain was tested as an expression host for one or more recombinant His-tagged proteins (E. coli alanyl-tRNA synthetase and/or E. coli glutamyl-tRNA synthetase).

(i) Major Ni-NTA contaminant SlyD (CBD-tagged) is removed by incubation with chitin beads. As a proof of principle, we chose first to address the most predominant Ni-NTA contaminant, SlyD. SlyD is a cytoplasmic protein originally identified in E. coli as a host factor required by bacteriophage phiX174 to induce cell lysis. The gene slyD (sensitivity to lysis) was first identified in a genetic selection for E. coli C strains resistant to lysis gene E of phage phiX174 (23). The protein was then isolated and characterized in a further study as a persistent contaminant in immobilized metal affinity chromatography that migrates at the apparent molecular mass of 27 kDa and was also called WHP for wondrous histidine-rich protein (44). With 196 residues and an actual mass of 21 kDa, SlyD is a peptidyl prolyl cis/trans-isomerase (PPIase or rotamase) that provides chaperone activity for proline-limited protein folding (17, 20, 39). The protein is divided into two domains, the N-terminal part (residues 1 to 146) containing the chaperone activity with the peptidyl-prolyl isomerase domain and the C-terminal part of 50 amino acids, particularly rich in histidine and cysteine residues and thus conferring high affinity for divalent cations such as Zn^{2+} and Ni^{2+} (44). SlyD has been reported as the major contaminant after Ni-NTA purification in several studies (3, 10, 19, 31, 35).

The strain BL21(DE3) *slyD-CBD* was constructed using the Hamilton allele exchange method (14) and then tested as a host for overexpression of alanyl-tRNA synthetase (AlaRS). AlaRS tagged at its N terminus with 6 histidines was expressed from the pQE30 vector in the control strain BL21(DE3) and in BL21(DE3) *slyD-CBD*, each carrying the mini-F plasmid pFOS1-*lacIq* as described in Materials and Methods. The cells were sonicated, and the clarified lysates were loaded onto a 5-ml HisTrap column using an AKTA FPLC system, followed by washing with 25 mM imidazole and elution using a gradient of 25 to 250 mM imidazole. Figure 1 shows Coomassie blue-stained gels with the protein profiles of the elution fractions collected from both strains, in which AlaRS is sequentially eluted. Comparable backgrounds of *E. coli* proteins are found

upon fractionation of both cell lysates, except for one protein that migrates at about 35 kDa in BL21(DE3) slyD-CBD (Fig. 1A and B). This protein was presumed to be SlyD-CBD since wild-type SlyD tends to migrate at 27 kDa and the CBD tag is 7.7 kDa. In fact, a Western blot analysis performed with anti-CBD antibodies on the same BL21(DE3) slyD-CBD fractions confirms the position of SlyD-CBD at about 35 kDa (Fig. 1C). These data indicate that SlyD-CBD is visually detectable as a contaminant of the Ni-NTA fractions and is eluted with 80 to 120 mM imidazole. This result is consistent with the data presented by Bolanos-Garcia et al. (5). Elution fractions 4 to 24 were pooled and then incubated with 2-ml chitin beads for 2 h at 4°C. The resulting flowthrough (FT) fraction and the chitin beads were analyzed by Western blotting, which shows that a significant amount of SlyD-CBD is removed from the protein pool after incubation with chitin beads (Fig. 2). The analysis for in vivo SlyD-CBD activity is presented in "(iv) In vivo activity analysis of each CBD-tagged candidate protein."

(ii) Mutation of 6 surface exposed histidines eliminates the binding of GlmS to nickel. GlmS is a 67-kDa L-glutamine:D-fructose-6-phosphate aminotransferase involved in an essential step of bacterial cell wall biosynthesis. The enzyme utilizes D-fructose-6-phosphate and L-glutamine to form D-glucos-amine-6-phosphate, which is a precursor for components of the peptidoglycan or lipopolysaccharide. Among 609 residues, GlmS contains 24 histidines, of which 15 are organized in four clusters of at least 3 histidines exposed on the surface, giving the protein a high potential for interacting with Ni²⁺ and Co²⁺ cations according to the crystal structure described by Mouilleron et al. (28).

When we replaced the *glmS* allele with the *glmS-CBD* allele on the chromosome of BL21(DE3), the resulting BL21(DE3) glmS-CBD strain was nearly nonviable (Fig. 3B). However, complementation of the GlmS-CBD defect was demonstrated using tryptone medium (LT) with addition of 200 µg/ml of glucosamine (GlcN) or N-acetylglucosamine (GlcNAc) (Fig. 3B) (42, 43). GlmS is actually composed of two domains in which the N-terminal domain holds the glutamine amidotransferase activity while the C-terminal domain contains the ketose/aldose isomerase activity. Respective active site residues are located at each extremity of the protein (position 2 for the amidotransferase and position 604 [of 609 residues] for the isomerase) (Fig. 3A). Since our attempt to place the CBD tag at the N-terminal end of GlmS was also unsuccessful, we assumed that the addition of the CBD tag at the N or C terminus of GlmS disturbs the respective active sites located at the termini of the protein. Rather than using a tag to eliminate GlmS from His-tagged protein fractions, we chose to alter GlmS affinity for divalent cations by generating a GlmS mutant in which the most exposed histidines would be replaced by alanines. After examining the three-dimensional (3D) structure of the GlmS dimer, the isomerase active form of GlmS (28), we selected six surface-exposed histidines for mutagenesis (Fig. 3A): histidines 62 and 65 are located in the center of a cluster of 4 exposed histidines, histidines 432 and 436 occupy a central position in another cluster, and finally, histidines 466 and 467 are part of a large cluster of 6 histidines highly surface exposed and generated by the dimer interface (data not shown). The selected histidines (positions 62, 65, 432, 435, 466,



FIG. 1. Purification of alanyl-tRNA synthetase (AlaRS), expressed in BL21(DE3) and BL21(DE3) slyD-CBD, by Ni²⁺ affinity chromatography. (A) Coomassie blue-stained SDS-PAGE gel of elution fractions obtained from a cell extract of BL21(DE3) slyD-CBD expressing His-tagged AlaRS. (B) Coomassie blue-stained SDS-PAGE gel of elution fractions obtained from a cell extract of BL21(DE3) slyD-CBD expressing His-tagged AlaRS. (C) Western blot analysis of fractions shown in panel B probed with anti-CBD antibody. Lane M, molecular mass marker in kDa; lane FT, flowthrough after Ni-NTA; lanes 1 to 23, eluted fractions from the IMAC column using a gradient from 25 mM to 250 mM imidazole. AlaRS, wild-type SlyD (SlyDwt), and SlyD-CBD are indicated by a dash.

and 467) are all poorly conserved residues among GlmS homologs (data not shown).

We first examined the activity of the resulting $GlmS_{6His-Ala}$ protein in a complementation assay using the strain BL21(DE3) *glmS-CBD*. As shown in Fig. 3B, in the absence of GlcN or GlcNAc, $GlmS_{6His-Ala}$ is able to restore the growth of BL21(DE3) *glmS-CBD* with the same efficiency as wild-type GlmS. Both proteins, GlmS and $GlmS_{6His-Ala}$, were expressed from the low-copy-number plasmid pMAK705 (14). Next, to test the affinity of $GlmS_{6His-Ala}$ for Ni-NTA resin, we performed a standard purification on a Ni-NTA column, using BL21(DE3) lysate as a control compared to BL21(DE3) overexpressing $GlmS_{6His-Ala}$ or GlmS from pMAK705. The cell lysates were loaded on a 1-ml HisTrap column using an AKTA FPLC system, followed by standard washing and elution steps (20 to 400 mM imidazole), and the elution fractions were analyzed by Coomassie blue SDS-PAGE and by MS. As an additional control, we included an intermediate mutant protein, $GlmS_{2His-Ala}$, in which only the histidines at positions 62 and 65 were replaced by alanines. Figure 3C shows that the band migrating at the expected position of GlmS (67 kDa) is detected in the sample overexpressing GlmS (lane "wt"), while the same band is less intense for the $GlmS_{2His-Ala}$ sample (lane labeled "2_{ala}"). Most importantly, overexpression of the $GlmS_{6His-Ala}$ mutant protein does not result in binding and elution of a 67-kDa protein from the Ni-NTA resin [compare lane "6_{Ala}" with lane "–" where cell extract was analyzed from BL21(DE3) carrying an empty plasmid]. Additionally, MS



FIG. 2. Purification of alanyl-tRNA synthetase (AlaRS), expressed in BL21(DE3) and BL21(DE3) *slyD-CBD*, on Ni²⁺ affinity chromatography followed by chitin bead treatment. The left panel represents SDS-PAGE analysis, and the right panel shows a Western blot analysis performed with anti-CBD antibodies. Lane M, molecular mass marker in kDa; lane L, cell lysate load; lane P, pool of Ni-NTA fractions 4 to 24; lane FT_C, flowthrough of chitin beads; lane B, boiled chitin beads. AlaRS, wild-type SlyD (SlyDwt), and SlyD-CBD are indicated by arrows.

analysis of the Ni-NTA elution fractions supports the SDS-PAGE observations. When analyzing Ni-NTA binding proteins in the control samples, we found 38 times more spectral counts of GlmS when wild-type GlmS is overexpressed from a plasmid than for chromosomal expression in the empty-plasmid control strain. In contrast, overexpression of mutant GlmS_{6His-Ala} from a plasmid results in only a modest increase of GlmS peptide in the imidazole elution fraction (4 spectral counts). These data indicate that the replacement of 6 surface-exposed histidines by alanines significantly decreases the affinity of GlmS for Ni-NTA resin, while the removal of only 2 histidines has a limited effect. The complementation studies show that this mutagenesis does not compromise the ability of GlmS to supply GlcNAc (Fig. 3B).

(iii) Construction of NiCo21(DE3) and NiCo22(DE3). Replacement of the *glmS* gene with the *glmS*_{6Ala} allele was the last step in the construction of the NiCo21(DE3) and NiCo22(DE3) protein expression strains. NiCo21(DE3) additionally contains the *slyD*, *can*, and *anA* genes tagged with the *CBD* ORF, while NiCo22(DE3) additionally contains the *slyD*, *can*, *amA*, and *aceE* genes tagged with the *CBD* ORF. The *can* gene (*yadF*) encodes a β -class *carbonic anhydrase* (Can), a zinc metalloenzyme which interconverts carbon dioxide (CO₂) and bicarbonate. This protein of 25 kDa exhibits four zinc-binding sites that confer a significant affinity for metal chelating resins. Moreover, Can expression increases in high-density cultivation, during slow growth, or during stress and starvation, in other words, conditions typically encountered during recombinant-protein overexpression (26).

The *arnA* gene encodes a bifunctional enzyme of 74.3 kDa, a UDP-L-Ara4N formyltransferase/UDP-GlcA *C*-4"-decarboxylase, involved in the modification of the lipid A required for lipopolysaccharide biosynthesis. Also this modification of lipid A with 4-amino-4-deoxy-L-arabinose confers to Gram-negative bacteria a resistance to the cationic antimicrobial peptides and antibiotics such as polymyxin (6). ArnA (also called YfbG) is found as a recurring contaminant in IMAC, presumably due to the several clusters of surface-exposed histidines detected in the 3D structure of the active hexamer (12).

The *aceE* gene encodes subunit E1 of the pyruvate dehydrogenase multienzyme complex formed from 12 dimers of subunit E1, 24 subunits of AceF, and 6 LpdA dimers (2). The major role of pyruvate dehydrogenase in the tricarboxylic acid (TCA) cycle is the production of acetyl-coenzyme A (CoA) from pyruvate (37). Although AceE (99.7 kDa) is not essential for viability, its inactivation reportedly leads to disturbance of carbon metabolism (21). The protein displays three magnesium-binding sites that might explain its affinity for metal chelating resins.

For each intermediate strain, we examined the growth rate at diverse temperatures (20°C, 30°C, and 37°C) in order to detect any effect on cell viability. We confirmed the expression and tested the activity of each CBD tagged protein *in vivo*. In Fig. 4A, the anti-CBD Western blot demonstrates the stable expression of the CBD-tagged proteins in NiCo21(DE3) and NiCo22(DE3). However, we observed that after addition of the *aceE-CBD* allele, NiCo22(DE3) grew more slowly in liquid culture and on agar plates than BL21(DE3) and NiCo21(DE3) (Fig. 4B). This suggests that the addition of a CBD tag at the C terminus of AceE affects the activity of the individual AceE protein and/or the formation of the pyruvate dehydrogenase complex.

During NiCo strain construction and thereafter, genome stability was verified by PCR and sequence analysis of the loci containing the CBD-tagged genes. When identical DNA sequence tags are present at either 3 or 4 times in a genome, it



FIG. 3. BL21(DE3) *glmS-CBD* and GlmS mutant characterization. (A) Representation of the GlmS protein showing the two enzymatic domains, the two active sites, and the locations of the histidines (His) selected for replacement by alanines. Numbers indicate the amino acid residues. (B) Bacterial growth assay of BL21(DE3) and BL21(DE3) *glmS-CBD* and complementation assays performed with BL21(DE3) *glmS-CBD* carrying pMAK705 (empty vector), pMAK*glmS*, or pMAK*glmS6Ala*. The assay was carried out by spotting 5 μ l of each overnight culture, diluted from 10⁻² to 10⁻⁷, on LT medium with or without *N*-acetylglucosamine (GlcNAc; 200 μ g/liter). The plates were then incubated 24 h at 30°C. (C) SDS-PAGE of elution fractions obtained after loading cell lysates containing GlmS or mutant GlmS onto a 1 ml HisTrap HP column. This affinity test of GlmS, GlmS_{2Ala}, and GlmS_{6Ala} for Ni-NTA was performed using cell lysate from 200 ml culture of each strain grown at 30°C and induced with 0.5 mM IPTG for 17 h at 20°C. Lane M, molecular mass marker in kDa; lane –, BL21(DE3) expressing GlmS from the chromosome (empty plasmid); lane wt, BL21(DE3) overexpressing GlmS from pMAK705; lane 6Ala, BL21(DE3) expressing GlmS_{6Ala} from pMAK705; lane 2, BL21(DE3) overexpressing GlmS_{2Ala} from pMAK705. The position of GlmS is indicated by an arrow. The table indicates the number of spectra of GlmS identified by mass spectrometry analysis in each elution fraction. The presented ratio was estimated after normalizing the total protein spectral intensity value obtained for GlmS to that for the control sample "–."

is certainly possible for recombination to occur, resulting in deletion of intervening sequence to occur. In the case of the NiCo strains, this type of event would result in cell death (and loss of such mutants from the culture) since intervening sequences are extensive and contain essential genes.

(iv) In vivo activity analysis of each CBD-tagged candidate protein. To assess the activity of the SlyD-CBD protein, we used an *in vivo* approach involving the lysis protein E originally expressed from the bacteriophage phiX174 genome. SlyD is required to induce lysis of *E. coli* C after infection by the phage phiX174 or in other *E. coli* strains (background B, C, and K-12) when protein E is expressed from a multicopy plasmid (38, 46). Wild-type *E. coli* is sensitive to lysis in both situations, while a *slyD* knockout strain is shown to be resistant. The lysis is actually the consequence of the inhibition, by protein E, of MraY, a conserved phosphotransferase involved in the formation of intermediates for the peptidoglycan biosynthesis (4, 25, 47). The lysis phenotype in BL21(DE3) and in the strains carrying *slyD-CBD* was examined by monitoring OD₆₀₀ following induction of expression of the protein E cloned on pMS119 plasmid. Figure 5A shows that all the strains display a sensitivity to lysis after induction of the protein E, indicating that SlyD-CBD is active. Lysis depends on lysis protein E, since it does not occur with empty-vector pMS119.



FIG. 4. Growth curves and detection of CBD-tagged proteins. (A) Western blot analysis (anti-CBD) of total cell extracts from three independent clones (1, 2, 3) of NiCo21(DE3) and NiCo22(DE3). The CBD-tagged proteins are indicated by arrows. (B) Growth of BL21(DE3), NiCo21(DE3), and NiCo22(DE3) performed at 30°C in LB medium.

The Can protein is essential for growth under normal atmospheric conditions, even though cells depleted of Can are able to survive when supplied with CO_2 or when cynT, a paralog gene of *can*, is activated upon induction with cyanate or azide (15, 26). The isolated strain BL21(DE3) *slyD-CBD can-CBD* did not show evidence of a growth defect on plates or under standard liquid culture conditions, suggesting that Can expressed with a C-terminal CBD tag is fully functional (data not shown).

Given that the enzymatic modifications carried out by ArnA confer polymyxin resistance to *E. coli*, we tested the activity of ArnA-CBD by restreaking the strains on LB medium supplemented with polymyxin B. The assays were performed at a concentration of 2 μ g/ml of polymyxin B as described in the literature (7). Figure 5B shows that this is the MIC for NiCo21(DE3) at 30°C, but at 37°C, the strain is fully sensitive to the same concentration. This result suggests that ArnA-CBD is generally less active than wild-type ArnA. Interestingly, NiCo22(DE3), which also expresses ArnA-CBD but contains one more CBD-tagged protein (AceE-CBD), is resistant to polymyxin B at both 30°C and 37°C temperatures. The noticeable phenotype of NiCo22(DE3) is its lower growth rate than those of NiCo21(DE3) and BL21(DE3) (Fig. 4B). We consequently hypothesize that the activity of ArnA-CBD may be

compromised and that slow growth allows for this protein to carry out its functions in cell envelope formation.

Depletion of AceE or elimination of its pyruvate dehydrogenase activity results in E. coli dependent on acetate for aerobic growth on glucose medium (21). We therefore performed a plate assay on minimal medium supplemented with glucose (0.2%) with and without potassium acetate (2 mM) to access AceE-CBD activity. As controls, we used two E. coli aceE mutants, the strains CGSC5477 and CGSC4823 (9, 16), which both show a growth defect in the absence of acetate. Despite the fact that NiCo22(DE3) exhibits a lower growth rate than BL21(DE3) and NiCo21(DE3), its growth is not acetate dependent (Fig. 5C). These results suggest that the CBD tag at the C terminus of AceE does not significantly affect AceE activity. If AceE-CBD activity was abolished, NiCo22(DE3) growth could be supported in the absence of acetate by a higher level of expression of the protein PoxB. PoxB is a nonessential pyruvate oxidase whose role of converting pyruvate into acetate and CO₂ contributes significantly to growth under aerobic conditions. Abdel-Hamid et al. identified some aceE mutants able to grow without acetate due to an increased activity of endogenous PoxB throughout the growth cycle (with yet a 30% decrease in the growth rate compared to the level for the wild-type strain) (1). Although the nature of the constitutive expression has not been identified, they proposed that the *poxB* promoter lost its dependence on RpoS, a sigma factor (RpoS, σ^{38} , and σ^{S}) required to induce genes in stationary phase (1). To address this uncertainty, we sequenced the poxB promoter of BL21(DE3), NiCo21(DE3), and NiCo22(DE3) and found no mutations in the promoter region of these strains, suggesting that PoxB is still well regulated by RpoS in NiCo22(DE3).

(v) Recombinant protein expression by NiCo21(DE3) and NiCo22(DE3). NiCo21(DE3) and NiCo22(DE3). were used to express the *E. coli* His-tagged alanyl-tRNA synthetase (AlaRS) or the His-tagged glutamyl-tRNA synthetase (GluRS) using two conditions of expression (standard shake flask cultures and batch fermentation). The purification procedure was performed on immobilized nickel using either nickel beads (Superflow; Qiagen) or a HisTrap column managed by an AKTA FPLC system. In both cases, Ni-NTA elution fractions were incubated with chitin beads and the protein profiles of the resulting flowthrough were compared to that of the Ni-NTA elution pool (no chitin incubation). Proteins were characterized by Coomassie blue-stained SDS-PAGE gels by Western blot analysis and by MS.

Glutamyl-tRNA synthetase (GluRS-6His) was expressed in each expression strain using a high-density batch fermentation process carried out at 30°C in rich medium monitored for glucose concentration and oxygen (see Materials and Methods). Figure 6 shows the purification of the GluRS-6His on Ni-NTA resin (Superflow; Qiagen) followed by incubation with chitin beads. The profiles of proteins eluted after Ni-NTA (lanes E) are similar for the three strains (Fig. 6A). After incubation with chitin beads (lanes FT chitin), the collected flowthrough is void of the main contaminants for NiCo21(DE3) and NiCo22(DE3) samples. The Western blot shown in Fig. 6B confirms the removal of the CBD-tagged proteins after treatment on chitin beads (compare lane E and lane FT chitin). Finally, MS analyses of the samples taken after Ni-NTA elu-



FIG. 5. Activity assays for SlyD-CBD, ArnA-CBD, and AceE-CBD. (A) SlyD-CBD activity is confirmed by lysis sensitivity of BL21(DE3), NiCo21(DE3), and NiCo22(DE3) upon induction of protein E. Cultures of these strains harboring the plasmid pMS119-*pE* or pMS119 were induced at an OD₆₀₀ of 0.15 to 0.55 with 500 μ M IPTG, indicated by an arrow, and monitored for optical density. (B) Bacterial growth assays with BL21(DE3), NiCo21(DE3), and NiCo22(DE3) were carried out by spotting 5 μ l of overnight culture, diluted from 10⁻² to 10⁻⁷, on LB medium with or without 2 μ g/ml polymyxin B. The plates were incubated for 24 h at 30°C or 37°C. (C) Bacterial growth assays with BL21(DE3), NiCo21(DE3), and two *aceE* mutants, CGSC5477 and CGSC4823, were carried out by spotting 5 μ l of overnight culture, diluted 10⁻² to 10⁻⁷, on minimal medium with 0.2% glucose and supplemented with 2 mM potassium acetate when indicated.

tion and after chitin flowthrough are consistent with our gel analyses, showing that, first, the CBD-tagged proteins identified in the elution after Ni-NTA are absent in the chitin flow-through and, second, that $GlmS_{6His-Ala}$ is significantly reduced in the NiCo21(DE3) and NiCo22(DE3) samples compared with GlmS identified in BL21(DE3) (Fig. 6C).

To demonstrate the utility of the NiCo strains for poorly expressed proteins, GluRS was expressed in BL21(DE3), NiCo21(DE3), and NiCo22(DE3) and each cell lysate was then mixed with a "mock" lysate prepared from the same strain grown with an empty expression vector (pET21a). A mixture of 1 ml GluRS lysate to 15 ml empty-vector lysate was chosen so that GluRS would be present at a very low concentration relative to host proteins (see lysate load [L] lanes in Fig. 7). The lysate mixture (1:15) corresponding to all three strains was subjected to standard Ni-NTA chromatography (wash with 8 CVs of 20 mM imidazole and elution with 1 CV of 250 mM imidazole). The Ni-NTA elution samples resulting from NiCo21(DE3) and NiCo22(DE3) were additionally incubated with chitin resin for 30 min. Lane 3 in Fig. 7 shows the protein profile of the Ni-NTA elution sample from the BL21(DE3) lysate mixture, while lanes 6, 7, 10, and 11 show the improvement in purity obtained when NiCo strain elution samples were further processed by chitin incubation. Visual inspection of the SDS-PAGE gel image shows that the NiCo samples (Fig. 7, lanes 7 and 11) have significantly less contaminants than the BL21(DE3) sample. To confirm the improvement in purity, we measured target protein purity using a Caliper LabChip GXII protein assay. Strikingly, the target protein is 90.74% and 87.48% pure after chitin incubation when purified from NiCo21(DE3) and NiCo22(DE3), respectively. When GluRS is expressed in BL21(DE3), the purity is only 56.01% when a standard Ni-NTA purification procedure is followed.



FIG. 6. Purification of the glutamyl-tRNA synthetase (GluRS) expressed in NiCo21(DE3) and NiCo22(DE3). Strains BL21(DE3), NiCo21(DE3), and NiCo22(DE3) expressing GluRS(6His) from pET21a were propagated in a high-density fermentor as indicated in Materials and Methods. One gram of each cell pellet was processed to purify GluRS(6His) using 1 ml Ni-NTA superflow resin followed by 1 h of batch incubation on chitin beads (1 ml). (A) Samples collected during the procedure were analyzed by SDS-PAGE. Lane M, molecular mass marker in kDa; lane L, cell lysate load; lane FT, flowthrough; lane W, first wash; lane B, boiled chitin beads; lane E, Ni-NTA elution. (B) Western blot analysis (anti-CBD) of the samples shown in panel A. GluRS, wild-type SlyD (SlyDwt), SlyD-CBD, ArnA-CBD, Can-CBD, and AceE-CBD are indicated by a symbol or an arrow. (C) The table indicates the number of spectra of each protein identified by mass spectrometry analysis in the elution fractions after Ni-NTA and in the flowthrough after chitin bead treatment.

In addition, we performed purification analyses using imidazole elution gradients and found that contaminants are equally problematic compared to single-step elution. Histagged alanyl-tRNA synthetase (AlaRS) was purified by loading cell lysates on a 5-ml HisTrap column, followed by standard washing with 20 mM imidazole and elution with a 20 to 400 mM imidazole gradient using an AKTA FPLC system. Elution fractions with high concentrations of target protein consistently displayed multiple contaminating proteins. SlyD-CBD and ArnA-CBD were eluted at about 80 to 140 mM imidazole, while AceE-CBD was eluted all along the elution gradient (see Fig. S2 in the supplemental material). Only Can-CBD appears to be washed out before the elution step, as confirmed by Western blotting. Importantly, incubating the target protein pool with chitin beads resulted in the removal of CBD-tagged contaminant proteins (Fig. S2).

NiCo21(DE3) and NiCo22(DE3) were initially designed to remove the most common contaminants found after Ni-NTA



FIG. 7. Utility of the NiCo strains for poorly expressed proteins. Coomassie blue-stained SDS-PAGE gel of samples taken during the purification of GluRS on Ni-NTA superflow resin (2 ml) followed by 30 min of batch incubation on chitin beads (2 ml). The purifications were performed from a mixed cell lysate containing 1 ml GluRS lysate and 15 ml of empty-vector lysate from the same strain. Lanes 1, 4, and 8 represent lysate mixtures loaded (L) onto Ni-NTA columns. Each mixture contains a low concentration of GluRS which becomes the prominent band in the 250 mM imidazole elution sample (E). ft, Ni-NTA flowthrough; FT_C , sample taken of the void volume after incubating the Ni-NTA elution sample with chitin beads.

purification. In order to determine the efficiency of protein purification on the alternative cobalt resin, we performed a comparison assay between Ni-NTA (Superflow; Qiagen) and cobalt (Talon superflow; Clontech) resins using the standard protocols recommended by the respective manufacturers. The amounts of purified His-tagged protein (glutamyl-tRNA synthetase) were equivalent from both resins, but the cobalt resin often gave elution fractions less contaminated by native E. coli proteins (data not shown). Although SlyD was reported as having a very weak affinity for cobalt affinity resin (24), we found after mass spectrometry analysis that the most abundant contaminant on cobalt resin is still SlyD (data not shown). Among the E. coli proteins copurified with His-tagged GluRS on cobalt, we also identified the elongation factor EF-Tu and the ferric uptake regulator protein (Fur) known to bind Ni-NTA resin.

DISCUSSION

In this work, we present the characterization of two protein expression strains designed to improve purity of target protein after immobilized metal affinity chromatography (IMAC). BL21(DE3) is the most widely used *E. coli* strain for protein overexpression and often provides the highest yield of target protein relative to endogenous proteins. For this reason, BL21(DE3) was chosen as the parent for the NiCo strains. NiCo21(DE3) and NiCo22(DE3) were both engineered to express the most common contaminants (SlyD, Can, ArnA, GlmS, and AceE) either with a tag for removal by a rapid chromatography flowthrough step or with mutations to decrease affinity for divalent cations. We have demonstrated that all the CBD-tagged proteins (SlyD, Can, ArnA, and AceE) are efficiently removed from the Ni-NTA elution pool after batch incubation with chitin beads and that the modified $GImS_{6His-Ala}$ protein has lost its affinity for nickel resin. NiCo21(DE3) and NiCo22(DE3) are thus preferred expression strains for obtaining recombinant His-tagged target proteins with reduced levels of host protein contamination.

We have demonstrated that the addition of a CBD tag at the C-terminal end of SlyD and Can protein does not affect their function in vivo. Importantly, addition of the CBD tag may have favorably altered the affinity of Can for nickel resin since we routinely observed Can-CBD in the 20 mM imidazole flowthrough or wash fractions (see Fig. S2 in the supplemental material), whereas Bolanos-Garcia et al. report that wild-type Can is eluted from Ni-NTA by 55 to 80 mM imidazole (5). In NiCo21(DE3), ArnA-CBD was observed to be partially active at 30°C but apparently inactive at 37°C according to the polymyxin B sensitivity phenotype. However, in the slow-growing strain NiCo22(DE3), the function of ArnA-CBD appears normal at 37°C. One explanation for this inconsistency is that at a higher growth rate, cell envelope formation (and resistance to polymyxin B) demands an ArnA protein with a higher specific activity. Therefore, our results suggest that the CBD tag only compromises and does not abolish the function of ArnA in vivo.

In the final strain NiCo22(DE3), the CBD tag modification of AceE seems to be responsible for the lower growth rate of this strain. However, the lack of acetate dependence for aerobic growth on glucose medium suggests that AceE-CBD is still active. On the other hand, we designed our acetate dependency assay based on data collected from experiments performed with E. coli K-12, since the two available control aceE mutants are K-12 strains (21). K-12 and B strains are known to contain major differences in their respective metabolic pathways, particularly in pyruvate and acetate production (33, 34). For example, in high-density cultures, E. coli K-12 excretes high levels of acetate in the presence of excess glucose concentrations, which affects its growth rate, while E. coli BL21 is much less sensitive to glucose concentration and produces lower levels of acetate (32). It seems that not only are the lower levels of production of pyruvate and acetate in E. coli BL21 the result of more-active carbon metabolism (TCA cycle, glyconeogenesis, glyoxylate shunt, and anaplerotic pathways), but glucose transport may be better controlled in BL21 than in K-12 (30, 33). In conclusion, the acetate assay indicates that AceE-CBD is most likely active.

One advantage of using chitin beads is their compatibility with diverse buffer conditions (HEPES, sodium phosphate, Tris-HCl, 50 mM to 2 M NaCl, 6 M urea, pH 6 to 9, 0.1 to 1% Triton X-100, and high imidazole concentrations) that allow direct loading of the elution fractions after IMAC without performing dialysis. In the experiment using cobalt resin, SlyD-CBD was directly cleared from the elution sample after 30 min batch incubation on chitin beads (data not shown). We routinely use 1-ml chitin beads to remove the CBD-tagged proteins from the Ni-NTA elution pool prepared from 1 g of cell pellet. For large cultures (2 liters or more) or especially when the recombinant His-tagged protein is expressed at a low level, batch incubation in a larger volume of chitin beads may help to efficiently remove all the CBD tagged proteins (1 ml chitin beads for every 0.1 liter of culture is another recommendation). However, increasing the incubation time with the beads (e.g., overnight) did not show any improvement (data not shown).

We show that mutating surface-exposed histidines on GlmS is an effective strategy for eliminating binding to Ni-NTA. However, we have not followed this as a general approach, since we expect that many proteins would not tolerate removal of multiple histidines.

In summary, NiCo21(DE3) is indistinguishable from BL21(DE3) with respect to growth characteristics and the potential for recombinant protein expression. The NiCo strains consistently produced a high yield of recombinant protein in all of our studies, and the most common IMAC contaminants were removed by simply exposing the target protein pool to chitin beads, which are compatible with commonly used IMAC buffers. Therefore, we propose that the NiCo strains are superior alternatives to BL21(DE3) as protein production hosts, and we expect that these strains will perform well under all conditions routinely employed for the propagation of BL21(DE3).

ACKNOWLEDGMENTS

We thank Mern Sibley for introduction of the *fhuA2* (T1 phage resistance) allele to BL21(DE3) and for providing the *aceE* mutant strains, Elisabeth Raleigh and Bill Jack for critical reading of the manuscript, Chris Taron for guidance throughout the project, and Don Comb and New England BioLabs for research support.

REFERENCES

- Abdel-Hamid, A. M., M. M. Attwood, and J. R. Guest. 2001. Pyruvate oxidase contributes to the aerobic growth efficiency of *Escherichia coli*. Microbiology 147:1483–1498.
- Angelides, K. J., S. K. Akiyama, and G. G. Hammes. 1979. Subunit stoichiometry and molecular weight of the pyruvate dehydrogenase multienzyme complex from *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. 76:3279–3283.
- Arnold, F. H. 1991. Metal-affinity separations: a new dimension in protein processing. Biotechnology (N. Y.) 9:151–156.
- Bernhardt, T. G., W. D. Roof, and R. Young. 2000. Genetic evidence that the bacteriophage ΦX174 lysis protein inhibits cell wall synthesis. Proc. Natl. Acad. Sci. U. S. A. 97:4297–4302.
- Bolanos-Garcia, V. M., and O. R. Davies. 2006. Structural analysis and classification of native proteins from *E. coli* commonly co-purified by immobilised metal affinity chromatography. Biochim. Biophys. Acta 1760:1304– 1313.
- Breazeale, S. D., A. A. Ribeiro, and C. R. Raetz. 2002. Oxidative decarboxylation of UDP-glucuronic acid in extracts of polymyxin-resistant *Escherichia coli*. Origin of lipid a species modified with 4-amino-4-deoxy-L-arabinose. J. Biol. Chem. 277:2886–2896.
- Bretscher, L. E., M. T. Morrell, A. L. Funk, and C. S. Klug. 2006. Purification and characterization of the L-Ara4N transferase protein ArnT from Salmonella typhimurium. Protein Expr. Purif. 46:33–39.
- Chen, M., K. Xie, N. Nouwen, A. J. Driessen, and R. E. Dalbey. 2003. Conditional lethal mutations separate the M13 procoat and Pf3 coat functions of YidC: different YIDC structural requirements for membrane protein insertion. J. Biol. Chem. 278:23295–23300.
- Dietrich, J., and U. Henning. 1970. Regulation of pyruvate dehydrogenase complex synthesis in *Escherichia coli* K 12. Identification of the inducing metabolite. Eur. J. Biochem. 14:258–269.
- Finzi, A., J. Cloutier, and E. A. Cohen. 2003. Two-step purification of Histagged Nef protein in native condition using heparin and immobilized metal ion affinity chromatographies. J. Virol. Methods 111:69–73.
- Furste, J. P., et al. 1986. Molecular cloning of the plasmid RP4 primase region in a multi-host-range *tacP* expression vector. Gene 48:119–131.
- Gatzeva-Topalova, P. Z., A. P. May, and M. C. Sousa. 2005. Structure and mechanism of ArnA: conformational change implies ordered dehydrogenase mechanism in key enzyme for polymyxin resistance. Structure 13:929–942.
- Graslund, S., et al. 2008. The use of systematic N- and C-terminal deletions to promote production and structural studies of recombinant proteins. Protein Expr. Purif. 58:210–221.
- Hamilton, C. M., M. Aldea, B. K. Washburn, P. Babitzke, and S. R. Kushner. 1989. New method for generating deletions and gene replacements in *Escherichia coli*. J. Bacteriol. 171:4617–4622.
- 15. Hashimoto, M., and J. Kato. 2003. Indispensability of the Escherichia coli

carbonic anhydrases YadF and CynT in cell proliferation at a low CO2 partial pressure. Biosci. Biotechnol. Biochem. **67**:919–922.

- Henning, U., G. Dennert, R. Hertel, and W. S. Shipp. 1966. Translation of the structural genes of the *E. coli* pyruvate dehydrogenase complex. Cold Spring Harbor Symp. Quant. Biol. 31:227–234.
- Hottenrott, S., T. Schumann, A. Pluckthun, G. Fischer, and J. U. Rahfeld. 1997. The *Escherichia coli* SlyD is a metal ion-regulated peptidyl-prolyl cis/trans-isomerase. J. Biol. Chem. 272:15697–15701.
- Howell, J. M., T. L. Winstone, J. R. Coorssen, and R. J. Turner. 2006. An evaluation of *in vitro* protein-protein interaction techniques: assessing contaminating background proteins. Proteomics 6:2050–2069.
- Kadokura, H., H. Tian, T. Zander, J. C. Bardwell, and J. Beckwith. 2004. Snapshots of DsbA in action: detection of proteins in the process of oxidative folding. Science 303:534–537.
- Knappe, T. A., B. Eckert, P. Schaarschmidt, C. Scholz, and F. X. Schmid. 2007. Insertion of a chaperone domain converts FKBP12 into a powerful catalyst of protein folding. J. Mol. Biol. 368:1458–1468.
- Langley, D., and J. R. Guest. 1977. Biochemical genetics of the alphaketo acid dehydrogenase complexes of *Escherichia coli* K12: isolation and biochemical properties of deletion mutants. J. Gen. Microbiol. 99:263– 276.
- Lengeler, J., A. M. Auburger, R. Mayer, and A. Pecher. 1981. The phosphoenolpyruvate-dependent carbohydrate: phosphotransferase system enzymes II as chemoreceptors in chemotaxis of *Escherichia coli* K 12. Mol. Gen. Genet. 183:163–170.
- Maratea, D., K. Young, and R. Young. 1985. Deletion and fusion analysis of the phage ΦX174 lysis gene E. Gene 40:39–46.
- McMurry, J. L., and R. M. Macnab. 2004. Cobalt-based protein purification resin does not bind *E. coli* SlyD, a common contaminant in Ni-NTA IMAC. Advantage Business Media, Rockaway, NJ.
- Mendel, S., J. M. Holbourn, J. A. Schouten, and T. D. Bugg. 2006. Interaction of the transmembrane domain of lysis protein E from bacteriophage ΦX174 with bacterial translocase MraY and peptidyl-prolyl isomerase SlyD. Microbiology 152:2959–2967.
- Merlin, C., M. Masters, S. McAteer, and A. Coulson. 2003. Why is carbonic anhydrase essential to *Escherichia coli*? J. Bacteriol. 185:6415–6424.
- Miller, J. H. 1992. A short course in bacterial genetics: a laboratory manual and handbook for *Escherichia coli* and related bacteria. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Mouilleron, S., M. A. Badet-Denisot, and B. Golinelli-Pimpaneau. 2006. Glutamine binding opens the ammonia channel and activates glucosamine-6P synthase. J. Biol. Chem. 281:4404–4412.
- Muller, K. M., K. M. Arndt, K. Bauer, and A. Pluckthun. 1998. Tandem immobilized metal-ion affinity chromatography/immunoaffinity purification of His-tagged proteins—evaluation of two anti-His-tag monoclonal antibodies. Anal. Biochem. 259:54–61.
- Negrete, A., W. I. Ng, and J. Shiloach. 2010. Glucose uptake regulation in E. coli by the small RNA SgrS: comparative analysis of E. coli K-12 (JM109 and MG1655) and E. coli B (BL21). Microb. Cell Fact. 9:75.
- 31. Parsy, C. B., C. J. Chapman, A. C. Barnes, J. F. Robertson, and A. Murray. 2007. Two-step method to isolate target recombinant protein from co-purified bacterial contaminant SlyD after immobilised metal affinity chromatography. J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. 853:314–319.
- Phue, J. N., B. Kedem, P. Jaluria, and J. Shiloach. 2007. Evaluating microarrays using a semiparametric approach: application to the central carbon metabolism of *Escherichia coli* BL21 and JM109. Genomics 89:300–305.
- 33. Phue, J. N., S. J. Lee, J. B. Kaufman, A. Negrete, and J. Shiloach. 2010. Acetate accumulation through alternative metabolic pathways in *ackA* (-) *pta* (-) *poxB* (-) triple mutant in *E. coli* B (BL21). Biotechnol. Lett. 32:1897–1903.
- 34. Phue, J. N., and J. Shiloach. 2005. Impact of dissolved oxygen concentration on acetate accumulation and physiology of *E. coli* BL21, evaluating transcription levels of key genes at different dissolved oxygen conditions. Metab. Eng. 7:353–363.
- Porath, J. 1988. High-performance immobilized-metal-ion affinity chromatography of peptides and proteins. J. Chromatogr. 443:3–11.
- Pryor, K. D., and B. Leiting. 1997. High-level expression of soluble protein in *Escherichia coli* using a His6-tag and maltose-binding-protein doubleaffinity fusion system. Protein Expr. Purif. 10:309–319.
- Quail, M. A., D. J. Haydon, and J. R. Guest. 1994. The *pdhR-aceEF-lpd* operon of *Escherichia coli* expresses the pyruvate dehydrogenase complex. Mol. Microbiol. 12:95–104.
- Roof, W. D., H. Q. Fang, K. D. Young, J. Sun, and R. Young. 1997. Mutational analysis of *slyD*, an *Escherichia coli* gene encoding a protein of the FKBP immunophilin family. Mol. Microbiol. 25:1031–1046.
- Roof, W. D., S. M. Horne, K. D. Young, and R. Young. 1994. slyD, a host gene required for ΦX174 lysis, is related to the FK506-binding protein family of peptidyl-prolyl cis-trans-isomerases. J. Biol. Chem. 269:2902–2910.
- Swaim, C. L., B. P. Anton, S. S. Sharma, C. H. Taron, and J. S. Benner. 2008. Physical and computational analysis of the yeast *Kluyveromyces lactis* secreted proteome. Proteomics 8:2714–2723.

- 41. Tchaga, G. S., and G. G. Jokhadze. February 2007. Polynucleotides encoding
- Yong Y. C. S. Trentmann, and J. W. Lengeler. 1989. Alternative route for biosynthesis of amino sugars in *Escherichia coli* K-12 mutants by means of a
- Wu, H. C., and T. C. Wu. 1971. Isolation and characterization of a glucos-amine-requiring mutant of *Escherichia coli* K-12 defective in glucosamine-6-phosphate synthetase. J. Bacteriol. 105:455–466.
 Wulfing, C., J. Lombardero, and A. Pluckthun. 1994. An *Escherichia coli*

protein consisting of a domain homologous to FK506-binding proteins (FKBP) and a new metal binding motif. J. Biol. Chem. 269:2895–2901.
 Yeliseev, A., L. Zoubak, and K. Gawrisch. 2007. Use of dual affinity tags for

- expression and purification of functional peripheral cannabinoid receptor.
- A. Soung, K. D., and R. Young. 1982. Lytic action of cloned ΦX174 gene E. J. Virol. 44:993–1002.
- 47. Zheng, Y., D. K. Struck, and R. Young. 2009. Purification and functional characterization of Φ X174 lysis protein E. Biochemistry **48**:4999–5006.