Andrew J. Link[1]*
Keith Robison[1]**
George M. Church[1,2]

[1]Department of Genetics,
[2]Howard Hughes Medical Institute,
Harvard Medical School,
Boston, MA, USA

Comparing the predicted and observed properties of proteins encoded in the genome of *Escherichia coli* K-12

Mining the emerging abundance of microbial genome sequences for hypotheses is an exciting prospect of "functional genomics". At the forefront of this effort, we compared the predictions of the complete *Escherichia coli* genomic sequence with the observed gene products by assessing 381 proteins for their mature *N*-termini, *in vivo* abundances, isoelectric points, molecular masses, and cellular locations. Two-dimensional gel electrophoresis (2-DE) and Edman sequencing were combined to sequence Coomassie-stained 2-DE spots representing the abundant proteins of wild-type *E. coli* K-12 strains. Greater than 90% of the abundant proteins in the *E. coli* proteome lie in a small isoelectric point and molecular mass window of 4–7 and 10–100 kDa, respectively. We identified several highly abundant proteins, YjbJ, YjbP, YggX, HdeA, and AhpC, which would not have been predicted from the genomic sequence alone. Of the 223 uniquely identified loci, 60% of the encoded proteins are proteolytically processed. As previously reported, the initiator methionine was efficiently cleaved when the penultimate amino acid was serine or alanine. In contrast, when the penultimate amino acid was threonine, glycine, or proline, cleavage was variable, and valine did not signal cleavage. Although signal peptide cleavage sites tended to follow predicted rules, the length of the putative signal sequence was occassionally greater than the consensus. For proteins predicted to be in the cytoplasm or inner membrane, the *N*-terminal amino acids were highly constrained compared to proteins localized to the periplasm or outer membrane. Although cytoplasmic proteins follow the *N*-end rule for protein stability, proteins in the periplasm or outer membrane do not follow this rule; several have *N*-terminal amino acids predicted to destabilize the proteins. Surprisingly, 18% of the identified 2-DE spots represent isoforms in which protein products of the same gene have different observed p*I* and M_r, suggesting they are post-translationally processed. Although most of the predicted and observed values for isoelectric point and molecular mass show reasonable concordance, for several proteins the observed values significantly deviate from the expected values. Such discrepancies may represent either highly processed proteins or misinterpretations of the genomic sequence. Our data suggest that AhpC, CspC, and HdeA exist as covalent homomultimers, and that IcdA exists as at least three isoforms even under conditions in which covalent modification is not predicted. We enriched for proteins based on subcellular location and found several proteins in unexpected subcellular locations.

1 Introduction

A major task accompanying the completion of the genome will be testing the predictions derived from the genomic sequence. Are the open reading frames (ORFs) authentic? Under what conditions and at what levels are ORFs expressed? Where are the products of the genes localized in the cell? Do the physical properties of the proteins agree with those predicted by the ORFs? While cDNA sequencing can find genes, it does not readily answer the above questions about the final gene products, the proteins [1–3]. The term "proteome" describes the protein complement expressed by a genome [4, 5]. The systematic comparison of protein properties, such as isoelectric point, molecular mass, cellular location, and mature *N*-terminal sequence, to the values expected from the genomic sequence can reveal protein processing events. For example, post-translational modifications and proteolysis play significant roles in regulating protein activity but are difficult to predict from DNA sequence [6–8]. Surprisingly, protein abundance under wild-type conditions is often overlooked when

Correspondence: Dr. George M. Church, Department of Genetics, Harvard Medical School, 200 Longwood Ave., Boston, MA 02115, USA (Tel: +617-432-7562; Fax: +617-432-7266; E-mail: church@salt2.med.harvard.edu)

Nonstandard abbreviations: **ATZ,** anilinothiazolinone amino acid; **cDNA,** complementary DNA; **cp*I*,** calculated p*I*; **FE,** fold enrichment; **M-abd,** corrected molar yield of the protein sequence tag; **mPTH$_{aa}$,** the highest background-subtracted PTH amino acid molar yield of a protein sequence signal during an Edman sequencing run; **MW,** calculated molecular mass; **N-abd,** protein cellular abundance; M_r, observed relative molecular mass; **NM,** *N*-terminal sequence tag not significantly matching a sequence in the databases; **nPTH$_{aa}$,** mean relative recovery of a PTH amino acid during Edman sequencing; **O-abd,** observed protein abundance on 2-DE master image; **O.D.,** optical density; **op*I*,** observed relative p*I*; **ORF,** open reading frame; **RY,** repetitive yield

Keywords: *Escherichia coli* / Functional genomics / Proteome / *N*-terminal sequencing / Two-dimensional polyacrylamide gel electrophoresis

* Current address: Department of Molecular Biotechnology, University of Washington, Seattle, WA 98195, USA
** Current address: Millennium Pharmaceuticals, 640 Memorial Drive, Cambridge, MA 02139, USA

© VCH Verlagsgesellschaft mbH, 69451 Weinheim, 1997

0173-0835/97/0808-1259 $17.50+.50/0

physically and biochemically characterizing the expressed gene products of the genome.

Here we describe the proteome of *Escherichia coli*, a gram-negative, motile rod belonging to the *Enterobacteriaceae* family. This facultative anaerobe is part of the normal flora of the lower intestine of humans and warm-blooded animals [9]. *E. coli* strains are classified by either serological reactions of antibodies to variable surface molecules or virulence factors that define five virotypes [10, 11]. Pathogenic strains of *E. coli* cause a large variety of human diseases including infections of the blood, central nervous system, and digestive and urinary tracts [10, 11]. Different sets of virulence genes control the pathogenicity. However, most strains of *E. coli* are avirulent. *E. coli* is the best characterized free-living, single-celled organism. Many biological models (*e.g.* metabolic pathways, genetic regulation, signal transduction, cell wall structure, secretion, *etc.*) have been derived from the study of this bacterium. The *E. coli* K-12 strain is probably the most widely used strain in molecular biology. The 4638858 bp genome of the K-12 strain, MG1655, has been completely sequenced and 4269 ORFs have been identified (F. Blattner, http://www.genetics.wisc.edu/). Of the identiefed ORFs, 58% are hypothetical genes without an assigned function.

It is possible to resolve most of *E. coli*'s proteins with two-dimensional gel electrophoresis (2-DE) [12, 13]. The separation is based on the independent physical properties related to isoelectric point and molecular mass. For a given protein, these two properties, as well as abundance, can be precisely measured from the 2-DE gel. Quantification of *E. coli*'s proteins by 2-DE analysis had primarily used radiolabeled cells [14–17] and comigration of purified proteins with total cell extracts [18–20]. Currently, protein overexpression, peptide mass mapping, Edman sequencing, and amino acid composition are being used to identify *E. coli* proteins on 2-DE gels [17, 21, 22].

Edman sequencing offers several advantages when used to identify 2-DE spots. The *N*-terminal sequence of proteins expressed under *in vivo* conditions verifies the genomic ORF and identifies the protein's mature *N*-terminus. By comparing the observed and predicted sequences, amino-terminal processing events such as initiator methionine and signal peptide cleavage can be identified. Although analyses of crude *E. coli* extracts suggest that most *N*-termini are unblocked [23], these studies were not systematic and were potentially biased toward abundant proteins. Since the molar yields from Edman sequencing of unblocked proteins reflect the quantity of protein, the sequencing yields of 2-DE spots should estimate the quantity of proteins.

In this project, we surveyed the proteome of *E. coli* K-12 and compared the observed properties of the proteins to the predictions of the completed genomic sequence. The proteins were characterized without biases based on the proteins' identity or *N*-terminal accessibility. To survey for major differences in the proteome under different growth and environmental conditions, *E. coli* was harvested at either growth phase in glucose-minimal media or stationary phase in rich media. Total-cell and fractionated protein extracts were purified by 2-DE, electrotransferred to polyvinylidene difluoride (PVDF) membranes, and stained. From the 2-DE gel, the observed isoelectric points (op*I*), molecular masses (M_r), and stained intensities of the proteins were measured. To identify and measure the abundance of the proteins, 364 2-DE spots were systematically sequenced and quantified using limited *N*-terminal Edman sequence analysis. For identified proteins, we compared the observed physical properties of p*I*, M_r, cell location, and *N*-terminal sequence to the properties derived from the genomic sequence. The unique proteins were ranked and sorted based on their abundance and function. These results complement the genomic sequence and provide the foundation for comparison against the proteomes of other *E. coli* strains grown under a variety of environmental conditions. A complete set of the data and figures are available at "http://twod.med.harvard.edu/". The *N*-terminal protein sequence tags have been deposited in Swiss_Prot.

2 Materials and methods

2.1 Strains

Escherichia coli K-12 strain EMG2 (F′ λ+), ATCC 23716, was used for SDS/heat total-cell extracts and fractionated extracts enriched for periplasmic, inner, or outer membrane proteins. To try and match identified spots with an existing *E. coli* 2-DE database [17], strain W3110 (F-*mcr*A *mcr*B IN(*rrn*D-*rrn*E)1 λ–), ATCC 27325, was used for preparing sonicated total-cell extracts. Using total-cell extracts prepared by either the SDS/heat or sonication protocols from both strains grown under similar conditions, we observed no obvious spot differences between the 2-DE gels of EMG2 and W3110.

2.2 Growth conditions

Exponential-phase *E. coli* cells were grown in glucose-MOPS minimal media [24] at 37°C under aerobic conditions to an $O.D._{600}$ between 0.2 and 0.25. Stationary-phase cells were grown in LB media (2% w/v tryptone, 1% w/v yeast extract, and 1% w/v NaCl) at 37°C under aerobic conditions for 12–15 h. All cell cultures were inoculated with a 1% volume of an overnight bacterial culture grown in LB media.

2.3 Preparation of total cellular protein extracts

Total-cell extracts for 2-DE were prepared using either SDS/heat or sonication to lyse the cells [25]. For SDS/heat lysed extracts, cell cultures were centrifuged and the cells washed in wash buffer (10 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$). The washed cell pellet was gently resuspended in lysis buffer A (50 mM Tris-HCl, pH 6.8, 2% w/v SDS, 5% v/v β-mercaptoethanol, and 5% v/v glycerol) and heated at 95°C for 4 min. The extract was quickly placed on ice, and incubated for 20 min in the presence of DNase/RNase solution (2 mg/mL DNase, 2 mg/mL RNase) to a final concentration of 0.4 mg/mL for each enzyme. The lysate was thoroughly mixed with solubilizer solution (9 M urea, 4% NP-40, 2% Servalyt pH 8–10 carrier ampholytes, 1% DTT) so that the final concentration of SDS in the extract was < 0.3% w/v.

The extract was centrifuged at 14 000 × *g* for 3 min and the supernatant transferred to a fresh tube and frozen at −70°C. Sonicated extracts for 2-DE were prepared by resuspending the washed cell pellet in wash buffer and transferring the cells to a microfuge tube. The cells were sonicated using a double step microtip (Branson Ultrasonics, Inc.) at 35 W power for eight 5 s bursts with 20 s intermittent cooling on ice. DNase/RNase solution was added to a final concentration of 0.4 mg/mL each and incubated on ice for 20 min. The lysate was placed at room temperature and solid urea and 1 × volume of solubilizer solution were added so that the final urea concentration in the extract was 9 M. The extract was centrifuged at 14 000 × *g* for 3 min, and the supernatant was transferred to a fresh tube and frozen at −70°C. Protein extracts for fractionation of low molecular mass proteins were prepared by centrifuging a bacterial cell culture and washing the cell pellet in wash buffer. After complete resuspension in lysis buffer B (50 mM Tris-HCl, pH 8.0, 0.5% w/v SDS, 10 mM EDTA, 5% v/v β-mercaptoethanol), the extract was immediately incubated at 95°C for 4 min with occasional mixing. If necessary, the viscosity of the extract was reduced by repeatedly forcing the extract through a 21-gauge needle. Unlike the protein extracts used for 2-DE, these protein extracts were not clarified by centrifugation. The protein extract was frozen at −70°C. Protein concentrations of 2-DE extracts were assayed using a modified Bradford assay [26].

2.4 Preparation of fractionated 2-DE extracts

Periplasmic proteins were enriched using an osmotic shock protocol [27]. Cell cultures were centrifuged at 4700 × *g* for 10 min. The cell pellet was gently resuspended in ice-cold 20% sucrose solution (20% sucrose, 30 mM Tris-HCl, pH 8.0, and 2 mM EDTA) and gently shaken at room temperature for 10 min. The cell suspension was centrifuged at 13 000 × *g* for 10 min. The cell pellet was rapidly resuspended in ice-cold deionized H_2O and gently shaken in an ice bath for 10 min. The cell suspension was centrifuged at 13 000 × *g* for 10 min. The supernatant was transferred to a new tube, frozen on dry ice, and lyophilized. The lyophilized proteins were resolubilized in solubilizer solution and stored at −70°C.

2.4.1 Enrichment of inner and outer membrane proteins

The inner and outer membrane proteins were enriched using isopycnic centrifugation of the cell membranes [28]. Cultured cells were centrifuged at 4700 × *g* for 10 min. The cells were resuspended in ice-cold sucrose wash solution (0.75 M sucrose), and a lysozyme solution (2 mg/mL lysozyme) was added to a final concentration of 0.1 mg/mL. The cells were gently converted to spheroplasts by slowly adding 1 × volume of ice-cold spheroplast solution (1.5 mM EDTA, pH 7.5). The spheroplasts were transferred to centrifuge tubes, and DNase/RNase solution (10 mg/mL DNase, 10 mg/mL RNase) was added to a final concentration of 0.4 mg/mL each. The cells were sonicated using a double step microtip at 35 W power for five 15 s bursts with 60 s intermittent cooling on ice. The cell lysate was centrifuged at 4200 × *g* for 15 min to pellet unlysed cells. The supernatant was transferred to an ultracentrifuge tube, and the membranes pelleted by centrifuging at 365 000 × *g* for 2 h at 4°C. The membrane pellet was washed 3 × with ice-cold sucrose solution A (25% w/w sucrose, 5 mM EDTA, pH 7.5). The membranes were resuspended in sucrose solution A by forcing the pellet through a 21-gauge needle. The resuspended mix of inner and outer membranes was loaded onto a 30–60% w/w sucrose step gradient (5 mM EDTA, pH 7.5) and centrifuged at 120 000 × *g* for 22 h at 4°C. The outer and inner membrane fractions were gently removed from the gradient and transferred to an ultracentrifuge tubes and diluted to 10% w/w sucrose using spheroplast solution. The fractions were centrifuged at 365 000 × *g* for 2 h and washed 3 × with spheroplast solution. The outer and inner membrane proteins were solubilized by the method used to prepare total cell SDS/heat extracts.

2.4.2 Enrichment for low molecular mass proteins

Low molecular mass proteins were isolated using Bio-Rad's preparative SDS-Prep Cell 491 (Hercules, CA). A 14%T, 2.7%C separating gel was cast in the large diameter annular tube assembly (37 mm) in 0.375 M Tris-HCl, pH 8.8, to a height of 100 mm. The separating gel was overlaid with a 4%T, 2.7%C stacking gel cast in 0.125 M Tris-HCl, pH 6.8, to a height of 20 mm. The cathode, anode, and elution buffers were Laemmli running buffer (0.025 M Tris, 0.192 M glycine, and 0.1% w/v SDS) [29]. A total-cell protein extract was loaded and the preparative gel run at a constant current of 50 mA at 4°C. Using an elution flow rate of 0.285 mL/min, 10 mL fractions were collected after the bromophenol blue tracking dye had migrated to the bottom of the separating gel. Fractions (0.2% of total volume) were assayed on a 16.5%T, 6%C Tricine SDS-PAGE gel. Selected protein fractions were precipitated using methanol-chloroform [30], lyophilized, and resolubilized in solubilizer solution.

2.5 2-DE

2.5.1 IEF

The first-dimensional isoelectric focusing (IEF) was performed using the Iso-Dalt IEF apparatus (Large Scale Biology, Rockville, MD) with 150 × 1.2 mm IEF gels (3.3%T, 5.7%C, 9 M urea, 2% v/v NP-40, and 2% w/v carrier ampholytes) cast in glass tubes (1.5 mm outer diameter × 200 mm). The carrier ampholytes consisted of either a 1:1 blend of BDH 4–8:Serva 5–7 or a 4:1 blend of Serva 5–7:Serva 3–10 (Serva, Westbury, NY). The gels were prefocused for 1 h at 200 V using 0.01 N H_3PO_4 anode buffer and 0.02 N NaOH cathode buffer. Protein samples were loaded and focused at 800 V for 14 h at room temperature. To generate PVDF membranes for *N*-terminal sequence analysis of 2-DE spots, 250 µg of total protein were typically loaded in a volume < 40 µL. The IEF gel was extracted from the glass rod, equilibrated for 20 min in equilibration buffer (10% v/v glycerol, 2% w/v SDS, 0.125 M Tris-HCl, pH 6.8, 8.6 mM DTT, and 0.1% w/v *m*-cresol purple), and either used for separation in the second dimension or stored frozen in equilibration buffer at −70°C.

2.5.2 NEPHGE

Nonequilibrium pH gradient gel electrophoresis (NEPHGE) was performed essentially as previously described [31]. The apparatus, solutions, and methods described for first-dimensional IEF were used with the following modifications. The Serva 3–10 carrier ampholytes (Serva) replaced the carrier ampholyte blends used for IEF. The running buffers and electric field were reversed, and the gels were not prefocused. The samples were loaded at the acidic end of the gel and immediately underwent electrophoresis at 800 V for 2.5 h.

2.5.3 SDS-PAGE

The second-dimension SDS-PAGE was performed using the SE600 Hoefer apparatus (Hoefer, San Francisco, CA) equipped for running four gels simultaneously (160 × 155 x 1.5 mm). The gels (11.5% or 12.5%T, 2.7%C) were cast in 0.375 M Tris-HCl, pH 8.8, and 0.1% SDS. No stacking gel was employed. The IEF rod gel was transferred to the SDS-PAGE gel using a loading lectern (Large Scale Biology) and secured in position with 0.5 mL of agarose solution (0.5% w/v agarose dissolved in 10% v/v glycerol, 2% w/v SDS, 0.125 M Tris-HCl, pH 6.8, and 8.6 mM DTT) [31]. The Laemmli running buffer was maintained at 20°C. All gels were run at constant current. Tricine-SDS-PAGE [32] was used to resolve low molecular mass proteins in the size range 2–20 kDa. A separating gel (16.5%T, 6%C) was cast (160 × 110 × 1.5 mm) in 1.0 M Tris-HCl, pH 8.45, and 13.3% w/v glycerol. The separating gel was overlaid with a spacer gel (10%T, 3%C) cast (160 × 20 × 1.5 mm) in 1.0 M Tris-HCl, pH 8.45. The spacer gel was overlaid with a stacker gel (4%T, 3%C) cast (160 × 20 × 1.5 mm) in 0.74 M Tris-HCl, pH 8.45. The gels were run at constant current (60 mA/gel) using 0.1 M Tris, 0.1 M Tricine, 0.1% w/v SDS cathode buffer and 0.2 M Tris-HCl, pH 8.9, anode buffer.

2.6 Electrotransfer of 2-DE gels and protein detection

Protein electrotransfer to PVDF membrane used a previously described method [33] with the following modifications. The second-dimensional SDS-PAGE gels were immediately transferred to PVDF using a "Transphor" TE62 apparatus (Hoefer Scientific) in transfer buffer (10 mM CAPS, pH 11.0, and 10% v/v methanol) at a constant current of 600 mA (≈ 1.4 mA/cm^2) for 2 h. During the electrotransfer, the transfer buffer was maintained at 10°C. After transfer, the PVDF membranes were rinsed with deionized H_2O and the proteins were stained with either colloidal gold (Bio-Rad) or Coomassie-Blue solution. Staining with colloidal gold was performed as previously described [34]. Coomassie detection consisted of soaking the PVDF membranes in 50% methanol and 0.1% w/v Coomassie Brilliant Blue R-250 for 10 min. After removing the staining solution, destaining solution (50% v/v methanol and 5% v/v acetic acid) was applied until maximum contrast was achieved. The membranes were rinsed several times with deionized H_2O, air dried, and stored at room temperature.

2.7 Determination of observed 2-DE (x, y) coordinates and spot intensity

Two-dimensional gels electrotransferred to PVDF were digitized using a reflective scanner (Abaton Scan 300/GS, Everex Systems Inc., Freemont, CA), and the images imported to a Sun SPARCstation IPX (Sun Microsystems Inc., Mountain View, CA). The observed spots were analyzed using the Milligen/BioImage 2-DE software package version 4.6 (Millipore, Ann Arbor, MI) to determine the (x, y) coordinate and "spot intensity". The (x, y) coordinates correspond to the centroid within a spot's boundary. The "spot intensity" is the background-corrected integrated intensity (O.D. × mm^2).

2.8 *N*-terminal amino acid sequencing

For *N*-terminal sequence tagging, 1–18 2-DE spots were cut out using a single-edge razor blade from duplicate PVDF membranes and pooled. The spots were sequenced and quantitated with 12 cycles of Edman sequencing and reverse-phase HPLC. Sequencing was performed on a modified ABI 477 sequenator (Applied Biosystems, Foster City, CA). The ABI 477 was converted to gas-phase delivery of trifluoroacetic acid (TFA) instead of normal pulse-liquid delivery of TFA (P. Tempst, personal communication). Specifically, gaseous TFA was delivered for 800 s during the cleavage step of the reaction cycle by bubbling argon at 1 psi through liquid TFA installed at bottle position X1. The reaction cartridge was inverted [35], and spots were loaded directly into the reaction cartridge without a biobrene-coated glass filter disk. The anilinothiazolinone amino acid (ATZ_{aa}) cleaved from the protein was extracted from the reaction cartridge using a 1:1 mixture of *n*-heptane and ethyl acetate [36]. The extraction of the ATZ_{aa} to the conversion flask used a pulse-delivery protocol [35]. The phenylthiohydantoin amino acids (PTH_{aa}) were sequentially detected using an on-line ABI 120a HPLC optimized for subpicomole PTH analysis [37].

2.9 Identification and quantification of PTH_{aa} yields

The raw, background-, and lag-corrected PTH_{aa} yields were calculated using the Applied Biosystems 900A software package (Applied Biosystems). A 6 pmole PTH_{aa} standard (Applied Biosystems) was used to calibrate the retention times and yields of PTH_{aa}. Amino acid sequences were determined by an analysis of a spreadsheet of peak integration of the PTH_{aa} yields and visual inspection of the raw chromatograms. Since no attempt was made to derivatize cysteine residues prior to sequence analysis, cysteine residues in the protein sequence were not identifiable in this project [38].

2.10 Estimating protein abundance by protein sequencing

The PTH_{aa} yields of the protein sequence tag were used to quantitate the abundance of the protein on the 2-DE gel (M-abd). To calculate the M-abd, the PTH_{aa} yields were corrected for the unequal recovery of PTH_{aa} and sample loss during a sequencing run [38]. Using the number of cell equivalents loaded onto the 2-DE gel,

the M-abd was converted to the abundance of protein in the cell (N-Abd).

2.10.1 Relative recovery

To calculate the relative recovery of PTH_{aa}, the observed recovery of the PTH_{aa} from protein sequencing runs of 80 different 2-DE spots (1401 cycles) was used. For each of the 80 runs, the log of the background-corrected PTH_{aa} yield was plotted versus the cycle number. From an exponential equation fitted to the data points using linear least-squares regression, a theoretical PTH_{aa} yield was calculated for each cycle. For every called PTH_{aa}, the ratio of the "observed PTH_{aa} yield/theoretical PTH_{aa} yield" was calculated. The relative recoveries of PTH_{aa} from 1401 cycles were averaged to generate the mean relative recovery ($nPTH_{aa}$) for the 19 detectable amino acids. Listed below is the mean relative recovery and standard deviation observed for the 19 PTH_{aa} (mean relative recovery, standard deviation): $nPTH_{Ile}$ (1.43, 0.33), $nPTH_{Phe}$ (1.39, 0.33), $nPTH_{Leu}$ (1.37, 0.56), $nPTH_{Val}$ (1.35, 0.34), $nPTH_{Tyr}$ (1.29, 0.30), $nPTH_{Pro}$ (1.20, 0.39), $nPTH_{Asp}$ (1.19, 0.38), $nPTH_{Ala}$ (1.14, 0.38), $nPTH_{Gly}$ (1.10, 0.40), $nPTH_{Gln}$ (1.04, 0.29), $nPTH_{Asn}$ (1.04, 0.29), $nPTH_{Glu}$ (0.98, 0.27), $nPTH_{Arg}$ (0.82, 0.35), $nPTH_{Thr}$ (0.81, 0.24), $nPTH_{Lys}$ (0.81, 0.56), $nPTH_{Met}$ (0.80, 0.28), $nPTH_{Ser}$ (0.55, 0.18), $nPTH_{His}$ (0.43, 0.21), $nPTH_{Trp}$ (0.15, 0.30). Since cysteine residues were not derivatized prior to sequencing, $nPTH_{Cys}$ was not calculated.

2.10.2 Calculation of M-abd

Calculation of M-abd from Edman sequencing of a 2-DE spot was based on the background-subtracted PTH_{aa} yields. The highest background-subtracted called PTH_{aa} yield among the 12 cycles (n) was used. This was corrected for 100% injection and is termed $mPTH_{aa}$. PTH_{Ile}, with the highest mean relative recovery, was assumed to have 100% recovery. To adjust for the variability in recovery of the other amino acids, a factor based on the $nPTH_{Ile}$ and $nPTH_{aa}$ was used. The yield for the $mPTH_{aa}$ in cycle (n) was corrected for sample loss during the sequencing run using the repetitive yield (RY) [39]. For the project, the RY value was an average calculated from 80 sequencing runs (mean RY = 91.6% with a standard deviation of 6%). For calculating the M-abd of 2-DE proteins from fractionated extracts, the sequencing yields were corrected for the increased number of cell equivalents (FE) loaded onto the 2-DE gel. The quantification of cell equivalents was based on the number of cells used to prepare the 2-DE extracts, the final volume of the extract, and the volume of extract loaded onto the IEF gel. The 2-DE gels with total-cell extracts had an estimated 9×10^8 cell equivalents loaded. The periplasmic 2-DE gels had an estimated 2.7×10^{10} cell equivalents (30-fold enrichment, FE = 30, compared to the total-cell extract with an FE = 1), while the inner membrane, outer membrane, and low molecular mass 2-DE protein gels had an estimated 9×10^9 cell equivalents loaded (FE = 10 compared to the total cell extracts). The sequencing yield was corrected for the number of 2-DE spots pooled and sequenced. The formula used for calculating the M-abd for a protein from the *N*-terminal sequence was (1):

$$\text{M} - \text{abd} = \frac{\left[\text{mPTH}_{aa} \times \left(\frac{\text{nPTH}_{Ile}}{\text{nPTH}_{aa}}\right)\right]}{[(\text{RY}^n) \times (\text{FE}) \times (\text{number_spots_pooled})]} \quad (1)$$

The units for M-abd were picomoles/gel. Using the number of cell equivalents loaded onto the 2-DE gels and Avogadro's number, the M-abd was converted to the abundance of protein molecules per cell with Eq. (2):

$$\text{N} - \text{abd} = \frac{(\text{M} - \text{abd})}{(9 \times 10^8)} \times \left(\frac{6.23 \times 10^{23}}{10^{12}}\right) \quad (2)$$

The N-abd was expressed in molecules/cell.

2.11 Identifying expressed ORFs

The *N*-terminal sequence tags were used as query sequences for searching the completed *E. coli* K-12 genomic sequence (strain MG1655) (http://www.genetics.wisc.edu/) and the Swiss_Prot protein database using the programs TBLASTN and BLASTP, respectively [40]. Several criteria were used to determine the significance of the matches. Using these programs, a Poisson score of $P(N) < 0.1$ was generally indicative of a significant match to a sequence in the database [40]. For $P(N)$ scores > 0.1, an *N*-terminal tag matching the 5′ end of an ORF was required as additional evidence that the protein tag identified an expressed ORF. The significance of gaps in the lineup of the query sequence and possible database matches were analyzed. Since the PTH_{aa} Cys, Trp, His, Arg, and Ser were difficult to call in protein sequencing [38], any ambiguous residues in the query sequence aligning with one of these amino acids in the database was considered a match. Finally, the predicted isoelectric point and molecular mass calculated for the translated open reading frame was compared to the observed p*I* and M_r obtained from analysis of the 2-DE spot mobility. To resolve multiple protein signals from a single spot, a regular expression defining the degenerate sequence was searched against the databases, allowing up to two mismatches.

2.12 Identifying blocked *E. coli* proteins by tandem mass spectrometry

2-DE spots blocked to *N*-terminal sequence analysis were identified by tandem mass spectrometry using an LCQ ion trap mass spectrometer (Finningan-Mat, San Jose, CA) and methods previously described [41].

2.13 Theoretical cp*I* and MW derived from genomic sequence

Protein and DNA sequence analysis use the Genetics Computer Group (GCG) sequence analysis software package version 9.0 (GCG, University of Wisconsin Biotechnology Center, Madison WI) running on a Digital VAXstation 4000–90 (Digital Equipment, Maynard, MA). The predicted isoelectric point (cp*I*) and molecular mass (MW) for identified proteins were calculated using the

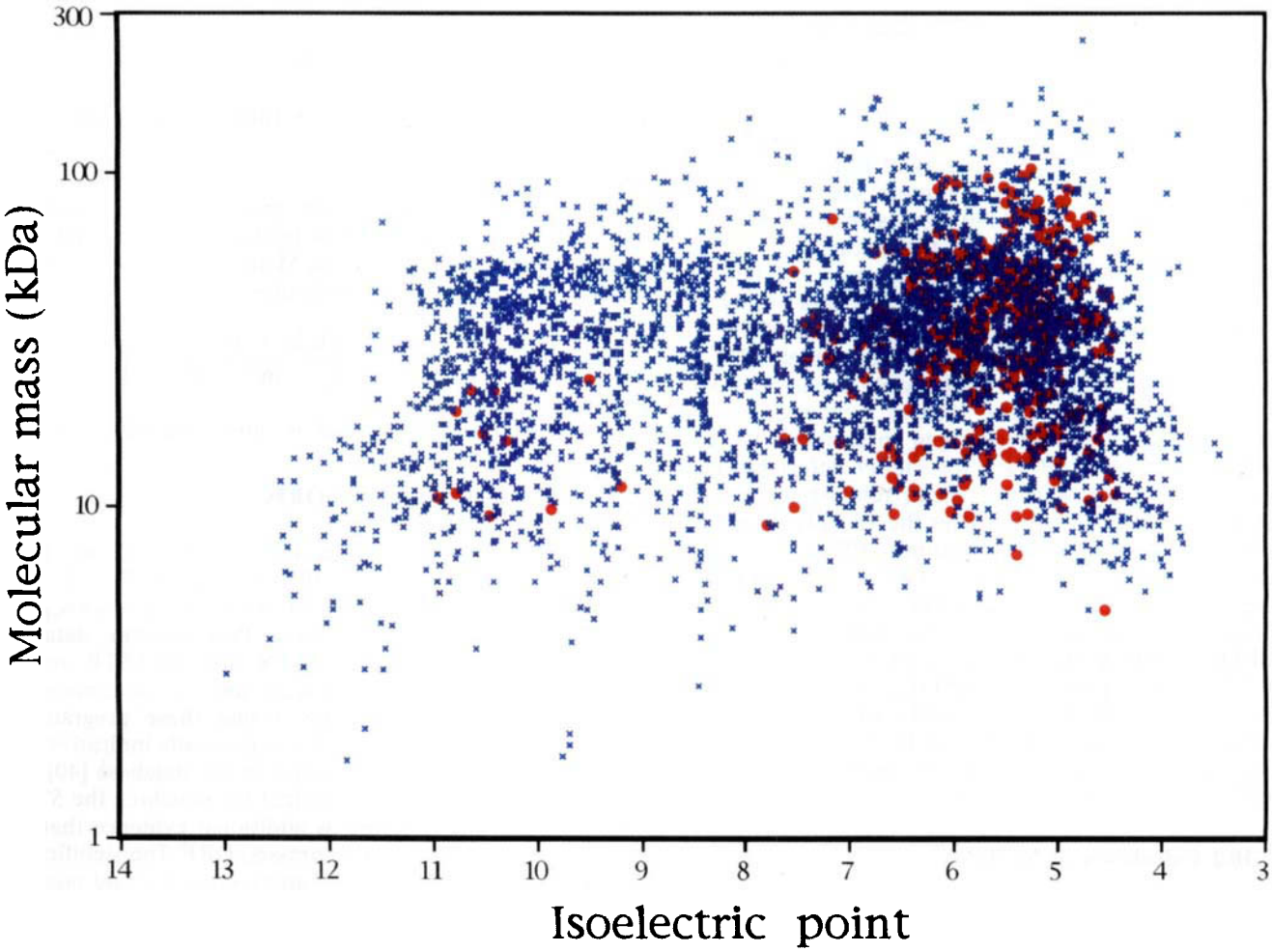

Figure 1. A two-dimensional plot of the observed p*I* and M_r of 364 *E. coli* proteins analyzed in this project superimposed on the predicted values from 4269 conceptually translated ORFs of the complete *E. coli* K-12 genome (F. Blattner, "http://www.genetics.wisc.edu"). The red dots represent the op*I* and M_r of the excised 2-DE spots. The blue "*x*"s are the predicted cp*I* and MW of the conceptually translated ORFs.

completed *E. coli* K-12 genomic sequence (GB: U00096)) starting from the codon coding for the observed *N*-terminal amino acid to the first downstream stop codon. For calculating the cp*I*, the p*K* values used for the charged amino acid residues, amino-terminus, and carboxyl-terminus are Tyr: 10.95, Cys: 8.30, Glu: 4.25, Asp: 3.91, Lys: 10.79, Arg: 12.50, His: 6.50, NH_2-terminus: 8.56, and COOH-terminus: 3.56.

2.14 Determination of observed p*I*, M_r, and abundance

The cp*I* and MW of identified proteins were used as internal calibration markers for determining the observed isoelectric points and molecular masses. The MWs were plotted *versus* the "*y*" coordinates, and a polynomial equation describing their relationship was fitted by linear least-squares regression. Using this equation, the spots' "*y*" coordinates were converted to observed M_r. A similar procedure was used for calculating the op*I* except the cp*I* and "*x*" coordinates were fitted by linear least-squares regression to an exponential equation. The observed abundance of a protein (O-abd) was derived from the "spot intensity" divided by the observed molecular mass. For two or more proteins comigrating as an unresolved 2-DE spot, the "spot intensity" was multiplied by the fraction of each proteins's M-abd (M-abd_i) contributing to the total M-abd (M-abd_{spoti}) from the spot. Equation (3) was used to calculate the observed abundance.

$$O - abd_i = \left(\frac{\text{"spot intensity"}}{M_r}\right)\left(\frac{M\ abd_i}{\sum M - abd_{spoti}}\right) \quad (3)$$

3 Results

3.1 Conceptual *E. coli* proteome

The cp*I* and MW of 4269 *E. coli* predicted proteins from the complete genomic sequence was calculated and plotted to show the distribution of the total proteome on a theoretical 2-DE gel (Fig. 1). To simulate protein mobility during 2-DE, the *y*-axis was drawn on a linear scale to approximate protein mobility during isoelectric focusing gel electrophoresis, and the *x*-axis was drawn on a logarithmic scale to represent migration during SDS-PAGE. The proteins' cp*I* ranged from 3.43k (MsyB) to 12.98 (LeuL) and the MW ranged from 1,723 Da

(TrpL) to 251 390 Da (Orf_2050038). Many of the 4269 *E. coli* proteins (58%) are based on the conceptual translation of novel ORFs in the genomic sequence, and it is unknown which of the hypothetical ORFs are authentic bacterial genes. Moreover, the predicted cp*I*, MW, and sequence do not take into account the effects of protein processing such as site-directed proteolysis and post-translational modifications. Finally, the conceptual 2-DE gel fails to show the *in vivo* abundance or subcellular location of the proteins in the cell under wild-type conditions.

3.2 Observed *E. coli* proteome

In order to compare the predicted and observed genomic output, a set of protein extracts was made from wild-type *E. coli* cells harvested either growing in glucose-minimal media or at early stationary phase in rich media. Two conditions were chosen to observe possible differences in the proteome when *E. coli* is harvested at two different growth and environmental states. To identify subcellular location and to enrich for less abundant proteins, fractionated protein extracts were made using cellular fractionation protocols [27, 28]. To enrich for low molecular mass proteins, a total-cell extract was fractionated using preparative SDS-PAGE, and the proteins eluting from the gel collected. The different total-cell and fractionated extracts were run on 2-DE gels to resolve the proteins. Since we found no single first- and second-dimensional running condition could simultaneously resolve and display all the proteins in the proteome [12, 13]; therefore multiple first- and second-dimensional conditions were combined (see Section 6: Appendix, Tables and Figs. 6–15, pp. 1275–1313). For several of the extracts and 2-DE running conditions, we systematically sequenced most of the 2-DE spots judged to have a sufficient quantity of protein (Tables 6 and 12–15). For the other conditions, we sequenced only 2-DE spots considered unique compared to the other more extensively characterized conditions (Tables 7–11). The images of the ten master membranes showing the spots targeted for sequencing and the data in Tables 6–15 are available at the WWW location "http://twod.med.harvard.edu/".

3.3 Analysis of the proteome

For each of the extracts and 2-DE conditions, a master membrane was chosen based on its overall quality (Figs. 6–15). To determine the selected spot's *N*-terminal sequence, protein identity, and cellular protein abundance, we did 12 cycles of Edman degradation and PTH_{aa} analysis on the selected spots. Based on these sequences, three classes of sequence tags were observed (Tables 6–10). In the first class, a single amino acid sequence was obtained from the spot. This result was found for 79% ($N = 364$) of the spots and suggests that these spots are each a homogeneous protein species. In the second class, 17% of the spots had multiple amino acid sequences, indicating the spots are a mixture of at least two proteins. These spots were given spot numbers corresponding to the number of sequences detected. For this class of spots, 65% of the sequences could be discriminated by the relative intensity of the sequencing signals. The third class of spots (4%, $n = 364$), had no detectable sequencing signal. In these cases, either the protein was blocked to Edman degradation because the *N*-terminus of the protein was modified or there was insufficient protein available for *N*-terminal sequence analysis. A spot was considered blocked if no amino acid sequence could be detected and the intensity of the spot was 2–10 × higher than spots of similar op*I* and M_r that generated a sequence signal. By these criteria, spots M153, M154, R151, and R136 were considered to be *N*-terminally blocked.

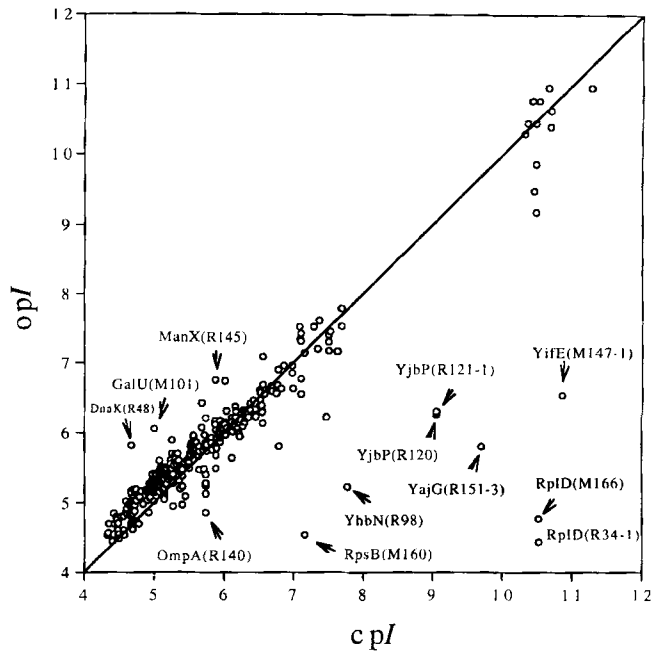

Figure 2. Comparison between the predicted and observed p*I* for identified proteins. For each protein, the deviation of op*I* from cp*I* was calculated. The arrows point to proteins for which this deviation is > 2 SD of the mean deviation for all proteins. The parentheses enclose the protein and the spot number (protein:spot number). Note that the cp*I* is based on the observed *N*-terminal sequence of the identified protein and not its predicted *N*-terminus.

The blocked proteins (M153, M154, R151, and R136) were identified by either Edman or tandem mass spectrometry sequencing of internal peptides of the spots. Spots M153 and M154 were identified as Elongation Factor-Tu by *in situ* digestion of the spot with trypsin, followed by HPLC purification and Edman sequencing released peptides [34]. Spot R151, enriched in the outer membrane fraction, was identified by tandem mass spectrometry as three unique proteins Pal, Slp and YajG. Using the same mass spectrometry method, spot R136, enriched in the inner membrane fraction, was identified as AtpF.

A total of 429 *N*-terminal sequences were generated from 364 2-DE spots (Tables 6–15). To identify the corresponding gene, the protein sequence tags were used as query strings to search the completed *E. coli* genomic sequence as well as nucleic acid and protein databases. We found significant matches for 91% ($n = 429$) of the *N*-terminal sequence tags (Table 6–15). The sequence tags were compared to the predicted gene sequences to

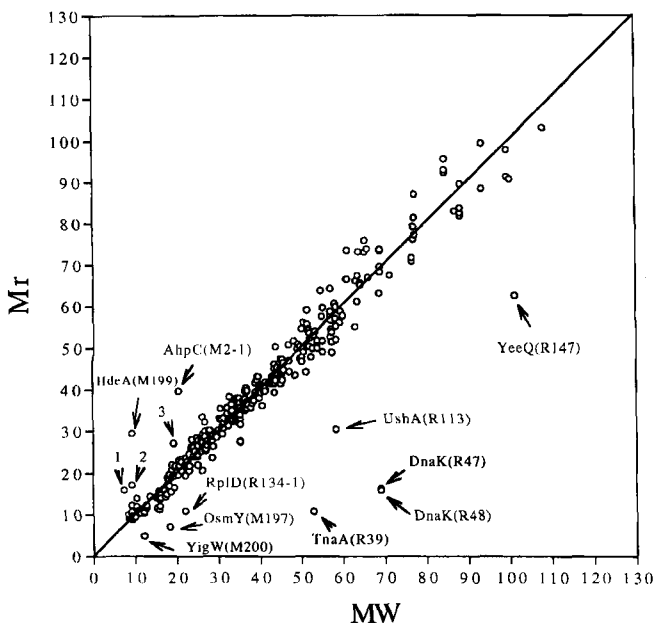

Figure 3. Comparison between predicted and observed molecular mass for identified proteins. For each protein, the deviation of M_r from MW was calculated. The arrows point to proteins for which this deviation is > 2 SD of the mean deviation for all proteins. The parentheses enclose the protein and the spot number (protein:spot number). Note that the MW is based on the observed *N*-terminal sequence of the identified protein and not its predicted *N*-terminus. (1) CspC(M39), (2) HdeA(R51), (3) YjbP(R120, R121-1).

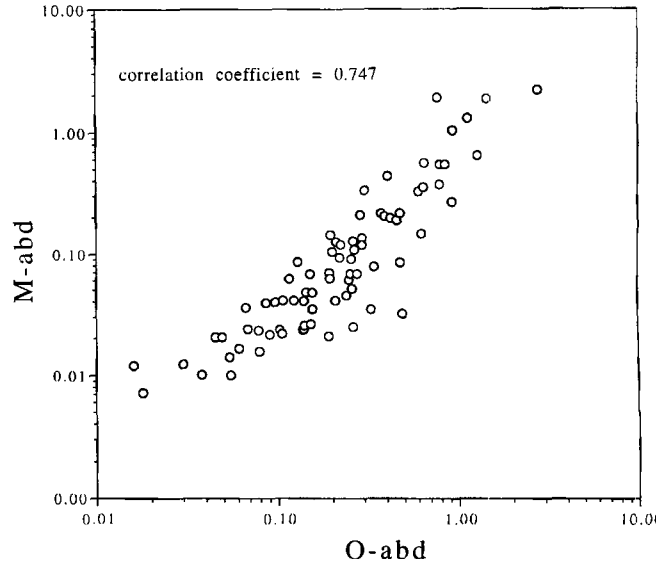

Figure 4. Comparison of protein sequencer yields and intensity of Coomassie-stained 2-DE spots. "M-abd" is the corrected sequencing tag molar yield.

identify the mature *N*-terminus of the proteins (Tables 6–15). The predicted and observed isoelectric points and molecular masses of the proteins were compared (Tables 6–15). For each sequence tag, we calculated the observed abundance (O-abd) and quantity of protein in the cell (N-abd) (Tables 6–15). Of the *N*-terminal sequence tags with significant matches to the genomic sequence, 238 of the sequences were unique (Table 1).

3.4 Comparing the observed and predicted *N*-terminal sequences

In this project, we identified the *N*-termini of proteins derived from 223 unique genes in *E. coli* from which seven classes of *N*-termini were observed (Table 1A-G). Proteins in the first class agree with the predicted Met initiation codon (Table 1A). Proteins in the second class began at the second residue of the conceptual protein (Table 1B). A comparison of these two classes showed that when the second residue was Ser, Ala, Thr, Gly, or Pro, the mature protein usually began with the second residue. This second class is indicative of proteins targeted by the methionine aminopeptidase PepM [42, 43]. While all initial Met residues were cleaved when the second residue was Ala and Ser, the removal of the Met residue was variable when the penultimate amino acid was Thr, Gly, or Pro. Several spots produced staggered *N*-terminal sequences from the same gene (Table 1). For these spots, resolution of the two sequences indicated that one protein species began with the initiator methionine residue while the others began with the second amino acid. Proteins in the third class began 19–44 residues downstream of the predicted initiator methionine. Moreover, the genomic sequence predicts small, apolar residues at −1 and −3 with respect to the observed *N*-terminal residue (Table 1C). These characteristics are indicative of proteolytic cleavage of a signal peptide sequence by the signal peptidase, Lep, from a precursor protein [44–46]. Proteins localizing in the periplasm and outer membrane are expected to have a signal sequence that helps direct their export across the inner membrane [46]. We found the average length of the putative signal sequences was 23–24 amino acids (SD = 4.3, n = 54), and the precursor proteins HdeB, YdcG, and YebL had a putative signal sequence > 3 SD longer than the average signal sequence length. Unexpectedly, we found proteins in the enriched periplasmic and outer membrane fractionations that did not have the canonical signal sequences (underlined in Tables 9, 13, and 15) [47, 48]. Proteins in the fourth class began with Met although the genomic sequence predicted a Val or Leu *N*-terminus (Table 1D). Despite having non-ATG start codons, these proteins still began with methionine. Proteins in the fifth class were blocked to Edman sequencing and were identified by sequencing internal peptides of the spots (Table 1E). The sixth class of spots began at sites near the 5′ end of the identified gene which did not follow the consensus of a signal peptide cleavage site (Table 1F). This class probably reflects misinterpretation of the translation initiation start site. Finally, *N*-terminal sequences in the seventh class began in the internal region of a gene (Table 1G). This class probably represents proteolytic fragments of precursor proteins, but unusual *N*-terminal start sites or sequence misinterpretations cannot be ruled out. Similar to the sequence specificity of a trypsin-like protease, we found that 7 of the 13 internal sites start after an Arg or Lys residues in the conceptual protein sequence (Table 1G).

For sequence tags that identified a unique *E. coli* gene, the observed protein sequence agreed with the genomic sequence for 97.2% (n = 2749) of the predicted amino acid residues (Table 1). Of the discrepancies, 48% (n =

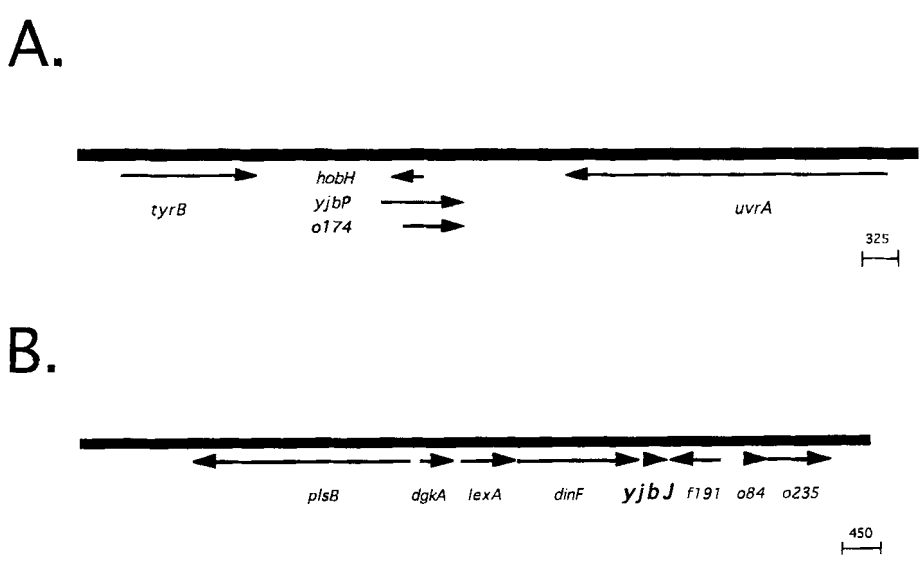

Figure 5. Unexpected *E. coli* genes expressing abundant proteins. The solid line represents the *E. coli* chromosome. (A) *E. coli* genomic region at minute 92 with three overlapping ORFs. Bracketed by *tyrB* and *uvrA*, three conceptual proteins are predicted. *hobH* has been identified based on the expression of a LacZ fusion to the ORF [67]. *yjbP* is an authentic, expressed gene based on the *N*-terminal sequence tag from spot R120. The expression of the third ORF, *o174*, has not been characterized. (B) The relatively small ORF, *yjbJ*, is a highly expressed gene based on several sequence tags (Table 12). The ORFs *f191*, *084*, and *0235* are predicted from the genomic sequence analysis of the region [66].

77) are because no PTH_{aa} signal was detected in the sequencing cycle or because of instrument failure. Cys, Trp, His, Arg, and Ser were usually the PTH_{aa} not identified. These amino acids are difficult to call because multiple PTH derivatives are formed during Edman sequencing, reducing the detectable signal [38]. For other discrepancies, no resequencing of either the protein or genomic region was attempted to resolve the differences between the predicted and observed sequences.

We compared the *N*-terminal amino acid frequencies for proteins predicted to localize in the periplasm or outer membrane (with signal sequences) to proteins predicted to localize in the cytoplasm or inner membrane (without signal sequences); see Table 2. The data show the *N*-terminal amino acid frequency depends on the cellular location of the proteins. The *N*-terminal amino acid frequency for proteins predicted to localize in the cytoplasm and inner membrane is more constrained compared to proteins localizing in the periplasm and outer membrane. The *N*-end rule states that the amino-terminal residue is one determinant of a protein's stability in the cell [49]. Our analysis of *N*-termini detected several violations of the *N*-end rule in proteins predicted to localize in either the periplasm or outer membrane (Table 2). Table 3 shows over 60% of *E. coli*'s proteins are processed by either excision of the initiator Met residue or cleavage of a signal sequence showing proteolytic processing is a highly active process in the cell.

3.5 Comparing the observed and predicted isoelectric point and molecular mass

The observed isoelectric point and molecular mass of each identified protein were compared to the predicted values from the conceptual protein sequence (Figs. 2 and 3). A diagonal line was drawn through the plot to show the deviation of a protein's op*I* or M_r value from the predicted values. The average difference between the op*I* and the cp*I* was 4.9% (SD = 6.4%, $n = 381$). Comparing molecular mass, the average difference between the M_r and MW was 9.4% (SD = 17.5%, $n = 381$). Proteins that deviate by more than two standard deviations from the predicted values are shown (Figs. 2 and 3).

Protein isoforms were identified when the same *N*-terminal sequence tag was detected in two or more spots with different mobilities given identical extraction and 2-DE conditions. Of the 223 *E. coli* genes identified here, 40 (18%) had protein isoforms (Table 3). The op*I* and M_r for each protein isoform is listed to show the relative mobility deviations of the isoforms. The differences are relatively small. The average op*I* difference was 0.26 p*I* units (SD = 0.41, $n = 95$) and the average mass difference was 2.0 kDa (SD = 5.2 kDa, $n = 95$). No attempts were made to identify the causes of the protein isoforms or their biological significance.

3.6 Cellular abundance of identified proteins

To see how well the yield from the *N*-terminal sequence analysis correlated with the protein abundance on the master membrane, we compared the sequencing yields to the observed intensities of the 2-DE spots (Fig. 4). This comparison indicates reasonable concordance between the measured sequencing yields and the observed intensity of the spots on the master gels. The estimated protein abundance in the cell (*N*-abd) was calculated using the sequencing yields and the number of cell equivalents loaded onto the 2-DE gels (Tables 6–15). We ranked and sorted the unique proteins based on their cellular abundance and predicted function (Table 4). For Table 4, the abundances of isoforms were combined to show the total protein expression from the gene, and the proteins were sorted into biological-role categories adapted from Riley and Labedan [50]. The number of genes predicted to be involved in each biological role is also listed to show the fraction of genes in each category that were identified [50]. Using the protein abundances, we calculated the apportionment of the identified proteins in the proteome to different cellular functions and structures (Table 5).

In the project, 364 spots were analyzed by *N*-terminal protein sequencing. To compare the distribution of the selected spots to the predicted complete proteome, the op*I* and M_r of all the excised spots are plotted along with the calculated values of the predicted proteins encoded in the complete *E. coli* genome (Fig. 1).

4 Discussion

The combination of 2-DE and limited protein sequencing provided a powerful set of tools to purify, identify, and characterize *E. coli*'s proteome. Using either total-cell or cellular-fractionated extracts, 2-DE displayed the observed isoelectric points, molecular masses and abundances of the proteins. *N*-terminal sequence analysis identified and quantified the proteins without any prior knowledge of their function in *E. coli*. The sequence tags directly linked the proteins to existing DNA and protein databases and identified the mature amino-termini. Ranking the proteins' abundances indicated the relative allocation of metabolic resources for expressing each gene product.

4.1 Technical implications

4.1.1 Different protein extractions and 2-DE conditions

Different protein extraction protocols were needed to assure that the Coomassie intensity on the 2-DE gels reflected the protein abundance *in vivo* and that no major proteins were overlooked. SDS/heat and sonication extraction protocols were used to eliminate biases that one of the methods might impart against certain classes of proteins. Most of the spots analyzed from the sonicated extract (Table 7) were chosen because they were absent from the SDS/heat extracts (Table 6). From the SDS/heat gel, 153 spots were analyzed and yielded 185 unique sequences. From the sonicated gel, 36 apparently novel spots were analyzed and produced 34 additional unique *N*-terminal sequences. The sonicated extracts gave improved recovery of proteins with a molecular mass greater than 60 kDa as reflected by the fact that most of the unique spots selected for analysis from sonicated extracts had observed masses > 60 kDa. We observed two classes of proteins whose observed abundance depended upon the extraction conditions. An example of the first class is MetE, a highly abundant protein in the sonicated cell extract that was underrepresented in the SDS/heat extracts. In the preparation of SDS/heat extracts, thermal denaturation may cause MetE and other high molecular mass proteins to precipitate. An example of the second class is the outer membrane protein OmpA, which was only observed using the heat/SDS protocol [51, 52]. Unlike the sonication protocol, the heat/SDS extracts effectively solubilized and denatured hydrophobic proteins.

Because *E. coli* proteins possess vast differences is solubility, p*I*, and MW, the analysis of the complete proteome required multiple protein extraction and gel conditions to assure the proteome was faithfully displayed. In this project, IEF and NEPHGE was combined with SDS-PAGE and tricine SDS-PAGE to display the proteome. We observed that IEF gel electrophoresis had a p*I* range from approximately 3.5 to 7.5. Therefore, NEPHGE was used to display and resolve proteins with p*I* > 7.5 [13]. Laemmli SDS-PAGE gels failed to adequately resolve proteins with a mass below 15 kDa; tricine SDS-PAGE thus replaced the conventional Laemmli SDS-PAGE in the second dimension to resolve low mass proteins [32].

4.1.2 Protein detection and identification

One concern was whether all the proteins from the 2-DE gel would electrotransfer, bind, and stain with Coomassie-Blue on the PVDF. To address this question, an untransferred Coomassie-stained 2-DE gel resolving proteins in a total cell extract was compared to a Coomassie-stained PVDF membrane and the gel after transfer. All the proteins transferred and bound to the PVDF. While most of the lower molecular mass proteins appeared to transfer quantitatively, proteins with observed masses > 50 kDa could be detected in the gel after transfer. High yield electroblotting onto PVDF gives a transfer efficiency between 48–91%, depending on the protein [53].

This project primarily used *N*-terminal sequencing to identify and quantify *E. coli* proteins. To expedite sequencing, 12 cycles (55 min/cycle) of protein sequencing were performed on two spots per day with one commercial sequencer. It was not uncommon for the sequences to have cycles with ambiguous PTH_{aa} signals. The 12 cycles of sequence tagging allowed for some uncertainty in the query string when searching the databases for a significant match. Even if 8 of the 12 cycles failed, a 4 amino acid query string would still be statistically adequate to identify a unique ORF in the 4.76 Mbp *E. coli* genome. Other groups have used a similar approach for identifying 2-DE spots by combining amino acid analysis and 3–4 cycles of Edman degradation [54]. These methods are an adaption of the 'sequence tag' approach combining peptide mass fingerprinting and sequence tagging [55].

Although 2-DE is capable of displaying thousands of proteins [12], 17% (n = 364) of the 2-DE spots sequenced had more than one protein sequence signal, indicating that the spot contained more than one protein species. In 65% of the cases (n = 63), the signals could be resolved using the different intensities of the sequencing signals. For the remaining multiple sequencing signals, compilations of the multiple sequences were used as queries with TBLASTN or BLASTP. The significant matches were used to identify the proteins and resolve the mixed sequences. The combination of these two approaches identified at least one of the proteins in 94% (n = 63) of the spots with multiple sequencing signals. Amino acid composition [5, 56] and mass spectrometry of proteolytic peptides [21, 57] can also be used to identify *E. coli* proteins separated by 2-DE. These two methods are extremely promising in terms of speed and sensitivity, but both would have difficulty identifying comigrating proteins. However, tandem mass spectrometry has the ability to identify protein mixtures from 2-DE spots even if the proteins are blocked to *N*-terminal sequencing [41].

A limitation of protein sequence analysis was purifying sufficient quantities of protein for sequence analysis. Using a sequencer optimized for picomole detection sensitivity and assuming at least a 30% initial yield of the protein [33], approximately 150 ng, or 3–4 pmoles, of a 40 kDa protein is typically required to detect a sequence signal. Approximately 50–250 ng of protein can be detected by Coomassie-Blue staining when electroblotted onto PVDF membrane [58]. Therefore, the detection of a protein after Coomassie staining became the standard for selecting spots for *N*-terminal sequence analysis. It was empirically determined that approximately 250 μg of total protein could be loaded onto the IEF gels used here. Attempts to increase the amount loaded actually decreased the number of proteins displayed. To generate sufficient amounts of material for sequencing, less abundant proteins required pooling spots from duplicate PVDF membranes or protein enrichment prior to 2-DE. Extended focusing using immobilized pH gradients in the first dimension can overcome some of these shortcomings [59, 60]. Also, protein sequencing by mass spectrometry has an increased sensitivity compared to Edman degradation and HPLC/UV detection of PTH_{aa} [61, 62].

Although the complete *E. coli* genome sequence is known, 9% ($n = 429$) of the sequencing tags did not significantly match a sequence in the genome. There are two possible explanations. First, the gene encoding the sequencing tag may be unique to the *E. coli* strain used for making the protein extracts (EMG2 or W3110) and not found in the genome of the strain MG1655 used for the genome project. Second, these sequence tags may be random noise in the sequencing output and misinterpreted as an actual protein sequence signal.

4.1.3 Estimating protein cellular abundance by Edman sequencing

Using 2-DE and radiolabeled *E. coli* proteins, the relative amounts of individual proteins in 2-DE gel spots appears to reflect their *in vivo* abundances [15]. In this project, the molar yield of the sequencing tag for a 2-DE spot was used to measure the abundance of the protein. This is the first example of using protein sequencing to systematically quantitate the abundance of proteins in the cell. To convert sequence yields to protein abundance, the sequencing signal was corrected for the unequal recovery of PTH_{aa} and sample loss during sequencing. An alternative method calculates the initial yield by plotting the logarithm of the background-corrected PTH_{aa} versus the cycle number and extrapolating back to cycle zero. In this project, the latter method consistently gave an underestimation of the sequencing yields. Calculations of protein abundance using 2-DE and *N*-terminal sequence yields does have caveats. Since the electrotransfer efficiency ranges from 48–91% [53] and sequencing initial yields range from 30–80% [33], the molar yield of a sequence tag has a potential range of 12–69%. This method of quantitating the abundance of proteins in the cell, like others such as radiolabeling and immunoaffinity methods [15, 63], assumes that protein extraction yields, precipitation during 2-DE, electrotransfer, and detection are uniform for all proteins in the proteome.

4.2 Biological implications

In this project, the proteins were selected based on abundance as indicated by their 2-DE gel Coomassie-staining intensity. Figure 1 illustrates that most of the abundant *E. coli* proteins analyzed in this project are in a window of p*I* 4–7 and molecular mass 10–100 kDa. Proteins with either extreme p*I* or mass may be difficult to keep folded or solvated in the cell or may be energetically expensive to express abundantly. As a caveat, the very basic window of the 2-DE gels where most of the ribosomal protein are known to migrate was not extensively analyzed in this project [17].

4.2.1 Identification of expressed ORFs

Authenticating a predicted gene requires analysis of the actual gene products expressed *in vivo* from wild-type cells. By matching the protein sequence tag to the genomic sequence (Table 1), the expression of 223 genes in the *E. coli* genome was confirmed. Predicting genes in the *E. coli* genome has generally been based on intrinsic (codon or oligonucleotide biases, ORF length, *etc.*) and extrinsic (sequence similarity) approaches [64]. Short ORFs, unusual translation initiation motifs, and unusual codon usage [65] confound the identification of authentic genes. Figure 5 shows two examples of *N*-terminal sequence tags authenticating genes that would otherwise have been difficult to predict by genomic sequence analysis alone.

Figure 5A shows three predicted *E. coli* gene products from overlapping ORFs on the chromosome. The sequence tag of spot R120 confirms that one of the reading frames, *yjbP*, encodes a 230 amino acid protein with a 23 amino acid signal sequence. An ORF with 55.3% identity to *yjbP* (two frameshifts required) was found adjacent to a chloramphenicol acetytransferase gene (*cat*) in *Proteus mirabilis*. A second ORF, *o174*, overlaps *yjbP* and probably reflects a sequencing error at the 5′ region of the gene [66]. On the opposite strand, a third ORF, *hobH*, encodes a putative DNA-binding protein [67]. If *hobH* is actually expressed, this sequence of *E. coli* DNA would surprisingly encode two proteins on opposite strands.

When annotating genomic sequence, small genes (< 100 codons) are difficult to distinguish from random stretches of sequence with no stop codons [66, 68]. We specifically enriched for low molecular mass proteins in the proteome to identify such small proteins (Fig. 12). One sequence tag confirmed the authenticity of a 69 amino acid gene yjbJ (Fig. 5B). Although YjbJ is one of the most abundant proteins observed in *E. coli* at early-stationary phase, the protein has been uncharacterized until now. A null mutation of the gene has no identifiable phenotype (unpublished observations). The ORF *yggX* encoding a 91 amino acid protein was also verified as an authentic gene (Table 12). Despite a 10-fold enrichment for low molecular mass proteins, we did not identify any proteins under 8 kDa, and the master gel enriched for this class of proteins shows few 2-DE spots under 8 kDa (Fig. 12).

4.2.2 *N*-terminal amino acid sequence analysis of mature proteins

Because intact proteins were *N*-terminally sequenced, the sequence tags determined the mature amino terminus of the proteins. An immediate observation from the success of *N*-terminal sequence tagging is that most of *E. coli*'s proteins are not blocked. A blocked protein was recognized when no PTH_{aa} signal was detected while the PTH_{aa} background increased steadily. We identified the proteins EF-Tu, AtpF, Pal, Slp, and YajG as proteins with blocked *N*-termini. EF-Tu has been previously shown to be acetylated at the *N*-terminal serine residue [69]. Pal, Slp, and YajG are all suspected to be lipoproteins, which generally have an *N*-terminal cysteine residue linked to a diglyceride and a fatty acid [70]. AtpF is a component of the inner membrane ATP complex, and its *N*-terminal structure is not known [71].

We compared the *N*-terminal processing events identified to known amino-terminal processing models. For cytoplasmic proteins, the current model is that the aminopeptidase PepM excises the initiator methionine depending on the side-chain length of the second amino acid [42, 43]. When the second amino acid is Ala, Cys, Gly, Pro, Ser, Thr, or Val, the initiator Met is excised. This model was verified for proteins with Ala and Ser as the second amino acid. Although the model predicts the excision of the Met residue for a "Met-(Thr, Val, Gly, Pro)-" amino terminus [43], violations of the rule were observed for these penultimate amino acids (Table 1). The cleavage of the Met was variable with Thr, Gly, and Pro, and it was not observed for proteins with a Val at the second position. These results indicate that excision of the initial Met residue preceding a Thr, Gly, and Pro either does not occur with 100% efficiency or is slow compared to excision when the second residue is Ala or Ser. These inconsistencies suggest that protein structures other than the second residue are involved in the excision specificity.

We compared the signal sequence cleavage sites with predicted consensus sequences. *E. coli* signal peptide sequences tend to be 15–30 amino acids long with a short positively charged amino-terminal region, a central hydrophobic region, and a carboxy-terminal region encompassing the cleavage site [46, 72]. The sequence pattern at the cleavage site by the signal peptidase Lep is generally a small, apolar residues at −1 (Ala, Gly, or Ser) and −3 (Ala, Ser, Gly, Thr, Val, or Leu) and a helix-breaking residue adjacent to the central hydrophobic core at −4 to −6 (Pro, Gly, or Ser) [45, 73, 74]. We found this pattern was conserved at all −1 and −3 sites (Table 1C). The helix-breaking residue at −4 to −6 was less conserved; however, a helix-breaking residue could be found in the −4 to −11 window (Table 1C). We found the length of the signal sequences for the precursor proteins HdeB, YdcG, and YebL was significantly longer (> 3 SD) than the observed average. These could be misinterpretation of the predicted translation start sites of the precursor proteins or evidence that signal sequences can be longer than previously predicted. Illustrating the value of *N*-terminal sequencing of mature proteins, we identified 13 novel proteins with putatively cleaved signal sequences (Table 1C). Many of these proteins were found in the enriched periplasmic protein extracts, supporting the prediction that they are exported across the inner membrane.

Thirteen sequence tags matched predicted internal regions of genes in the *E. coli* genome (Table 1G). These fragments could be an *in vitro* artifact of the protein extraction, although cellular proteins were rapidly extracted and fractionated under highly denaturing conditions to prevent spurious proteolysis. Alternatively, these fragments may result from an *in vivo* event. In either case, if proteolysis fragmented the protein into two pieces, both fragments might be detectable; however, no second fragments were detected. Seven of the thirteen internal proteolytic sites are on the carboxyl side of Arg or Lys residues. These could define the recognition sites for an enzyme(s) resembling mammalian trypsin. None of the putative cleavage sites matches any known *E. coli* protease recognition sequences, although little is known of the target specificity of *E. coli* proteases [8, 75, 76]. Finally, the proteins could also be translation products that initiate at internal sites of the genes.

Table 4 shows the nonrandom distribution of the mature *N*-terminal amino acid frequencies compared to the net amino acid composition of *E. coli's* proteins. This was first observed from crude *E. coli* extracts [23] and is similar for eukaryotic cells [77]. An unusual observation is that cytoplasmic and inner membrane proteins (without signal sequences) have a different *N*-termini distribution compared to proteins in the periplasm or outer membrane (with signal sequences); see Table 2. One model, the *N*-end rule, relates a protein's half-life to the identity of its amino-terminal residue [49]. In *E. coli* proteins, Phe, Leu, Trp, Tyr, Arg, and Lys are destabilizing *N*-terminal residues [78]. Arg and Lys residues are known as secondary destabilizing residues because their destabilizing activity requires the post-translational conjugation of Leu or Phe residues to the *N*-terminus [78, 79]. While cytoplasmic and inner membrane proteins in this study follow the *N*-end rule, 15% ($n = 53$) of the periplasmic or outer membrane proteins have an *N*-terminal amino acid residue predicted to destablize the protein. These results support the model that the *N*-end rule applied only to cytoplasmic or inner membrane proteins in *E. coli*, and not to proteins exported across the inner membrane [49].

4.2.3 Analysis of protein abundance

A number of models have tried to predict the expression and stability of a gene product in *E. coli* based on transcriptional and translational signals such as promoter strength, codon frequencies, amino acid usage, and *N*- and *C*-terminal amino acids [65, 78, 80–92]. Due to the large number of environmental and cellular variables controlling the expression of a gene, these predictions are unreliable. We measured the actual abundance of proteins in the cell using the yields of the sequencing tags and ranked and sorted the proteins in Table 4 based on their abundance and function to see the distribution of the abundant proteins to cellular processes.

We found several proteins that are unexpectedly abundant (Table 4). The product of the *HdeA* gene, a putative periplasmic protein, was highly abundant in both the growth and stationary phases. Originally identified as gene for which expression is induced in an *hns* deletion background, HdeA has no identified function [93, 94]. CspC is a suppressor of a chromosome partitioning mutation *muk*A [95] and has 80% sequence similarity to CspA, a cold-shock induced protein with DNA-binding sequence motifs [96]. AhpC, an alkyl hydroperoxide reductase involved in protecting the cell from oxidative stress, was highly abundant in both growth and stationary phases [97, 98]. In *Salmonella typhimurium*, the reductase activity is due to a complex of AhpC and AhpF [97, 99, 100]. Surprisingly, AhpF was not identified in this project, and no abundant spots appeared on the 2-DE gels in its expected position. Finally, we identified 38 novel proteins in *E. coli* that are relatively abundant and yet have no published function or phenotype.

Although originally identified based on induction during environmental stress, several proteins are relatively abundant in the cell during noninducing conditions. Originally observed as heat-shock-induced proteins, the molecular chaperones MopA, DnaK, Tig, MopB, and HtpG are all abundant in *E. coli* under conditions not expected to induce their expression. These chaperones assist in the assembly of oligomeric protein structures and the export of proteins [101–103].

We used two different growth and environmental conditions to survey major differences in the proteome at different cell states. Although the *E. coli* proteome was not comprehensively analyzed at early-stationary phase in rich media, several highly abundant proteins in this condition were identified that were not observed in cells growing in glucose-minimal media. Among these, TnaA, or tryptophanase, catalyzes both the degradation and synthesis of tryptophan [104]. Another, MglB or galactose-binding protein, is involved in the transport of galactose into the cell [105]. *MglB* expression is strongly inhibited by glucose, which would explain its absence in cells harvested while growing in the glucose-minimal media [106, 107]. A third protein, Dps, was originally identified as a starvation-inducible protein which forms stable complexes with DNA, rendering the DNA resistant to oxidative damage [63]. These results show that the proteome is dynamic, and *E. coli* changes the contents of the proteome in response to changing environmental conditions [17].

The molar amounts of different enzymes in biochemical pathways were highly unequal (Table 4). In the glycolytic pathway, GapA and Eno are eight-fold more abundant than the least abundant enzyme identified in the pathway, PykF. In the tricarboxylic acid cycle, IcdA is 65-fold more abundant than the least abundant identified enzyme FumA in the cycle. There are several possible reasons why enzymes are not present in equimolar amounts. The enzymes in a pathway may have different efficiencies, requiring more of one enzyme than another. Some of the enzymes may have additional functions in the cell. It has been postulated that functionally related enzymes in a sequential pathway associate to form a complex in the cell to channel metabolites and coordinate the catalysis of complex reactions [108–110]. Measuring the abundances of proteins in the cell is one method of estimating the stoichiometry of proteins in putative multienzyme complexes.

We summed the protein abundances in the different protein categories to see the apportionment of the abundant proteins to various cellular processes (Table 5). Over 25% of the cell's proteome was devoted to the synthesis of macromolecules, especially protein synthesis. A large fraction of the proteome was devoted to the biosynthesis of small precursor molecules for protein and DNA synthesis. Proteins involved in energy metabolism and the transport of molecules into the cell were the next largest fraction. The importance of proteins involved in the detoxification of harmful molecules is evident because they constitute a significant fraction of the proteome. Although many cellular activities have large numbers of genes assigned to them, relatively few of these are abundant proteins.

Besides the fact that we only sequenced a subset of the approximately 1850 observable *E. coli* 2-DE spots [17], there are several reasons why additional proteins in other functional categories were not identified. First, proteins in many catagories probably were not sufficiently abundant to be detected by the methods used in this project. Second, proteins that failed to solubilize during extraction or that precipated during 2-DE would escape analysis. These would include hydrophobic proteins or proteins covalently interacting with insoluble cell structures such as the peptidoglycan [111]. Third, proteins only expressed under specific growth or environmental conditions would go undetected. For example, proteins from glucose-repressed genes would not be present in protein extracts from *E. coli* grown in glucose-rich media. Fourth, unstable proteins with a short half life would go undetected [112].

4.2.4 Predicted *vs.* observed isoelectric point and molecular mass

The observed isoelectric points and molecular masses of the identified proteins were compared to those predicted from conceptual translation of the genomic sequence to identify either potential DNA sequencing errors or proteins with unusual 2-DE mobilities. Few proteins deviated significantly from their expected values (Figs. 2 and 3). However, it is not known how many of the proteins were modified, yet migrated close to their predicted p*I* and MW. Of the 223 genes identified, 18% had gene-products migrating as protein isoforms, indicting they were post-translationally modified. A single charge difference can change the mobility of a protein during IEF [12]. Artifactual charge heterogeneity is recognized when each protein on a 2-DE gel forms a series of spots having equal spacing at the same molecular mass [12]. Since no overall spot heterogeneity was seen on the representative gels, sample preparation does not appear to cause the deviation between observed and predicted isoelectric point.

Some of the discrepancies between the observed and predicted isoelectric point or molecular mass may have been due to errors in the DNA sequence (Figs. 2 and 3). For example, although the initial conceptual translation of the *S. typhimurium ahpC* gene predicted a cp*I* of 10.20, we found the *E. coli* AhpC homologue had an op*I* of 5.16 [100]. Subsequent correction of a frameshift error left the *S. typhimurium* AhpC with a cp*I* similar to the observed value of the homologous *E. coli* protein [113]. A comparison of the predicted p*I* for the sequences of AroK, PckA, and TolC found in different protein databases showed large discrepancies for each protein. Using the completed genomic sequence of *E. coli*, the predicted p*I* for these proteins now agrees with the op*I*.

An interesting example of a large discrepancy between the predicted and observed p*I* is RplD, encoding 50S ribosomal subunit protein L4 [114]. While the cp*I* is 10.52 for the conceptually translated sequence, op*I*s of 4.44 and 4.78 were found for the protein in two different gel conditions (Table 7 and 12). From both gels, the M_r agrees with the predicted MW of the protein. We confirmed the protein's identity by sequencing internal peptides of the 2-DE spots using tandem mass spectrometry (Table 7 and 12). Other groups have observed RplD migrating on 2-DE gels at the expected location [17, 115]. The cause of this large discrepancy is unknown.

In one of the best-characterized post-translational events, phosphorylation controls the activity of isocitrate dehydrogenase, or IcdA [6]. IcdA converts isocitrate to 2-oxoglutarate in the tricarboxylic acid cycle [116]. With growth on acetate, IcdA is phosphorylated, inhibiting its enzymatic activity [117, 118]. In this project, we identified three IcdA isoforms from cells growing in glucose-minimal media, conditions in which the enzyme should have been in the active form. Two other enzymes in the tricarboxylic acid cycle, Mdh and LpdA, also showed several isoforms during growth in minimal media. Protein isoforms of the known phosphoproteins MopA and PtsH [119, 120] were identified in the project, but it is unknown if phosphorylation of the proteins is caused by the different mobilities.

We found examples of proteins that appear to retain homomultimeric interactions even under highly reducing conditions. AphC, CspC, and HdeA migrate with an M_r/MW ratio of 1.9, suggesting that they are homodimers (Fig. 3). For HdeA, another spot with an M_r/MW of 3.0 was found, suggesting a trimer (Fig. 3). For both AphC and HdeA, 2-DE spots of the proteins were also found with the M_r close to the predicted MW. This multimer formation is unexpected, considering the proteins were extracted and resolved under highly denaturing conditions and suggests that these two proteins have a strong self-affinity or are covalently linked. The yeast thiol-specific antioxidant (TS) with a 40% amino acid identity to AphC, has been shown to maintain a dimeric structure even under denaturing and reducing conditions [121]. Several proteins have an M_r much lower than expected although they begin with the predicted *N*-terminal sequence (Fig. 3). These spots may represent truncated or proteolytically processed proteins.

The exact source of each mobility deviation is difficult to predict with the experimental tools applied in this project. Although 20 amino acids are typically used for protein synthesis, over 200 amino acid derivatives occur in both prokaryotic and eukaryotic organisms [122]. Covalent protein modifications identified in *E. coli* include phosphorylation [123, 124], acetylation [69, 125], methylation [7, 126], adenylylation [127], and lipid coupling [128]. SDS-PAGE lacks the sensitivity and accuracy to measure the precise molecular mass changes resulting from potential covalent modifications. Mass spectrometry has the potential for high-throughput identification of protein modifications that alter 2-DE mobilities, but the quantities of proteins required preclude its application to most proteins derived from 2-DE gels [129, 130]. Alternatively, *C*-terminal sequencing would identify the carboxyl-terminus of proteins for comparison with the DNA sequence.

4.2.5 Subcellular protein location

Other groups have used 2-DE to display the proteins in the different *E. coli* subcellular locations but have not identified or quantified the proteins [131–134]. Subcellular location is often simply predicted from the protein sequence [135]. In this study, subcellular fractionation prior to 2-DE allowed identification of the subcellular location of proteins and enriched for low abundance proteins not detectable in total-cell extracts. The periplasmic protein extract from *E. coli* harvested at early stationary phase in rich media was the most extensively analyzed subcellular fraction. Of the *N*-terminal sequence-tagged proteins, 85% (n = 76) were not detected in total-cell extracts. We found that binding proteins involved in the transport of molecules across the inner membrane into the cytoplasm are the most abundant class of proteins found in the periplasm (Table 13). The second most abundant class are "scavenging" enzymes, which degrade complex molecules into simpler components [48].

Unexpectedly, we found cytoplasmic proteins in the periplasmic extracts (underlined in Tables 9 and 13). The simplest explanation is unwanted cell lysis that contaminates the periplasmic extracts with cytoplasmic proteins. These "contaminating" proteins may be associated with the inner membrane and expelled when the cell is osmotically shocked. A more intriguing possibility is that these cytoplasmic proteins do partially localize in the periplasmic space. There are examples of *E. coli* proteins exported across the inner membrane without a signal sequence [136–138]. Thus, this suggestion is not entirely improbable. Clearly, additional independent experiments are required to conclusively prove that these proteins are localizing in unexpected cellular locations [139–141]. Although predicted to be exclusively in the cytoplasm, superoxide dismutase (SodB) was the fourth most abundant protein identified in the periplasmic extract. A second dismutase, SodA, was also identified in the periplasm. If SodB and SodA were simply contaminating the subcellular fraction, other more abundant cytoplasmic proteins would be expected in the periplasmic extracts. This was not the case, suggesting SodB and SodA function in the periplasm. These enzymes remove harmful

free radicals from the cell [142]. Perhaps the two dismutases function in the periplasm to protect periplasmic proteins and lipids from oxidative damage. Other proteins unexpectedly found in the periplasmic extract are DnaK and PpiB which assist in protein folding to prevent irreversible aggregation [143, 144]. They could be performing similar functions in the periplasm. Alternatively, the unexpected cytoplasmic proteins in the periplasmic, inner membrane, and outer membrane extracts could have second unknown biological functions.

4.3 Concluding remarks

We have presented the advantages and limitations of systematically identifying and characterizing the protein content of *E. coli*. This project exploited the ability of 2-DE to rapidly resolve most of the cell's proteins, and *N*-terminal protein sequencing to identify each protein. The resulting database of protein sequence, observed isoelectric point and molecular mass, cellular abundance, and subcellular location helps in interpreting the *E. coli* genomic sequence and aids in the study of genes and cellular processes. Since the *N*-terminal residues of 70–80% of eukaryotic proteins are likely to be modified [145], *N*-terminal sequencing as employed in this project would fail at a high frequency in eukaryotes. An attempt to identify human liver proteins by *N*-terminal sequencing failed to detect a PTH_{aa} signal for 43% of the 102 sequenced spots [146]. Methods for deblocking *N*-acetyl groups from eukaryotic proteins could improve the success rate [147, 148].

The authors wish to thank M. Temple, and K. Landry for help with computer analyses, P. Tempst, J. Rush, L. Riviere, M. Fleming, and C. Elicone for assistance and advice with protein sequencing, and L. Hayes and A. Dongre for assistance with mass spectrometry. We are especially grateful to E. Malone for her critical review of this manuscript. This work was funded by DOE grant DE-FG02-87ER60565.

Received Otcober 29, 1996

5 References

[1] Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Merril, C. R., Wu, A., Olde, B., Moreno, R. F., Kerlavage, A. R., McCombie, W. R., Venter, J. C., *Science* 1991, *252*, 1651–1656.

[2] McCombie, W. R., Adams, M. D., Kelley, J. M., FitzGerald, M. G., Utterback, T. R., Khan, M., Dubnick, M., Kerlavage, A. R., Venter, J. C., Fields, C., *Nature Genetics* 1992, *1*, 124–131.

[3] Waterston, R., Martin, C., Craxton, M., Huynh, C., Coulson, A., Hillier, L., Durbin, R., Green, P., Shownkeen, R., Halloran, N., Metzstein, M., Hawkins, T., Wilson, R., Berks, M., Du, K., Thomas, K., Thierry-Mieg, J., Sulson, J., *Nature Genetics* 1992, *1*, 114–123.

[4] Wasinger, V. C., Cordwell, S. J., Cerpa-Poljak, A., Yan, J. X., Gooley, A. A., Wilkins, M. R., Duncan, M. W., Harris, R., Williams, K. L., Humphery-Smith, I., *Electrophoresis* 1995, *16*, 1090–1094.

[5] Wilkins, M. R.,. Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J. C., Yan, J. X., Gooley, A. A., Hughes, G., Humphery-Smith, I., Williams, K. L., Hochstrasser, D. F., *Bio/Technology* 1996, *14*, 61–65.

[6] LaPorte, D. C., Koshland, D. E., *Nature* 1984, *305*, 286–290.

[7] Stock, A., in: Zaapia, V., Galletti, P., Porta, R. (Eds.), *Advances in Post-Translational Modifications of Proteins and Ageing*, Plenum Press, New York 1988, pp. 387–399.

[8] Maurizi, M. R., *Experientia* 1992, *48*, 178–201.

[9] Holt, J. G., Krieg, N. R., Sneath, P. H. A., Staley, J. T., Williams, S. T., *Bergey's Manual of Determinative Bacteriology*, Williams and Wilkens, Baltimore 1994.

[10] Schaechter, M., in: Lederberg, J. (Ed.), *Encyclopedia of Microbiology*, Academic Press, San Diego 1992, pp. 115–124.

[11] Salyers, A. A., Whitt, D. D., *Bacterial Pathogenesis: A Molecular Approach*, ASM Press, Washington, DC 1994.

[12] O'Farrell, P. H., *J. Biol. Chem.* 1975, *250*, 4007–4021.

[13] O'Farrell, P. Z., Goodman, H. M., O'Farrell, P. H., *Cell* 1977, *12*, 1133–1142.

[14] Neidhardt, F. C., Bloch, P. L., Pedersen, S., Reeh, S., *J. Bacteriol.* 1977, *129*, 378–387.

[15] Pedersen, S., Bloch, P. L., Reeh, S., Neidhardt, F. C., *Cell* 1978, *14*, 179–190.

[16] Herendeen, S. L., VanBogelen, R. A., Neidhardt, F. C., *J. Bacteriol.* 1979, *139*, 185–194.

[17] VanBogelen, R. A., Abshire, K. Z., Pertsemlidis, A., Clark, R. L., Neidhardt, F. C., in: Neidhardt, F. C., Curtiss, R.., Gross, C., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Riley, M., Schaechter, M., Umbarger, H. E. (Eds.), *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ASM Press, Washington, DC 1996, pp. 2067–2117.

[18] Bloch, P. L., Phillips, T. A., Neidhardt, F. C., *J. Bacteriol.* 1980, *141*, 1409–1420.

[19] Phillips, T. A., Bloch, P. L., Neidhardt, F. C., *J. Bacteriol.* 1980, *144*, 1024–1033.

[20] VanBogelen, R. A., Sankar, P., Clark, R. L., Bogan, J. A., Neidhardt, F. C., *Electrophoresis* 1992, *13*, 1014–1054.

[21] Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., Watanabe, C., *Proc. Natl. Acad. Sci. USA* 1993, *90*, 5011–5015.

[22] Pasquali, C., Frutiger, S., Wilkins, M. R., Hughes, G. J., Appel, R. D., Bairoch, A., Schaller, D., Sanchez, J.-C., Hochstrasser, D. F., *Electrophoresis* 1996, *17*, 547–555.

[23] Waller, J. P., *J. Mol. Biol.* 1963, *7*, 483–496.

[24] Neidhardt, F. C., Bloch, P. L., Smith, D. F., *J. Bacteriol.* 1974, *199*, 736–747.

[25] Phillips, T. A., *DNA Protein Eng. Tech.* 1988, *1*, 5–9.

[26] Ramagli, L. S., Rodriguez, L. V., *Electrophoresis* 1985, *6*, 559–563.

[27] Neu, H. C., Heppel, L. A., *J. Biol. Chem.* 1965, *240*, 3685–3692.

[28] Osborn, M. J., Munson, R., *Methods Enzymol.* 1974, *31*, 642–653.

[29] Laemmli, U. K., *Nature* 1970, *227*, 680–685.

[30] Wessel, D., Flügge, U. I., *Anal. Biochem.* 1984, *138*, 141–143.

[31] Anderson, L., *Two-Dimensional Electrophoresis*, Large Scale Biology Press, Washington, DC 1988.

[32] Schägger, H., Von Jagow, G., *Anal. Biochem.* 1987, *166*, 368–379.

[33] Matsudaira, P., *J. Biol. Chem.* 1987, *262*, 10035–10038.

[34] Tempst, P., Link, A. J., Riviere, L. R., Fleming, M., Elicone, C., *Electrophoresis* 1990, *11*, 537–553.

[35] Speicher, D. W., in: Hugli, T. E. (Ed.), *Techniques in Protein Chemistry*, Academic Press, San Diego 1988, pp. 24–35.

[36] Speicher, D. W., Mozdzanowski, J., Beam, K., Chen, D., *J. Protein Chem.* 1990, *9*, 254–255.

[37] Tempst, P., Riviere, L., *Anal. Biochem.* 1989, *183*, 290–300.

[38] Hunkapiller, M. W., Moyer-Granlund, K., Whiteley, N. W., in: Shively, J. E. (Ed.), *Methods of Protein Microcharacterization*, Humana Press, Clifton, NJ 1986, pp. 315–327.

[39] Hunkapiller, M. W., Granlund-Moyer, K., Whiteley, N. W., in: Shively, J. E. (Ed.), *Methods of Protein Microcharacterization*, Humana Press, Clifton, NJ 1986, pp. 223–247.

[40] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J., *J. Mol. Biol.* 1990, *215*, 403–410.

[41] Link, A. J., Eng, J., Yates III, J. R., *Am. Lab.* 1996, *28*, 27–30.

[42] Ben-Bassat, A., Bauer, K., Chang, S.-Y., Myambo, K., Boosman, A., Chang, S., *J. Bacteriol.* 1987, *169*, 751–757.

[43] Hirel, P.-H., Schmitter, J.-M., Dessen, P., Fayat, G., Blanquet, S., *Proc. Natl. Acad. Sci. USA* 1989, *86*, 8247–8251.

[44] von Heijne, G., *Eur. J. Biochem.* 1983, *133*, 17–21.

[45] von Heijne, G., *Nucleic Acids Res.* 1986, *14*, 4683–4690.

[46] Murphy, C. K., Beckwith, J., in: Neidhardt, F. C., Curtiss, R., Gross, C., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik,

B., Riley, M., Schaechter, M., Umbarger, H. E. (Eds.), *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ASM Press, Washington, DC 1996, pp. 967–978.
[47] Nikaido, H., in: Neidhardt, F. C., Curtiss, R., Gross, C., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Riley, M., Schaechter, M., Umbarger, H. E. (Eds.), *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ASM Press, Washington, DC 1996, pp. 29–47.
[48] Oliver, D. B., in: Neidhardt, F. C., Curtiss, R., Gross, C., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Riley, M., Schaechter, M., Umbarger, H. E. (Eds.), *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ASM Press, Washington, DC 1996, pp. 88–103.
[49] Bachmair, A., Finley, D., Varshavsky, A., *Science* 1986, *234*, 179–234.
[50] Rileyl, M., Labedan, B., in: Neidhardt, F. C., Curtiss, R., Gross, C., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Riley, M., Schaechter, M., Umbarger, H. E. (Eds.) *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ASM Press, Washington, DC 1996, pp. 2118–2202.
[51] Schnaitman, C. A., *Arch. Biochem. Biophys.* 1973, *157*, 541–552.
[52] Schnaitman, C. A., *J. Bacteriol.* 1974, *118*, 442–453.
[53] Mozdzanowski, J., Hemback, P., Speicher, D. W., *Electrophoresis* 1992, *13*, 59–64.
[54] Wilkins, M. R., Ou, K., Appel, R. D., Sanchez, J. C., Yan, J. X., Golaz, O., Farnsworth, V., Cartier, P., Hochstrasser, D. F., Williams, K. L., Gooley, A. A., *Biochem. Biophys. Res. Commun.* 1996, *221*, 609–613.
[55] Mann, M., Wilm, M., *Anal. Chem.* 1994, *66*, 4390–4399.
[56] Shaw, G., *Proc. Natl. Acad. Sci. USA* 1993, *90*, 5138–5142.
[57] Pappin, D. J. C., Hojrup, P., Bleasby, A. J., *Curr. Biol.* 1993, *3*, 327–332.
[58] Gültekin, H., Heermann, K. H., *Anal. Biochem.* 1988, *172*, 320–329.
[59] Bjellqvist, B., Pasquali, C., Ravier, R., Sanchez, J.-C., Hochstrasser, D. F., *Electrophoresis* 1993, *14*, 1357–1365.
[60] Bjellqvist, B., Sanchez, J.-C., Pasquali, C., Ravier, F., Paquet, N., Frutiger, S., Hughes, G. J., Hochstrasser, D. J., *Electrophoresis* 1993, *14*, 1375–1378.
[61] Figeys, D., Ducret, A., Yates, J. R., Aebersold, R., *Nature Biotechnol.* 1996, *14*, 1579–1583.
[62] Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., Mann, M., *Nature* 1996, *379*, 466–469.
[63] Almirón, M., Link, A. J., Furlong, D., Kolter, R., *Genes Develop.* 1992, *6*, 2646–2654.
[64] Borodovsky, M., Koonin, E. V., Rudd, K. E., *Trends Biol. Sci.* 1994, *19*, 309–313.
[65] Gold, L., *Annu. Rev. Biochem.* 1988, *57*, 199–233.
[66] Blattner, F. R., Burland, V., Plunkett, G., Sofia, H. J., Daniels, D. L., *Nucleic Acids Res.* 1993, *21*, 5408–5417.
[67] Herrick, J., Kern, R., Guha, S., Landoulsi, A., Fayet, O., Malki, A., Kohiyama, M., *EMBO J.* 1994, *13*, 4695–4703.
[68] Johnston, M., Andrews, S., Brinkman, R., Cooper, J., Ding, H., Dover, J., Du, Z., Favello, A., Fulton, L., Gattung, S., Geisel, C., Kirsten, J., Kucaba, T., Hilier, L., Jier, M., Johnston, L., Langston, Y., Latreille, P., Louis, E. J., Macri, C., Mardis, E., Menezes, S., Mouser, L., Nhan, M., Rifkin, L., Riles, L., St. Peter, H., Trevaskis, E., Vaughan, K., Vignati, D., Wilcox, L., Wohldman, P., Waterston, R., Wilson, R., Vaudin, M., *Science* 1994, *265*, 2077–2082.
[69] Arai, K., Clark, B. F. C., Duffy, L., Jones, M. D., Kaziro, Y., Laursen, R. A., L'Italien, J., Miller, D. L., Nagarkatti, S., Nakamura, S., Nielsen, K. M., Petersen, T. E., Takahashi, K., Wade, M., *Proc. Natl. Acad. Sci. USA* 1980, *77*, 1326–1330.
[70] Hantke, K., Braun, V., *Eur. J. Biochem.* 1973, *256*, 3125–3129.
[71] Harold, F., Maloney, P. C., in: Neidhardt, F. C., Curtiss, R., Gross, C., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Riley, M., Schaechter, M., Umbarger, H. E. (Eds.), *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ASM Press, Washington, DC 1996, pp. 283–306.
[72] von Heijne, V., *Biochim. Biophys. Acta* 1988, *947*, 307–333.
[73] Shen, L. M., Lee, J.-I., Cheng, S., Jutte, H., Kuhn, A., Dalbey, R. E., *Biochemistry* 1991, *30*, 11775–117801.
[74] Dalbey, R. E., von Heijne, G., *Trends Biol. Sci.* 1992, *17*, 474–478.
[75] Lazdunski, A. M., *FEMS Microbiol. Rev.* 1989, *63*, 265–276.
[76] Miller, C. G., in: Neidhardt, F. C., Curtiss, R., Gross, C., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Riley, M., Schaechter, M., Umbarger, H. E. (Eds.), *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ASM Press, Washington, DC 1996, pp. 938–954.
[77] Tsunasawa, S., Stewart, J. W., Sherman, F., *J. Biol. Chem.* 1985, *260*, 5382–5391.
[78] Tobias, J. W., Shrader, T. E., Rocap, G., Varshavsky, A., *Science* 1991, *254*, 1374–1377.
[79] Shrader, T. E., Tobias, J. W., Varshavsky, A., *J. Bacteriol.* 1993, *175*, 4364–4374.
[80] Grantham, R., Gautier, C., Guoy, M., Jacobzone, M., Mercier, R., *Nucleic Acids Res.* 1981, *9*, r43–r51.
[81] Ikemura, T., *J. Mol. Biol.* 1981, *151*, 389–409.
[82] Guoy, M., Gautier, C., *Nucleic Acids Res.* 1982, *10*, 7055–7074.
[83] Gribskov, M., Devereux, J., Burgess, R. R., *Nucleic Acids Res.* 1984, *12*, 539–549.
[84] McLachlan, A. D., Staden, R., Boswell, D. R., *Nucleic Acids Res.* 1984, *12*, 9567–9575.
[85] Mulligan, M. E., Hawley, D. K., Entriken, R., McClure, W. R., *Nucleic Acids Res.* 1984, *12*, 789–800.
[86] Deuschle, U., Kammerer, W., Gentz, R., Bujard, H., *EMBO J.* 1986, *5*, 2987–2994.
[87] Folley, L. S., Yarus, M., *J. Mol. Biol.* 1989, *209*, 359–378.
[88] Guttman, G. A., Hatfield, G. W., *Proc. Natl. Acad. Sci. USA* 1989, *86*, 3699–3703.
[89] McCarthy, J. E. G., Gualerzi, C., *Trends Genet.* 1990, *6*, 78–85.
[90] Parsell, D. A., Silber, K. R., Sauer, R. T., *Genes Develop.* 1990, *4*, 277–286.
[91] Barrick, D., Villanueba, K., Childs, J., Kalil, R., Schneider, T. D., Lawrence, C. E., Gold, L., Stormo, G. D., *Nucleic Acids Res.* 1994, *22*, 1287–1295.
[92] Lobry, J. R., Gautier, C., *Nucleic Acids Res.* 1994, *22*, 3174–3180.
[93] Yoshida, T., Ueguchi, C., Mizuno, T., *J. Bacteriol.* 1993, *175*, 7747–7748.
[94] Yoshida, T., Ueguchi, C., Yamada, H., Mizuno, T., *Mol. Gen. Genet.* 1993, *237*, 113–122.
[95] Yamanaka, K., Mitani, T., Ogura, T., Niki, H., Hiraga, S., *Mol. Microbiol.* 1994, *13*, 301–312.
[96] Lee, S. J., Xie, A., Jiang, W., Etchegaray, J. P., Jones, P. G., Inouye, M., *Mol. Microbiol.* 1994, *11*, 833–839.
[97] Jacobson, F. S., Morgan, R. W., Christman, M. F., Ames, B. N., *J. Biol. Chem.* 1989, *264*, 1488–1496.
[98] Storz, G., Jacobson, F. S., Tartaglia, L. A., Morgan, R. W., Silveira, L. A., Ames, B. N., *J. Bacteriol.* 1989, *171*, 2049–2055.
[99] Christman, M. F., Morgan, R. W., Jacobson, F. S., Ames, B. N., *Cell* 1985, *41*, 753–762.
[100] Tartaglia, L. A., Storz, G., Brodsky, M. H., Lai, A., Ames, B. N., *J. Biol. Chem.* 1990, *265*, 10535–10540.
[101] Hemmingsen, S. M., Woolford, C., van der Vies, S. M., Tilly, K., Dennis, D. T., Georgopoulos, C. P., Hendrix, R. W., Ellis, R. J., *Nature* 1988, *333*, 330–334.
[102] Guthrie, B., Wickner, W., *J. Bacteriol.* 1990, *172*, 5555–5562.
[103] Zeilstra-Ryalls, J., Fayer, O., Georgopoulos, C., *Annu. Rev. Microbiol.* 1991, *45*, 301–325.
[104] Newton, W. A., Morino, Y., Snell, E. E., *J. Biol. Chem.* 1965, *240*, 1211–1218.
[105] Lin, E. C. C., in: Neidhardt, F. C., Curtiss, R., Gross, C., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Riley, M., Schaechter, M., Umbarger, H. E. (Eds.), *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ASM Press, Washington, DC 1996, pp. 307–342.
[106] Benner-Lugar, D., Boos, W., *Mol. Gen. Genet.* 1988, *214*, 579–587.
[107] Hogg, R. W., Voelker, C., Carlowitz, I. V., *Mol. Gen. Genet.* 1991, *229*, 453–459.
[108] Mowbray, J., Moses, V., *Eur. J. Biochem.* 1976, *66*, 25–36.
[109] Srivastava, D. K., Bernhard, S. A., *Science* 1986, *234*, 1081–1086.
[110] Srere, P. A., *Annu. Rev. Biochem.* 1987, *56*, 89–124.
[111] Braun, V., *Biochim. Biophys. Acta* 1975, *415*, 335–377.
[112] Mizusawa, S., Gottesman, S., *Proc. Natl. Acad. Sci. USA* 1983, *83*, 358–362.
[113] Chae, H. Z., Robison, K., Poole, L. B., Church, G., Storz, G., Rhee, S. G., *Proc. Natl. Acad. Sci. USA* 1994, *91*, 7017–7021.

[114] Noller, H. F., Nomura, M., in: Neidhardt, F. C., Curtiss, R., Gross, C., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Riley, M., Schaechter, M., Umbarger, H. E. (Eds.), *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ASM Press, Washington, DC 1996, pp. 167–186.
[115] Chittum, H., Champney, W. S., *J. Bacteriol.* 1994, *176*, 6192–6198.
[116] Cronan, J. E., LaPorte, D., in: Neidhardt, F. C., Ingraham, J. L., Low, B. L., Magasanik, B., Schaechter, M., Umbarger, H. E. (Eds.), *Escherichia coli and Salmonella: Cellular and Molecular Biology*, American Society for Microbiology, Washington, DC 1996, pp. 206–216.
[117] Garnak, M., Reeves, H. C., *Science* 1979, *203*, 1111–1112.
[118] Thorsness, P. E., Koshland, D. E., *J. Biol. Chem.* 1987, *262*, 10422–10425.
[119] Postma, P. W., Lengeler, J. W., Jacobson, G. R., in: Neidhardt, F. C., Ingraham, J. L., Low, B. L., Magasanik, B., Schaechter, M., Umbarger, H. E. (Eds.), *Escherichia coli and Salmonella: Cellular and Molecular Biology*, American Society for Microbiology, Washington, DC 1996, pp. 1149–1174.
[120] Yu, M., Goldberg, S., Goldberg, A. L., *Nature* 1992, *357*, 167–169.
[121] Chae, H. Z., Uhm, T. B., Rhee, S. G., *Proc. Natl. Acad. Sci. USA* 1994, *91*, 7022–7026.
[122] Krishna, R. G., Wold, F., *Adv. Enzymol.* 1993, *65*, 265–298.
[123] Cortay, J.-C., Rieul, C., Duclos, B., Cozzone, A. J., *Eur. J. Biochem.* 1986, *159*, 227–237.
[124] Dadssi, M., Cozzone, A. J., *Inter. J. Biochem.* 1990, *22*, 493–499.
[125] Terhorst, C., Wittmann-Liebold, B., Moller, W., *Eur. J. Biochem.* 1972, *25*, 13–19.
[126] Ferro-Luzzi Ames, G., Niakido, K., *J. Biol. Chem.* 1979, *254*, 9947–9950.
[127] Foster, P. L., *Methods Enzymol.* 1991, *204*, 114–124.
[128] Tokunaga, M., Tokunaga, H., Wu, H. C., *Proc. Natl. Acad. Sci. USA* 1982, *79*, 2255–2259.
[129] Siuzdak, G., *Proc. Natl. Acad. Sci. USA* 1994, *91*, 11290–11297.
[130] Yates III, J. R., Eng, J. K., McCormack, A. L., Schieltz, D., *Anal. Chem.* 1995, *67*, 1426–1436.
[131] Johnson, W. C., Silhavy, T. J., Boos, W., *Appl. Microbiol.* 1975, *29*, 405–413.
[132] Ferro-Luzzi Ames, G., Nikaido, K., *Biochemistry* 1976, *15*, 616–623.
[133] Sato, T., Ito, K., Yura, T., *Eur. J. Biochem.* 1977, *78*, 557–567.
[134] Ferro-Luzzi Ames, G., Prody, C., Kustu, S., *J. Bacteriol.* 1984, *160*, 1181–1183.
[135] Nakai, K., Kanehisa, M., *Proteins* 1991, *11*, 95–110.
[136] Kuwajima, G., Kawagishi, I., Homma, M., Asaka, J., Kondo, E., Macnab, R. M., *Proc. Natl. Acad. Sci. USA* 1989, *86*, 4953–4957.
[137] Derman, A. I., Puziss, J. W., Bassford, P. J., Beckwith, J., *EMBO J.* 1993, *12*, 879–888.
[138] Fath, M. J., Kolter, R., *Microbiol. Rev.* 1993, *57*, 995–1017.
[139] Manoil, C., Beckwith, J., *Proc. Natl. Acad. Sci. USA* 1985, *82*, 8129–8133.
[140] Durrenberger, M., Bjornsti, M.-A., Uetz, T., Hobot, J. A., Kellenberger, E., *J. Bacteriol.* 1988, *170*, 4757–4768.
[141] Thornton, M., Armitage, M., Maxwell, A., Dosanjh, B., Howells, A. J., Norris, V., Sigee, D. C., *Microbiology* 1994, *140*, 2371–2382.
[142] Storz, G., Tartaglia, L. A., Farr, S. B., Ames, B. N., *Trends Genet.* 1990, *6*, 363–368.
[143] Hayano, T., Takahashi, N., Kato, S., Maki, N., Suzuki, M., *Biochemistry* 1991, *30*, 3041–3048.
[144] Craig, E. A., Gambill, B. D., Nelson, R. J., *Microbiol. Rev.* 1993, *57*, 402–414.
[145] Brown, J. L., Roberts, W. K., *J. Biol. Chem.* 1976, *251*, 1009–1014.
[146] Hochstrasser, D. F., Frutiger, S., Paquet, N., Bairoch, A., Ravier, F., Pasquali, C., Sanchez, J.-C., Tissot, J.-D., Bjellqvist, B., Vargas, R., Appel, R. D., Hughes, G. J., *Electrophoresis* 1992, *13*, 992–1001.
[147] Wellner, D., Panneerselvam, C., Horecker, B. L., *Proc. Natl. Acad. Sci. USA* 1990, *87*, 1947–1949.
[148] Farries, T. C., Harris, A., Auffret, A. D., Aitken, A., *Eur. J. Biochem.* 1991, *196*, 679–685.
[149] Koonin, E. V., Tatusov, R. L., Rudd, K. E., in: Neidhardt, F. C., Curtiss, R., Gross, C., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Riley, M., Schaechter, M., Umbarger, H. E. (Eds.), *Escherichia coli and Salmonella: Cellular and Molecular Biology*, ASM Press, Washington, DC 1996, pp. 2203–2217.

6 Appendix: Tables and Figures 6–15

Table 1. Predicted *versus* observed *N*-terminal start sites

(A) ***N*-termini with methionine starts**

Locus	Predicted *N*-terminus[a)]	Observed *N*-terminus[b)]	Multiple *N*-termini[c)]
accC	MLDKIVIANRGEIA	**MLDKIVIANRGE**	
adk	MRIILLGAPGAGKG	**MRIILLGAPGAG**	
arcA	MQTPHILIVEDELVT	**MQTPHILIVEDE**	
aroG	MNYQNDDLRIKEIK	**MNYQNDDLRIKE**	
asd	MKNVGFIGWRGMVG	**MKNVGFIGWRGM**	
aspC	MFENITAAPADPIL	**MFENITAAPADP**	
atpA	MQLNSTEISELIKQ	**MQLNSTEISELI**	
bcp	MNPLKAGDIAPKFS	**MNPLKAGDIAPK**	
dapA	MFTGSIVAIVTPMD	**MFTGSIVAIVTP**	
dapD	MQQLQNIIETAFER	**MQQLQNIIETAF**	
dkaA	MQEGQNRKTSSL	**MQEGQNRK(PS)FXL**	
fabI	MGFLSGKRILVTGV	**MGFLSGKRILVT**	B
frr	MISDIRKDAEVRMD	**MISDIRKDAEVR**	
ftsZ	MFEPMELTNDAVIK	**MFEPMELTNDAV**	
gdhA	MDQTYSLESFLNHV	**MDQTYSLEXFLN**	
glyA	MLKREMNIADYDAE	**MLKREMNIADYD**	
guaB	MLRIAKEALTFDDV	**MLRIAKEALTFD**	
guaC	MRIEEDLKLGFKDV	**MRIEEDLKLGFL**	
hemX	MTEQEKTSAVVEET	**MTEQEKTSAAXE**	
hmpA	MLDAQTIATVKATI	**MLDAQTIATVKA**	
htpG	MKGQETRGFQSE	**MKGQETXGFQ**	
hupA	MNKTQLIDVIAEKA	**MNKTQLIDVIAE**	

Table 1. continued

(A) *N*-termini with methionine starts

Locus	Predicted *N*-terminus[a)]	Observed *N*-terminus[b)]	Multiple *N*-termini[c)]
hupB	MNKSQLIDKIAAGA	MNKSQLIDKIAA	
icdA	MESKVVVPAQGKKI	MESKVVVPAQGK	
ilvI	MEMLSGAEMVVRSL	MEMLSGAEMVVR	
kdsA	MKQKVVSIGDINVA	M(GM)QKVVSIGDIN	
mdh	MKVAVLGAAGGIGQ	MKVAVLGAAGGI	
mopB	MNIRPLHDRVIVKR	MNIRPLHDRVIV	
mreB	MLKKFRGMFSNDLS	MLKKFRGMFSND	
nadE	MTLQQQIIKALGEN	MTLQQQIIKALG	
nfnB	MDIISVALKRHSTK	MDIISVALKRHS	
nuoB	MDYTLTRIDPNGEN	MDYTLTRIDPNG	
nuoI	MTLKELLVGFGTQV	MTLXELLVGFGT	
nusA	MNKEILAVVEAVSN	MNXEILAVVEXV	
panB	MKPTTISLLQKYKQ	MKPTTISLLQXY	
pckA	MRVNNGLTPQELEA	MRVNNGLTPQEL	
ppiB	MVTFHTNHGDIVIK	MVTFHTNHGDIV	
proS	MRTSQYLLSTLKET	MRTSQYLLSTLK	
ptsH	MFQQEVTITAPNGL	MFQQEVTITAPN	
ptsI	MISGILASPGIAF	MISGILAXPGI	
purC	MQKQAELYRGKAKT	MQKQAELYRGKA	
purH	MQQRRPVRRALLSV	MQQRRPVRRALL	
pykF	MKKTKIVCTIGPKT	MKKTKIVATIGP	
pyrI	MTHDNKLQVEAIKR	MTHDNKLQVEAI	B
rfaD	MIIVTGGAGFIGSN	MIIVTGGAGFIG	
rho	MNLTELKNTPVSEL	MNLTELKNTPVS	
rpiA	MTQDELKKAVGWA	MTQDELKKAVG	
rplC	MIGLVGKKVGMT	MIGLVGKKVG	
rplD	MELVLKDAQSALTV	MELVLKDAQSAL	
rplI	MQVILLDKVANLGS	MQVILLDKVANL	
rplY	MFTINAEVRKEQGK	MFTINAEVRREQ	
rpoA	MQGSVTEFLKPRLVD	MQGSVTEFLKPRL	
rpsA	MTESFAQLFEESLK	MTESFAQLFEES	B
rpsF	MRHYEIVFMVHPDQ	MRHYEIVFMVXP	
rpsJ	MQNQRIRIRLKAFD	MQNQRIXIRLLA	
rpsP	MVTIRLARHGAKKR	MVTIRLAR(EA)GA(VP)	
sdhA	MKLPVREFDAVVIG	MKLPVREFDAVV	
sdhB	MRLEFSIYRYNPD	MRLEFSIYRYN	
sucC	MNLHEYQAKQLFAR	MNLHEYQAKQLF	
thrC	MKLYNLKDHNEQVS	MKLYNLKDHNEQ	
tig	MQVSVETTQGLGRR	MQVSVETTQGLG	G
tnaA	MENFKHLPEPFRIR	MENFKHLPEPFR	
tpiA	MRHPLVMGNWKLNG	MRXPLVMGNXKL	
trpA	MERYESLFAQLKER	MERYESLFAQLK	
upp	MKIVEVKHPLVKHK	MKIVEVKHPLVK	
valS	MEKTYNPQDIEQPL	MEFTYNPQDIEQ	
ybdQ	MYKTIIMPVDVFEM	MYKTIIMPVDVF	
ychF	MGFKCGIVGLPNVG	XXFKXGIVGLPN	B
ydfG	MIVLVTGATAGFGE	MIVLVTGATAGF	
ygaG	MPLLDSFTVDH	(A)PLLDSFTV	B
ygiN	MLTVIAEIRTRPGQ	MLTVIAEIRTRP	
yjbJ	MNKDEAGGNWKQFK	MNKDEAGGNXKQ	

(B) Initiator methionine processed starts

Locus	Predicted *N*-terminus	Observed *N*-terminus	Multiple *N*-termini
accA	M*SLNFLDFEQPIAEL	SLNFLDFEQPIA	
aceE	M*SERFPNDVDPIETR	SERFPNDVDPIE	

Table 1. continued

(B)	Initiator methionine processed starts		
Locus	**Predicted *N*-terminus**	**Observed *N*-terminus**	**Multiple *N*-termini**
aceF	M*AIEIKVPDIGADEV	AIEIKVPDIGAD	
ahpC	M*SLINTKIKPFKNQA	SLINTKIKPFKN	
aldA	M*SVPVQHPMYIDGQF	SVPVQ(KG)PMYXD	
argD	M*AIEQTAITRATFDE	AIEQTAI(FS)RATF	
argG	M*TTILKHLPVGQRIG	TTILKHLPVGQR	
argI	M*SGFYHKHFLKLLD	SGFYXXXFLKGL	
asnS	M*SVVPVADVLQGR	AVVPVADVLQ	
atpD	M*ATGKIVQVIGAVVD	ATGKIVQVIGAV	
crr	M*GLFDKLKSLVSDDK	GLFDKLKSLVSD	
cspC	M*AKIKGQVKWFNESK	AKIKGQVKXFNE	
cysK	M*SKIFEDNSLTIGHT	SKIFEDNSLTIG	
dnaK	M*GKIIGIDLGTTNSC	GKIIGIDLGTTN	
dps	M*STAKLVKSKATNLL	STAKLVKSKATN	
efp	M*ATYYSNDFRAGLKI	ATYYSNDFRAGL	
eno	M*SKIVKIIGREIIDS	SKIVKIIGREII	
fabD	M*TQFAFVFPGQGS	TQFAFVFPGQ	
fabI	M*GFLSGKRILVTGVA	GFLSGKRILVTG	A
fba	M*SKIFDFVKPGVITG	SKIFDFVKPGVI	
fklB	M*TTPTFDTIEAQASY	TTPTFDTIEAQA	
fliC	M*AQVINTNSLSLITQ	AQVINTNSLSLI	
folE	M*PSLSKEAALVHEAL	PSLSKEAALVTE	
fumA	M*SNKPFHYQAPFPL	SNKPFHYQAPF	
fusA	M*ARTTPIARYRNIGI	ARTTPIARYRNI	
galU	M*AAINTKVKKAVIPV	AAINTKVKKAVI	F
gapA	M*TIKVGINGFGRIGR	TIKVGINGFGRI	
gcvT	M*AQQTPLYEQHTLCG	AQQTPLYEQHTL	
glnA	M*SAEHVLTMLNEH	SAEHVLTMLN	
glnS	M*SEAEARPTNFIRQI	SEAEARPTNFIR	
gltD	M*SQNVYQFIDLQRVD	SQNVYQFIDLQR	
glyS	M*SEKTFLVEIGTEEL	SEKTFLVEIGTE	
gpmA	M*AVTKLVLVRHGESQ	AVTKLVLVRHGE	
hisD	M*SFNTIIDWNSCT	SFNTIIDPNX(FYEK)T	
hns	M*SEALKILNNIRTLR	SEALKILNNIRT	
ilvC	M*ANYFNTLNLRQQLA	ANYFNTLNLRQQ	
leuA	M*SQQVIIFDTTLRDG	SQQVIIFDTTLR	
leuB	M*SKNYHIAVLPGDGI	SKNYHIAVLPGD	
leuC	M*AKTLYEKLFDAHVV	AKTLYEKLFDAH	
manX	V*TIAIVIGTHGWAAET	TIAIVIGTHGWA	
metE	M*TILNHTLGFPRVGL	TILNHTLGFPRV	
minD	M*ARIIVVTSGKGGVG	ARIIVVTSGKGG	
mopA	M*AAKDVKFGNDARVK	AAKDVKFGNDAR	
nuoG	M*ATIHVDGKEYEVNG	ATIHVDGKEYEV	
osmC	M*TIHKKGQAHWEGDI	TIHKKGQAHIEG	
pgk	M*SVIKMTDLDLAGKR	SVIKMTDLDLAG	
ppa	M*SLLNVPAGKDLPED	SLLNVPAGKDLP	G
purA	M*GNNVVVLGTQWGDE	GNNVXXLGTQXA(VL)	
purM	M*TDKTSLSYKDAGV	TDKTSLSXXDQ	
pyrB	M*ANPLYQKHIISIN	ANPLYQKHIIS	
pyrC	M*TAPSQVLKIRRPDD	TAPSQVLKIRRP	
pyrG	M*TTNYIFVTGGVVSS	TTNYIFVTGGVV	
pyrI	M*THDNKLQVEAIKRG	THDNKLQVEAIK	A
rplA	M*AKLTKRMRVIREKV	AKLTKRMRVIIE	
rplF	M*SRVAKAPVVVPAGV	SRVAKAPVVVPA	
rplL	M*SITKDQIIEAVAAM	SITKDXIIEXV	

Table 1. continued

(B) Initiator methionine processed starts

Locus	Predicted *N*-terminus	Observed *N*-terminus	Multiple *N*-termini
rpsA	M*TESFAQLFEES	TESFAQLFE	A
rpsB	M*ATVSMRDMLKAGVH	ATVSMRDMLKAG	
serA	M*AKVSLEKDKIKFLL	AKVSLEKDKIKF	
serC	M*AQIFNFSSGPAMLP	AQIFNFSSGPAM	
sodA	M*SYTLPSLPYAYDAL	SYTLPSLPYAYD	
sodB	M*SFELPALPYAKDAL	SFELPALPYAKD	
sseA	M*STTWFVGADWLAEH	STTXFVGADDXA	
sspA	M*AVAANKRSVMTLFS	AVAANKRSVMTL	
sucB	M*SSVDILVPDLPESV	SSVDILVPDLPE	
sucD	M*SILIDKNTKVICQG	SILIDKNTKVIC	
talB	M*TDKLTSLRQYTTVV	TDKLTSLRQYTT	
tpx	M*SQTVHFQGNPVTVA	SQTVHFQGNPVT	
trpB	M*TTLLNPYFGEFGGM	TTLLNPYFGEFG	
tsf	M*AEITASLVKELRER	AEITAGLVKELR	
uspA	M*AYKHILIAVDLSPE	AYKHILIAVDLS	
ychF	M*GFKCGIVGLPNVGK	(AS)FKXGIVGLPNV	A
yfiA	M*TMNITSKQMEITPA	TMNITSKQMEIF	
ygaU	M*GLFNFVKDAGEKLW	GLFNFVKDAGEK	
ygaG	M*PLLDSFTVDHT	PLLDSFTVD	A
ygfZ	M*AFTPFPPRQPTASA	AFTPFPPRQPTA	
yggX	M*SRTIFCTFLQREAA	SRTIFXTFLQIE	
yhbG	M*ATLTAKNLAKAYKG	ATLTAKNLAXAY	
yifE	M*AESFTTTNRYFDNK	AESFTTTNRYFD	
yjgF	M*SKTIATENAPAAIG	SKTIATENAPAA	
yjjK	M*AQFVYTMHRVGK	A(DE)FVYTMXRV(LI)(GA)	

(C) *N*-termini with cleaved signal peptides

Locus	Predicted *N*-terminus	Observed *N*-terminus	Multiple *N*-termini
agp	MNKTLIAAAVAGIVLLASNAQA*QTVPEGYQLQQVLM	QTVPEGYQLQQV	
araF	MHKFTKALAAIGLAAVMSQSAMA*ENLKLGFLVKQPEE	ENLKLGFLVKQP	
argT	MKKSILALSLLVGLSTAASSYA*ALPETVRIGTDTTYA	PLPETVRIGTDT	
artI	MKKVLIAALIAGFSLSATA*AETIRFATEASYPP	AETIRFATEASY	
artJ	MKKLVLAALLASFTFGASA*AEKINFGVSATYPP	AEKINFGVSATY	
btuB	MIKKASLLTACSVTAFSAWA*QDTSPDTLVVTANR	QDTSPDTLVVTA	
cpdB	MIKFSATLLATLIAASVNA*ATVDLRIMETTDLH	ATVDLRIMETTD	
cysP	MAVNLLKKNSLALVASLLLAGHVQA*TELLNSSYDVSREL	TELLNSXYDVSR	
dppA	MRISLKKSGMLKLGLSLVAMTVAASVQA*KTLVYCSEGSPEGF	KTLVYXSEGSPE	
dsbA	MKKIWLALAGLVLAFSASA*AQYEDGKQYTTLEK	AQYEDGKQYTTL	
dsbC	MKKGFMLFTLLAAFSGFAQA*DDAAIQQTLAKMGIKSSDIQ	DDAAIQQTLAKM	
eco	MKTILPAVLFAAFATTSAWA*AESVQPLEKIAPYP	AESVQPLEKIAP	
fkpA	MKSLFKVTLLATTMAVALHAPITFA*AEAAKPATAADSKA	AEAAKPATAADS	
fliY	MKLAHLGRQALMGVMAVALVAGMSVKSFA*DEGLLNKVKERGTL	DEGLLNKVKERG	
glnH	MKSVLKVSLAALTLAFAVSSHA*ADKKLVVATDTAFV	ADKKLVVATDTA	
hdeA	MKKVLGVILGGLLLLPVVSNA*ADAQKAADNKKPVN	ADAQKAADNKKP	
hdeB	MGYKMNISSLRKAFIFMGAVAALSLVNAQSALA*ANESAKDMTCQEFI	ANESAKDMTHQE	
hisJ	MKKLVLSLSLVLAFSSATAAFA*AIPQNIRIGTDPTY	AIPQNIRIGTDP	
hlpA	VKKWLLAAGLGLALATSAQA*ADKIAIVNMGSLFQ	ADKIAIVNMXSL	
imp	MKKRIPTLLATMIATALYSQQGLA*ADLASQC	ADLAS	
livJ	MNIKGKALLAGCIALAFSNMALA*EDIKVAVVGAMSGP	EDIKVAVVGAMS	
livK	MKRNAKTIIAGMIALAISHTAMA*DDIKVAVVGAMSGP	DDIKVAVVGAMS	
lolA	MMKKIAITCALLSSLVASSVWA*DAASDLKSRLDKVS	DAASDLKSRLDK	
malE	MKIKTGARILALSALTTMMFSASALA*KIEEGKLVIWINGD	KIEEGKLVIXIN	
mdoG	MMKMRWLSAAVMLTLYTSSSWA*FSIDDVAKQAQSLA	FXIDDVAKQAXS	
mglB	MNKKVLTLSAVMASMLFGAAAHA*ADTRIGVTIYKYDD	ADTRIGVTIYKY	
ompA	MKKTAIAIAVALAGFATVAQA*APKDNTWYTGAKLG	APKDNTWYTGAK	

Table 1. continued

(C) *N*-termini with cleaved signal peptides

Locus	Predicted *N*-terminus	Observed *N*-terminus	Multiple *N*-termini
ompC	MKVKVLSLLVPALLVAGAANA*AEVYNKDGNKL	AEVYNKDGN	
ompF	MMKRNILAVIVPALLVAGTANA*AEIYNKDGNKVDLY	AEIYNKDGNKVD	G
oppA	MTNITKRSLVAAGVLAALMAGNVALA*ADVPAGVTLAEKQT	ADVPAGVTLAEKQ	
osmY	MTMTRLKISKTLLAVMLTSAVATGSAYA*ENNAQTTNESAGQK	ENNAQTTNESAG	
potD	MKKWSRHLLAAGALALGMSAAHA*DDNNTLYFYNWTEY	DDNNTLYFYNXT	
potF	MTALNKKWLSGLVAGALMAVSVGTLA*AEQKTLHIYNW	AENKTLXIYNV	
pstS	MKVMRTTVATVVAATLSMSAFSVFA*EASLTGAGATFPAP	EASLTGAGATFP	
rbsB	MNMKKLATLVSAVALSATVSANAMA*KDTIALVVSTLNNP	KDTIALVVSTLN	
sbp	MNKWGVGLTFLLAATSVMA*KDIQLLNVSYDPTR	KDIQLLNVSYDP	
sufI	MSLSRRQFIQASGIALCAGAVPLKASA*AGQQQPLPVPPLLE	AGQQQPLPVPPL	
surA	MKNWKTLLLGIAMIANTSFA*APQVVDKVAAVVNN	APQVVDKVAAVV	
tolC	MKKLLPILIGLSLSGFSSLSQA*ENLMQVYQQARLSN	ENLMQVYQQARL	
ushA	MKLLQRGVALALLTTFTLASETALA*YEQDKTYKITVLHT	YEQDKTYKITVL	
xylF	MKIKNILLTLCTSLLLTNVAAHA*KEVKIGMAIDDLRL	KEVKIGMAIDDL	
yacK	MQRRDFLKYSVALGVASALPLWSRAVFA*AERPTLPIPDLLTT	AERPTLPIPTLL	
yaeT	MAMKKLLIASLLFSSATVYG*AEGFVVKDIHFEGL	AEGFVVKDIHFE	G
ybiS	MNMKLKTLFAAAFAVVGFCSTASA*VTYPLPTDGSRLVG	VTYPLPTDGSRL	
yceI	MKKSLLGLTFASLMFSAGSAVA*ADYKIDKEGQHAFV	ADYKIDKEGQHA	
ydcG	MDRRRFIKGSMAMAAVCGTSGIASLFSQAAFA*ADSDIADGQTQRFD	ADSDIADGQTQR	
yebL	MKCYNITLLIFITIIGRIMLHKKTLLFAALSAALWGGATQAADA*AVVASLKPVGFIAS	AVVASLKPVGFI	
yphF	MPTKMRTTRNLLLMATLLGSALFARA*AEKEMTIGAIYLDT	AEKEMTIGAIYL	
yhbN	MKFKTNKLSLNLVLASSLLAASIPAFA*VTGDTDQPIHIESD	VTGDTDQPIHIE	
yhjJ	MQGTKIRLLAGGLLMMATAGYVQA*DALQPDPAWQQGTL	DALQPDPAXQQG	
ytfJ	MTLRKILALTCLLLPMMASA*HQFETGQRVPPIGI	HQFETGQRVPPI	
ytfQ	MWKRLLIVSAVSAAMSSMALA*APLTVGFSQVGSES	APLTVGFSQVGS	
yjbP	MRKITQAISAVCLLFALNSSAVALA*SSPSPLNPGTNVAR	SSPSPLNPGTNV	C
yjbP	MRKITQAISAVCLLFALNSSAVA*LASSPSPLNPGTNV	LASSPSPLNPGT	C

(D) *N*-termini with non-ATG start sites[d)]

Locus	Predicted *N*-terminus	Observed *N*-terminus	Multiple *N*-termini
carA	LIKSALLVLEDGTQ	MIKSALLVLEDG	
pnp	LLNPIVRKFQYGQH	MLNPIVRKFQYG	
prsA	VPDMKLFAGNATPE	(PA)PDMKLFAGNAT	
yacI	VLEEYRKHVAERAA	MLEEYRKHVAER	
yihK	VIEKLRNIAIIAHV	MIEKLRNIAIIA	

(E) Blocked *N*-termini

Locus	Predicted *N*-terminus	Observed *N*-terminus	Multiple *N*-termini
atpF	MNLNATILGQAIAF		
pal	MQLNKVLKGLMIALPVMAIAA**CSSNKNASNDGS		
slp	MNMTKGALILSLSFLLAA**CSSIPQNIKGNN		
tufA/B	M**SKEKFERTKPHVNV		
yajG	MFKKILFPLVALFMLAG**CAKPPTTIEVSP		

(F) Misinterpreted 5′ *N*-termini

Locus	Predicted *N*-terminus	Observed *N*-terminus	Multiple *N*-termini
aroK	MRFQFMSCRRSLSEAGLSLTNSLSSTEKM*AEKRNIFLVGPMG	AEKRNIFLVGPM	
cysI	MSM*SEKHPGPLVVEGKL	SEKHPGPLVVEG	
lpdA	MM*STEIKTQVVVLGAG	STEIKTQVVVLG	
ynaF	MNSVITQKVSSGVTLYADTKTGGF*MNRTILVPIDISDS	MNRTILVPIDIS	
yeaD	MKLKDCV*MIKKIFALPVIEQI	MIKKIFALPVIE	
yhfO	MMYGVYRA*MKLPIYLDYSATTP	MKLPIYLDYSAT	
yiaE	MERS*MKPSVILYKALP	M(NI)PSVI(NVD)YTAIP	
yigW	MKKFAAVIAVMALCSAPV*MAAEQGGFSGPSATQS	AAAEQGGFSGPSA	
galU	MAAINT*KVKKAVIPVAGLGT	ANLKAVIPVAGL	B
galU	MAAIN*TKVKKAVIPVAGLGT	TINLKAVIPVAG	B

Table 1. continued

(G)	*N*-termini from internal sites		
Locus	Predicted *N*-terminal start sites[e]	Observed *N*-terminus	Multiple *N*-termini
ahpC	(67)Y*AVSTDTHFTHKA(104)	AVSTDTHFTHKA	
gadA	(378)F*KLKDGEDPGYTL(72)	LLKDGEDPGYTL	
ompF	(36)K*AVGLHYFSK(313)	AVGLHYFSK	C
ppa	(92)L*KMTDEAGEDAKL(68)	KMTDEAGEDAKL	B
rplM	(34)R*RLRGKHKAEYTP(96)	RLRGKHKAEYTP	
tig	(35)K*KVRIDGFRKGKV(381)	KVRIDGFRKGKV	A
tig	(42)R*KGKVPMNIVAQ(374)	KGKVPMNIVAQ	A
tig	(43)R*KGKVPMNIVAQR(373)	KGKVPMNIVAQR	A
tig	(44)K*GKVPMNIVAQRY(372)	GKVPMNIVAQRY	A
tufA/B	(308)Y*ILSKDEGGRHTP(70)	ILSKDEGGRHTP	
yaeT	(350)R*KIRFEGNDTSKD(448)	KIRFEGNDTSXD	C
yeeQ	(104)A*ADIVVHPGETV(976)	ADIVVHPGTTT	
yhjW	(228)A*RVDESSDNNSLL(245)	RVDESSDNNSLL	

a) The predicted *N*-terminal sequence is from the completed genome of *E. coli* K-12 strain MG1655. An "*" in the protein sequence shows the observed start site based on the *N*-terminal sequence tag. A "**" in the protein sequence shows the predicted *N*-terminus of the mature protein based on published literature.

b) The observed *N*-termini are based on the data in Tables 6–15. An "X" indicates no amino acid was identified for that position during Edman sequencing. Parentheses indicate two amino acids were observed for that position during Edman sequencing. An underlined amino acid indicates a discrepancy between the predicted and observed protein sequence.

c) For genes that have more than one unique *N*-terminal sequence tag, the class (A–G) of the other observed *N*-terminal sequence is indicated.

d) The genes *manX* and *hlpA* have non-ATG initiator codons (gtg), but because their gene products have processed *N*-termini, they are listed under other classes.

e) For observed *N*-termini matching the internal region of an *E. coli* gene, the number of amino acids between the predicted *N*- and *C*-terminus of the conceptual protein and the observed *N*-terminus is shown in parenthesis.

Table 2. Distribution of *E. coli* *N*-terminal amino acid residues

NH_2-terminal residue[a]	PepM specificity[b]	*N*-end rule stability[c]	Waller[e]	Observed in this project		
				Total[f]	w/o signal Seq.[g]	w/ signal Seq.[h]
M (2.3)	+	+	37	37	49	0
A (7.5)	+	+	26	25	17	51
S (7.3)	+	+	26	15	20	2
T (6.0)	+	+	7	7	9	2
K (5.8)	–	(–)[d]	1	2	0	9
E (6.2)	–	+	(2)[i]	2	0	9
G (7.1)	+	+	0	3	4	0
D (5.2)	–	+	(2)[i]	3	0	11
Q (4.3)	–	+	0	1	0	4
H (2.2)	–	+	0	1	0	2
Y (3.2)	–	–	0	1	0	2
P (5.1)	+	+	0	1	1	0
R (5.2)	–	(–)[d]	0	0	0	0
W (1.3)	–	–	0	0	0	0
C (1.8)	+	+	0	0	0	0
F (3.9)	–	–	0	0	0	2
I (5.5)	–	+	(1)[j]	0	0	0
N (4.5)	–	+	0	0	0	0
L (9.1)	–	–	(1)[j]	0	0	2
V (6.5)	+	+	0	1	0	4

a) Number in parenthesis is the frequency of amino acid usage for the predicted *E. coli* proteins in the nonredundant protein database [149].

b) Amino acids following the initiator methionine residue that determine the specificity of the PepM aminopeptidase [43]. "+" indicates cleavage of the initiator methionine.

c) *N*-terminal amino acids that stabilize (+) or destabilize (–) the protein [78].

d) Secondary destabilizing *N*-terminal amino acids [78].

e) Data from Waller [23] using a crude protein extract.

f) Overall frequency of the *N*-terminal amino acids for all proteins with identified ORFs. A protein was counted once for all isoforms with the identical *N*-terminal amino acid sequence.

g) *N*-terminal amino acid frequency for proteins without a recognized signal peptide sequence.

h) *N*-terminal amino acid frequency for proteins with a recognized signal peptide sequence.

i) Aspartic and glutamic acid are combined.

j) Leucine and isoleucine are combined.

Table 3. *E. coli* protein isoforms

(A) Growth phase in minimal media

Protein	Spot ID	op*I*[a)]	Molecular mass (kDa)	N-abd (molecules/cell)	Functional name
AhpC	M2-1	5.16	39.7	2260	Alkyl hydroperoxide reductase, C22 subunit
	M31-1	5.15	21.6	6040	
	M97-2	5.04	21.9	300	
		4.90	***20.6***		
CysK	M10-1	6.17	33.8	2820	O-Acetylserine sulfhydrolase A
	M99	6.01	33.7	240	
		5.94	***34.4***		
Eno	M15	5.60	44.4	2160	Enolase
	M75-1	5.40	44.6	200	
	M76-1	5.51	44.4	1660	
		5.23	***45.5***		
GalU	M109-1	5.34	38.3	260	Uridine diphosphoglucose pyrophosphorylase
	M101-1,2	6.06	31.4	260	
		5.00	***32.8***		
GapA	M108	7.42	35.7	1660	Glyceraldehyde-3-phosphate dehydrogenase A
	M23	7.31	35.4	1840	
	M34	6.77	35.5	740	
	M67-2	6.55	35.7	220	
		7.11	***35.4***		
GlnA	M28	5.44	54.1	1040	Glutamine synthetase
	M27	5.38	54.5	420	
		5.23	***51.8***		
GltD	M127	5.81	50.3	620	Glutamate synthase, small subunit
	M128-2	5.70	50.3	180	
		5.52	***51.9***		
GlyA	M16	6.41	43.8	3180	Serine hydroxymethyltransferase
	M92-2	6.26	44.1	300	
		6.46	***45.3***		
HdeA	M196	4.50	12.3	8460	*hns* deletion induced protein
	M199	4.50	29.5	1060	
		4.49	***9.2***		
IcdA	M33-1	5.25	44.8	8180	Isocitrate dehydrogenase
	M78	5.19	46.5	260	
	M79	5.17	44.8	1480	
		5.00	***45.8***		
IlvC	M71	5.45	51.2	1420	Ketol-acid reductoisomerase
	M77-1	5.44	49.9	160	
	M72	5.38	51.7	480	
		5.07	***53.9***		
LpdA	M84	6.15	49.9	380	Lipoamide dehydrogenase
	M86	6.08	50.1	160	
		6.10	***50.6***		

Table 3. continued
(A) Growth phase in minimal media

Protein	Spot ID	opI[a)]	Molecular mass (kDa)	N-abd (molecules/cell)	Functional name
Mdh	M9	5.94	34.1	2640	Malate dehydrogenase
	M111	5.69	34.4	260	
		5.56	***32.3***		
MetE	M186	6.08	93.0	11520	Tetrahydropteroyltriglumate methyltransferase
	M185	6.04	95.7	8640	
	M176	5.96	92.3	4260	
		5.97	***84.6***		
MopA	M14	5.02	57.7	2180	Molecular chaperone
	M46	4.97	58.7	360	
		4.67	***57.2***		
OmpA	M8	5.85	33.1	2300	Outer membrane protein 3a
	M118-2	5.40	27.7	140	
	M52-1	5.24	27.5	3380	
	M151	5.13	27.6	1480	
		5.74	***35.2***		
OmpF	M154	4.71	36.4	17600	Outer membrane protein 1a
	M51	4.61	36.7	2160	
		4.48	***37.1***		
OppA	M20	6.31	55.1	760	Oligopeptide transport protein/binding protein
	M82	6.16	55.1	380	
		6.17	***58.4***		
PotD	M47	4.92	35.1	1040	Spermidine/putrescine transport protein
	M149	4.84	35.3	260	
		4.69	***36.5***		
Ppa	M31-2	5.15	21.6	5480	Inorganic pyrophosphatase
	M97-1	5.04	21.9	500	
		4.90	***19.6***		
RpsB	M161	7.14	32.2	4540	30S ribosomal subunit protein S2
	M160	4.54	30.0	3540	
		7.16	***26.7***		
RpsF	M141	5.45	14.2	560	30S ribosomal subunit protein S6
	M142	5.38	14.2	1120	
	M143	5.25	14.2	940	
		5.17	***15.2***		
YaeT	M60	4.98	81.8	180	Novel protein
	M61	4.96	82.6	80	
		4.72	***88.4***		
YacI	M164	5.33	99.4	1120	Novel Protein
	M181	5.28	98.6	140	
		5.14	***93.5***		

Table 3. continued

(B) Early stationary phase in rich media

Protein	Spot ID	opI[a)]	Molecular mass (kDa)	N-abd (molecules/cell)	Functional name
Agp	R73	5.31	46.1	60	Periplasmic acid glucose-1-phosphatase
	R87	5.45	45.1	380	
		5.35	***43.6***		
DnaK	R130-2	4.71	73.5	20	Molecular chaperone protein
	R129	4.67	73.9	140	
		4.67	***69.0***		
HdeA	R35-1	4.55	10.7	5460	*hns* Deletion induced protein
	R51	4.55	17.2	1020	
		4.49	***9.2***		
Hns	R29	5.90	11.6	1020	Histone-like protein HLP-II
	R25-2	5.48	14.3	1780	
		5.25	***15.4***		
MalE	R102-2	5.60	40.9	20	Periplasmic maltose binding protein
	R56-1	5.25	40.9	720	
	R82	5.15	40.8	240	
		5.07	***40.7***		
MglB	R60	5.21	32.7	1320	Periplasmic galactose binding protein
	R103-2	5.09	35.0	20	
	R110	5.05	32.9	40	
		5.14	***33.4***		
OmpA	R146	6.21	36.8	500	Outer membrane protein 3a
	R140	4.86	31.7	2500	
		5.74	***35.2***		
OppA	R54	6.22	56.8	1300	Oligopeptide transport protein/binding protein
	R117	6.08	57.0	400	
		6.17	***58.4***		
OsmY	R45	5.23	16.5	780	Hyperosmotically inducible periplasmic protein
	R46	4.97	16.6	160	
		5.40	***18.2***		
PotD	R57	4.73	37.3	380	Spermidine/putrescine transport protein
	R105	4.62	37.8	40	
		4.69	***36.5***		
PtsH	R22-1	5.85	9.3	4180	Phosphohistidinoprotein-hexose phophotransferase
	R23-1	5.28	9.5	2100	
		5.74	***9.1***		
SodB	R76	5.68	22.8	900	Superoxide dismutase, iron
	R68-2	5.49	23.4	20	
		5.88	***21.1***		
TalB	R89	5.21	34.3	100	Transaldolase B
	R101-1	5.08	36.5	60	
		4.94	***35.1***		

Table 3. continued
(B) Early stationary phase in rich media

Protein	Spot ID	op*I*[a)]	Molecular mass (kDa)	N-abd (molecules/cell)	Functional name
UshA	R91	5.50	60.6	80	UDP-sugar hydrolase (5'-nucleotidase)
	R113	5.20	30.4	20	
		5.37	*58.2*		
XylF	R103-1	5.09	35.0	100	Xylose binding protein
	R101-2	5.08	36.5	20	
		4.93	*33.3*		
YjbJ	R21	5.40	9.3	12620	Novel protein
	R23-2	5.28	9.5	60	
	R50	4.95	10.0	440	
		5.27	*8.3*		
YjbP	R121-1	6.31	27.2	20	Novel protein
	R120	6.27	27.1	20	
		9.05	*19.2*		

a) Underlined isoelectric points and molecular masses are the predicted values calculated from the conceptual protein sequence.

Table 4. Analyzed *E. coli* proteins ranked by abundance and sorted by function

Functional classification[a)]		Growth, minimal		Stationary, rich		
		N-abd[b)]	Rank[c)]	N-abd	Rank	Functional name
I Small molecule metabolism						
a. Degradation (152)						
	GalU	520	98			Uridine diphosphoglucose pyrophosphorylase
	TnaA			7160	4	Tryptophanase
b. Energy metabolism						
1. Glycolysis (17)						
	GapA	4460	14			Glyceraldehyde-3-phosphate dehydrogenase
	Eno	4020	16	(140)	77	Enolase
	Pgk	2990	24	1460	31	Phosphoglycerate kinase
	Fba	1480	46	(60)	93	Fructose-biphosphate aldolase
	TpiA	1120	62	(60)	98	Triosephosphate isomerase
	GpmA	960	68	(40)	104	Phosphoglyceromutase
	PykF	500	101			Pyruvate kinase II
2. Pyruvate dehydrogenase (4)						
	AceE	1140	60	(400)	55	Pyruvate dehydrogenase (decarboxylase component)
	LpdA	540	97	(1880)	25	Lipoamide dehydrogenase
	AceF			(1120)	35	Dihydrolipoamide acetyltransferase
3. Tricarboxylic acid cycle (16)						
	IcdA	9900	5			Isocitrate dehydrogenase
	Mdh	2900	28	2380	21	Malate dehydrogenase
	SdhA	1900	41	(1360)	33	Succinate dehydrogenase, flavoprotein subunit
	SucD	900	73			Succinyl-CoA synthetase, alpha subunit
	SdhB	480	104			Succinate dehydrogenase, iron sulfur subunit
	SucB	420	113			2-Oxoglutarate dehydrogenase
	SucC	340	129			Succinyl-CoA synthetase, beta subunit
	FumA	160	160			Fumerase A
4. Pentose phosphate pathway (8)						
	RpiA	740	84			Ribosephosphate isomerase
	TalB	380	119	(160)	73	Transaldolase B
5. Entner-Doudoroff pathway (3)						
6. Respiration (104)						
	NuoB			(520)	50	NADH dehydrogenase I chain B
	HmpA			(100)	84	Flavohemoprotein
	NuoI			(60)	95	NADH dehydrogenase I chain I
	NuoG			(60)	96	NADH dehydrogenase I chain G

Table 4. continued

Functional classification[a]	Growth, minimal		Stationary, rich		Functional name
	N-abd[b]	Rank[c]	N-abd	Rank	
7. Fermentation (25)					
AldA	280	136			Aldehyde dehydrogenase family
8. ATP-proton interconversion (9)					
AtpD	1210	57	(4820)	12	Membrane-bound ATP synthase, beta subunit
AtpA	820	78	(2000)	24	Membrane-bound ATP synthase, alpha subunit
AtpF[d]			(5200)	11	Membrane-bound ATP synthase, b subunit
c. Central inter. metabolism (76)					
Ppa	5980	12	(200)	68	Pyrophosphatase
CysI	640	87			Sulfite reductase, alpha subunit
GltD	800	81			Glutamate synthase, small subunit
GadA			3820	15	Glutamate decarboxylase
Agp			(440)	54	Periplasmic acid glucose-1-phosphatase
UshA			(80)	90	UDP-sugar hydrolase
GcvT			(40)	103	T protein of glycine cleavage system
PckA			(20)	121	Phosphoenolpyruvate carboxylase
SseA			(20)	124	Rhodanese similarity
d. Amino acid biosynthesis					
1. Glutamate family (20)					
GlnA	1460	47			Glutamine synthetase
ArgG	380	117			Argininosuccinate synthetase
ArgD	200	151			Acetylornithine gamma-aminotransferase
ArgI	140	166			Ornithine carbamoyltransferase
GdhA	140	167			NADP-specific glutamate dehydrogenase
2. Aspartate family (28)					
MetE	24420	2			Tetrahydropteroyltriglumate methyltransferase
Asd	900	72			Aspartate-semialdehyde dehydrogenase
DapA	840	76			Dihydrodipicolinate synthase
ThrC	560	95			Threonine synthase
AspC	460	106	(40)	101	Aspartate aminotransferase
DapD	420	110			Succinyldiaminopimelate transaminase
3. Serine family (9)					
GlyA	3480	19			Serine hydroxymethyltransferase
CysK	3060	22			O-acetylserine sulfhydrolase
SerC	2500	35			3-Phosphoserine aminotransferase
SerA	380	118			D-3-Phosphoglycerate dehydrogenase
4. Aromatic family (24)					
AroG	600	91			DAHP synthetase
TrpB	360	123			Tryptophan synthase, B protein
AroK	340	124			Shikimate kinase I
TrpA	280	139			Tryphtophan synthase, A protein
5. Histidine family (9)					
HisD	220	148			Histidine biosynthesis enzyme
6. Pyruvate family (2)					
7. Branched-chain family (24)					
IlvC	2060	38			Keto-acid reductoisomerase
LeuB	400	115			3-Isopropylmalate dehydrogenase
IlvI	360	121			Acetolactate synthase III
LeuC	280	137			alpha-Isopropylmalate synthase
LeuA	160	162			alpha-Isopropylmalate isomerase subunit
e. Polyamine biosynthesis (7)					
f. Purine, pyrimidine, nucleotides					
1. Purine biosynthesis (21)					
PurC	1680	42			SAICAR synthetase
Adk	1640	44	(60)	92	Adenylate kinase
PurA	940	69			Adenylosuccinate synthetase
PurH	640	89			AICAR formyltransferase
GuaB	600	92			IMP dehydrogenase
PurM	180	157			Phosphoribosylaminoimidazole synthetase
PrsA	160	164			Phosphoribosylpyrophosphate synthetase
GuaC	160	161			GMP synthetase
2. Pyrimidine biosynthesis (9)					
PyrB	2080	37			Aspartate carbamoyltransferase, catalytic subunit
PyrI	1200	58			Aspartate carbamoyltransferase
CarA	620	90			Carbamoyl-phosphate synthetase, glutamine subunit
PyrC	340	128			Dihydro-orotase
3. 2'-Deoxyribonucleotide metab. (10)					
4. Salvage of nucleotides, miscellaneous (27)					
Upp	2580	32			Uracil phosphoribosyltransferase
PyrG	480	103			CTP synthetase
CpdB			(100)	82	2':3'-Cyclic-nucleotide 2'-phosphodiesterase

Table 4. continued

Functional classification[a]	Growth, minimal		Stationary, rich		Functional name
	N-abd[b]	Rank[c]	N-abd	Rank	
g. Biosynthesis of small molecules (98)					
PanB	1080	64			Ketopentoate hydroxymethyltransferase
FolE	420	111			GTP cyclohydrolase
HemX			(40)	105	Uroporphyrinogen III methylase
h. Fatty acid biosynthesis (26)					
FabD	500	100			Malonyl-CoA-transacylase
FabI	440	108			Malonyl CoA-acyl carrier protein transacylase
AccC	180	156			3-Ketoacyl-acyl carrier protein reductase
AccA	180	155			Acetyl-CoA carboxylase, alpha subunit
i. Broad regulatory functions (51)					
SspA	480	105			Stringent starvation protein
ArcA	200	150			Negative regulator of genes in aerobic pathway
NadE	100	175			Nitrogen reulatory-like complementation protein
II. Macromolecules					
a. Synthesis of macromolecules					
1. Ribosomal proteins (57)					
RpsB	8100	9			30S ribosomal protein S2
RplD	4960	13	680	41	50S ribosomal protein L4
RplI	3400	21	1340	34	50S ribosomal protein L9
RpsF	2620	30			30S ribosomal protein S6
RplM	2580	31			50S ribosomal protein L13
RpsA	1360	50	(100)	86	30S ribosomal protein S1
RpsP	1220	56			30S ribosomal protein S16
RplC	180	158	8500	2	50S ribosomal protein L3
RplA			8480	3	50S ribosomal protein L1
RplF			5680	10	50S ribosomal protein L6
RplY			2700	20	50S ribosomal protein L25
RpsJ			1360	32	30S ribosomal protein S10
RplL			540	49	50S ribosomal protein L7/L12
2. Ribosome maturation (15)					
3. Aminoacyl tRNA synthetases (49)					
GlnS	520	99			Glutamine tRNA synthetase
GlyS	400	114			Glycine tRNA synthetase, beta chain
ValS	120	174			Valine tRNA synthetase
AsnS	80	179			Asparagine tRNA synthetase
4. Nucleoproteins (7)					
HupA	1380	49			DNA-binding protein HU-alpha
HupB	860	74	2040	23	DNA-binding protein HU-1
Hns	820	79	2800	19	Histone-like protein HLP-II
5. DNA replication, modification, recombination (91)					
6. RNA synthesis, transcription, modification (28)					
Pnp	1000	66			Polynucleotide phosphorylase
RpoA	850	75	580	48	RNA polymerase, alpha subunit
ProS	800	82			RNA synthesis factor
NusA	340	127			Transcription termination factor
Rho	300	135			Transcription termination factor
7. Protein translation (27)					
TufA/B[d]	41400	1			Protein chain elongation factor EF-Tu
Tsf	2670	29	1100	36	Protein chain elongation factor EF-Ts
FusA	2520	34	(20)	117	Protein chain elongation factor EF-G
Efp	1600	45			Elongation factor P
Frr	840	77			Ribosome-releasing factor
PpiB	780	83	(60)	97	Peptidyl-prolyl cis-trans isomerase B
FklB	340	126			Peptidyl-prolyl cis-trans isomerase
FkpA			(200)	67	Peptidyl-prolyl cis-trans isomerase
DsbA			(140)	76	Protein disulfide isomerase I
8. Polysaccharides (6)					
9. Phospholipids (12)					
b. Degradation of macromolecules (68)					
Eco			(260)	60	Serine protease inhibitor
c. Cell envelope					
1. Membranes. lipoproteins, porins (31)					
OmpF	21740	3			Outer membrane protein 1a
OmpA	7300	10	(3000)	18	Outer membrane protein 3a
OmpC	6020	11			Outer membrane protein 1b
HlpA	580	94			Lipoprotein 28
Slp[d]			(600)	43	Outer membrane protein

Table 4. continued

Functional classification[a]	Growth, minimal N-abd[b]	Growth, minimal Rank[c]	Stationary, rich N-abd	Stationary, rich Rank	Functional name
2. Surface polysaccharides, etc. (44)					
RfaD	220	149			ADP-l-glycero-D-mannoheptose-6-epimerase
KdsA	200	152			3-Deoxy-D-manno-octulosinic acid 8-phosphate synthase
3. Surface structures (55)					
FliC	4220	15			Flagellin, filament structural protein
Tpx	2480	36			Fimbria associated protein
DsbC			(100)	83	Protein disulfide isomerase II
4. Murein sacculus and peptidoglycan (37)					
MreB	140	168			Rod shape-determining protein
Pal[d]			(600)	42	Peptidoglycan-associated lipoprotein
III. **Processes**					
a. Transport/binding proteins (250)					
LivJ	10140	4			Branched chain amino acids binding protein
RbsB	(3940)	17	5900	8	D-ribose periplasmic binding protein
Crr	3540	18	500	51	Glucose phosphotransferase system enzyme III
HisJ	3400	20	(320)	59	Histidine-binding protein
FliY	1920	40	(80)	89	Extracellular solute-binding protein
GlnH	1440	48	(460)	54	Glutamine-binding protein
PotD	1300	53	(420)	55	Spermidine/putrescine transport protein
Sbp	(1240)	55			Periplasmic sulfate-binding protein
ArtJ	1180	59			Arginine binding protein
Bcp	920	70			Bacterioferritin comigratory protein
ArgT	900	71	(100)	80	Lysine, arginine, ornithine binding protein
MalE	660	86	(980)	38	Periplasmic maltose-binding protein
CysP	(460)	107			Thiosulfate binding protein
PtsI	360	122			Phosphotransferase system enzyme I
ArtI	340	125	(160)	71	Arginine binding protein
LivK	240	145			Branched chain amino acids binding protein
ManX	220	147	(3140)	17	Phoshotransferase system enzyme II
PotF	200	153			Polyamine binding protein
PtsH			6280	7	Phosphohistidinoprotein-hexose phosphotransferase
MglB			4720	13	Galactose-binding protein
XylF			(120)	78	Xylose binding protein transport system
BtuB			(100)	81	Receptor for vitamin B12
AraF			(80)	88	L-Arabinose-binding protein
PstS			(20)	122	Phosphate-binding periplasmic protein
b. Chaperones (7)					
MopA	2540	33	(480)	52	Molecular chaperone groEL
DnaK	1300	52	(240)	63	Molecular chaperone heat shock protein
MopB	980	67	5720	9	Molecular chaperone groES
HtpG	640	88			Heat shock protein C 62.5
c. Cell division (37)					
Tig	1070	65	600	44	Trigger factor
FtsZ	320	132			GTPase involved in cell division
MinD	300	133			Cell division inhibitor
TolC	120	173	(240)	66	Outer membrane protein
SufI			(20)	125	Suppressor of a ftsI mutation
d. Chemotaxis (12)					
e. Protein/peptide secretion (30)					
DppA	(2980)	24	(40)	102	Dipeptide transport protein
OppA	1140	61	(2360)	22	Perimplasmic oligopeptide binding protein
LolA			(60)	95	Translocation of lipoproteins to outer membrane
f. Osmotic adaption (16)					
OsmY	(2960)	26	940	39	Hyperosmotically inducible periplasmic protein
MdoG	280	138			Periplasmic membrane-derived oligosaccharide synthesis
OsmC			3180	16	Osmotically inducible protein
g. Detoxification (10)					
AhpC	8590	7	1580	28	alkyl hydroperoxide reductase, C22 subunit
SodA	2980	25	(20)	123	Superoxide dismutase, manganese
SodB	2020	39	(920)	40	Superoxide dismutase, iron
Tpx			(160)	74	Oxidoreductase
h. Cell killing (6)					
IV. **Other**					
a. Phage-related functions (25)					
b. Colicin-related functions (12)					
c. Plasmid-related functions (7)					

Table 4. continued

Functional classification[a]	Growth, minimal N-abd[b]	Growth, minimal Rank[c]	Stationary, rich N-abd	Stationary, rich Rank	Functional name
d. Drug/analog sensitivity (46)					
NfnB	260	140	(120)	118	Sensitivity to nitrofurantoin
e. Radiation sensitivity (5)					
g. Adaptions/ atypical conditions (11)					
HdeA	(9520)	6	6480	6	*hns* Deletion induced protein
CspC	8300	8			Cold shock-like protein, suppressor of a *mukB* mutation
UspA	1660	43	1580	30	Universal stress protein
SurA	(400)	116	(160)	72	Survival protein precursor
DksA	260	142			Suppressor of *dnaK* deletion
Imp	60	185			Organic solvent tolerance protein
Dps			6720	5	Global regulator during starvation conditions
HdeB			1840	26	*hns* Deletion induced protein
V. Novel proteins					
YjgF	1320	51			
YdfG	1300	54			
YacI	1260	63			
YigW	(600)	93			
YeaD	560	96			
YgaG	440	109			
YaeT	380	120			
YiaE	160	163			
YifE	340	130			
YihK	340	131			
YchF	260	144			
YebL	240	146			
YhfO	180	159			
YbiS	160	165			
YjjK	140	170			
YjbJ			13120	1	
YgaU			4340	14	
YfiA			600	46	
YajG[k]			(600)	45	
YaeT			(330)	58	
YeeQ			(260)	61	
YgiN			260	62	
YggX			200	69	
YbdQ			180	70	
YbiS			(160)	75	
YhbG			(120)	79	
YnaF			(80)	91	
YebL			(60)	99	
YhbN			(60)	100	
YdcG			(40)	111	
YgfZ			(40)	112	
YtfQ			(40)	116	
YceI			(40)	110	
YacK			(40)	109	
YhjJ			(40)	113	
YtfJ			(40)	115	
YjbP			(40)	114	
YeaD			(20)	126	
YphF			(20)	127	
YhjW			(20)	128	
VI. Unidentified proteins					
M33-2	2900	28			
M18-2	820	80			
M162-2	700	85			
M26	480	102			
M167-2	420	112			
M75-2	300	134			
M107-2	260	141			
M146-2	260	143			
M146-1	200	154			
M109-2	140	169			
M124-1	120	172			
M29-2	100	176			
M70	100	177			
M147-2	100	178			
M63-2	80	180			
M77-2	80	181			
M114	80	182			
M131	80	183			
M135	80	184			

Table 4. continued

Functional classification[a)]	Growth, minimal		Stationary, rich		Functional name
	N-abd[b)]	Rank[c)]	N-abd	Rank	
M124-2	60	186			
M133	60	187			
R17-2			1600	27	
R31-2			1580	29	
R32-1			980	37	
R34-2			580	47	
R38			340	57	
R56-2			(240)	64	
R22-2			240	65	
R53			100	85	
R32-2			100	87	
R134-2			(40)	106	
R41-2			40	107	
R42-2			40	108	
R97-2			(20)	119	
R121-2			(20)	120	

a) The numbers in parentheses represent the estimated number of genes predicted to be involved in the biological process [50].
b) For genes with protein isoforms, the isoform abundances have been combined. Proteins identified only from fractionated extracts are enclosed with parentheses to indicate that the values may not reflect the total abundance in the cell.
c) The descending order is based on abundance of the proteins analyzed in the given growth and environmental condition.
d) The proteins TubA/B, AtpF, Pal, Slp, and YajG were blocked to *N*-terminal sequence analysis. There *N*-abd was estimated by directly correlating the Coomassie-intensity of the blocked spots to the intensity and *N*-abd of sequenceable spots with similar op*I* and M_r.
Abbreviations: DAHP, phospho-2-dehydro-3-deoxyheptonate aldolase; SAICAR, phosphoribosylaminoimidazole-succinocarboxamide; AICAR, phosphoribosylaminoimidazolecarboxamide

Table 5. Distribution of *E. coli* proteins among functional categories

Functional classification	% Abundance in the proteome[a)]
I. Small molecule metabolism	
a. Degradation + Energy metabolism	12
c. Central intermediatory metabolism	2
c. Other (biosynthensis, regulation, *etc.*)	19
II. Macromolecules	
a. Synthesis of macromolecules	26
c. Cell envelope	14
c. Degradation	—[b)]
III. Processes	
a. Transport/binding proteins	8
b. Chaperones	2
c. Detoxification	4
d. Other (cell division, osmotic adaption, *etc.*)	2
IV. Miscellaneous	6
V. Novel proteins	2
VI. Unidentified proteins	3

a) Based on *E. coli* growing in glucose-minimal media using the cellular abundances of proteins in Table 4.
b) No proteins were identified in this category.

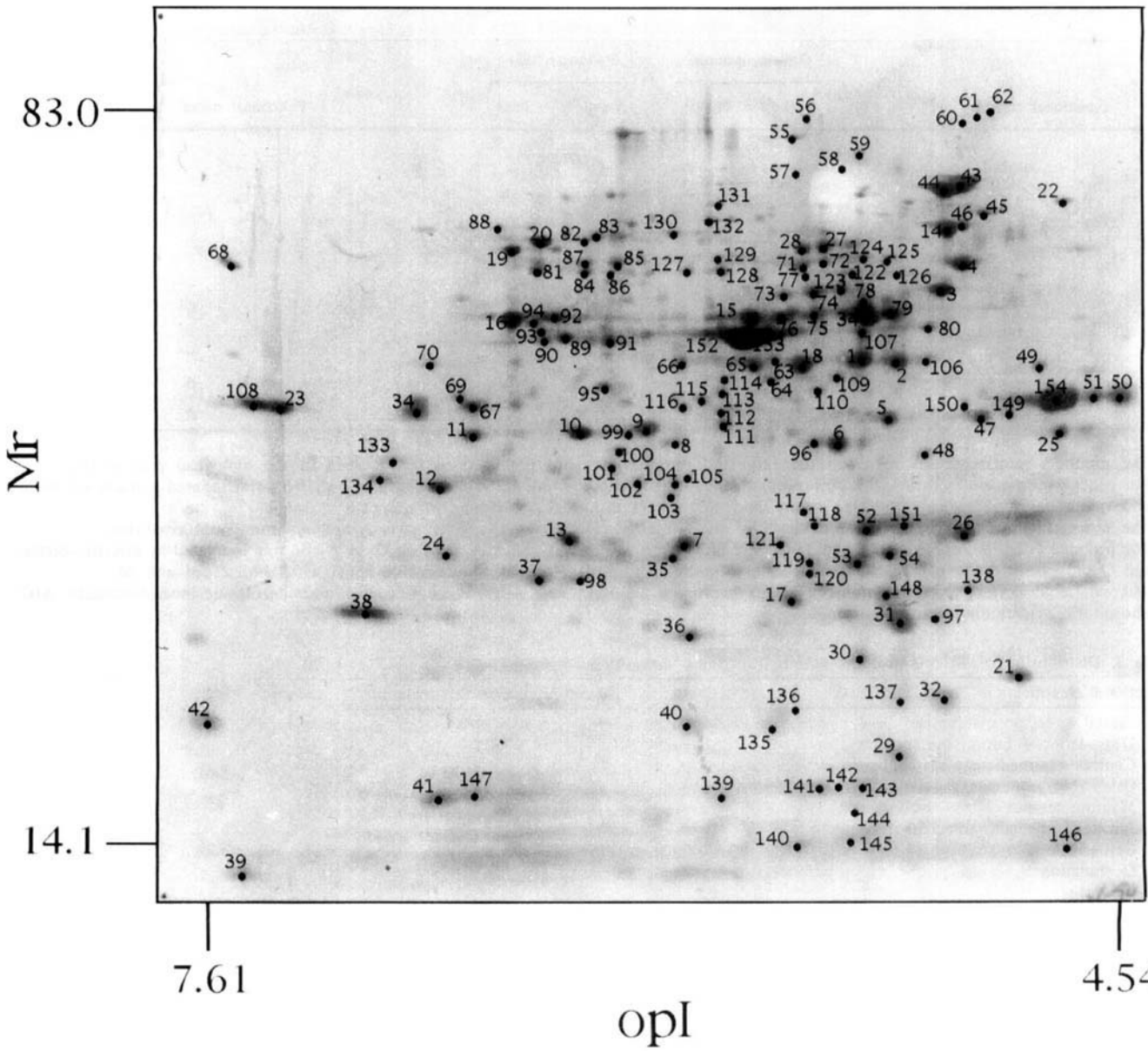

Figure 6. Spots M1–M154: Master gel of total-protein extract fractionated by 2-DE, electrotransferred to PVDF, and stained with Coomassie Blue. The protein extract was prepared using the heat/SDS method from EMG2 growing in glucose-minimal media. The IEF used a 4:1 carrier ampholyte blend of Serva 5–7:Serva 3–10, and the SDS-PAGE used a 12.5%T, 2.7%C gel. The Laemmli anode buffer in the second dimension was supplemented with 50 mM sodium acetate. The largest and smallest op*I* and M_r of the identified 2-DE spots from the gel are shown on the op*I* and M_r axes. The data for spots M1–M154 are found in Table 6.

Table 6. Growth-phase, minimal-media *E. coli* proteins from an SDS/heat extract

		Predicted *versus* observed sequence, p*I*, MW						Abundance	
Spot ID	*N*-terminal sequence tag[a]	Sequence tag match to genomic ORF[b]	Locus[c]	*cpI*[d]	*opI*[e]	MW[f]	M_r[g]	O-abd[h]	N-abd[i]
M100	SLNFLDFEQPIA	M*SLFLDFEQPIA	*accA*	5.88	6.04	35.1	32.5	0.14	180
M68-2	(Mas) (NL) (LD) (TK) (EI) (LV) (KI) (NA) (TN) (PR) (GV) (Esil)	MLDKIVIANRGE	*accC*	7.09	7.52	49.3	50.8	0.05	180
M7	MRIILLGAPGAG	MRIILLGAPGAG	*adk*	5.53	5.81	23.6	26.4	0.56	1640
M97-2	(Sm) (LA) (IL) (NE) (VT) (PK) (AIL) (GK) (KP) (DF) (LKd) (PNl)	M*SLINLTKIKPFKN	*ahpC*	4.90	5.04	20.6	21.9	0.14	300
M2-1	(MSD) (QDL) (IG) (NSK) (VT) (ARKT)	M*SLINTK	*ahpC*	4.90	5.16	20.6	39.7	0.32	2250
M31-1	SL (IL)N (TV) (PK) (IA) (GK) (KP) (DF) (LK) (NP)	M*SLINTKIKPFKN	*ahpC*	4.90	5.15	20.6	2.16	0.68	6040
M125	(MS) VPVQ (Kg) P (Mr) YX (Dvd)	M*SVPVQHPMYIDG	*aldA*	4.91	5.19	52.1	52.3	0.08	280
M118-1	(AM) (QP) (TX) (PD) (ND) (IT) (LX) IY) (VY) (EG) DE	MQTPHILIVEDE	*arcA*	5.08	5.40	27.3	27.7	0.10	200
M91-2	(AT) (IT) (EL) (LQ) (NTV) (AP) (YI) (FS) (RG) (AE) (FT) (GF)	M*AIEQTAITRATF	*argD*	6.20	6.07	43.6	41.5	0.08	200
M74	TTILKHLPVGqR	M*TTILKHLPVGQR	*argG*	5.14	5.41	49.8	47.4	0.17	380
M113	(SAQ)GFYXKXFLKgl	M*SGFYHKHFLKL	*argI*	5.51	5.70	36.8	36.9	0.09	140
M53-2	(DA) (EL) (GP) (LE) (LT) (NV) (KR) (VI) (KG) (EX) (RD) (GX)	MKKSILALSLLVGLSTAASSAY*ALPETV RIGTDT	*argT*	5.05	5.27	25.8	25.3	0.21	900
M67-1	(Mast) (NI) (YK) (QV) (NG) (DI) (DX) (Lg) (Rf) (IX) (KX) (Ea)	MNYQNDDLRIKE	*aroG*	6.58	6.55	38.0	35.7	0.22	600
M136-1	(As) (Ea) (Ke) (Rk) (Nr) (In) (Fi) (Lf) (Vl) (Gv) (Pg) (Mp)	MRFQFMSCRASLSEAGLSLTNSLSSTEK M*AEKRNIFLVGPM	*aroK*	5.11	5.51	19.4	16.6	0.16	140
M136-2	(As) (Ea) (Ke) (Rk) (Nr) (In) (Fi) (Lf) (Vl) (Gv) UPg) (Mp)	MRFQFMSCRASLSEAGLSLTNSLSSTEK M*AEKRNIFLVGP	*aroK*	5.11	5.51	19.4	16.6	0.22	200
M120	AETIRFATEA (GS) Y	MKKVLIAALIAGFSLSATA*AETIRFA TEASY	*artI*	5.20	5.42	25.0	24.6	0.30	340
M37	AEKINFGVSATY	MKKLVCLAALLASFTFGASA*A EKINFGVSATY	*artJ*	6.29	6.32	24.9	24.1	0.48	1180
M64	(Mq) KNVGFIGGRGM	MKNVGFIGWRGM	*asd*	5.32	5.54	40.0	38.0	0.14	900
M122	A (GVL) (VI) PVAD (PV) (LV) QXR	M*SVVPVADVLQ	*asnS*	5.04	5.33	52.4	47.8	0.11	80
M66-2	(SM) (KFq) (IE) (FN) (DI) (FT) (VGa) (KA) (Pi) (GA) (VD) (IPSy)	MFENITAAPADP	*aspC*	5.58	5.83	43.6	39.4	0.08	460
M87	MQLNSTEISELI	MQLNSTEISELI	*atpA*	6.01	6.15	55.2	51.6	0.12	820
M3	ATGKIT	M*ATGKIV	*atpD*	4.74	5.03	50.2	47.8	0.32	1210
M29-1	(Msq) (NDM) (Prt) (Lk) (KI) (AD) (GV) (Dks) (IX) (AL) (Pg) (Kr)	MNPLKAGDIAPK	*bcp*	4.93	5.15	17.6	15.0	0.54	920
M89	MIKSALLVLEDG	MIKSALLVLEDG	*carA*	6.35	6.22	41.4	41.8	0.18	620
M21	GLFDKLKSLVSD	M*GLFDKLKSLVSD	*crr*	4.56	4.81	18.1	18.5	0.66	3540
M39	AKIKGQVKXFNE	M*AKIKGQVKWFNE	*cspC*	7.53	7.45	7.3	16.0	0.61	8300
M10-1	(ST) (KI) (IA) (FX) (EV) (Di) (Ng) (ST) (LX) (Tg) (IX) (Ga)	M*SKIFEDNSLTIG	*cysK*	5.94	6.17	34.4	33.8	0.44	2820
M99	SKIFEDNSLTIi	M*SKIFEDNSLTI	*cysK*	5.94	6.01	34.4	33.7	0.12	240
M133	sEL (Lr) NXsID (lv) X (lv)	MAVNLLKKNSLALVASLLLAGHVQA* TELLNSSYDVSR	*cysP*	6.99	6.85	35.1	31.7	0.18	60
M104	MQQLQNIIETAF	MQQLQNIIETAF	*dapD*	5.53	5.85	29.9	30.3	0.13	420
M137	(Sm) Q (Ei) N (QT) NRK (PS) FXL	MQEGQNRKTSSL	*dksA*	4.89	5.15	17.5	17.2	0.44	260
M43	GKIIGIDLGTTN	M*GKIIGHIDLGTTN	*dnaK*	4.67	4.97	69.0	68.3	0.37	1300
M85	(Kms) (Ts) LVYXSEGsP	MRISLKKSGMLKLGLSLVAMTVAASVQ A*KTLVYCSEGSP	*dppA*	5.94	6.05	57.4	51.4	0.13	660
M148	ATYYSNDFRAGL	M*ATYYSNDFRAGL	*efp*	4.73	5.19	20.5	23.2	0.47	1600
M75-1	(Sma) (KNE) (IN) V (Vk) (Vi) (il) (fg) (ta) (eaqgd) ii	M*SKIVKIIGREII	*eno*	5.23	5.40	45.5	44.6	0.11	200
M15	SKIVKIIGR	M*SKIVKIIGR	*eno*	5.23	5.60	45.5	44.4	0.35	2160
M76-1	(Sgam) (KN) (IN) (Vtge) (KX) (IX) (Il) (Gd) (Rtd) (Eq) (IX) (Ivl)	M*SKIVKIIGREII	*eno*	5.23	5.51	45.5	44.4	0.14	1660
M48	TQFAFVFPGQ	M*TQFAFVFPGQ	*fabD*	4.79	5.08	32.3	32.4	0.19	500
M102-1	(GM) (FG) (LF) (SL) (GS) (KG) (RK) (IR) (LI) (VL) (TL) (GT)	MGFLSGKRILVT	*fabI*	5.72	5.97	27.9	30.3	0.08	200
M102-2	(GM) (FG) (LF) (SL) (GS) (KG) (RK) (IR) (LI) (VL) (TL) (GT)	M*GFLSGKRILVTG	*fabI*	5.72	5.97	27.7	30.3	0.09	240

Table 6. continued

Spot ID	*N*-terminal sequence tag[a]	Predicted *versus* observed sequence, p*I*, MW						Abundance	
		Sequence tag match to genomic ORF[b]	Locus[c]	cp*I*[d]	op*I*[e]	MW[f]	M_r[g]	O-abd[h]	N-abd[i]
M66-1	(SM) (KFq) (IE) (FN) (DI) (FT) (VGa) (KA) (Pi) (GA) (VD) (IPSy)	M*SKIFDFVKPGVI	*fba*	5.75	5.83	39.0	39.4	0.14	1480
M138	TTPTFDT (Ig) EAQ (ALgr)	M*TTPTFDTIEAQA	*fklB*	4.68	4.95	22.1	23.7	0.25	340
M53-1	(DA) (EL) (GP) (LE) (LT) (NV) (KR) (VI) (KG) (EX) (RD) (GX)	MKLAHLGRQALMGVMAVAL VAGMSVKSFA*DEGLLNKVKERG	*fliY*	5.14	5.27	26.1	25.3	0.45	1920
M38-2	(SM) (YI) (TS) (LD) (PI) (sr) (LK) (PD) (Ya) (AE) (YV) (DR)	MISDIRKDAEVR	*frr*	6.86	6.96	20.6	21.9	0.40	840
M49	(Mq) FEPMELTNDAV	MFEPMELTNDAV	*ftsZ*	4.47	4.76	40.3	39.3	0.12	320
M88	(Si) NKPFHYQAPF	M*SNKPFHYQAPF	*fumA*	6.52	6.47	60.2	57.5	0.03	160
M55	ARTTPIARYRNI	M*ARTTPIARYRNI	*fusA*	5.13	5.47	77.4	81.4	0.06	1080
M101-1	(At) (Ni) (LN) (KL) (AK) (VA) (IV) (PI) (VP) (AV) (GA) (LG)	MAAINT*KVKKAVIPVAGL	*galU*	5.00	6.06	32.3	31.4	0.08	120
M101-2	(At) (Ni) (LN) (KL) (AK) (VA) (IV) (PI) (VP) (AV) (GA) (LG)	MAAIN*TKVKKAVIPVAG	*galU*	5.00	6.06	32.4	31.4	0.09	140
M109-1	(Akst) (AI) I (NE) (TG) K (VL) K (Ik) AV (IN)	M*AAINTKVKKAVI	*galU*	5.00	5.34	32.8	38.3	0.07	260
M67-2	(Mast) (NI) (YK) (QV) (NG) (DI) (DX) (Lg) (Rf) (IX) (KX) (Ea)	M*TIKVGINGFGRI	*gapA*	7.11	6.55	35.4	35.7	0.08	220
M23	TIKVGIDGFGRI	M*TIKVGINGFGRI	*gapA*	7.11	7.31	35.4	35.4	0.60	1840
M34	TIKVGINGFG	M*TIKVGINGFG	*gapA*	7.11	6.77	35.4	35.5	0.64	740
M108	TIKVGINGFGRI	M*TIKVGINGFGRI	*gapA*	7.11	7.42	35.4	35.7	0.35	1660
M114	(AQ) Xk (Td) PLYEVXt (tl)	M*AQQTPLYEQHTL	*gcvT*	5.35	5.69	40.1	38.1	0.09	80
M94	MDQTYsL (Ei) XFLN	MDQTYSLESFLN	*gdhA*	6.37	6.33	48.6	43.4	0.09	140
M27	(Sk) (AG) EHVLTMLN	M*SAEHVLTMLN	*glnA*	5.23	5.38	51.8	54.5	0.15	420
M28	SAEHVLTMLNE	M*SAEHVLTMLNE	*glnA*	5.23	5.44	51.8	54.1	0.14	1040
M127	SQNVYQFIDLQR	M*SQNVYQFIDLQR	*gltD*	5.52	5.81	51.9	50.3	0.08	620
M128-2	(Sq) (QS) (VN) (VD) (YI) (LGq) (VF) (IP) DL (QP) (RE)	M*SQNVYQFIDLQR	*gltD*	5.52	5.70	51.9	50.3	0.03	180
M16	(Mqh) LKRE (Mh) NIADYD	MLKREMNIADYD	*glyA*	6.46	6.41	45.3	43.8	0.45	3180
M92-2	(AM) (KL) (VX) (RS) (Le) (ME) (NK) (DI) (AK) (Id) (KY) (FX)	MLKREMNIADY	*glyA*	6.46	6.26	45.3	44.1	0.11	300
M57	(SM) EKTFLVEIGTe	M*SZKTFLVEIGTE	*glyS*	5.18	5.47	76.7	70.8	0.03	80
M13	AVTKLVLVRHGE	M*AVTKLVLVRHGE	*gpmA*	6.11	6.21	28.4	26.7	0.45	960
M19	(Mq)LRIAKEALTFD	MLRIAKEALTFD	*guaB*	6.38	6.41	52.0	53.5	0.16	600
M69	MRIEEDLKLg(Fv)l	MRIEEDLKLGFK	*guaC*	6.51	6.61	37.4	36.4	0.10	160
M126	SFNTII (Dg) PnX (PYEK) t	M*SFNTIIDWNSCT	*hisD*	5.05	5.15	46.0	49.9	0.06	220
M52-1	(AX) (IP) (PK) (QD) (NX) (IT) (RX) (IY) (GX) (TX) (DA) (PK)	MKKLVLSLSLVLAFSSATAALA*AIPQNI RIGTDP	*hisJ*	5.01	5.24	26.2	27.5	0.35	3400
M139	SEALKILNNIRT	M*SEALKILNNIRT	*hns*	5.25	5.70	15.4	14.1	0.74	820
M33-1	(MAQHS) (EK) (SHQ) (KLQ) (VIF) (VGT) (VS) (PRE) (AS) (QHM) (GAS) (KV)	MESKVVVPAQGK	*icdA*	5.00	5.25	45.8	44.8	0.56	8180
M78	MEXKVVVPAQGK	MESKVVVPAQGK	*icdA*	5.00	5.19	45.8	46.5	0.08	240
M79	MESKVVVPAQGK	MESKVVVPAQGK	*icdA*	5.00	5.17	45.8	44.8	0.32	1480
M77-1	(AS) (NE) (LY) (FS) (VQ) (TG) (LV) (NPG) (LQ) (rX) (IX) (rX)	M*ANYFNTLNLRQQ	*ilvC*	5.07	5.44	53.9	49.9	0.04	160
M71	ANYFNTLNLRQQ	M*ANYFNTLNLRQQ	*ilvC*	5.07	5.45	53.9	51.2	0.14	1420
M72	ANYFNTLNLRQQ	M*ANYFNTLNLRQQ	*ilvC*	5.07	5.38	53.9	51.7	0.08	480
M62	ADLAS	MKKRIPTLLATMIATALYSQQ GLA*ADLAS	*imp*	4.70	4.92	87.1	83.0	0.01	60
M134	(AM) (GM)QK(VG)VSI(gp)D(di)N	MKQKVVSIGDIN	*kdsA*	6.78	6.90	30.8	30.4	0.24	200
M129	SQQ(VLR) (GI)I(FGE)D(TG)TI(Er)	M*SQQVIIFDTTLR	*leuA*	5.56	5.71	57.1	52.5	0.05	160
M107-1	(SM) (KE) (NI) (YK) (VH) (IT) (AD) (VP) (LA) (PQ) (GX) (DK)	M*SKNYHIAVLPGD	*leuB*	5.03	5.26	39.4	52.4	0.12	400
M81	AKTLYEKLFDA(AH)	M*AKTLYEKLFDAH	*leuC*	6.28	6.32	49.8	50.3	0.12	280
M18-1	(ETMA) (DQS) (IX) (KF) (VNH) (AF) (VS) (VX) (GX) (AP) (Mh)SG	MNTKGKALLAGLIALAFSNMALA*E DIKVAVVGAMSG	*livJ*	5.17	5.44	36.8	39.4	0.51	10160
M106	(MSD)D(Ig)KV(At)V(Vf)GAMs	MKRNAKTIIAGMIALAISHTAMA*DDIK VAVVGAMS	*livK*	4.94	5.07	37.0	39.8	0.13	240

Table 6. continued

Spot ID	*N*-terminal sequence tag[a]	Predicted *versus* observed sequence, p*I*, MW						Abundance	
		Sequence tag match to genomic ORF[b]	Locus[c]	cp*I*[d]	op*I*[e]	MW[f]	M_r[g]	O-abd[h]	N-abd[i]
M84	STEIKTQVVVL(IG)	MM*STEIKTQVVVLG	*lpdA*	6.10	6.15	50.6	49.9	0.07	380
M86	STEIKTQCCCIg	MM*STEIKTQVVVLG	*lpdA*	6.10	6.08	50.6	50.1	0.04	160
M10-2	(ST) (KI) (IA) (FX) (EV) (Di) (Ng) (ST) (LX) (Tg) (IX) (Ga)	M*TIAIVIGTHGWA	*manX*	5.87	6.17	34.9	33.8	0.03	220
M9	(Mq)KVAVLGAAGGI	MKVAVLGAAGG	*mdh*	5.56	5.94	32.3	34.1	0.42	2640
M111	(Mra)KVAVLGAAGGI	MKVAVLGAAGGI	*mdh*	5.56	5.69	32.3	34.4	0.20	260
M117	ARIIVVTSGKgg	M*ARIIVVTSGKGG	*minD*	5.08	5.44	29.5	28.6	0.11	300
M14	AAKDVKFGNDAR	M*AAKDVKFGNDAR	*mopA*	4.67	5.02	57.2	57.7	0.33	2180
M46	AAKDVKFGNDAR	M*AAKDVKFGNDA	*mopA*	4.67	4.97	57.2	58.7	0.09	360
M144	MNIRPLHDRVIV	MNIRPLHDRVIV	*mopB*	4.99	5.28	10.4	14.0	0.26	980
M110-1	(MX) (LM) (KL) (XK) (FX) (RF) (GR) (MG) (FM) (SF) (LS) (DL)	MLKKFRGMFSND	*mreB*	5.03	5.39	37.0	37.1	0.04	60
M110-2	(MX) (LM) (KL) (XK) (FX) (RF) (GR) (MG) (FM) (SF) (LS) (DL)	MLKKFRGMFSN	*mreB*	5.03	5.39	37.0	37.1	0.04	80
M112	MTLQqqIIKALg	MTLQQQIIKALG	*nadE*	5.33	5.70	30.6	35.5	0.11	100
M98	MD(It)ISVALKR(hp)s	MDIISVALKRHS	*nfnB*	6.17	6.17	23.9	24.1	0.22	260
M22	MNXEILAVVEXV	MNKEILAVVEAV	*nusA*	4.35	4.70	54.9	63.7	0.08	340
M118-2	(AM) (QP) (TX) (PD) (ND) (IT) (LX) (IY) (VY) (EG)DE	MKKTAIAIAVALAGFATVAQA*APKDNT WYTGAK	*ompA*	5.74	5.40	35.2	27.7	0.07	140
M8	APKDNTXYTGAK	MKKTAIAIAVALAGFATVAQA*APKDNT WYTGAK	*ompA*	5.74	5.85	35.2	33.1	0.31	2300
M52-2	(AX) (IP) (PK) (QD) (NX) (IT) (RX) (IY) (GX) (TX) (DA) (PK)	MKKTAIAIAVALAGFATVAQA*APKDNT WYTGAK	*ompA*	5.74	5.24	35.2	27.5	0.34	3380
M151	APKDNTRYTGAK	MKKTAIAIAVALAGFATVAQA*APKDNT WYTGAK	*ompA*	5.74	5.13	35.2	27.6	0.48	1480
M50	AE(Vi)YNKDGN	MKVKVLSLLVPALLVAGAANA*A EVYNKDGN	*ompC*	4.33	4.54	38.3	36.7	0.50	6020
M25	AVGLHYFSK	(36)K*AVGLHYFSK(313)	*ompF*	4.43	4.70	35.3	34.0	0.24	1980
M51	AEI(Yi)NKDGNKVD	MMKRNILAVIVPALLVAGTANA*AEI YNKDGNKVD	*ompF*	4.48	4.61	37.1	36.7	0.25	2160
M154	AEIYNKDGNKVD	MMKRNILAVIVPALLVAGTANA*AEI YNKDGNKVD	*ompF*	4.48	4.71	37.1	36.4	1.35	17600
M20	ADVPAGVTLAEK	MTNITKRSLVAAGVLAALMAGNVA LA*ADVPAGVTLAEK	*oppA*	6.17	6.31	58.4	55.1	0.21	760
M82	ADVPAGVtLAEK	MTNITKRSLVAAGVLAALMAGNVA LA*ADVPAGVTLAEK	*oppA*	6.17	6.16	58.4	55.1	0.05	380
M1	SVIKMT	M*SVIKMT	*pgk*	4.93	5.27	41.0	39.8	0.56	2990
M59	MLNPIVRKFQYG	MLNPIVRKFQYG	*pnp*	4.96	5.28	77.1	76.2	0.03	200
M47	DDNNTLYFYNXT	MKKWSRHLLAAGALALGMSAA HA*DDNNTLYFYNWT	*potD*	4.69	4.92	36.5	35.1	0.29	1040
M149	(AS) (Drh)NNTLY(Fi)YNX(It)	MKKWSRHLLAAGALALGMSAA HA*DDNNTLYFYNWT	*potD*	4.69	4.84	36.5	35.3	0.11	260
M116	(Ap) ENK (TD) LXIYNv	MTALNKKWLSGLVAGALMAVSVGTLA* AEOKTLHIYNW	*potF*	5.35	5.82	38.2	35.7	0.10	200
M31-2	SL(IL)N(TV) (PK) (IA) (GK) (KP) (DF) (LK) (NP)	M*SLLNVPAGKDLP	*ppa*	4.90	5.15	19.6	21.6	0.62	5480
M97-1	(Sm) (LA) (IL) (NE) (VT) (PK) (AIL) (GK) (KP) (DF) (LKd) (PNl)	M*SLLNVPAGKDLP	*ppa*	4.90	5.04	19.6	21.9	0.24	500
M40	MVTFHTNHGDIV	MVTFHTNHGDIV	*ppiB*	5.64	5.80	18.2	16.1	0.53	780
M135	MVTFsT (NI) pXXI	MVTFHTNHGDI	*ppiB*	5.64	5.53	18.2	16.0	0.09	80
M96	(PA)PDM(KL)LFAGNAT	MPDMKLFAGNAT	*prsA*	5.13	5.40	34.2	33.1	0.15	160
M45	MISGilAXPGI	MISGILASPGI	*ptsI*	4.60	4.91	63.6	61.1	0.08	360
M76-2	(Sgam) (KN) (IN) (Vtge) (KX) (IX) (Il) (Gd) (Rtd) (Eq) (IX) (Ivl)	M*GNNVVVLGTOWG	*purA*	5.22	5.51	47.2	44.4	0.08	940
M54	MQKQAELYRG(Kr)A	MQKQAELYRGKA	*purC*	4.88	5.18	27.0	25.8	0.39	1680
M130	M(QG)Q(Gd)RPVL(dr)ALL	MQQRRPVRRALL	*purH*	5.60	5.86	57.3	56.6	0.06	160
M83	MKKTKIVaTIg(PG)	MKKTKIVCTIGP	*pykF*	5.98	6.12	50.7	56.0	0.07	500
M11	ANPLYQKHIISI	M*ANPLYQKHIIS	*pyrB*	6.75	6.56	34.3	33.5	0.34	2080

Table 6. continued

Spot ID	*N*-terminal sequence tag[a]	Predicted *versus* observed sequence, p*I*, MW						Abundance	
		Sequence tag match to genomic ORF[b]	Locus[c]	cp*I*[d]	op*I*[e]	MW[f]	M_r[g]	O-abd[h]	N-abd[i]
M95	TAPSQVLKIRRP	**M*TAPSQVLKIRRP**	*pyrC*	6.15	6.09	38.7	37.4	0.11	340
M42-1	(MT) (TH) (HD) (DN) (Nk) (KL) (LQ) (QV) (VE) (EA) (AI) (Ik)	**MTHDNKLQVEAI**	*pyrI*	7.37	7.61	17.1	16.0	0.55	760
M42-2	(MT) (TH) (HD) (DN) (Nk) (KL) (LQ) (QV) (VE) (EA) (AI) (Ik)	**M*THDNKLQVEAIK**	*pyrI*	7.37	7.61	17.0	16.0	0.32	440
M150	(Ms)IIVTGGAGFIG	**MIIVTGGAGFIG**	*rfaD*	4.64	4.97	34.9	36.2	0.09	220
M68-1	(Mas) (NL) (LD) (TK) (EI) (LV) (KI) (NA) (TN) (PR) (GV) (Esil)	**MNLTELKNTPVS**	*rho*	7.29	7.52	47.0	50.8	0.08	300
M41	MQVILLDKVANL	**MQVILLDKVANL**	*prlI*	6.56	6.68	15.8	14.0	1.05	3400
M2-2	(MSD) (QDL) (IG) (NSK) (VT) (ARKT)	**MQGSVT**	*rpoA*	4.81	5.16	36.5	39.7	0.12	850
M44-1	(MX) (TE) (ES) (SF) (FA) (AQ) (QL) (LF) (FE)	**MTESFAQLF**	*rpsA*	4.71	5.03	61.2	66.5	0.23	820
M44-2	(MX) (TE) (ES) (SF) (FA) (AQ) (QL) (LF) (FE)	**M*TESFAQLFE**	*rpsA*	4.71	5.03	61.0	66.5	0.15	540
M141	(AMS)RHYEIVFMVXP	**MRHYEIVFMVHP**	*rpsF*	5.17	5.45	15.2	14.2	0.59	560
M142	MRHYEIVFMVHP	**MRHYEIVFMVHP**	*rpsF*	5.17	5.38	15.2	14.2	0.60	1120
M143	MRHYEIVFMYHP	**MRHYEIVFMVHP**	*rpsF*	5.17	5.25	15.2	14.2	0.67	940
M24	(Mqk)RLEFSIYRYN	**MRLEFSIYRYN**	*sdhB*	6.68	6.66	26.8	25.6	0.36	480
M92-1	(AM) (KL) (VX) (RS) (Le) (ME) (NK) (DI) (AK) (Id) (KY) (FX)	**M*AKVSLEKDKIKF**	*serA*	6.33	6.26	44.0	44.1	0.15	380
M65	A(QD)IFNFSSGPAM	**M*AQIFNFSSGPAM**	*serC*	5.33	5.59	39.7	39.3	0.27	2500
M38-1	(SM) (YI) (TS) (LD) (PI) (sr) (LK) (PD) (Ya) (AE) (YV) (DR)	**M*SYTLPSLPYAYD**	*sodA*	6.98	6.96	23.0	21.9	1.45	2980
M36	SFELPALPYAKD	**M*SFELPALPYAKD**	*sodB*	5.88	5.80	21.1	20.7	0.56	2020
M119	AVAANKRSVMTL	**M*AVAANKRSVMTL**	*sspA*	5.10	5.42	24.2	25.3	0.19	480
M128-1	(Sq) (QS) (VN) (VD) (YI) (LGq) (VF) (IP)DL(QP) (RE)	**M*SSVDILVPDLPE**	*sucB*	5.62	5.70	43.9	50.3	0.06	420
M63-1	(Ma) (Nd) (Li) (Hfk) (Ev) (Yf) (QX) (Av)KQLF	**MNLHEYQAKQLF**	*sucC*	5.28	5.53	41.4	39.7	0.08	340
M12	SILIDKNTKVIX	**M*SILIDKNTKVIC**	*sucD*	6.77	6.68	29.6	30.0	0.42	900
M5	TDKLT(AS)	**M*TDKLTS**	*talB*	4.94	5.18	35.1	34.9	0.29	380
M73	MKLYNLKD(HD)NE(QE)	**MKLYNLKDHNEQ**	*thrC*	5.16	5.50	47.1	47.0	0.11	560
M4	MQVSVE	**MQVSVE**	*tig*	4.64	4.98	48.2	51.7	0.33	1070
M123	ENLMQVYQ(Qg)A(Rg)L	MKKLLPILIGLSLSGFSSLSQA***ENLMQ VYQQARL**	*tolC*	5.09	5.29	51.5	50.3	0.08	120
M35	MRXPLVMGNXKL	**MRHPLVMGNWKL**	*tpiA*	5.89	5.85	27.0	25.4	0.23	1120
M32	SQTVHFQGNPV(TS)	**M*SQTVHFQGNPVT**	*tpx*	4.59	5.02	17.8	17.4	0.68	2480
M121	(SAmgd)ERYESLFA(Qg)LK	**MERYESLFAQLK**	*trpA*	5.23	5.51	28.7	26.4	0.15	280
M91-1	(AT) (IT) (EL) (LQ) (NTV) (AP) (YI) (FS) (RG) (AE) (FT) (GF)	**M*TTLLNPYFGEFG**	*trpB*	6.04	6.07	42.9	41.5	0.15	360
M6	AEITAX	**M*AEITAS**	*tsf*	5.07	5.33	30.3	33.0	0.49	2670
M152	XXXXXXXXXXXX		*tufA/B*	5.30	5.68	43.2	42.0	1.83	
M153	XXXXXXXXXXXX		*tufA/B*	5.30	5.62	43.2	41.8	1.84	
M17	(Mqth) (Ka) (It)VEVKHPLVK	**MKIVEVKHPLVK**	*upp*	5.23	5.47	22.5	22.9	0.49	2580
M145	AYKHILIAVDLS	**M*AYKHILIAVDLS**	*uspA*	5.02	5.22	15.9	14.1	0.75	1660
M56	(Ma) (Lr)EEYRK(Hr)VA(EG)	**VLEEYRKEVAE**	*yacI*	5.14	5.43	93.5	88.4	0.02	120
M60	AEGFVVKDIXFE	MAMKKLLIASLLFSSATVYG***AEGFVVK DIHFE**	*yaeT*	4.72	4.98	88.4	8.18	0.01	180
M61	AEGFVVKDIHF(Ev)	MAMKKLLIASLLFSSATVYG***AEGFVVK DIHFE**	*yaeT*	4.72	4.96	88.4	82.6	0.01	80
M103	(VA)TYPLP(Te)DGsrL	**MNMKLKTLFAAAFAVVGFCSTASA*V TYPLPTDGSRL**	*ybiS*	5.74	5.86	30.9	29.4	0.13	160
M80-1	(ASX) (FX) (KF) (KX) (gX) (Ig) (VI) (GV) (LG) (PL) (NP) (VN)	**M*GFKCGIVGLPNV**	*ychF*	4.70	5.07	39.5	43.2	0.05	120
M80-2	(ASX) (FX) (KF) (KX) (gX) (Ig) (VI) (GV) (LG) (PL) (NP) (VN)	**MGFKCGIVGLPN**	*ychF*	4.70	5.07	39.7	43.2	0.06	140
M105	(As)VVASLKPVGFI	MKCYNITLLIFITIIGRIMLHKKTLL FAALSAALWGGATQAADA***AV VASLKPVGFI**	*yebL*	5.58	5.80	31.2	30.8	0.08	240
M30-1	(AP) (DPL) (YL) (DL) (DSI) (dfs) (kft) (etv) (vd)	**M*PLLDSFTVD**	*ygaG*	5.13	5.26	19.3	19.5	0.26	260

Table 6. continued

Spot ID	*N*-terminal sequence tag[a]	Predicted *versus* observed sequence, p*I*, MW						Abundance	
		Sequence tag match to genomic ORF[b]	Locus[c]	cp*I*[d]	op*I*[e]	MW[f]	M_r[g]	O-abd[h]	N-abd[i]
M30-2	(AP) (DPL) (YL) (DL) (DSI) (dfs) (kft) (etv) (vd)	MPLLDSFTV	*ygaG*	5.13	5.26	19.4	19.5	0.18	180
M90	(MS) (Kg)LPIYLDYsAT	MMYGVYRA*MKLPIYLDYSAT	*yhfO*	6.34	6.30	45.1	41.5	0.09	180
M115	M (NI) PS (IV) I (NVD) (LY) taip	MERS*MKPSVILYKALP	*yiaE*	6.86	5.76	40.1	36.4	0.13	160
M147-1	(AM) (ET) (SVLN) (FNq) (TQ) (TLV) (TC) (NL) (RK) (PRY) (FRE)D	M*AESFTTTNRYFD	*yifE*	10.85	6.54	16.2	14.1	0.52	340
M58	x(Qi) (Ev)s(Lp)RNIA(Vi)IA	MIEKLRNIAIIA	*yihK*	4.97	5.40	65.4	73.1	0.01	60
M140	SKTIATENAPAA	M*SKTIATENAPAA	*yjgF*	5.25	5.44	13.5	14.5	0.70	1320
M132	(SA) (DE) FVYTMXRV (LI) (Ga)	M*AQFVYTMHRVGK	*yjjK*	5.24	5.67	64.9	59.4	0.03	140
M18-2	(ETMA) (DQS) (IX) (KF) (VNH) (AF) (VS) (VX) (GX) (AP) (Mh)SG		NM		5.44		39.4	0.04	820
M26	AP(KV)D(NK)(Td)m(YV)Vsr(kv)		NM		4.97		27.1	0.53	480
M29-2	(MSq) (NDM) (Prt) (Lk) (KI) (AD) (GV) (Dks) (IX) (AL) (Pg) (Kr)		NM		5.15		15.0	0.06	100
M33-2	(MAQHS) (EK) (SHQ) (KLQ) (VIF) (VGT) (VS) (PRE) (AS) (QHM) (GAS) (KV)		NM		5.25		44.8	0.20	2900
M63-2	(Ma) (Nd) (Li) (Hfk) (Ev) (Yf) (QX) (Av)KQLF		NM		5.53		39.7	0.02	80
M70	(MAi) (Ed)IL(av) (dk)e(en)X(ap)s(va)		NM		6.72		39.2	0.10	100
M75-2	(Sma) (KNE) (IN)V(Vk) (Vi) (il) (fg) (ta) (eaqgd)ii		NM		5.40		44.6	0.17	300
M77-2	(AS) (NE) (LY) (FS) (VQ) (TG) (LV) (NGP) (LQ) (rX) (IX) (rX)		NM		5.44		49.9	0.02	80
M107-2	(SM) (KE) (NI) (YK) (VH) (IT) (AD) (VP) (LA) (PQ) (GX) (DK)		NM		5.26		42.4	0.08	260
M109-2	(Akst) (AI)I(NE) (TG)K(VL)K(lk)AV(IN)		NM		5.34		38.3	0.04	140
M124-1	M(LIV) (QKPV)X(RF) (EK) (EQ) (MG) (VA)R(VA) (IQ)		NM		5.26		52.5	0.04	120
M124-2	M(LIV) (QKPV)X(RF) (EK) (EQ) (MG) (VA)R(VA) (IQ)		NM		5.26		52.5	0.02	60
M131	iGTE(gy)XPq(qgi)XXg		NM		5.71		62.9	0.04	80
M146-1	(ASM) (ND) (TV) (IK) (DKS) (MV) (TR)I(FNr)X(AV)		NM		4.69		14.6	0.24	200
M146-2	(ASM) (ND) (TV) (IK) (DKS) (MV) (TR)I(FNr)X(AV)		NM		4.69		14.6	0.33	260
M147-2	(AM) (ET) (SVLN) (FNq) (TQ) (TLV) (TV) (NL) (RK) (PRY) (FRE)D		NM		6.54		14.1	0.14	100
M93	XXXXXXXXXXXX				6.30		42.5	0.05	

a) The the observed *N*-terminal amino acid sequence of the spot. A single upper case letter indicates an unambiguous PTH_{aa} signal in the cycle. A lower case letter indicates that the PTH_{aa} signal is close to the level of background noise or lag. Parentheses indicate that a cycle had two or more PTH_{aa} signals. The order of amino acids in parentheses is from strongest to weakest PTH_{aa} signals. An upper case "X" indicates no PTH_{aa} signal could be detected above the background lag. A lower case "x" indicates instrument failure in the cycle and no amino acid call. The symbol "XXXXXXXXXXXX" indicates that no PTH_{aa} sequence signal was detected above the background lag during the entire sequencing run.

b) The observed protein sequence compared to the amino terminal sequence of the matching conceptual protein in the *E. coli* genome. The predicted sequence is shown and the observed sequence is in bold. An "*" in the protein sequence shows the observed start site based on the *N*-terminal sequence tag. An underlined amino acid indicates a discrepancy between the predicted sequence and the observed sequence tag. The predicted amino acid is shown. For query-sequence matching the internal region of an *E. coli* gene, the numbers of amino acids between the predicted *N*- and *C*-termini of the conceptual protein that match the start and stop sites of the sequence tag are shown in parentheses.

c) The *E. coli* gene symbol assigned to the ORF identified by the protein sequence. The symbol "NM" indicates that the query string was used to search the database, but no significant match was found. The table is sorted alphabetically by loci names.

d) The predicted isoelectric point calculated from the conceptual sequence of the protein.

e) The observed isoelectric point of the protein calculated from the (x) coordinates of the spot.

f) The predicted mass calculated from the conceptual sequence of the protein. The units are kDa.

g) The observed mass of the protein calculated from the (y) coordinates of the spot. The units are kDa.

h) The observed abundance on the master 2-DE gel.

i) The estimated cellular abundance of the protein species. The units are molecules/cell.

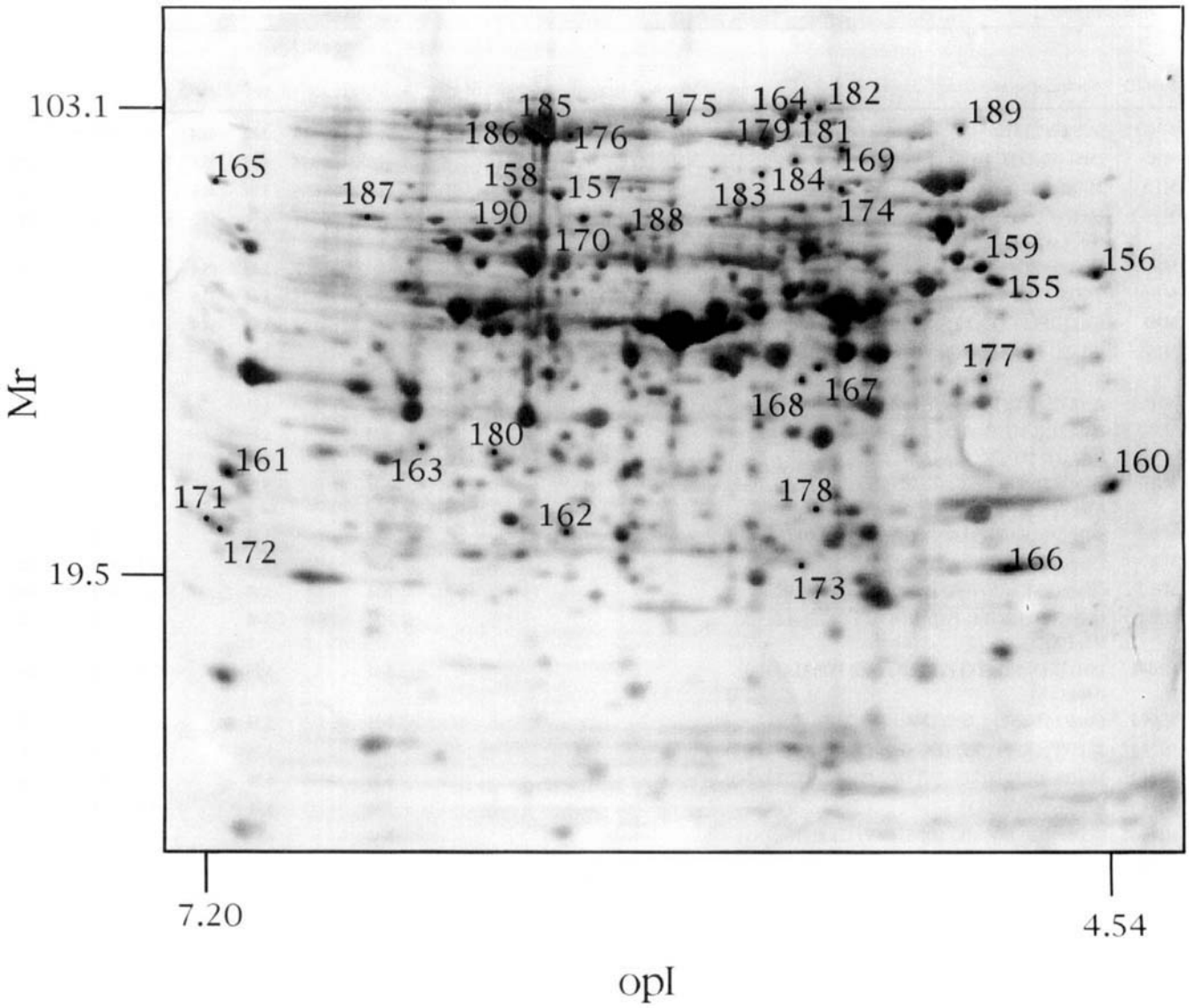

Figure 7. Spots M155–M190: Master gel of total-protein extract fractionated by 2-DE, electrotransferred to PVDF, and stained with Coomassie Blue. The protein extract was prepared using the sonication method from W3110 growing in glucose-minimal media. The IEF used a 4:1 carrier ampholyte blend of Serva 5–7:Serva 3–10, and the SDS-PAGE used a 12.5%T, 2.7%C gel. The largest and smallest op*I* and M_r of the identified 2-DE spots from the gel are shown on the op*I* and M_r axes. The data for spots M155–M190 are found in Table 7.

Table 7. Growth-phase, minimal-media *E. coli* proteins from a sonicated cell extract[a)]

		Predicted *versus* observed sequence, p*I*, MW						Abundance	
Spot ID	*N*-terminal sequence tag	Sequence tag match to genomic ORF	Locus	cp*I*	op*I*	MW	M_r	O-abd	N-abd
M175	SERFPNDVDPIE	**M*SERFPNDVDPIE**	*aceE*	5.50	5.65	99.5	97.9	0.14	1140
M165	SEKHPGPLVVEG	**MSM*SEKHPGPLVVEG**	*cysI*	7.63	7.16	63.9	73.2	0.07	640
M163	MFTGSIVAIVTP	**MFTGSIVAIVTP**	*dapA*	6.42	6.46	31.3	33.2	0.21	840
M156	AQVINTNSLSLI	**M*AQVINTNSLSLI**	*fliC*	4.34	4.57	51.2	44.2	0.26	4200
M171	(PA) (SD) LSKEAA (LT)VTE	**M*PSLSKEAALVHE**	*folE*	7.35	7.20	24.7	28.5	0.34	420
M179	AyTTPIARYRNI	**M*ARTTPIARYRNI**	*fusA*	5.13	5.40	77.4	87.0	0.22	2520
M172	ADKKLVVATDTA	MKSVLKVSLAALTLAFAVSSHA***ADKKLV VATDTA**	*glnH*	7.51	7.17	25.0	27.7	0.29	1440
M158	SEAEARPTNFIR	**M*SEAEARPTNFIR**	*glnS*	6.23	6.14	63.3	66.1	0.11	520
M183	SE(DKRa)XFLVEIGTE	**M*SEKTFLVEIGTE**	*glyS*	5.18	5.41	76.7	71.7	0.06	400
M174-2	M(RK) (TG) (SQ) (QE) (TY) (LX) (LG) (SF) (TQ) (SL) (KX)	**MKGQETRGFQ**	*htpG*	4.95	5.16	71.4	67.4	0.04	640
M190	MEMLSGAEMVVR	**MEMLSGAEMVVR**	*ilvI*	6.30	6.16	63.0	55.0	0.12	340
M167-1	(SAK)IE(EN) (GT)K(LV)VIA(IV) (ND)	MKIKTGARILALSALTTMMFSASALA***KIEEGKLVIWIN**	*malE*	5.07	5.27	40.7	36.2	0.12	660
M187	(ATDF)XIDDVAKQAK(SQD)	MMKMRWLSAAVMLTLYTSSSWA***F SIDDVAKQAQS**	*mdoG*	6.66	6.62	55.4	59.6	0.17	280
M185	TILNHTLGFP(CR)V	**M*TILNHTLGFPRV**	*metE*	5.97	6.04	84.6	95.7	0.30	8640
M186	TILNHTLGFPRV	**M*TILNHTLGFPRV**	*metE*	5.97	6.08	84.6	93.0	0.23	11520
M176	TILNHTLGFPRV	**M*TILNHTLGFPRV**	*metE*	5.97	5.96	84.6	92.3	0.23	4260
M168	(SGAM)LKKFRGMFSND	**MLKKFRGMFSND**	*mreB*	5.03	5.31	37.0	35.5	0.13	260
M178	MKPTTISLLQXY	**MKPTTIASLQKC**	*panB*	5.06	5.29	28.2	28.5	0.18	1080
M169	MLNPIVRKFQYG	**MLNPIVRKFQYG**	*pnp*	4.96	5.19	77.1	79.2	0.15	1000
M174-1	(M(RK) (TG) (SY) (QE) (TY) (LX) (LG) (SF) (TQ) (SL) (KX)	**MRTSQYLLSTLK**	*proS*	4.98	5.16	63.7	67.4	0.05	800
M188	MQQRRPVRRALL	**MQQRRPVRRALI**	*purH*	5.60	5.79	57.3	54.4	0.25	640
M177	(Ts)DKTslsXXdd	**M*TDKTSLSYKDAG**	*purM*	4.65	4.84	36.7	35.4	0.12	180
M170	(TM) (TG)NYIFVTGGVV	**M*TTNYIFVTGGVV**	*pyrG*	5.82	5.93	60.2	57.7	0.09	480
M173	MTQDELKKAVG	**MTQDELKKIAVG**	*rpiA*	5.12	5.33	22.9	20.7	0.25	740
M166	MELVLKDAQSAL	**MELVLKDAQSAL**	*rplD*	10.52	4.78	22.1	19.5	1.65	4960
M160	ATVSMRDMLKA	**M*ATVSMRDMLKAG**	*rpsB*	7.16	4.54	26.7	30.0	0.37	3540
M161	ATVSMRDMLKAG	**M*ATVSMRDMLKAG**	*rpsB*	7.16	7.14	26.7	32.2	0.43	4560
M157	MKLPVREFDAVV	**MKLPVREFDAVV**	*sdhA*	6.22	6.00	64.4	65.6	0.13	1900
M155-1	(GK) (GK) (KVG) (VPK) (PVM) (MPN) (NMI) (IVN) (VIA) (AVQ) (QAR) (RYQ)	**(42)R*KGKLVPMNIVAQ**(374)	*tig*	4.46	4.80	43.1	43.3	0.06	300
M155-2	(GK) (GK) (KVG) (VPK) (PVM) (MPN) (NMI) (IVN) (VIA) (AVQ) (QAR) (RYQ)	(43)R***KGKVPMNIVAQR**(373)	*tig*	4.46	4.80	43.1	43.3	0.10	500
M155-3	(GK) (GK) (KVG) (VPK) (PVM) (MPN) (NMI) (IVN) (VIA) (AVQ) (QAR) (RYQ)	(44)K***GKVPMNIVAQRY**(372)	*tig*	4.52	4.80	43.2	43.3	0.07	340
M159	KVRIDGFRKGKV	(35)K***KVRIDGFRKGKV**(381)	*tig*	4.61	4.84	44.3	45.4	0.18	2340
M182	MEfTYN(Pn)QDI(IE)Q	**MEKTYNPQDIEQ**	*valS*	5.07	5.24	108.2	103.1	0.05	120
M164	MLEEYRKHVAER	**VLEEYRKHVAER**	*yacI*	5.14	5.33	93.5	99.4	0.14	1120
M181	MLEEyrkXkaa	**VLEEYRKHYAE**	*yacI*	5.14	5.28	93.5	98.6	0.04	140
M189	AEGFVVK(D)IHFE	MAMKKLLIASLLFSSATVYG***AEGFVVK DIHFE**	*yaeT*	4.72	4.88	88.4	89.6	0.06	380
M162-1	M(IE)V(AL) (PV) (TE) (GE) (PA) (GT)A(MG) (AF)	**MIVLVTGATAGF**	*ydfG*	5.92	6.00	27.2	26.7	0.22	1300
M180	MIKKIFALPVIE	**MKLKDCV*MIKKIFALPVIE**	*yeaD*	6.32	6.23	32.7	32.9	0.16	560
M184	MIEKLXNIAIIA	**MIEKLRNIAIIA**	*yihK*	4.97	5.32	65.4	75.9	0.06	340
M162-2	M(IE)V(AL) (PV) (TE) (GE) (PA) (GT)A(MG) (AF)		*NM*		6.00		26.7	0.12	700
M167-2	(SAK)IE(EN) (GT)K(LV)VIA(IV) (ND)		*NM*		5.27		36.2	0.07	420

a) See Table 6 for explanation of column headings.

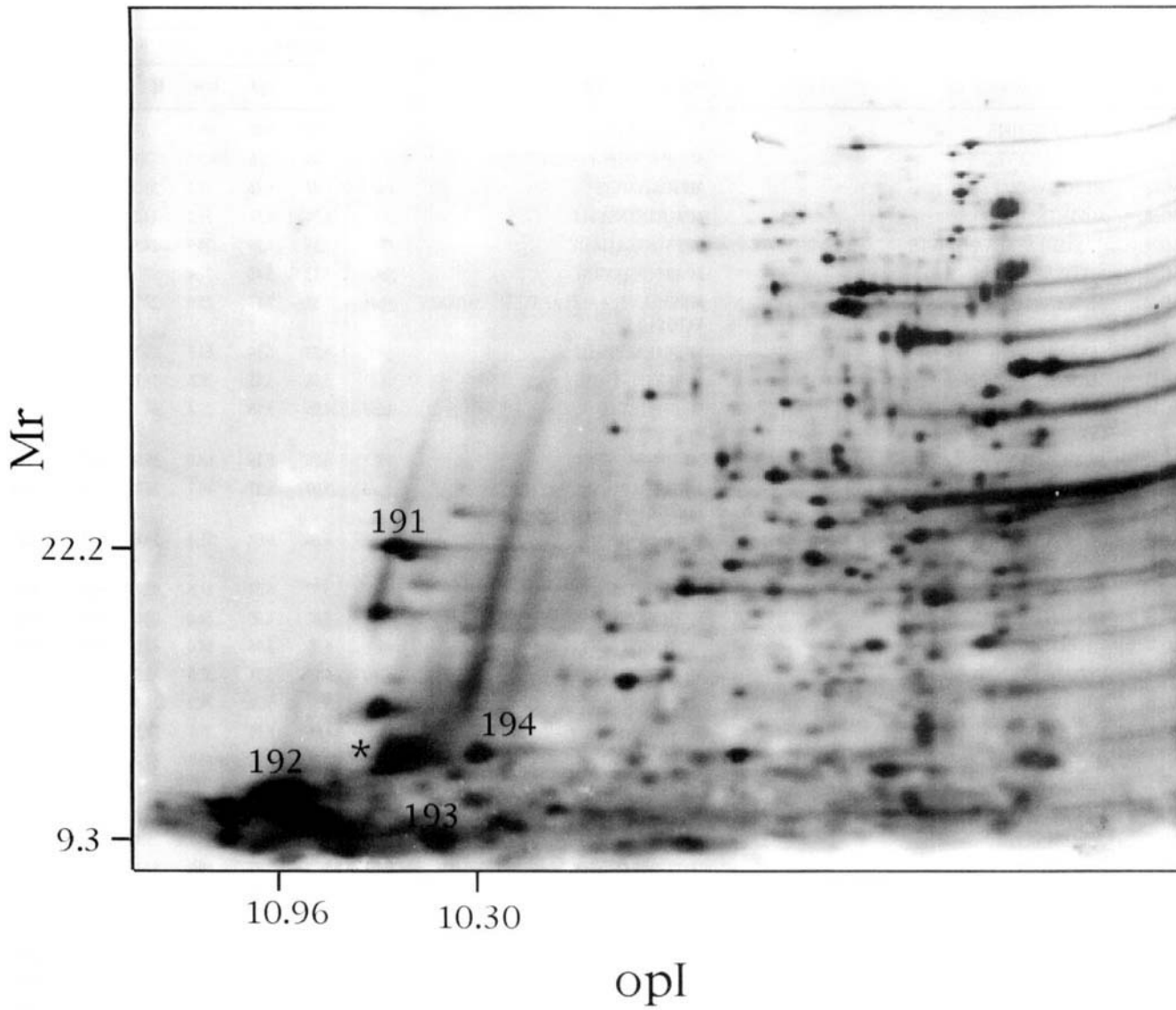

Figure 8. Spots M191–194: Master gel of total protein extract fractionated by NEPGHE 2-DE, electrotransferred to PVDF, and stained with colloidal gold. The protein extract was prepared using the heat/SDS method from EMG2 growing in glucose-minimal media. The first-dimensional NEPGHE used Serva 3–10 carrier ampholytes and the second-dimensional SDS-PAGE used a 12.5%T, 2.7%C gel. The Laemmli anode buffer in the second dimension was supplemented with 50 mM sodium acetate. The largest and smallest op*I* and M_r of the identified 2-DE spots from the gel are shown on the op*I* and M_r axes. The data for spots M191–M194 are found in Table 8. The spot marked with an "*" was identified as DnaseI, which was added to the protein extract.

Table 8. Growth-phase, minimal-media *E. coli* proteins analyzed using NEPHGE for the first dimension[a)]

		Predicted *versus* observed sequence, p*I* MW						Abundance	
Spot ID	*N*-terminal sequence tag	Sequence tag match to genomic ORF	Locus	cp*I*	op*I*	MW	M_r	O-abd	N-abd
M194	ADKIAIVNMYsL	MKKWLLAAGLGLALATSAQA*ADKIA IVNMGSL	*hlpA*	10.32	10.30	15.7	15.7	4.20	580
M193-1	MNK(TS)QLID(VPK)IA(AE)	MNKTQLIDVIAE	*hupA*	10.36	10.45	9.5	9.3	10.29	1380
M193-2	MNK(TS)QLID(VPK)IA(AE)	MNKSQLIDKIAA	*hupB*	10.48	10.45	9.2	9.3	6.31	860
M191	(DMS)I(GI)LVGdXv(Gp)	MIGLVGKKVG	*rplC*	10.69	10.63	22.2	22.2	1.91	180
M192-1	(AMsr) (LV) (RTK) (GI) (RK) (Lh) (KA) (ADr) (EA) (YGVi) (TAI) (VP)	(34)R*RLRGKHKAEYTP(96)	*rplM*	10.65	10.96	12.2	10.6	7.13	2580
M192-2	(AMsr) (LV) (RTK) (GI) (RK) (Lh) (KA) (ADr) (EA) (YGVi) (TAI) (VP)	MVTIRLARHGAK	*rpsP*	11.27	10.96	9.2	10.6	3.36	1220

a) See Table 6 for explanation of column headings.

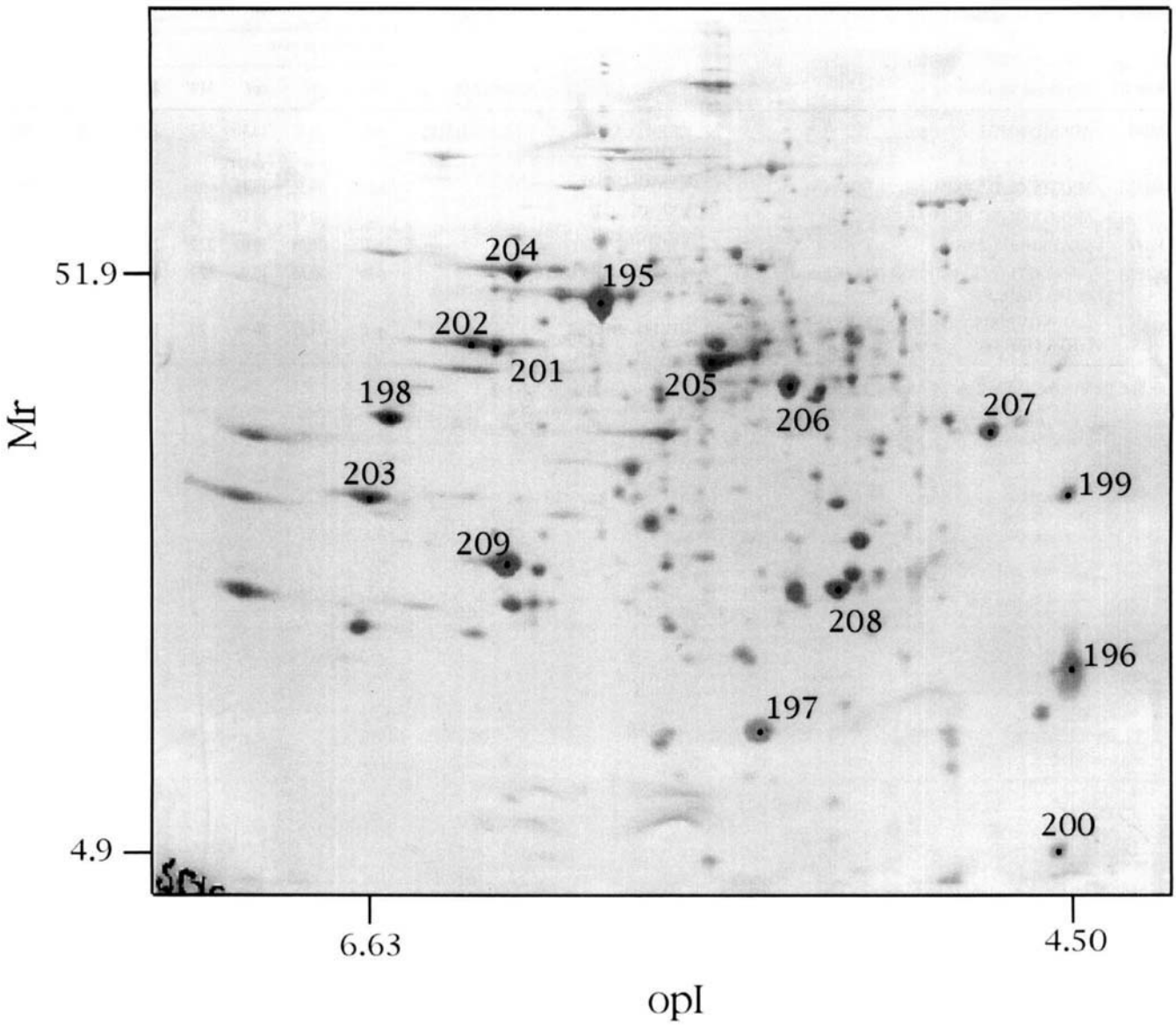

Figure 9. Spots M195–M209: Master gel of periplasmic protein extract fractionated by 2-DE, electrotransferred to PVDF, and stained with Coomassie Blue. The periplasmic protein extract was prepared from EMG2 growing in glucose-minimal media. The IEF used a 4:1 carrier ampholyte blend of Srva 5–7: Serva 3–10, and the SDS-PAGE used an 11.5%T, 2.7%C gel. The largest and smallest op*I* and M_r of the identified 2-DE spots from the gel are shown on the op*I* and M_r axes. The data for spots M195–M209 are found in Table 9.

Table 9. Periplasmic-enriched *E. coli* proteins from cells harvested in growth phase, minimal media[a)]

		Predicted *versus* observed sequence, p*I*, MW						Abundance	
Spot ID	*N*-terminal sequence tag	Sequence tag match to genomic ORF	Locus[b)]	cp*I*	op*I*	MW	M_r	O-abd	N-abd
M203	TELLNSXYDVSR	MAVNLLKKNSLALVASLLLAGHVQA* **TELLNSSYDVSR**	*cysP*	6.99	6.63	35.1	38.0	0.71	440
M195	LTLVYXSEGSPE	MRISLKKSGMLKLGLSLVAMTVAASVQ A*KTLVYCSEGSPE	*dppA*	5.94	5.87	57.4	48.9	1.21	2980
M208	DEGLLNKVKERG	MKLLAHLGRQALMGVMAVAL VAGMSVKSFA***DEGLLNKVKERG**	*fliY*	5.14	5.14	26.1	20.7	1.36	1060
M201	MLKREMNIADYD	**MLKREMNIADYD**	*glyA*	6.46	6.23	45.3	44.7	0.41	240
M196	ADAQKAADNKKP	MKKVLGVILGGLLLLPVVSNA*ADAQ **KAADNKKP**	*hdeA*	4.49	4.50	9.2	12.3	4.69	8460
M199	ADAQKAADNKKP	MKKVLGVILLGGLLLLPVVSNA*ADAQ **KAADNKKP**	*hdeA*	4.49	4.50	9.2	29.5	0.48	1060
M206	EDIKVAVVGAMS	MNTKGKALLAGLIALAFSNMALA*E **DIKVAVVGAMS**	*livJ*	5.17	5.27	36.8	40.4	0.81	1160
M204	xDVPAGVTLAEKQ	MTNITKRSLVAAGVLAALMAGNVA LA*ADVPAGVTLAEKQ	*oppA*	6.17	6.15	58.4	51.9	0.46	500
M197	ENNAQTTNESAG	MTMTRLKISKTLLAVMLTSAVATGSAY A*ENNAQTTNESAG	*osmY*	5.40	5.38	18.2	7.1	4.28	2960
M207	DDNNTLYFYNXT	MKKWSRHLLAAGALALGMSAA HA*DDNNTLYFYNWT	*potD*	4.69	4.71	36.5	35.7	0.63	460
M209	KDTIALVVSTLN	MNMKKLATLVSAVALSATVSANA MA*KDTIALVVSTLN	*rbsB*	6.31	6.21	28.5	23.7	1.88	3940
M198	KDIQLLNVSYDP	MNKWGVGLTFLLAATSVMA*K **DIQLLNVSYDP**	*sbp*	6.82	6.63	34.7	38.0	0.71	1240
M202	APQVVDKVAAVV	MKNWKTLLLGIAMIANTSFA*APQVVDK **VAAVV**	*surA*	6.54	6.33	45.1	45.1	0.50	400
M200	AAAEQGGFS(gd)P(sk) (ak) (tr)	MKKFAAVIAVMALCSAPV*MAA **EQGGFSGPSA**	*yigW*	4.57	4.54	12.1	4.9	6.30	600
M205	XXXXXXXXXXXX				5.49		42.9	1.08	

a) See Table 6 for explanation of column headings.
b) An underlined locus indicates the conceptual protein does not have a recognized signal peptide which is usually associated with proteins exported across the inner membrane.

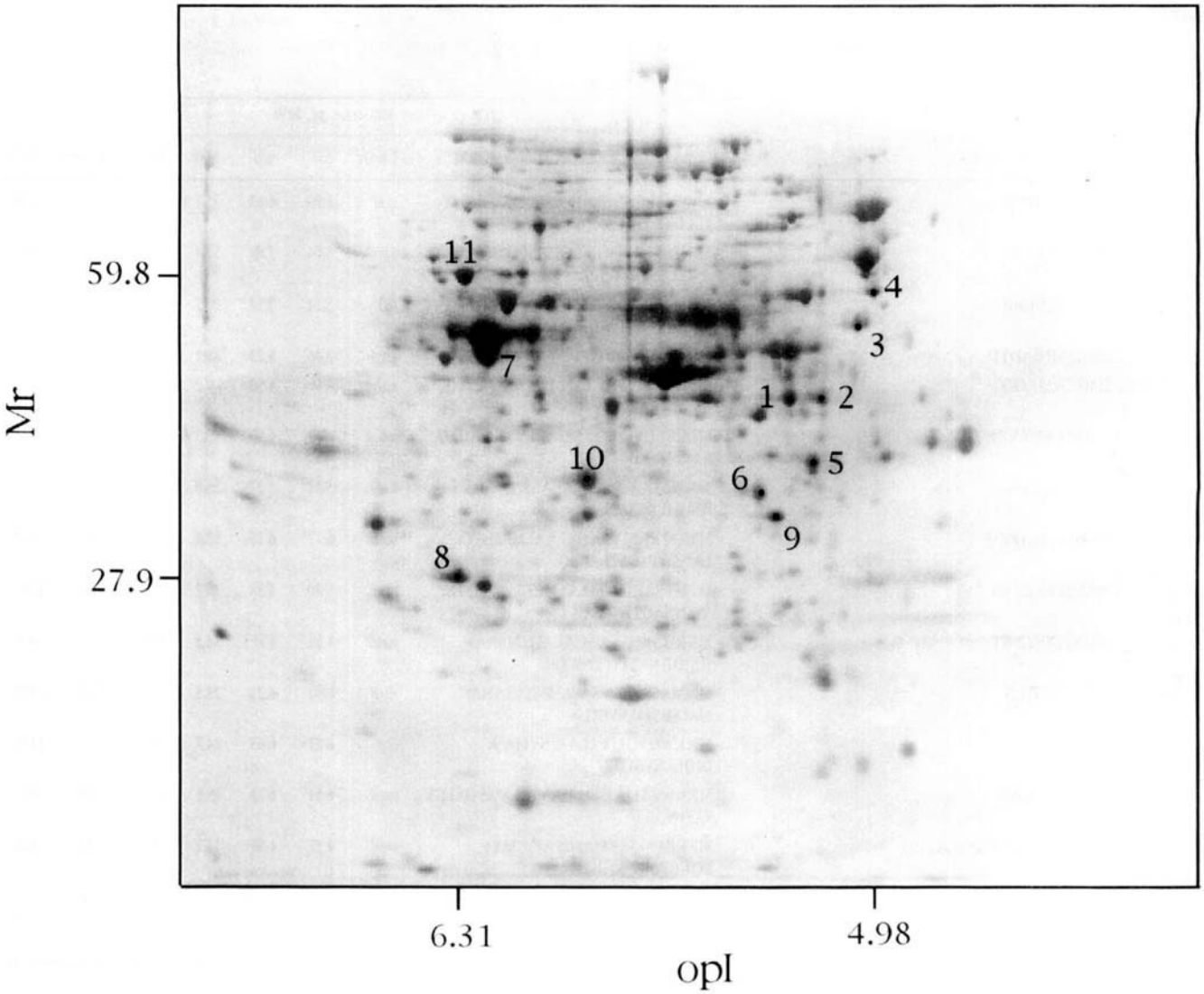

Figure 10. Spots R1–R11: Master gel of total-cell protein extract fractionated by 2-DE, electrotransferred to PVDF, and stained with Coomassie Blue. The protein extract was prepared using the heat/SDS method from EMG2 grown in rich media to early stationary phase. The IEF used a 4:1 carrier ampholyte blend of Serva 5–7: Serva 3–10, and the SDS-PAGE used an 11.5%T, 2.7%C gel. The largest and smallest op*I* and M_r of the identified 2-DE spots from the gel are shown on the op*I* and M_r axes. The data for spots R1–R11 are found in Table 10.

Table 10. Stationary-phase, rich-media *E. coli* proteins from an SDS/heat extract[a)]

		Predicted *versus* observed sequence, p*I*, MW						Abundance	
Spot ID	*N*-terminal sequence tag	Sequence tag match to genomic ORF	Locus	cp*I*	op*I*	MW	M_r	O-abd	N-abd
R2-1	(SM) (LQ) (IG) (NK) (TV) (KA) (IE) (KF) (PL) (FA) (KM) (NR)	**M*SLINTKIKPFKN**	*ahpC*	4.90	5.16	20.6	19.7	0.32	1580
R3	ATLGKIVQVIGAV	**M*ATGKIVQVIGAV**	*atpD*	4.74	5.03	50.2	47.8	0.32	1300
R10	MKVAVLGAAGGI	**MKVAVLGAAGGI**	*mdh*	5.56	5.96	32.3	35.6	0.42	23.80
R9	ADTRIGVTIYKY	**MNKKVLTLSAVMASMLFGAAAHA*ADT RIGVTIYKY**	*mglB*	5.14	5.49	33.4	32.4	0.45	4720
R11	ADVPAGVTLAEK	**MTNITKRSLVAAGVLAALMAGNVA LA*ADVPAGVTLAEK**	*oppA*	6.17	6.30	58.4	59.8	0.23	23.60
R1	SVIKMTDLDLAG	**M*SVIKMTDLDLAG**	*pgk*	4.93	5.27	41.0	39.8	0.56	1460
R8	KDTIALVVSTLN	**MNMKKLATLVSAVALSATVSANAMA*KD TIALVVSTLN**	*rbsB*	6.31	6.31	28.5	27.9	0.83	5900
R2-2	(SM) (LQ) (IG) (NK) (TV) (KA) (IE) (KF) (PL) (FA) (KM) (NR)	**MQGSVTEFLKPRL**	*rpoA*	4.81	5.16	36.5	39.7	0.12	560
R5	TDKLTSLRQY	**M*TDKLTSLR**	*talB*	4.94	5.18	35.1	34.9	0.29	900
R4	MQVSVETTQGLG	**MQVSVETTQGLG**	*tig*	4.64	4.98	48.2	51.7	0.33	600
R7	MENFKHLPEPFR	**MENFKHLPEPFR**	*tnaA*	6.16	6.21	52.8	49.8	0.80	7160
R6	xxxxAGLVKELR	**M*AEITASLVKELR**	*tsf*	5.07	5.33	30.3	33.0	0.49	1100

a) See Table 6 for explanation of column headings.

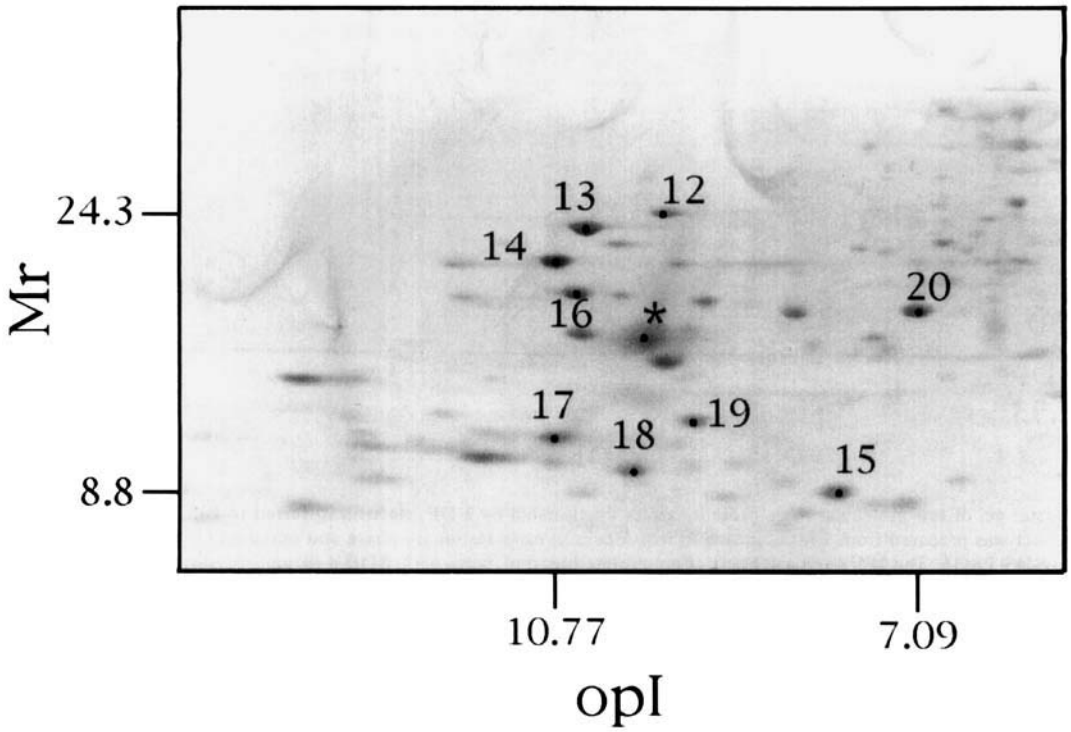

Figure 11. Spots R12–R20: Master gel of total-protein extract fractionated by NEPGHE 2-DE, electrotransferred to PVDF, and stained with Coomassie Blue. The protein extract was prepared using the heat/SDS method from EMG2 grown in rich media to early stationary phase. The first-dimensional NEPGHE used Serva 3–10 carrier ampholytes. The second dimension used a 16% Tricine SDS-PAGE. The largest and smallest op*I* and M_r of the identified 2-DE spots from the gel are shown on the op*I* and M_r axes. The data for spots R12–R20 are found in Table 11. The spot marked with an "*" was identified as DnaseI, which was added to the protein extract.

Table 11 Stationary-phase, rich-media *E. coli* proteins analyzed using NEPHGE for the first dimension[a]

		Predicted *versus* observed sequence, p*I*, MW						Abundance	
Spot ID	*N*-terminal sequence tag	Sequence tag match to genomic ORF	Locus	cp*I*	op*I*	MW	M_r	O-abd	N-abd
R18	MNK(TS)QLID(DNK)IAA	**MNKSQLIDKIAA**	*hupB*	10.48	9.87	9.2	9.8	0.39	2040
R17-2	(MK) (FE)T(IA) (AN) (AK) (KETA) (VFETKL) (EFRVE)R(RE) (QE)		MM		10.77		10.9	0.15	1600
R12	AKLTKRMRVIIE	**M*AKLTKRMRVIRE**	*rplA*	10.45	9.49	24.6	24.3	0.12	8480
R13	MIGLVGKKVGMT	**MIGLVGKKVGMT**	*rplC*	10.69	10.40	22.1	22.4	0.23	8500
R14	SRVAKAPVVVPA	**M*SRVAKAPVVVPA**	*rplF*	10.52	10.77	18.8	19.2	0.28	5680
R20	MQVILLDKVANN	**MQVILLDKVANL**	*rplI*	6.56	7.09	15.8	15.4	0.16	1340
R17-1	(MK) (FE)T(IA) (AN) (AK) (KETA) (VFETKL) (EFREVE)R(RE) (QE)	**MFTINAEVRKEQ**	*rplY*	10.42	10.77	10.7	10.9	0.26	2700
R19	MQNQRIXI(IR)LLA	**MQNQRIRIRLKA**	*rpsJ*	10.48	9.19	11.7	11.4	0.27	1360
R15	ILSKDEGGR(HR)T(PT)	(308)Y***ILSKDEGGRHTP**(70)	*tufA/B*	7.69	7.79	9.2	8.8	0.56	12960
R16	XXXXXXXXXXX				10.52		16.6	0.24	

a) See Table 6 for explanation of column headings.

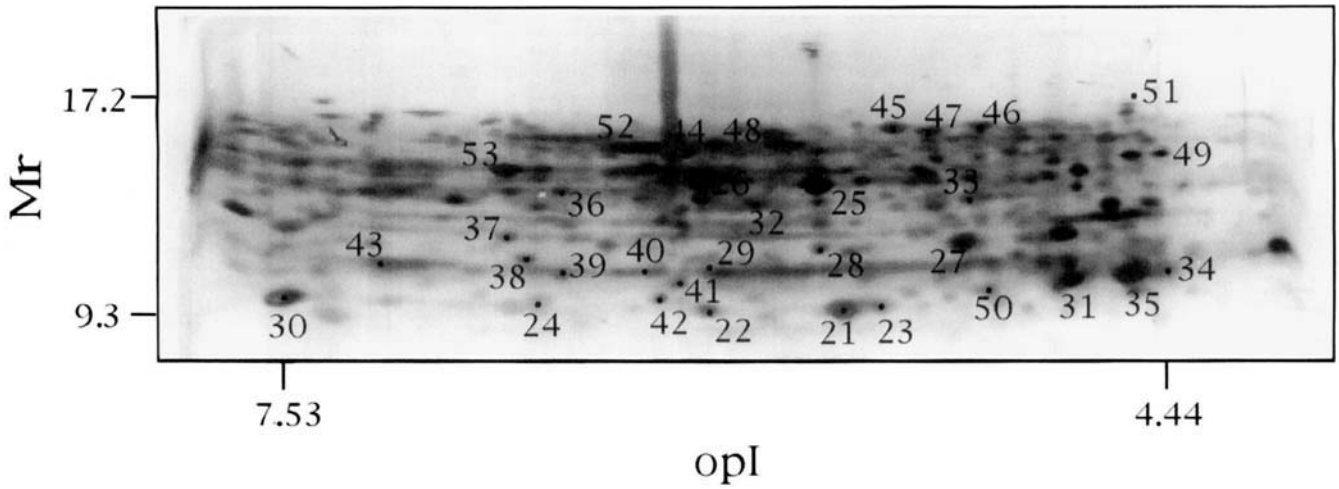

Figure 12. Spots R21–R53: Master gel of low molecular mass protein extract fractionated by 2-DE, electrotransferred to PVDF, and stained with colloidal gold. The protein extract was prepared from EMG2 grown in rich media to early stationary phase and enriched for low molecular mass protens by continuous elution SDS-PAGE. The IEF used a 1:1 carrier ampholyte blend of Serva 5–7: BDH 4–8, and the second dimension used a 16% Tricine SDS-PAGE. The largest and smallest op*I* and M_r of the identified 2-DE spots from the gel are shown on the op*I* and M_r axes. The data for spots R21–R53 are found in Table 12.

Table 12. Low-molecular-mass *E. coli* proteins from cells at early stationary phase in rich media[a)]

		Predicted *versus* observed sequence, p*I*, MW						Abundance	
Spot ID	*N*-terminal sequence tag	Sequence tag match to genomic ORF	Locus	cp*I*	op*I*	MW	M_r	O-abd	N-abd
R28	AVSTDTHFTHKLA	(67)Y*AVSTDTHFTHKA(104)	*ahpC*	5.64	5.47	12.7	11.7	0.18	920
R49	GLFDKLKSLVVP	M*GLFDKLKSLVSD	*crr*	4.56	4.61	18.1	16.1	0.11	500
R47	GKIIGIDLGTTN	M*GKIIGIDLGTTN	*dnaK*	4.67	5.12	69.0	16.4	0.16	320
R48	GKIIGIDLGTTN	M*GKIIGIDLGTTN	*danK*	4.67	5.82	69.0	16.0	0.09	500
R44	STAKLVKSKATN	M*STAKLVKSKATN	*dps*	6.02	5.97	18.6	15.6	0.50	6720
R24	LLKDGEDPGYTL	(378)F*KLKDGEDPGYTL(72)	*gadA*	6.52	6.57	9.9	9.5	0.16	3820
R35-1	(AMS) (DILV) (AKPT) (QK) (KD) (AX) (AI) (DI) (NEL) (KX) (KV)P	MKKVLGVILGGLLLLPVVSNA*ADAQ KAADNKKP	*hdeA*	4.49	4.55	9.2	10.7	0.98	4350
R51	ADAQKAADDKK(PK)	MKKVLGVILLGGLLLLPVVSNA*ADAQ KAADNKKP	*hdeA*	4.49	4.55	9.2	17.2	0.14	1020
R31-1	(AN) (NK) (ET) (SRn) (AVF) (GAVK) (DK) (MLD)TH(AEQ) (NEK)	MGYKMNISSLRKAFIFMGAVAALSLV NAQSALA*ANESAKDMTCQE	*hdeB*	4.78	4.70	9.1	10.4	0.23	1840
R25-2	(TS) (IE) (AH) (KL)K(IG) (QL) (AN) (HN)I(ER) (TG)	M*SEALKILNNIRT	*hns*	5.25	5.48	15.4	14.3	0.19	1780
R29	SEALKILNNIRT	M*SEALKILNNIRT	*hns*	5.25	5.90	15.4	11.6	0.27	1020
R27	MNILRPLHDRVIV	MNIRPLHDRVIV	*mopB*	4.99	5.02	10.4	12.0	0.52	5720
R25-1	(TS) (IE) (AH) (KL)K(IG) (QL) (AN) (HN)I(ER) (TG)	M*TIHKKGQAEWEG	*osmC*	5.71	5.48	15.0	14.3	0.34	3180
R45	ENNAQTTNEEX(G)	MTMLTRLKISKTLLAVMLTSAVATGSAY A*ENNAQTTNESAG	*osmY*	5.40	5.23	18.2	16.5	0.22	780
R46	E(ND) (NGD)AQRRNE(EV)X(GV)	MTMTRLKISKTLLAVMLTSAVATGSAY A*ENNAQTTNESAG	*osmY*	5.40	4.97	18.2	16.6	0.19	160
R42-1	(KTA) (MLD) (TR) (DI) (EGKQ) (AX) (GNP) (EP) (DR) (AL) (KG)L	(92)L*KMTDEAGEDAKL(68)	*ppa*	6.12	6.02	9.4	9.7	0.16	200
R22-1	(MSA) (FL) (QD) (EG) (VT) (TS) (IPD) (TG) (ALY) (PX) (NL)	MLFQQEVTITAPN	*ptsH*	5.74	5.85	9.1	9.3	0.36	4180
R23-1	(MA) (FLN) (QK) (QDP) (ET) (VA) (TG) (IG) (TNL)APN	MFQQEVTITAPN	*ptsH*	5.74	5.28	9.1	9.5	0.25	2100
R34-1	(MAPS) (KENLV) (IL) (VK) (LT) (KVA)D(AG)QA(AKL)L	MELVLKDAQSAL	*rplD*	10.52	4.44	22.1	10.9	0.26	680
R52	MQVILLDKVA(AN) (LN)	MQVILLDKVANL	*rplI*	6.56	6.14	15.8	15.8	0.33	660
R35-2	(AMS) (DILV) (AKPT) (QK) (KD) (AX) (AI) (DI) (NEL) (KX) (KV)P	M*SITKDQIIEAVA	*rplL*	4.42	4.55	12.2	10.7	0.10	540
R39	MENLFKHLPEX(PF)R	MENLFKHLPEPFR	*tnaA*	6.16	6.37	52.8	10.8	0.35	460
R30	ILSKDEGgRHTP	(308)Y*ILSKDEGGRHTP(70)	*tufA/B*	7.69	7.53	9.2	9.9	0.80	4980
R33	AYKHILIAVDLS	M*AYKHILIAVDLS	*uspA*	4.84	4.98	15.9	13.7	0.18	1580
R36	MYKTIIMPVDVF	MYKTIIMPVDVF	*ybdQ*	6.51	6.38	15.9	14.1	0.16	180
R37	TMNITSKQMEIF	M*TMNITSKQMEIT	*yfiA*	6.69	6.59	12.7	12.3	0.20	600
R26	GLFNFVKDAGEK	M*GLFNFVKDAGEK	*ygaU*	5.82	5.86	15.9	14.6	0.59	4340
R40	(SAGM) (Xnr)TIFX(TL)FLQi(Ed)	M*SRTIFCTFLORE	*yggX*	6.27	6.10	10.8	10.9	0.16	200
R41-1	(MSA) (LV) (TD) (VK) (IHP) (APL) (EX) (IL) (RM) (TD) (RX) (PV)	MLTVIAEIRTRP	*ygiN*	6.23	5.95	11.5	10.4	0.16	260
R21	MNKDEAGGNXKQ	MNKDEAGGNWKQ	*yjbJ*	5.27	5.40	8.3	9.3	0.83	12620
R23-2	(MA) (FLN) (QK) (QDP) (ET) (VA) (TG) (IG) (TNL)APN	MNKDEAGGNWKQ	*yjbJ*	5.27	5.28	8.3	9.5	0.01	60
R50	MNKDEAGGNNKK	MNKDEAGGNWKQ	*yjbJ*	5.27	4.95	8.3	10.0	0.29	440
R22-2	(MSA) (FL) (QD) (QD) (EG) (VT) (TS) (IPD) (TG) (ALY) (PX) (NL)		NM		5.85		9.3	0.02	240
R31-2	(AN) (NK) (ET) (SRn) (AVF) (GAVK) (DK) (MLD)TH(AEQ) (NEK)		NM		4.70		10.4	0.20	1580
R32-1	(VAMS) (GIN) (QX) (QF) (RF) (FKS) (GX) (GI) (KX) (VX) (IN)		NM		5.68		13.6	0.23	980
R32-2	(VAMS) (GIN) (QX) (QF) (RF) (FKS) (GX) (GI) (KX) (VX) (IN)		NM		5.68		13.6	0.02	100
R34-2	(MAPS) (KENLV) (IL) (VK) (LT) (KVA)D(AG)QA(AKL)L		NM		4.44		10.9	0.22	580
R38	(AM)PI(DK)NLTXY(YTE)G(AGI) (Ag)		NM		6.52		11.4	0.19	340
R41-2	(MSA) (LV) (TD) (VK) (IHP) (APL) (EX) (IL) (RM) (TD) (RX) (PV)		NM		5.95		10.4	0.02	40
R42-2	(KTA) (MLD) (TR) (DI) (EGKQ) (AX) (GNP) (EP) (DR) (AL) (KG)L		NM		6.02		9.7	0.02	40
R53	P(XT)VDIV(NDL)		NM		6.62		15.0	0.35	100
R43	XXXXXXXXXXX				7.00		11.2	0.55	

a) See Table 6 for explanation of column headings.

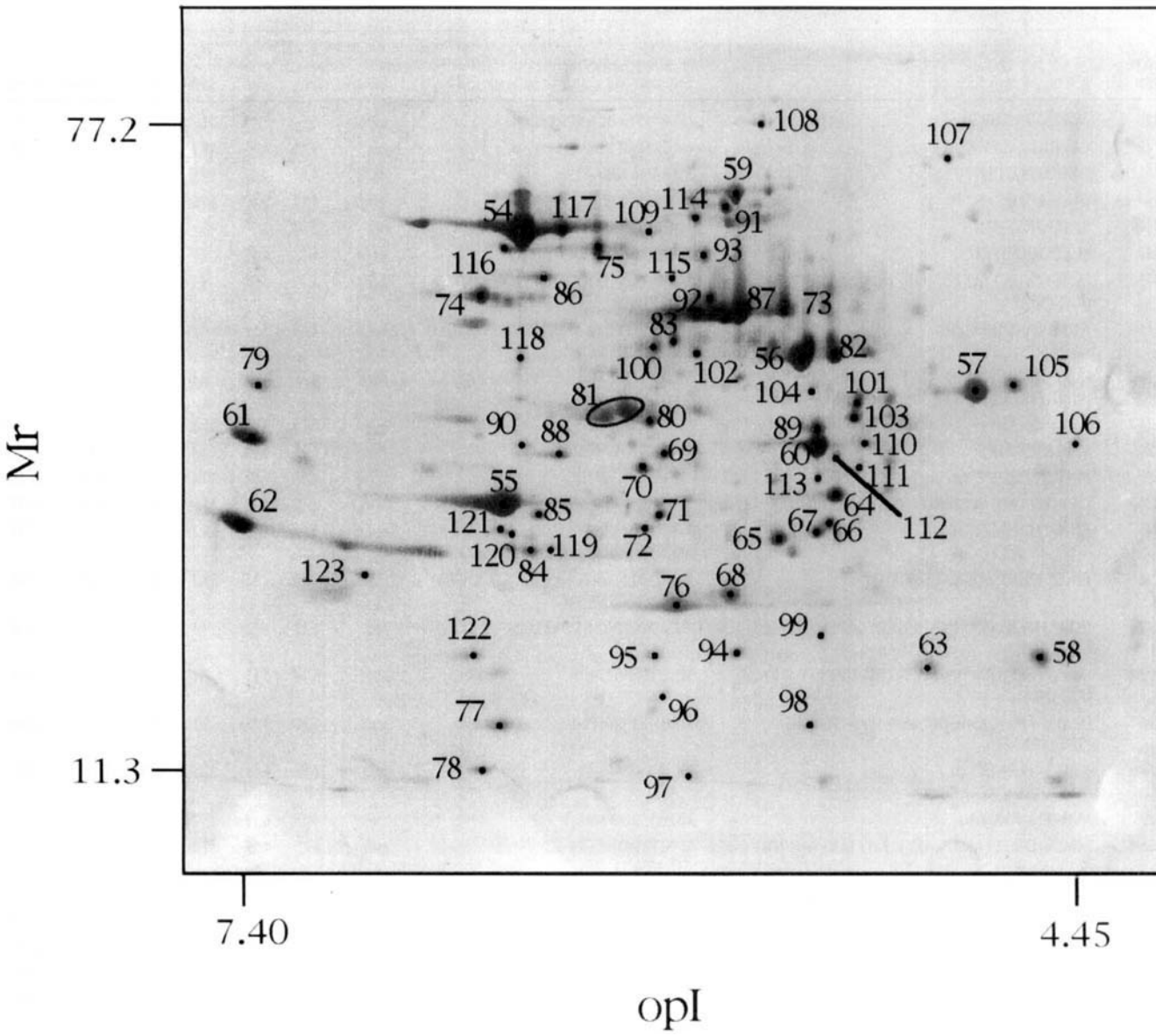

Figure 13. Spots R54–R123: Master gel of periplasmic protein extract fractionated by 2-DE, electrotransferred to PVDF, and stained with Coomassie Blue. The periplasmic protein extract is prepared from EMG2 grown in rich media to early stationary phase. The IEF used a 4:1 carrier ampholyte blend of Serva 5–7: Serva 3–10, and the SDS-PAGE used an 11.5%T, 2.7%C gel. The largest and smallest op*I* and M_r of the identified 2-DE spots from the gel are shown on the op*I* and M_r axes. The data for spots R54–R123 are found in Table 13.

Table 13. Periplasmic-enriched *E. coli* proteins from cells at stationary phase in rich media[a)]

		Predicted *versus* observed sequence, *pI*, MW						Abundance	
Spot ID	*N*-terminal sequence tag	Sequence tag match to genomic ORF	Locus[b)]	cp*I*	op*I*	MW	M_r	O-abd	N-abd
R71	MYIILLGAPGAG	**MRIILLGAPGAG**	*adk*	5.53	5.73	23.6	28.1	0.19	60
R73	QTVPEGYQLQQV	MNKTLIAAAVAGIVLLASNAQA***QTVPE GYQLQQV**	*agp*	5.35	5.31	43.6	46.1	0.48	60
R87	QTVPEGYQLYYV	MNKTLIAAAVAGIVLLASNAQA***QTVPE** GYQLQQV	*agp*	5.35	5.45	43.6	45.1	0.80	380
R80	ENLKLGFLVKQP	MHKFTKALAAIGLAAVMSQSA MA***ENLKLGFLVKQP**	*araF*	5.60	5.76	33.2	34.8	0.26	80
R66	ALPETVyIGTDT	MKKSILALSLLVGLSTAASSYA***ALPETV RIGTDT**	*argT*	5.05	5.16	25.8	27.6	0.21	100
R65	AETIRFATEASY	MKKVLIAALIAGFSLSATA***AETIRFA TEASY**	*artI*	5.20	5.33	25.0	26.7	0.38	160
R100-2	(SM) (FKv) (IE) (FN) (DI) (TF) (VAi) (KAD) (PI) (GA) (VD) (IP)	**MFENITAAPADP**	*aspC*	5.58	5.74	43.6	41.7	0.07	40
R59	ATVDLRIMETTD	MIKFSATLLATLIAASVNA***ATVDLRI METTD**	*cpdB*	5.37	5.47	68.9	63.1	0.26	100
R58	GLFDKLKSLV(ISV)D	**M*GLFDKLKSLVSD**	*crr*	4.56	4.55	18.1	19.6	0.42	140
R107	(Gs)KIIGIDLG(TH)TN	M***GKIIGIDLGTTN**	*dnaK*	4.67	4.81	69.0	69.6	0.06	20
R75	KTLVYXSEGSPE	MRISLKKSGMLKLGLSLVAMTVAASVQ A***KTLVYCSEGSPE**	*dppA*	5.94	5.94	57.4	53.7	0.24	40
R68-1	A(QF) (Qk) (EL) (DP) (GA) (XKL) (QP) (YX) (TA) (TX) (Ld)	MKKIWLALAGLVLAFSASA***AQYEDGK QYTTL**	*dsbA*	5.39	5.49	21.1	23.4	0.46	140
R84	DDAAIQQTLAkM	MKKGFMLFTLLAAFSGFAQA***DDAA IQQTLAKM**	*dsbC*	6.24	6.19	23.5	26.0	0.29	100
R77	AESVQPLEKIAP	MKTILPAVLFAAFATTSAWA***AESVQPLE KIAP**	*eco*	6.03	6.31	16.1	14.9	0.80	260
R92	SKIVKIIGREII	**M*SKIVKIIGREII**	*eno*	5.23	5.55	45.5	47.3	0.15	20
R100-1	(SM) (KFv) (IE) (FN) (DI) (TF) (VAi) (KAD) (PI) (GA) (VD) (IP)	**M*SKIFDFVKPGVI**	*fba*	5.75	5.74	39.0	41.7	0.12	60
R61	AEAAKPATAADS	MKLSLFKVTLLATTMAVALHAPITFA***A EAAKPATAADS**	*fkpA*	7.10	7.34	26.2	33.5	0.93	200
R67	DEGLLNKVKEXG	MKLAHLGRQALMGVMAVAL VAGMSVKSFA***DEGLLNKVKERG**	*fliY*	5.14	5.21	26.1	27.0	0.20	80
R108	(AS)RTTPIArYrNI	M***ARTTPIARYRNI**	*fusA*	5.13	5.38	77.4	77.2	0.05	20
R102-1	(ASk) (QI) (QE) (TER) (PG) (LK) (YX) (EV) (QI) (HX) (TI) (LN)	**M*AQQTRPLYEQHTL**	*gcvT*	5.35	5.60	40.0	40.9	0.10	40
R62	ADKKLVVATDTA	MKLSVLKVSLAALTLAFAVSS HA***ADKKLVVATDTA**	*glnH*	7.51	7.40	25.0	27.7	1.29	460
R85	AVTXLVLVyXxE	M***AVTKLVLVRHGE**	*gpmA*	6.11	6.16	28.4	28.2	0.21	40
R64	AIPQNIRIGTDP	MKKLVLSLSLVLAFSSATAALA***AIPQNI RIGTDP**	*hisJ*	5.01	5.15	26.2	29.4	0.41	320
R83	MLDAQTIATVKA	**MLDAQTIATVKA**	*hmpA*	5.63	5.67	43.9	42.3	0.22	100
R95	DAASDLKSRLDK	MMKKIAITCALLSSLVASSVWA* **DAASDLKSRLDK**	*lolA*	5.70	5.75	20.3	19.7	0.28	60
R56-1	(KA) (IN) (ES) (EN) (GX) (KX) (LI) (VX) (IX) (XA) (IX) (Nr)	MKIKTGARILALSALTTMMFSASALA* **KIEEGKLVIWIN**	*malE*	5.07	5.25	40.7	40.9	0.94	720
R82	KIEEGKLVIiIN	MKIKTGARILALSALTTMMFSASALA* **KIEEGKLVIWIN**	*malE*	5.07	5.15	40.7	40.8	0.61	240
R102-2	(ASk) (QI) (QE) (TER) (PG) (LK) (YX) (EV) (QI) (HX) (TI) (LN)	MKIKLTGARILALSALTTMMFSASALA* **KIEEGKLVIWIN**	*malE*	5.07	5.60	40.7	40.9	0.02	20
R81	MKVAVLGAAGGT	**MKLVAVLGAAGGI**	*mdh*	5.56	5.84	32.3	35.7	0.65	260
R103-2	(KAs) (ED) (VT) (KX) (IT) (GX) (MV) (ATr) (IX) (DY) (DX) (Ln)	MNKKLVLTLSAVMASMLFGAA AHA***ADTRIGVTIYKY**	*mglB*	5.14	5.09	33.4	35.0	0.02	20
R60	ADTRIGVTIYKY	MNKKVLTLSAVMASMLFGAAAHA***ADT RIGVTIYKY**	*mglB*	5.14	5.21	33.4	32.7	0.78	1320
R110	ADTRIGVTIY(KG)Y	MNKKVLTLSAVMASMLFGAAAHA***ADT RIGVTIYKY**	*mglB*	5.14	5.05	33.4	32.9	0.16	40
R119	(MA)DIIS(VN)ALXRXs	**MDIISVALKRHS**	*nfnB*	6.17	6.11	23.9	26.2	0.11	20
R54	ADVPAGVTLAEK	MTNITKRSLVAAGVLAALMAGNVA LA***ADVPAGVTLAEK**	*oppA*	6.17	6.22	58.4	56.8	1.46	1300

Table 13. continued

Spot ID	*N*-terminal sequence tag	Predicted *versus* observed sequence, p*I*, MW						Abundance	
		Sequence tag match to genomic ORF	Locus[b)]	cp*I*	op*I*	MW	*M*.	O-abd	N-abd
R117	ADVPAGVTLAEK	MTNITKRSLVAAGVLAALMAGNVA LA*ADVPAGVTLAEK	*oppA*	6.17	6.08	58.4	57.0	0.65	400
R94	ENNAQTTNESAG	MTMTRLKISKTLLAVMLTSAVATGSAY A*ENNAQTTNESAG	*osmY*	5.40	5.47	18.2	19.9	0.22	80
R114	(SAMt)RVNNGLTP(QG)E(LEGD)	MRVNNGLTPQEL	*pckA*	5.46	5.60	59.6	58.8	0.10	20
R57	DDNNTLYFYNXT	MKKWSRHLLAAGALALGMSAA HA*DDNNTLYFYNWT	*potD*	4.69	4.73	36.5	37.3	0.85	380
R105	DDNNTxYFYNt	MKKWSRHLLAAGALALGMSAA HA*DDNNTLYFYN	*potD*	4.69	4.62	36.5	37.8	0.26	40
R96	MVTFHTNHGDIV	MVTFHTNHGDIV	*ppiB*	5.64	5.73	18.2	17.0	0.34	60
R79	(ES)ASLTGAGATFP	MKVMRTTVATVVAATLSMSAFSVFA*E ASLTGAGATFP	*pstS*	7.51	7.30	34.4	37.8	0.26	20
R55	KDTIALVVSTLN	MNMKKLATLVSAVALSATVSANA MA*KDTIALVVSTLN	*rbsB*	6.31	6.30	28.5	28.8	2.80	1520
R123	SYTLPSLPyaYD	M*SYTLPSLPYAYD	*sodA*	6.98	6.85	23.0	24.6	0.19	20
R68-2	A(QF) (Yk) (EL) (DP) (GA) (XKL) (QP) (YX) (TA) TX) (Ld)	M*SFELPALPYAKD	*sodB*	5.88	5.49	21.1	23.4	0.05	20
R76	(SG) (FG)ELPALPYAKD	M*SFELPALPYAKD	*sodB*	5.88	5.68	21.1	22.8	1.14	900
R106	STTXFvGA(Da)DXA	M*STTWFVGADWLA	*sseA*	4.40	4.45	30.7	32.9	0.14	20
R115	AGQQQPLPVPPL	MSLSRRQFIQASGIALCAGAVPLKA SA*AGQQQPLPVPPL	*suffI*	5.52	5.68	49.1	49.9	0.08	20
R74	APQVVDKVAAVV	MKNWKTLLLGIAMIANTSFA*APQVVDK VAAVV	*surA*	6.54	6.39	45.1	47.4	0.39	160
R89	TDKLTSLrQxTT	M*TDKLTSLRQYTT	*talB*	4.94	5.21	35.1	34.3	0.29	100
R101-1	(Tsa) (DE) (KGV) (LX) (TI) (SG) (LM) (RAD) (QI) (DYPf) (TD) (TL)	M*TDKLTSLRQYTT	*talB*	4.94	5.08	35.1	36.5	0.25	60
R86	(Ms)ENFKHLPEPF	MENFKHLPEPF	*tnaA*	6.16	6.14	52.8	49.9	0.16	40
R72	MYHPLVMGNXKL	MRHPLVMGNWKL	*tpiA*	5.89	5.78	27.0	27.3	0.15	60
R63	SQTVHFQGNPVT	M*SQTVHFQGNPVT	*tpx*	4.59	4.87	17.8	19.0	0.48	160
R91	YEQDKTYKITVL	MKLLQRGVALALLTTFTLASETALA*Y EQDKTYKITVL	*ushA*	5.37	5.50	58.2	60.6	0.13	60
R113	(YA)EQDKT(YT)KITVL	MKLLQRGVALALLTTFTLASETALA*Y EQDKTYKITVL	*ushA*	5.37	5.20	58.2	30.4	0.09	20
R101-2	(Tsa) (DE) (KGV) (LX) (TI) (SG) (LM) (RAD) (QI) (DYPf) (TD) (TL)	MKIKNILLTLCTSLLLTNVAAHA*KEV KIGMAIDDL	*xylF*	4.93	5.08	33.3	36.5	0.06	20
R103-1	(KAs) (ED) (VT) (KX) (IT) (GX) (MV) (ATr) (IX) (DY) (DX) (Ln)	MKIKNILLTLCTSLLLTNVAAHA*KEV KIGMAIDDL	*xylF*	4.93	5.09	33.3	35.0	0.20	100
R116	AERPTLPIPtLL	MQRRDFLKYSVALGVASALPLWSRAV FA*AERPTLPIPDLL	*yacK*	6.55	6.29	53.4	53.9	0.11	40
R78	(MQ)YKTIIMPVDVF	MYKTIIMPVDVF	*ybdO*	6.51	6.38	15.9	11.3	0.63	120
R70	VTYPLPTDGSXL	MNMKLKTLFAAAFAVVGFCSTASA*V TYPLPTDGSRL	*ybiS*	5.74	5.79	30.9	31.2	0.29	160
R99	ADYKIDKFGQHA	MKKSLLGLTFASLMFSAGSAVA*ADY KIDKEGQHA	*yceI*	5.12	5.19	18.7	21.0	0.14	40
R109	ADSDIA(Dn)GQT(QTP)R	MDRRRFIKGSMAMAAVCGTSGIASLFS QAAFA*ADSDIADGQTQR	*ydcG*	5.83	5.76	59.4	56.1	0.09	40
R90	(Ma)IKKIFALPVIE	MKLKDCV*MIKKIFALPVIE	*yeaD*	6.32	6.22	32.7	32.8	0.14	20
R69	AVVASLXPVGFI	MKCYNITLLIFITIIGRIMLHKKTLL FAALSAALWGGATQAADA*AV VASLKPVGFI	*yebL*	5.58	5.71	31.2	32.2	0.19	60
R104	AFTPFPDRQPTA	M*AFTPFPPRQPTA	*ygfZ*	5.01	5.22	36.0	37.4	0.12	40
R98	VTGDTDQPIHiE	MKFKTNKLSLNLVLASSLLAASIPA FA*VTGDTDQPIHIE	*yhbN*	7.77	5.23	17.3	14.9	0.25	60
R93	(STDG)ALQPDPAXQQg	MQGTKIRLLAGGLLMMATAGYVQA* DALQPDPAWQQG	*yhjJ*	5.43	5.57	53.0	52.9	0.14	40
R118	(STr)VDESSDNNsLL	(228)A*RVDESSDNNSLL(245)	*yhjW*	7.47	6.23	41.3	40.4	0.07	20
R120	SsPSPLNPGTNV	MRKITQAISAVCLLFALNSSAVA LA*SSPSPLNPGTNV	*yibP*	9.05	6.27	19.2	27.1	0.14	20

Table 13. continued

		Predicted *versus* observed sequence, p*I*, MW						Abundance	
Spot ID	*N*-terminal sequence tag	Sequence tag match to genomic ORF	Locus[b]	cp*I*	op*I*	MW	M_r	O-abd	N-abd
R121-1	(SL)A(PS)SP(SL) (PN) (LP)N(TP) (GN)T	MRKITQAISAVCLLFALNSSAVA* LASSPSPLNPGT	*yibP*	9.05	6.31	19.2	27.2	0.05	20
R97-1	(MA) (NI) (RK) (TP) (IS) (LF) (VS) (PE) (IL) (DX) (IX) (SN)	MNSVITQKVSSGVT LYADTKTGGF*MNRTILVPIDIS	*ynaF*	5.89	5.88	16.0	12.1	0.27	80
R112	(ASGt)E(PK) (Eqr)M(TG)IGAIYL	MPTKMRTTRNLLLMATLLGSALFARA*A EKEMTIGAIYL	*yphF*	5.03	5.15	32.2	31.9	0.08	20
R122	HQFETGQRVPPI	MTLRKILALTCLLLPMMASA*HQ FETGQRVPPI	*ytfJ*	5.68	6.42	18.2	19.7	0.33	40
R88	APLTVGFSQVGS	MWKRLLIVSAVSAAMSSMA LA*APLTVGFSQVGS	*ytfQ*	5.68	6.08	32.1	32.2	0.49	40
R56-2	(KA) (IN) (ES) (EN) (GX) (KX) (LI) (VX) (IX) (XA) (IX) (Nr)		NM		5.25		40.9	0.30	240
R97-2	(MA) (NI) (RK) (TP) (IS) (LF) (VS) (PE) (IL) (DX) (IX) (SN)		NM		5.88		12.1	0.03	20
R121-2	(SL) A (PS) SP (SL) (PN) (LP) N (TP) (GN) T		NM		6.31		27.2	0.04	20
R111	XXXXXXXXXXXX				5.07		31.2	0.09	

a) See Table 6 for explanation of column headings.
b) An underlined locus indicates the conceptual protein does not have a recognized signal peptide, which is usually associated with proteins exported across the inner membrane.

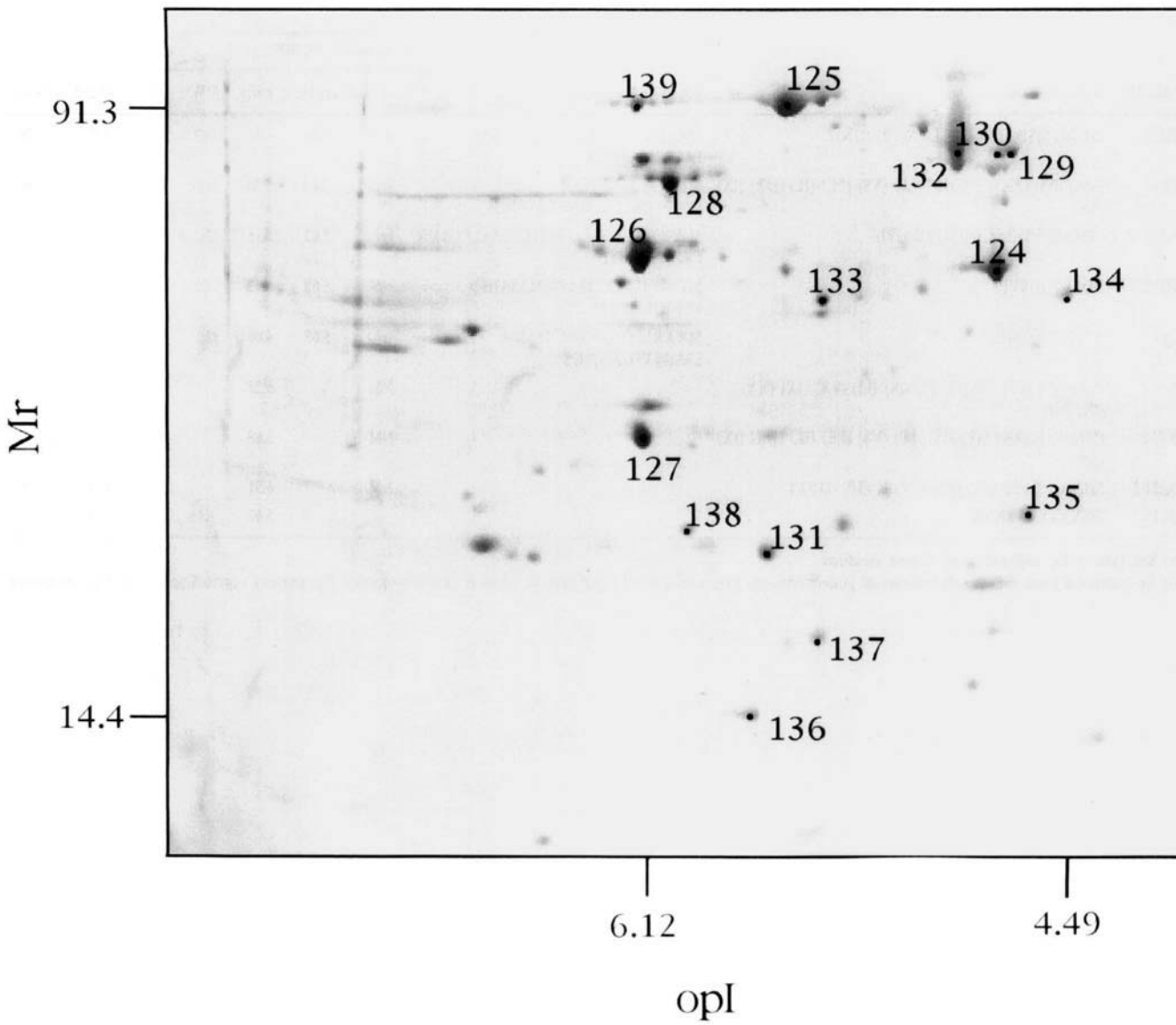

Figure 14. Spots R124–R139: Master gel of inner membrane protein extract fractionated by 2-DE, electrotransferred to PVDF, and stained with Coomassie Blue. The subcellular fractionated protein extract was prepared from EMG2 grown in rich media to early stationary phase. The IEF used a 4:1 carrier ampholyte blend of Serva 5–7: Serva 3–10, and the SDS-PAGE used an 11.5%T, 2.7%C gel. The largest and smallest op*I* and M_r of the identified 2-DE spots from the gel are shown on the op*I* and M_r axes. The data for spots R124–R139 are found in Table 14.

Table 14. Inner-membrane-enriched *E. coli* proteins from cells at stationary phase in rich media[a)]

		Predicted *versus* observed sequence, p*I*, MW						Abundance	
Spot ID	*N*-terminal sequence tag	Sequence tag match to genomic ORF	Locus[b)]	cp*I*	op*I*	MW	M_r	O-abd	N-abd
R125	SERFPNDVDPIE	M*SERFPNDVPIE	*aceE*	5.50	5.49	99.5	91.3	0.27	400
R132	AIEIKVPDIGAD	M*AIEIKVPDIGAD	*aceF*	4.92	4.86	66.0	73.9	0.27	1120
R126-2	(MS) (QT) (EL) (IN) (KS) T (QE) (IV) (VS) (VE) L (IG)	MQLNSTEISELI	*atpA*	6.01	6.12	55.2	49.0	0.21	1880
R124	ATGKIVQVIGAV	M*ATGKIVQVIGAV	*atpD*	4.74	4.73	50.2	46.7	0.28	4040
R136	XXXXXXXXXXXX		*atpF*	6.11	5.64	17.3	14.4	2.83	
R129	GKIIGIDLGTTN	M*GKIIGIDLGTTN	*dnaK*	4.67	4.67	69.0	73.9	0.04	140
R130-2	(MG) (TK) (EI) (SI) (GF) (AI) (QD) (LX) (GF) (ET) (TE) (NS)	M*GKIIGIDLGTTN	*dnaK*	4.67	4.71	69.0	73.5	0.01	20
R133	SKIVKIIGr(ER)EI	M*SKIVKIIGREI	*eno*	5.23	5.36	45.5	42.4	0.06	120
R134-1	(MG) (TG) (EQ) QE (TK) TSaAX (DE)	MTEQEKTSAVVE	*hemX*	4.49	4.49	43.0	42.9	0.02	40
R126-1	(MS) (QT) (EL) (IN) (KS) T (QE) (IV) (VS) (VE) L (IG)	MM*STEIKTQVVVLG	*lpdA*	6.10	6.12	50.6	49.0	0.21	1880
R127	(Ts) IAIVIGTHGXA	M*TIAIVIGTHGWA	*manX*	5.87	6.12	34.9	32.1	0.42	2880
R131	MDYTLTRIDPNG	MDYTLTRIDPNG	*nuoB*	5.49	5.58	25.1	27.9	0.28	520
R139	ATIHVDGKEY(YE)V	M*ATIHVDGKEYEV	*nuoG*	6.21	6.14	100.2	90.7	0.06	60
R137	MTLXELLVGFGT	MTLKELLVGFGT	*nuoI*	5.25	5.38	20.5	23.1	0.24	60
R130-1	(MG) (TK) (EI) (SI) (GF) (AI) (QD) (LX) (GF) (ET) (TE) (NS)	MTESFAQLFEES	*rpsA*	4.71	4.71	61.2	73.5	0.02	60
R128	MKLPVREFDAVV	MKLPVREFDAVV	*sdhA*	6.22	5.99	64.4	65.1	0.14	1360
R138	ATLTAKNLAXAY	M*ATLTAKNLAKAY	*yhbG*	5.89	5.92	26.7	28.8	0.08	120
R134-2	(MG) (TG) (EQ) QE (TK) TSaAX (DE)		NM		4.49		42.9	0.03	40
R135	XXXXXXXXXXXX				4.62		29.3	0.16	

a) See Table 6 for explanation of column headings.
b) An underlined locus indicates the conceptual protein is predicted to localize in the cytoplasm.

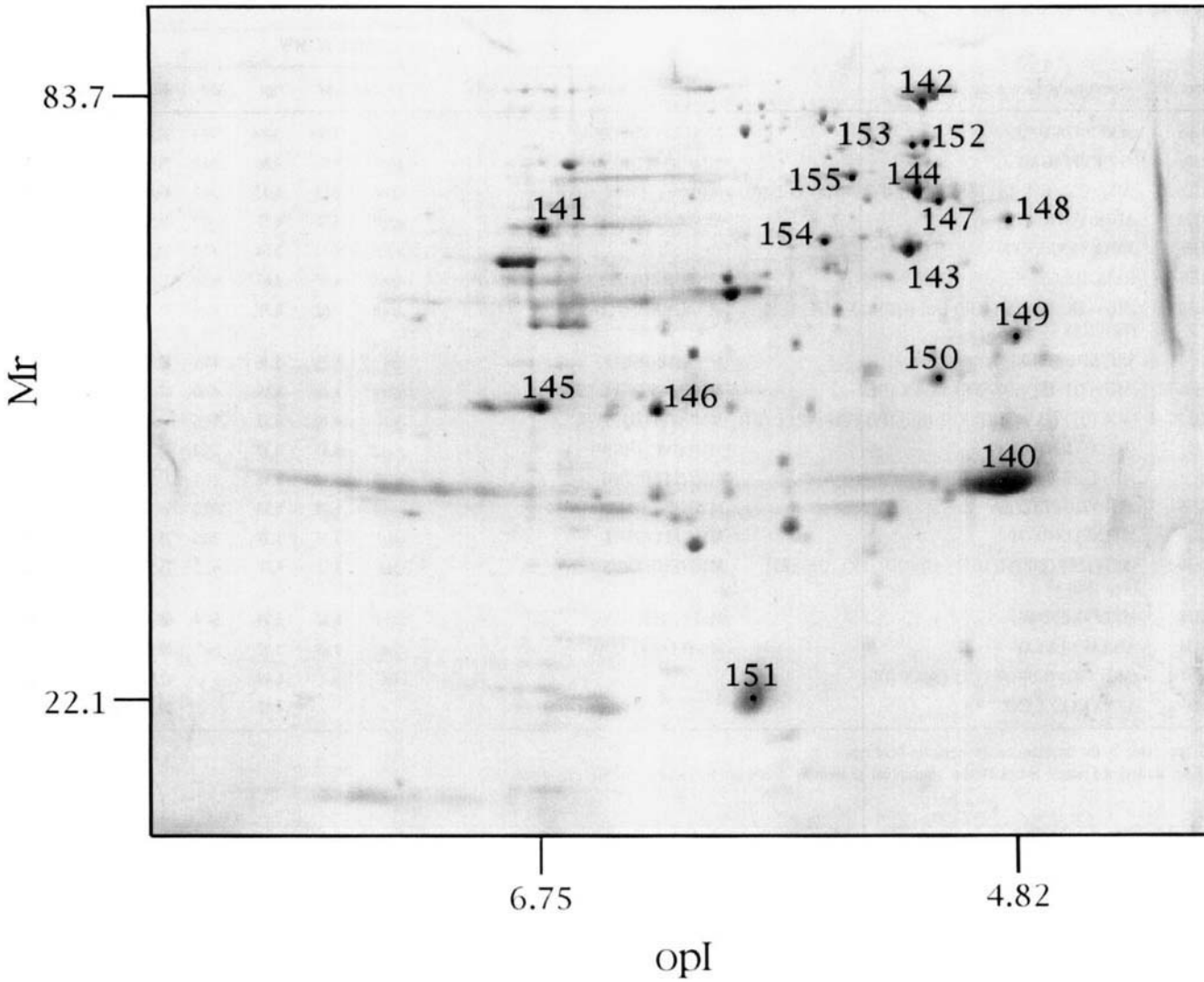

Figure 15. Spots R140–155: Master gel of outer membrane protein extract fractionated by 2-DE, electrotransferred to PVDF, and stained with Coomassie Blue. The subcellular fractionated protein extract was prepared from EMG2 grown in rich media to early stationary phase. The IEF used a 4:1 carrier ampholyte blend of Serva 5–7: Serva 3–10, and the SDS-PAGE used an 11.5%T, 2.7%C gel. The largest and smallest op*I* and M_r of the identified 2-DE spots from the gel are shown on the op*I* and M_r axes. The data for spots R140–R155 are found in Table 15.

Table 15. Outer-membrane-enriched *E. coli* proteins from cells at stationary phase in rich media[a)]

		Predicted *versus* observed sequence, p*I*, MW						Abundance	
Spot ID	*N*-terminal sequence tag	Sequence tag match to genomic ORF	Locus[b)]	cp*I*	op*I*	MW	M_r	O-abd	N-abd
R141	gQLNSTEISE(LG)I	**MQLNSTEISELI**	*atpA*	6.01	6.74	55.2	57.5	0.11	120
R143	ATGLIVQVIGAV	M*ATGKIVQVIGAV	*atpD*	4.74	5.19	50.2	54.5	0.14	780
R155	QDTSPDTLvVT(TA)	MIKKASLLTACSVTAFSAWA* **QDTSPDTLVVTA**	*btuB*	5.01	5.41	66.3	66.9	0.04	100
R152	GKIIGIDLGTtn	M*GKIIGIDLGTTN	*dnaK*	4.67	5.14	60.0	73.8	0.04	60
R145	TIAIVIGTHGXA	M*TIAIVIGTHGWA	*manX*	5.87	6.75	34.9	37.1	0.22	260
R144	AAKDVLFGNDAR	**M*AAKDVKFGNDAR**	*mopA*	4.67	5.17	57.2	64.2	0.11	480
R140	APKDNTXYTGA	MKKTAIAIAVALAGFATVAQA* **APKDNTWYTGA**	*ompA*	5.74	4.86	35.2	31.7	1.38	2500
R146	APKDNTXYTGG	KKTAIAIAVALAGFATVAQA* **APKDNTWTGA**	*ompA*	5.74	6.21	35.2	36.8	0.14	500
R151-1	XXXXXXXXXXXX		*pal*	6.79	5.81	18.8	22.1	1.04	
R153	(GA) TES (FG) AQLFEE	**MTESFAQLFEE**	*rpsA*	4.71	5.19	61.2	73.5	0.03	40
R151-2	XXXXXXXXXXXX		*slp*	6.79	5.81	22.2	22.1	1.04	
R154	ENLMQVYQQA (AR) (LR)	MKKLLPILIGLSLSGFSSLSQA* **ENLMQVYQQARL**	*tolC*	5.09	5.51	51.5	55.7	0.05	240
R142	AEGFVVKD(ID)HFE	MAMKKLLIASLLFSSATVYG* **AEGFVVKDIHFE**	*yaeT*	4.72	5.17	88.4	83.7	0.10	220
R148	KIRFEGNDTSXD	(350)R*KIRFEGNDTSKD(448)	*yaeT*	4.42	4.85	51.6	59.0	0.04	100
R151-3	XXXXXXXXXXXX		*yajG*	9.70	5.81	25.0	22.1	1.04	
R147	ADIVVHPGTTT	(104)A*ADIVVHPGETV(976)	*yeeQ*	5.44	5.09	101.3	62.5	0.07	260
R149	XXXXXXXXXXXX				4.82		43.4	0.11	
R150	XXXXXXXXXXXX				5.09		39.4	0.10	

a) See Table 6 for explanation of column headings.
b) An underlined locus indicates the conceptual protein does not have a recognized signal peptide, which is usually associated with proteins exported across the inner membrane.