The Proteome of *Shigella flexneri* 2a 2457T Grown at 30 and 37 °C*[S]

Li Zhu‡§, Ge Zhao‡§¶, Robert Stein‖, Xuexue Zheng‡, Wei Hu‡, Na Shang‡, Xin Bu‡, Xiankai Liu‡, Jie Wang, Erling Feng‡, Bin Wang‡‡, Xuemin Zhang**, Qinong Ye‡, Peitang Huang‡, Ming Zeng‡‡§§, and Hengliang Wang‡¶¶**

To upgrade the proteome reference map of *Shigella flexneri* 2a 2457T, the protein expression profiles of log phase and stationary phase cells grown at 30 and 37 °C were thoroughly analyzed using multiple overlapping narrow pH range (between pH 4.0 and 11.0) two-dimensional gel electrophoresis. A total of 723 spots representing 574 protein entries were identified by MALDI-TOF/TOF MS, including the majority of known key virulence factors. 64 hypothetical proteins and six misannotated proteins were also experimentally identified. A comparison between the four proteome maps showed that most of the virulence-related proteins were up-regulated at 37 °C, and the differences were more notable in stationary phase cells, suggesting that the expressions of these virulence factors were not only controlled by temperature but also controlled by the nutrients available in the environment. The expression patterns of some virulence-related genes under the four different conditions suggested that they might also be regulated at the post-transcriptional level. A further significant finding was that the expression of the protein ArgT was dramatically up-regulated at 30 °C. The results of semiquantitative RT-PCR analysis showed that expression of *argT* was not regulated at the transcriptional level. Therefore, we carried out a series of experiments to uncover the mechanism regulating ArgT levels and found that the differential expression of ArgT was due to its degradation by a periplasmic protease, HtrA, whose activity, but not its synthesis, was affected by temperature. The cleavage site in ArgT was between position 160 (Val) and position 161 (Ala). These results may provide useful insights for understanding the physiology and pathogenesis of *S. flexneri*. *Molecular & Cellular Proteomics 9:1209–1220, 2010.*

Shigella flexneri is a Gram-negative facultative pathogen that causes the majority of communicable bacterial dysenteries in developing countries. Seven years ago, the whole genome sequence of *S. flexneri* 2a strain 2457T was determined (1), and a proteome reference map was also constructed at the same time (2). Predicting the pI of all proteins encoded by the bacterial genome revealed that about 45.54% (1972 proteins) were basic. However, all of these basic proteins and the vast majority of virulence-related proteins were missed in the old proteome reference map because of the limitations of the experimental conditions used at that time. Now, the proteomics platform of two-dimensional polyacrylamide gel electrophoresis (2-DE)[1] has been optimized and further developed. The resolution and sensitivity of today's 2-DE technique has been greatly improved. As a result, the expression profiles obtained today are difficult to compare with old profiles. An upgrade of the reference proteome map is urgently required for further proteomics research of *S. flexneri*, particularly for comparative studies.

The virulence genes of *Shigella flexneri* are located on a large (~230-kb) plasmid acquired by horizontal transfer and encode a type III secretion system apparatus, substrates of this apparatus, and their dedicated chaperones. This enables *S. flexneri* to invade epithelial cells in the lower gut of humans (3, 4). The expression of virulence genes is induced under growth conditions similar to those found at the site of invasion. For example, a temperature of 37 °C is a particularly important environmental signal. Maurelli *et al.* (5) found that *S. flexneri* cultivated at 37 °C produced keratoconjunctivitis in guinea pigs and was able to penetrate and replicate in intestinal epithelial cells, but bacteria grown at 30 °C were phenotypically avirulent and non-invasive. The mechanism regulating virulence gene expression in *S. flexneri* has been well studied (for a review, see Ref. 6). A living cell is a highly complex entity, all parts of which are involved in its normal functions. We can expect that a large number of genes, other than virulence genes, will be simultaneously differentially expressed in response to a change in environmental temperature. To obtain a global view on protein expression level changes in *S. flexneri* induced by temperature up-shift, we compared the protein expression profiles of log phase and

From the ‡Beijing Institute of Biotechnology, State Key Laboratory of Pathogen and Biosecurity, 100071 Beijing, China, ‖Informatics and Biology, D-12169 Berlin, Germany, **National Center of Biomedical Analysis, 100850 Beijing, China, ‡‡National Institute for the Control of Pharmaceutical and Biological Products, 100050 Beijing, China, and ¶Shandong Eye Institute, 266071 Qingdao, China

Received, September 23, 2009, and in revised form, February 10, 2010

Published, MCP Papers in Press, February 17, 2010, DOI 10.1074/mcp.M900446-MCP200

[1] The abbreviations used are: 2-DE, two-dimensional polyacrylamide gel electrophoresis; CAI, codon adaptation index; PMF, peptide mass fingerprinting; LB, Luria-Bertani; ACTH, adrenocorticotropic hormone; NCBInr, National Center for Biotechnology Information non-redundant; IPTG, isopropyl 1-thio-β-D-galactopyranoside; COG, clusters of orthologous groups; ID, identification number.

stationary phase cells grown at 30 and 37 °C. In agreement with previous reports, most up-regulated genes at 37 °C are known virulence-related genes and heat shock proteins, and the differences were more notable in stationary phase cells. At the same time, we found that the level of the protein ArgT is dramatically increased at 30 °C.

EXPERIMENTAL PROCEDURES

Bacterial Strains and Growth Conditions—*Escherichia coli* strain DH5α, used for plasmid constructions, was grown in Luria-Bertani (LB) agar or LB broth (Difco). Wild-type *S. flexneri* 2a strain 2457T was grown in tryptic soy agar (Difco) containing 0.01% Congo red or in LB broth. When necessary, antibiotics were added at the following concentrations: ampicillin, 100 μg/ml; kanamycin, 50 μg/ml; and chloramphenicol, 30 μg/ml.

Preparation of Whole-cell Protein Extracts—The virulent *S. flexneri* serotype 2a strain 2457T was grown aerobically in 100 ml of LB medium at 37 °C. Cells were harvested in the log phase ($A_{600\ nm}$ = 1.0) and stationary phase (after about 9 h of culture). The preparation of whole-cell protein extracts was performed as described previously (7). The protein concentration of samples was measured using the PlusOne 2-D Quant kit (GE Healthcare), and 0.8-mg aliquots were stored at −80 °C.

2-DE and In-gel Protein Digestion—To obtain better separation, pH 4.0–5.0, pH 4.5–5.5, pH 5.0–6.0, pH 5.5–6.7, and pH 6–11 IPG strips (18 cm; GE Healthcare) were used in the IEF analysis. For each analysis, the samples were treated with the 2-D Clean-Up kit (GE Healthcare) according to the kit instructions and resuspended in 350 μl of rehydration buffer (7 M urea, 2 M thiourea, 4% (w/v) CHAPS, 50 mM DTT, 0.5% IPG Buffer (same pH range of the IPG strip)). The following 2-DE procedure and the in-gel protein digestion were carried out as described previously (7). Briefly, an 800-μg protein sample was used to rehydrate an 18-cm immobilized gradient strip for 12 h at 20 °C. Focusing started at 300 V (1 h), was increased to 1000 V (1 h), and was further increased to 8000 V (12 h). After focusing, the strips were equilibrated for 15 min in equilibrium buffer (2% SDS, 50 mM Tris-HCl, pH 8.8, 6 M urea, 30% glycerol, 0.002% bromphenol blue, 1% DTT). The strips were then overlaid onto 12.5% SDS-polyacrylamide gels for the second dimensional separation. The protein spots were carefully excised from the Coomassie-stained 2-DE gel, destained, washed, and then digested for 13 h with sequencing grade modified trypsin (Roche Applied Science). Peptides from digested proteins were used for MALDI-TOF/TOF analysis.

Image analysis was processed by ImageMaster 2D Platinum software (GE Healthcare). To facilitate the discrimination between real spots and artifacts, the spot detection parameters were adjusted as follows: smooth, 3; minimum area, 50; and saliency, 6. The images of two samples under different cultivation conditions were compared. The relative volume of each spot was determined from the spot intensities in pixel units and normalized to the sum of the intensities of all the spots on the gel.

MALDI-TOF/TOF MS—MALDI-TOF/TOF MS measurements were performed on a Bruker Ultraflex™ III MALDI-TOF/TOF MS (Bruker Daltonics) operating in reflectron mode with 20-kV accelerating voltage and 23-kV reflecting voltage. A saturated solution of α-cyano-4-hydroxycinnamic acid in 50% acetonitrile and 0.1% trifluoroacetic acid was used as the matrix. One microliter of the matrix solution and sample solution at a ratio of 1:1 were applied onto the Score384 target well. Routinely, a standard peptide calibration mixture in the mass range 800–3200 Da (Bruker Daltonics) was analyzed for external calibration of the mass spectrometer. The calibration mixture contained angiotensin II, angiotensin I, substance P, bombesin, ACTH clip 1–17, ACTH clip 18–39, and somatostatin 28. A series of eight samples was spotted around one external calibration mixture. The SNAP algorithm (signal to noise ratio threshold, 5; quality factor threshold, 30) in FlexAnalysis™ 3.4 was used to pick up the 150 most prominent peaks in the mass range *m/z* 700–4000. The subsequent MS/MS analysis was performed in a data-dependent manner, and the five most abundant ions fulfilling certain preset criteria (signal to noise ratio higher than 25 and quality factor higher than 50) were subjected to high energy CID analysis. The collision energy was set to 1 keV, and nitrogen was used as the collision gas.

Data Interpretation and Database Searching—Peptide mass fingerprints (PMFs) were searched by the program Mascot 2.1 (Matrix Science Ltd.) against the database of *S. flexneri* 2a 2457T (4540 entries, including all of the ORFs predicted on the chromosome of *S. flexneri* 2a 2457T (GenBank™ accession number GI:30043918), the virulence plasmid pCP301 (GenBank accession number GI:18462515) from *S. flexneri* 2a 301, and the large IncHI plasmid R27 (GenBank accession number GI:7800243) from *Salmonella typhi*) to eliminate redundancy resulting from multiple members of the same protein family, and the results were checked against the NCBInr database (version October 21, 2006, 4,072,503 sequences). For those proteins identified in the NCBInr database, the proteins of *S. flexneri* spp. were selected as the best hits from the homologue protein lists. The search parameters were as following: trypsin digestion with one missed cleavage; carbamidomethyl modification of cysteine as a fixed modification and oxidation of methionine as a variable modification; peptide tolerance maximum, ±0.2 Da; MS/MS tolerance maximum, ±0.6 Da; peptide charge, 1+; monoisotopic mass. Scores greater than 49 were significant ($p < 0.05$) for a local PMF search. For unambiguous identification of proteins, more than five peptides must be matched for a PMF search.

RNA Extraction and Semiquantitative RT-PCR—Under different growth conditions, the relative abundance of *argT* mRNA was determined by semiquantitative RT-PCR. Total RNA was extracted from 2457T cultivated at 30 and 37 °C using the RNeasy Mini kit (Qiagen) and treated with the RNase-free DNase set (Qiagen). RNA concentration and purity were evaluated by spectrophotometric analysis at 260 and 280 nm. After digestion by DNase I to eliminate contaminating DNA, reverse transcription was carried out with the reverse SuperScript Choice system (Invitrogen) using 2 μg of total RNA as the template. Specific target RT-PCR products were normalized to an established endogenous control, 16 S ribosomal RNA, the expression of which is relatively constant in bacteria (8). Negative controls were performed using RNA as template to confirm the absence of contaminating DNA in the RNA preparations.

Construction of Plasmids and Strains—*argT* genes of *S. flexneri* were amplified from the 2457T genome by PCR using primer a1 and primer a2. The PCR products were cloned into pProEX-HTb (Invitrogen, Ap^r) and pGEX-6p-2 (Amersham Biosciences, Ap^r) via BamHI and EcoRI sites, generating pProEX-HTb-argT (pProEX-argT) and pGEX-6p-2-argT (pGEX-argT), respectively. Similarly, primer a3 and primer a2 were used to amplify 2457T *argT* without signal peptide, and primer h1 and primer h2 were used to amplify 2457T *htrA*. The PCR products were cloned into pGEX-6p-2 via BamHI and EcoRI sites, generating pGEX-6p-2-argT2 (pGEX-argT2) and pGEX-6p-2-htrA^{2457T} (pGEX-htrA), respectively.

To introduce a point mutation in *htrA*, PCR was performed using *Pfu* DNA polymerase (TaKaRa), primers containing the desired mutation, and the appropriate plasmid as the template according to the QuikChange® site-directed mutagenesis kit (Stratagene). The PCR products were digested by DpnI and then transformed into DH5α competent cells. The desired point mutation was confirmed by sequencing. Primers S210A1 and S210A2 were used to generate pGEX-htrA$_{S210A}$.

An *htrA* deletion mutant of *S. flexneri* (2457TΔ*htrA*) was constructed by a modified form of lambda *red* recombination, originally

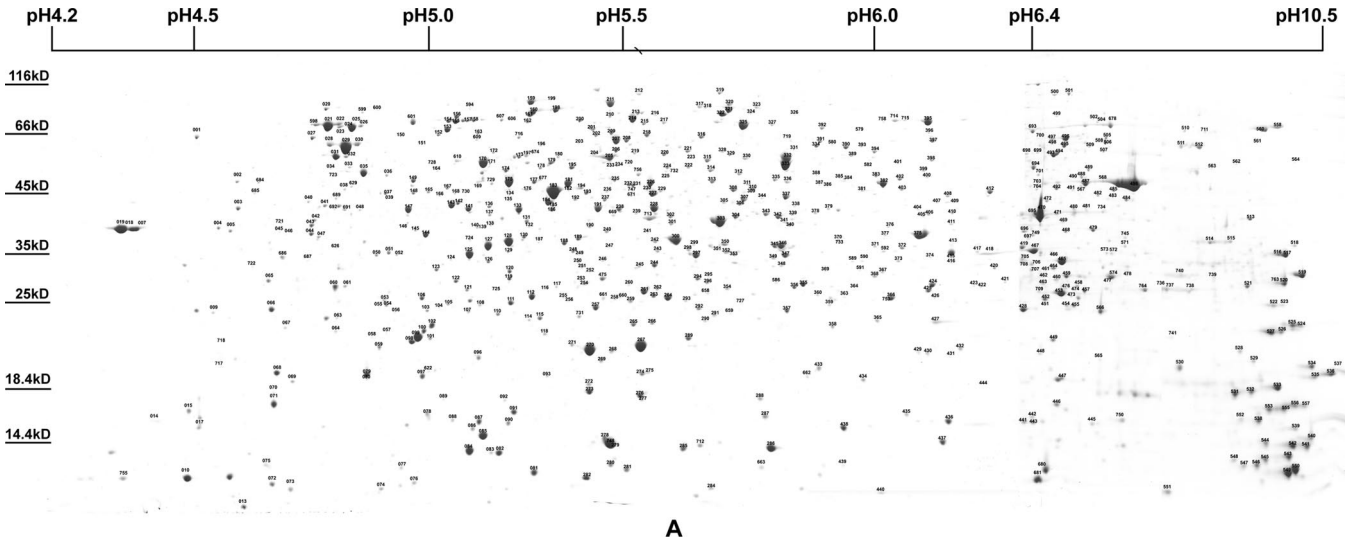

FIG. 1. **Two-dimensional gel electrophoresis and identified spots of whole-cell proteins at 37 °C in stationary phase *S. flexneri*.** The identified spots are labeled on the integrated 2-DE map. The whole set of four maps is shown in supplemental Fig. 1 and can be found at the Proteome 2D-PAGE Database.

described by Datsenko and Wanner (9). Briefly, a 500-bp upstream region and a 500-bp downstream region of *htrA* were independently amplified by PCR using primers h5p1 and h5p2 and primers h3p1 and h3p3, respectively. The resulting products were cloned into pET22b (Novagen) along with *kan* amplified from the plasmid pKD4. *S. flexneri* 2457T, carrying pKD46 (9) containing the *red* recombinase genes, was grown at 30 °C in the presence of 10 mM arabinose (to induce the recombinase genes) and was transformed with gel-purified h5-Kan-h3 PCR product amplified by h5p1 and h3p2 primers from the previously constructed pET22b based vector. Recombinants were selected on LB plates containing kanamycin (50 μg/ml). The primers used in this study are listed in supplemental Table 1.

Periplasmic Protein Extraction—Periplasmic protein was extracted by the improved osmotic shock method (10). About 5 ml of early stationary phase cells were harvested and washed twice with 1× PBS. The washed cell pellet was resuspended in 100 μl of cell lysis buffer (20% sucrose, 30 mM Tris-Cl, pH 8.0, 1 mM EDTA, 1 mg/ml lysozyme). The cell suspension was shaken gently in ice-cold water for 15 min followed by centrifugation at 6000 rpm for 15 min. After that, the recovered supernatant represents the periplasmic proteins. The pellet, representing the remaining whole-cell proteins, was resuspended in 500 μl of water.

Protein Expression and Purification—Expression of fusion proteins was induced by adding 1 mM IPTG when cultures had grown to an A_{600} of 0.6–0.9. After the cell growth continued for another 4 h at 30 *or* 37 °C, an aliquot of 1 ml was harvested. The induced ArgT and HtrA fusion proteins were purified from cultures growing at 20 °C for 16 h using Glutathione-Sepharose 4B beads (GE Healthcare). The proteins were eluted with elution buffer (20 mM glutathione, 100 mM Tris-HCl, pH 8.0, 120 mM NaCl).

Antibody Preparation—To generate polyclonal ArgT antibodies, purified GST-ArgT proteins were injected subcutaneously into two BALB/c female mice. Sera from the immunized mice were collected and purified by affinity chromatography according to the manufacturer's instructions (Pierce).

SDS-PAGE and Western Blot Analysis—All samples were boiled and loaded on 12% acrylamide gels for SDS-PAGE, and after electrophoresis the proteins were stained with Coomassie Blue. When necessary, proteins were transferred to PVDF membranes (Immobilon, Millipore) after electrophoresis. The membranes were blocked with 5% nonfat milk and then incubated for 1 h with primary antibodies raised against ArgT. The membranes were then washed, and horseradish peroxidase-conjugated goat anti-mouse IgG secondary antibodies (Santa Cruz Biotechnology) were added for 1 h. The ECL Plus kit from Pierce was used for detection.

Proteolysis Test—Proteolysis was performed as described previously (11). HtrA (0.5 μg) was incubated with ArgT (50 μg) in 50 mM Tris-Cl, pH 8.0 buffer in a final volume of 100 μl. 15-μl samples were taken every hour and then resolved by SDS-PAGE.

Coomassie-stained ArgT fragments cleaved by HtrA were cut out, and in-gel digestion was performed by trypsin (7) and Glu-C (12) as described previously. Peptides from digested fragments were resolved in 2 μl of 0.5% trifluoroacetic acid. Mass spectrometry was performed on a Bruker Ultraflex III MALDI-TOF/TOF mass spectrometer (Bruker Daltonics) according to the procedures described above.

RESULTS AND DISCUSSION

Proteome Reference Map of S. flexneri 2a 2457T

Two-dimensional Gel Electrophoresis and Protein Identification

Because bacteria only express a limited portion of their proteome under any one condition (13), we analyzed the proteome of *S. flexneri* at two different temperatures and at two different growth phases to obtain better proteome coverage. According to our previous report (2), most of the whole-cell proteins were scattered in the pI ranges of pH 4–7. So, variable IPG strips of pH 4.0–5.0, pH 4.5–5.5, pH 5.0–6.0, and pH 5.5–6.7 were used to resolve protein in the densely populated pH 4.0–7.0 zone. These basic proteins were also analyzed using IPG strips of pH 6.0–11.0. The separate graphs of the above different pH gradient range gels were merged to produce a single artificial gel map of pH 4.0–11.0. The two artificial maps of log phase cells displayed more than 1500 Coomassie-stained protein spots, whereas those of stationary phase cells only displayed about 1000 spots (Fig. 1 and

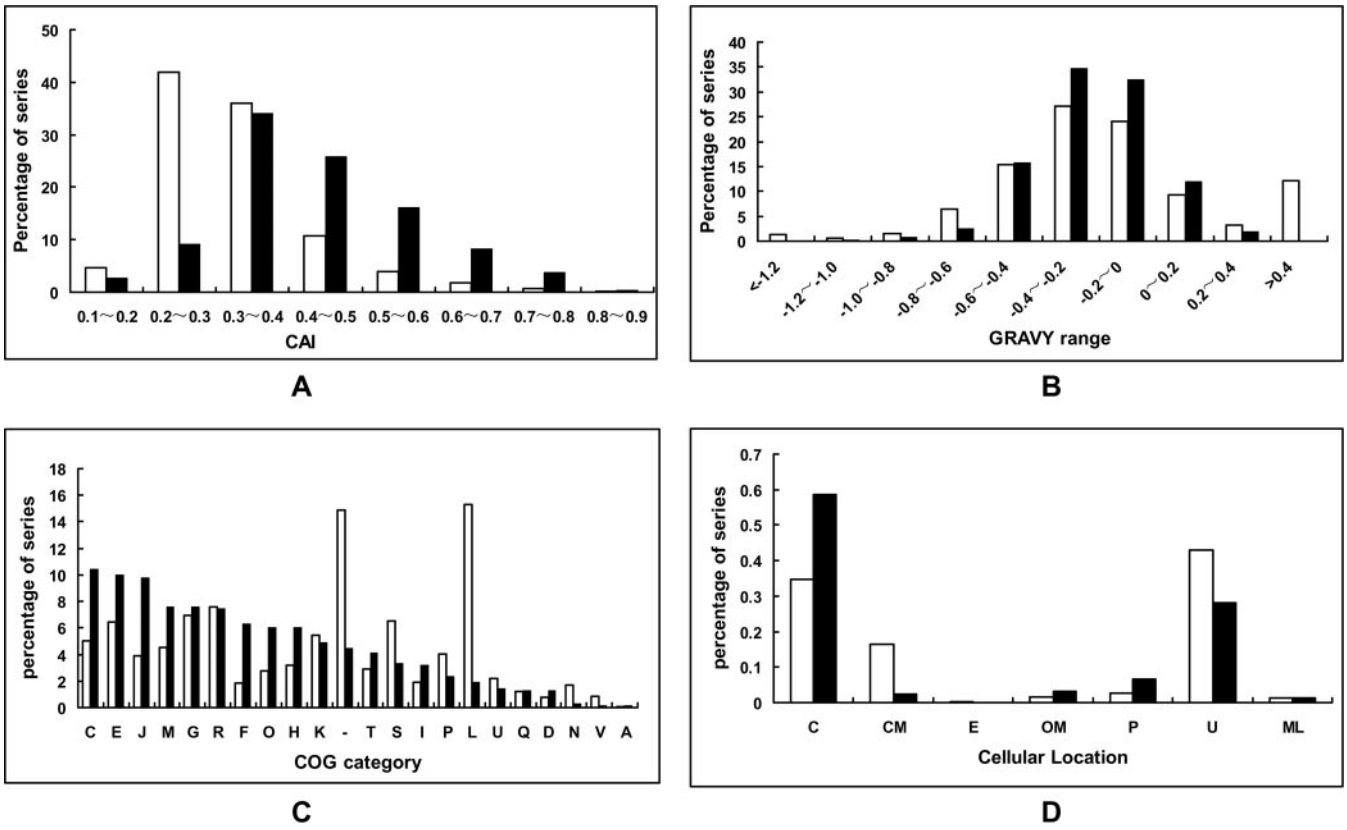

FIG. 2. **Distribution of proteins according to CAI (*A*), GRAVY value (*B*), cellular role categories (*C*), and localization (*D*).** The whole set of genes encoding proteins (*open bars*) and identified proteins (*black bars*) was compared. The single letter codes in *C* represent a particular functional category as follows: *J*, translation; *A*, RNA processing and modification; *K*, transcription; *L*, replication, recombination, and repair; *D*, cell cycle control, mitosis, and meiosis; *V*, defense mechanisms; *T*, signal transduction mechanisms; *M*, cell wall/membrane biogenesis; *N*, cell motility; *U*, intracellular trafficking and secretion; *O*, post-translational modification, protein turnover, chaperones; *C*, energy production and conversion; *G*, carbohydrate transport and metabolism; *E*, amino acid transport and metabolism; *F*, nucleotide transport and metabolism; *H*, coenzyme transport and metabolism; *I*, lipid transport and metabolism; *P*, inorganic ion transport and metabolism; *Q*, secondary metabolite biosynthesis, transport, and catabolism; *R*, general function prediction only; *S*, function unknown; _, not in COG. The meanings of abbreviations in *D* are as follows: *C*, cytoplasmic; *CM*, cytoplasmic membrane; *E*, extracellular; *OM*, outer membrane; *P*, periplasmic; *U*, unknown; *ML*, multiple localization sites.

supplemental Fig. 1). This indicated that the various metabolisms of cells were more active in the log phase. After destaining and in-gel trypsin digestion, a total of 723 spots representing 574 proteins were successfully identified by MALDI-TOF/TOF MS (supplemental Table 2). Maps and information for all identified proteins are available at the Proteome 2D-PAGE Database.

Predicted and Actual Proteome of S. flexneri 2a 2457T

The theoretical 2-DE map was constructed according to the genome annotation of *S. flexneri* 2a 2457T and the virulence plasmid pCP301 and compared with the identified maps. As shown in supplemental Fig. 2, the actual 2-DE map displayed a bimodal protein distribution pattern similar to the predicted map. The identified basic proteins (13.74%) were less abundant than predicted (45.54%). This is a common scenario in other bacteria (14). The possible explanations are that the basic proteins are inherently difficult to separate by 2-DE or that their levels are very low. Isoelectric point values and molecular weight values estimated by gel electrophoresis matched closely with predicted values, although some discrepancies were seen (supplemental Fig. 3). The differences of isoelectric point or molecular weight values suggested that these proteins had undergone post-translational modifications. The differences in molecular weight values are mainly due to the cleavage of signal peptides or other functional sequences.

Distribution of Proteins According to CAI, GRAVY Value, Cellular Role Categories, and Localization

CAI estimates the degree of synonymous codon adaptation in a coding region compared with the optimal usage (15). The CAI distributions of genes coding for the proteins identified were compared with those of all predicted proteins (Fig. 2). The proteins encoded by genes with a CAI value >0.4 account for 17.45% of the total proteins but 54.37% of the

TABLE I
Most abundant proteins (volume % >0.6) common to four different conditions

Spot ID	GI	Gene	Synonym	Score[a]	Sequence coverage	Molecular weight	pI	CAI	ENc[b]	COG[c]	Product
					%						
024	30040692	*rpsA*	S0971	587	47	61,235	4.89	0.781	29.24	J	30 S ribosomal subunit protein S1
029	30043782	*mopA*	S4564	806	65	57,464	4.85	0.792	27.9	O	Heat shock protein GroEL, chaperone Hsp60, peptide-dependent ATPase
099	30040307	*ahpC*	S0529	425	78	20,862	5.03	0.796	29.94	O	Alkyl hydroperoxide reductase, C22 subunit
127	30039969	*tsf*	S0163	296	71	30,518	5.22	0.766	31.95	J	Protein chain elongation factor EF-Ts
128	30040740	*ompA*	S1023	580	60	37,374	5.65	0.761	31.77	M	Outer membrane protein 3a (II*;G;d)
175	30040931	*icdA*	S1238	411	62	46,083	5.28	0.564	39.65	C	Isocitrate dehydrogenase
181	30042382	*eno*	S2988	199	43	45,683	5.32	0.843	27.47	G	Enolase
183	30042997	*tufB*	S3682	510	76	43,457	5.3	0.789	29.72	J	Protein chain elongation factor EF-Tu
332	30043222	*atpA*	S3954	570	37	55,416	5.8	0.667	32.42	C	Membrane-bound ATP synthase, F1 sector, α-subunit
465	30039968	*rpsB*	S0162	374	50	26,812	6.61	0.759	30.42	J	30 S ribosomal subunit protein S2
549	30042978	*hupA*	S3663	376	77	9,529	9.57	0.654	37.06	L	DNA-binding protein HU-α (HU-2)

[a] Score greater than 49 is significant. The detailed information about protein identification is summarized in supplemental Table 2.
[b] ENc is the effective number of codons, which is a measure of the extent of codon preference in a gene.
[c] The single letter codes represent a particular functional category as follows: J, translation; L, replication, recombination, and repair; M, cell wall/membrane biogenesis; O, post-translational modification, protein turnover, chaperones; C, energy production and conversion; G, carbohydrate transport and metabolism.

currently identified proteins, supporting the idea that proteins encoded by genes with a high CAI were abundant and easily identified. The overall hydrophobicity of a protein, expressed as the GRAVY index (15), was also calculated for all identified and predicted proteins. A comparison between identified proteins and total predicted proteins (Fig. 2) shows that both hydrophobic proteins (with very high GRAVY values) and the very hydrophilic proteins (with very low GRAVY values) are not present among the identified proteins. These results were similar to those of previous reports from *E. coli* (16), *Lactococcus lactis* (17), *Bacillus subtilis* (18), and *Bifidobacterium longum* (19).

Based on clusters of orthologous groups (COG) information, experimentally identified proteins were grouped into cellular roles and are summarized in Fig. 2. Proteins related to energy production and conversion (category C), translation (category J), and amino acid transport and metabolism (category E) were the three categories containing the most identified proteins. Interestingly, although proteins related to replication, recombination, and repair (category L) were predicted to be the most abundant category (18.02%), only a small proportion of these proteins (1.52%) was identified. However, this observation was not in agreement with similar proteomics data obtained in other bacteria (19, 20). When compared with the COG distributions of the theoretical proteomes of related microbes, it was found that only about 5% of the predicted proteins were grouped into category L in other proteobacteria, such as *E. coli* and *Salmonella*. The reason for this difference is that there is an astonishingly large number of transposable genetic elements in the genome of *Shigella* spp. (1, 21) that encode proteins related to recombination, such as transposase. The low rate of identification in this category suggested that these transposon-associated proteins were not necessary for cell grown in complex media. The cellular localizations of all identified proteins, predicted by PSORT Version 2.0, are also compared with those of all predicted proteins (Fig. 2). Briefly, 337 proteins identified are cytoplasmic, 13 proteins were predicted to reside in the cytoplasmic membrane, and 18 proteins were predicted to be located in the outer membrane. Proteins located in the outer membrane seem to be more easily identified than those located in the cytoplasmic membrane.

Abundance Analysis of Identified Proteins

The abundance of many proteins varied largely according to the four different conditions, such as the virulence factors shown in Fig. 3 and some growth phase-dependent proteins (such as Spot IDs 125, 205, and 270). Besides these, there were still some highly abundant proteins (volume % >0.6) common to the different conditions, suggesting that they play important functions. The information about these abundant proteins is summarized in Table I. Among these proteins, only one protein has a calculated CAI value of less than 0.65, indicating that proteins encoded by genes with a high CAI are abundant. Furthermore, we also located the identified proteins in the genome with respect to their abundance. As shown in supplemental Fig. 4, the distribution and organization of expressed genes in the genome does not show any distinct active regions, although it is known that most functional genes are located in gene clusters. Despite this, there are still several regions of ~100 kb without any genes expressed (so-called "expression deserts"; such as 1.86M–1.96M and 3.85M–3.95M regions in the genome) and one 20-kb region where abundant proteins are clustered (called an "expression oasis"; 4.33M–4.35M in the genome). Interestingly, this expression-abundant region is not the most highly expressed region (the highest peak in the outer circle in supplemental Fig. 4).

Hypothetical Genes and Misannotated Genes

The lack of homology to other known proteins makes hypothetical proteins interesting as they may have unique functions required in a particular organism. Altogether 64 products of hypothetical genes, whose functions have not been assigned up to now, were experimentally confirmed in this study. Proteome studies focusing on the presence of these proteins during growth phase, stress, or other processes may give insight into their cellular roles (19). Moreover, the expression of six misannotated genes was also confirmed. Among these, four genes (Spot IDs A123, 0673, zx70, and A153) are annotated as pseudogenes due to frameshift, and one gene (Spot ID A020) is unannotated in the genome sequencing project. In fact, previous proteome research of bacteria also provided evidence for expression of genes annotated as pseudogenes (20, 22), suggesting that expression of pseudogenes is common in bacteria. The last misannotated gene, previously predicted to encode a 4.3-kDa protein in the same region of the genome, actually encodes a 26-kDa protein homologous to protein GI: 110805376. All of these data will be helpful to correct the genome annotation of *S. flexneri*.

Comparative Proteomics Analysis of S. flexneri

Comparative Proteomics Analysis of S. flexneri in Log Phase and Stationary Phase

The results of this study are similar to a previous report (23). Briefly, compared with the protein expression profiles of cells in log phase, the expression of 2-oxoglutarate dehydrogenase (SucAB; Spot IDs X60 and 205) was reduced. The expressions of the pyruvate dehydrogenase (AceE; Spot ID 198) and fumarate reductase (FrdAB; Spot IDs 322 and 425), aspartase (AspA; Spot ID 170), and periplasmic asparaginase II (AnsB; Spot ID 300) were, however, greatly increased. These results are consistent with bacterial cells converting their aerobic metabolism into anaerobic fumarate metabolism during the stationary phase (24). The expression of the glycerol 3-phosphate (*glp*) regulon-encoded proteins (GlpABDKQ; Spot IDs 396, 344, 505, 179, and 133) was also increased in the stationary phase. These proteins might be related to stationary autophagy (dwarfing phenomenon), an important survival mechanism of bacteria (25).

Comparative Proteomics Analysis of S. flexneri Grown at 30 and 37 °C

Because proteins are the functional outputs of genes, analysis of the changes in the proteome of pathogens grown at conditions similar to their host environments will uncover those precisely regulated and selectively expressed genes. The 2-DE profiles of *S. flexneri* cultures grown at 30 and 37 °C are highly comparable (supplemental Fig. 1). A protein was regarded to be temperature-dependent if the spot density in one condition was significantly different (>2-fold) from that in the other condition.

Virulence-related proteins—The database search results obtained from MALDI-TOF/TOF MS showed that most up-regulated proteins at 37 °C were heat shock proteins and virulence factors (such as IpaABCD; Spot IDs 395, 678, 486, and 303) in agreement with previous reports (6). Locating the positions of these virulence factors on 2-DE maps may facilitate the study of their cellular roles in subsequent functional proteomics analyses (*i.e.* under invasion or other processes). All of the identified known virulence factors encoded by the virulence plasmid are listed in Table II. The abundance of some virulence factors under the different conditions is also shown in Fig. 3. Interestingly, although the abundance of IpaC is highest in the stationary phase at 37 °C, its CAI value is only 0.222, which is much lower than those of other abundant proteins encoded by the chromosome. Thus, evaluating synonymous codon usage is not accurate enough for the prediction of expression abundance, particularly for those exogenous genes acquired by horizontal transfer.

As shown in Fig. 3, the expression patterns of virulence factor IpaC and its protein chaperon IpgC are similar. Their abundances were greatly increased at 37 °C in the stationary phase but not in the log phase no matter what the growth temperature was. Strangely, this kind of expression pattern is not consistent with that of other virulence factors, such as IpaA and IpaD. The levels of IpaA and IpaD were not apparently increased in the stationary phase, although they were increased at 37 °C. Because the encoding genes of the above virulence factors are located in a single cluster in the virulence plasmid of *S. flexneri*, the difference of their expression patterns might result from a particular post-transcriptional regulation. This regulatory mechanism is probably related to the growth phase transition. Furthermore, the abundance of virulence protein IpaB was significantly lower than its group partners, suggesting that its distinct physical-chemical characteristics probably lead to dissolution difficulties. In addition, the levels of all of these virulence proteins were higher in the log phase than in the stationary phase at 30 °C. The reason for this might be that the superhelix conformations of the plasmid DNA were much looser under rapid growth and efficient metabolism.

Another finding is that the abundance of protein SepA (Spot ID 475), whose expression was not previously considered to be controlled by temperature (26), was actually higher at 37 °C than at 30 °C (Fig. 3). SepA is a secreted serine protease of the IgA1 protease family. Its function is not required for entry into epithelial cells or for intercellular dissemination, but it might play a role in tissue invasion (26). The molecular mass of SepA precursor is 146 kDa, and its mature form is about 110 kDa after autonomous proteolysis. However, the experimental molecular mass of identified spots was only about 26 kDa. Further analysis of the PMF database search revealed that most matched peptides were clustered at the C-terminal outer membrane autotransporter barrel domain (TIGR01414)

TABLE II
Known virulence factors identified

Spot ID	GI	Gene	Synonym	Score[a]	Sequence coverage	Matched/ searched	Peptide sequence confirmed	Peptide score[a]	Molecular weight	pI	CAI	ENc[b]	GRAVY
					%								
112	18462586	*phoN2/apy*	CP0004	153	45	11/64	VICGAHWQSDVEAGR	81	27,727	5.61	0.255	51.07	−0.34
574	18462543	*ospF*	CP0010	392	66	14/93	SEIPQMLSANER	70	28,095	7.62	0.144	51.89	−0.45
							NNFNILYNQIR	78					
							GDFVGDKFHISIAR	98					
							IAPGEYPASDVRPEDWK	59					
709	18462546	*ospD1*	CP0022	175	48	8/22	SINNYGLHPANNK	71	24,850	6.45	0.157	56.56	−0.37
496	18462548	*ospC2*	CP0063	488	33	19/36	SLNLSIEAPSGAR	25	55,744	6.58	0.195	46.17	−0.54
							HQSNANLSNNTLNIK	97					
							LGIGLHLIDFIR	70					
							FREFCYNK	34					
							VLEYFISNGLVDVNKR	105					
475	18462537	*sepA*	CP0070	528	18	24/42	MGDLRDTQGDAGVWAR	57	146,041	6.1	0.288	48.57	−0.36
							DYSSHSWYAGAEVGYR	144					
							AGLGYQFDLLANGETVLQDASGEKR	187					
395	18462581	*ipaA*	CP0125	400	59	22/56	DLNTVAVFPELLR	52	70,080	6.13	0.184	44.93	−0.72
							VYDNSYHENPENDAQSPTSQTNDLLSR	144					
303	18462578	*ipaD*	CP0126	647	68	21/34	IANSINDINEQYLK	95	36,696	5.65	0.196	46.76	−0.51
							YKDKPLYPANNTVSQEQANK	130					
							YSNANSIFDNLVK	98					
							VLSSTISSCTDTDKLFLHF	119					
486	56383093	*ipaC*	CP0127	699	56	19/33	LSEQVQHDSEIIAR	108	38,838	8.54	0.222	50.6	−0.46
							TSLSSPDISLQDKIDTQR	101					
							RTYELNTLSAQQK	57					
							TYELNTLSAQQK	79					
							ATMETSAVAGNISTSGGR	139					
504	18462580	*ipaB*	CP0128	579	31	16/21	NLEFSDKINTLLSETEGLTR	138	62,190	7.66	0.231	44.3	−0.05
							INQIQTR	56					
							NDLALFQSLQESR	97					
							KSDEYAAEVR	102					
017	18462568	*ipgC*	CP0129	246	79	12/43	GRIEEAEVFFR	36	17,916	4.58	0.26	48.54	−0.15
							NDYTPVFHTGQCQLR	85					
520	18462570	*ipgB1*	CP0130	189	65	18/43	SAQTSISHTALNEIASSGLR	61	23,783	9.56	0.158	55.78	−0.35
560	18462582	*icsB*	CP0132	749	65	41/70	GPIRVEDTEHGPILIAQK	109	56,613	9.28	0.2	49.28	−0.54
							NYRLENQK	28					
							NSIQHSFTGR	89					
							DTEINYLFEKR	89					
							YIESDFDLSR	92					
558	18462567	*ipgD*	CP0133	166	33	13/23	KLPLSSLELSYSER	50	60,256	9.42	0.16	46.86	−0.48
716	18462559	*mxiD*	CP0145	296	28	13/25	YVAQSDTVGSFFER	75	63,148	5.54	0.159	47.84	−0.14
							GEDIVIPGVATVVER	41					
							TFYVSLVGER	44					
755	18462536	*spa15*	CP0148	172	52	6/50	SNINLVQLVR	27	15,164	4.38	0.187	43.55	0.30
							VVIKDDYVHDGIVFAEILHEFYQR	79					
756	18462530	*spa47*	CP0149	504	54	22/71	LLTQLSFPNR	67	47,774	5.5	0.167	46.64	−0.07
							QTLIKPTAQFLHTQVGR	88					
							GYPVSVFDSLPR	97					
							SILDGHIYLSR	83					
308	56383102	*virA*	CP0181	668	67	24/50	IITFGIYSPHETLAEK	115	45,163	5.71	0.178	52.28	−0.22
							DTDFNSVWHDIYR	100					
							LNFMPEQR	44					
							AYEVSSSILPSHITCNGVGINK	64					
							IETSYLVHAGTLPSSEGLR	127					
L18	18462521	*icsA/virG*	CP0182	73	6	5/17	TVGGHNEHNLADR	43	116,572	5.33	0.231	51.2	−0.35
366	18462539	*phoN1*	CP0190	210	42	8/89	LLTNMIEDAGDLATR	90	27,309	6.9	0.24	52.43	−0.29
							AKDEFANNQK	72					
L24	56383114	*icsP/sopA*	CP0271	396	54	16/68	GWLLNNLDYR	71	35,689	9.08	0.225	51.46	−0.51
							LGLIAGYQESR	80					
							IPYIGLTANYR	48					
							HENFEFGAELK	71					

[a] Score greater than 49 is significant for a PMF search, and 21 is significant for an ion MS/MS search.
[b] ENc is the effective number of codons, which is a measure of the extent of codon preference in a gene.

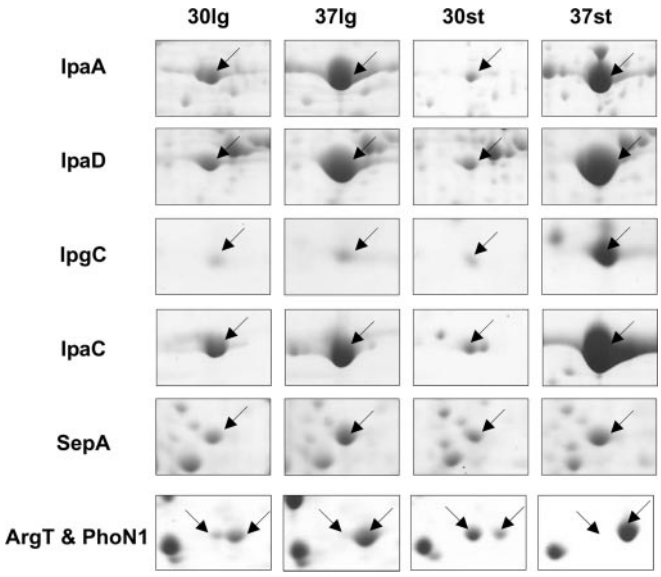

FIG. 3. **Enlarged images of 2-DE gels highlighting selected differentially expressed proteins.** *30lg* indicates grown to log phase at 30 °C, *37lg* indicates grown to log phase at 37 °C, *30st* indicates grown to stationary phase at 30 °C, and *37st* indicates grown into stationary phase at 37 °C. The arrows indicated the spots of corresponding proteins listed on the left.

of SepA, suggesting that this protein was cut off, and only its C terminus was left. Although this cleavage has been analyzed by traditional biochemical methods (26), it still indicated the capability of proteomics data mining for finding post-translational modifications.

There are still some known virulence factors that were not identified, such as *ospD3*, *ospE1/2*, and *ipaH*. In fact, the expression of virulence proteins is controlled by two kinds of plasmid-encoded transcriptional regulators: 1) *virB* and *virF*, which are active upon temperature up-shift, and 2) *mxiE*, which is active under conditions of secretion (27). The above unidentified virulence factors are those controlled only by regulator *mxiE*, which will not be active in our experimental conditions. It should also be noted that the identification of parts of known virulence factors was in agreement with a previous microarray analysis (4). Thus, these proteomics results not only provide a global view on expression of virulence proteins but also provide new evidence for the classification of effectors in the type III secretion system of *S. flexneri*.

De Novo Biosynthesis of Purine Nucleotides—Five of the down-regulated genes at 37 °C (purDEHMT; Spot ID 036/444/208/042/229, ratio 0.36/0.26/0.49/0.27/0.46 at stationary phase) were associated with the *de novo* biosynthesis of purine nucleotides. An additional four genes of this pathway (purABCU; Spot ID 177/335/106/467, ratio 0.69/0.55/0.58/0.67 at stationary phase) were also down-regulated, although the ratios were not significant. These genes are unlinked single units in the genome of *S. flexneri* but may act in a single unit of regulation controlled by the "PurR" repressor protein (28). Therefore, the temperature-dependent expression of these genes is possibly regulated at the level of transcription. Simultaneous detection of the differential expression of these scattered genes also exhibited the power and the confidence of comparative proteomics analysis. Because the bacterium divides more rapidly at higher temperature, it is surprising to see a decrease in abundance of Pur gene products. It might be a result of using rich growth medium and hence no need for *de novo* synthesis. The decreased expression of these proteins might be caused by the increased expression of virulence-related proteins at 37 °C that consume a lot of energy and material resources. Another possible explanation is that the enzyme activities of Pur proteins are higher at 37 °C than those at 30 °C. Then, if the same units of enzyme are needed for the survival of bacteria, the total abundance of these proteins will be increased.

Amino Acid Transporter ArgT—A further interesting finding is that the spots of protein ArgT (Spot ID 113) disappeared at 37 °C (Fig. 3), which has not been previously reported in related species. The *argT* gene is predicted to encode the lysine/arginine/ornithine-binding protein, a periplasmic protein responsible for binding the amino acid substrates during the process of amino acid transport into the cytoplasm (29). According to studies in *E. coli*, the expression of ArgT is up-regulated in the stationary growth phase (30) or under glucose-limiting conditions (31), suggesting its role in the physiological adaptation of bacteria to detrimental conditions. Our following experiments were focused on the regulatory mechanism of ArgT expression.

Regulation Mechanism of ArgT Expression

Differential Expression of ArgT Is Due to Proteolysis

In bacteria, regulation at the transcriptional level is the most efficient and cost-effective way to control gene expression. However, the results of semiquantitative RT-PCR analysis (Fig. 4*A*) showed that the quantity of *argT* mRNA at 37 °C is similar to that at 30 °C, suggesting that the regulation of ArgT expression did not occur at the transcriptional level.

It is well known that translation is coupled to transcription in bacteria. Because the regulation of ArgT expression did not occur at the transcriptional level, post-translational regulation, such as proteolysis, was the most plausible mechanism; that is the protein might be degraded at 37 °C. To test our hypothesis, we constructed an ArgT-overexpressing plasmid, pProEX-HTb-argT (pProEX-argT), and transformed it into competent cells of *S. flexneri* 2457T. After induction by IPTG at 37 and 30 °C, the cultures were collected and analyzed by SDS-PAGE. The results showed that even if 2457T cells were compelled to express ArgT by the exogenous plasmid the expression product could not be detected at 37 °C (Fig. 4*B*). Because transcription and translation of *argT* in the exogenous expression vector were not subject to the effect of temperature, we concluded that the differential expression of ArgT was due to protein proteolysis.

ArgT Undergoes Proteolysis in Periplasmic Space of S. flexneri

According to the genome annotation of *S. flexneri*, ArgT is located in the periplasmic space. The ArgT protein sequence contains a signal peptide to traverse the inner membrane and reach the periplasmic space. Hence, is ArgT degraded in the cytoplasm or in the periplasmic space? To answer this question, the periplasmic proteins of 2457T/pProEX-argT were extracted and probed by Western blot analysis with anti-ArgT antibody. ArgT protein was only detected in the periplasmic proteins extracted from cultures grown at 30 °C but not in the remaining whole-cell proteins at 30 °C or in the protein samples extracted from cultures grown at 37 °C (Fig. 5*A*).

On the other hand, 2457T strains carrying pGEX-argT and carrying pGEX-argT2, which expressed ArgT without the signal peptide, were induced to express at 30 °C and then incubated at 37 and 30 °C for 4 h. The total proteins of each culture were analyzed by Western blot with anti-ArgT (Fig. 5*B*). The results showed that ArgT protein was not degraded when the signal peptide was removed. Thus, we concluded that ArgT undergoes proteolysis in the periplasmic space of *S. flexneri*.

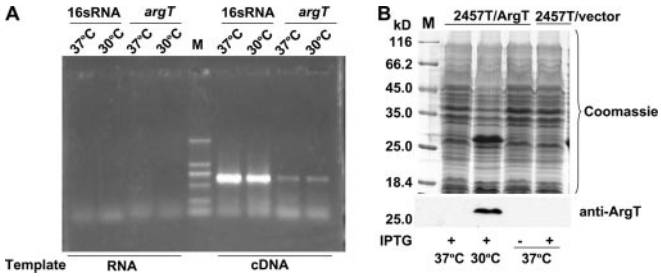

FIG. 4. **Differential expression of ArgT is due to proteolysis.** *A*, semiquantitative RT-PCR analysis of *argT*. *B*, overexpression of ArgT at 37 and 30 °C. 2457T/pProEX-argT (2457T carrying pProEX-HTb-argT) was induced by IPTG at 37 and 30 °C, respectively. *M*, protein marker (Fermentas).

ArgT Proteolysis Is Not Dependent on de Novo Protein Synthesis

Is the protease responsible for the cleavage of ArgT synthesized at 37 °C but not at 30 °C? To answer this question, we did the following experiment. 2457T/pProEX-argT cells were cultivated and induced to express ArgT by IPTG at 30 °C for 4 h, and then chloramphenicol was added to prevent new protein synthesis. After that, the temperature of the experimental group was increased to 37 °C for another 6 h, whereas the temperature of the control group was kept at 30 °C. Whole-cell protein samples were prepared every hour and analyzed by Western blot. As shown in Fig. 6, the overexpressed ArgT proteins at 30 °C were degraded rapidly after temperature up-shift to 37 °C, and this process was not dependent on any newly synthesized proteins. That is, ArgT was degraded at 37 °C by a protease synthesized at 30 °C but activated at 37 °C.

ArgT Was Degraded by HtrA Protease

Because ArgT was degraded in the periplasmic space of 2457T at 37 °C but not at 30 °C, it is reasonable to suppose that a temperature-activated periplasmic protease may be responsible for the degradation of ArgT. HtrA (also named DegP) attracted our attention because it is a heat shock-induced protein and plays crucial roles with regard to protein quality control in the periplasmic space, functioning as both a molecular chaperone and protease (32, 33). It has been reported that the proteolytic activity of HtrA increases with increasing temperature (11, 33).

Is ArgT specifically degraded by HtrA at 37 °C *in vivo*, or can another protease perform the same function? To answer this question, an *htrA* deletion mutant of *S. flexneri*

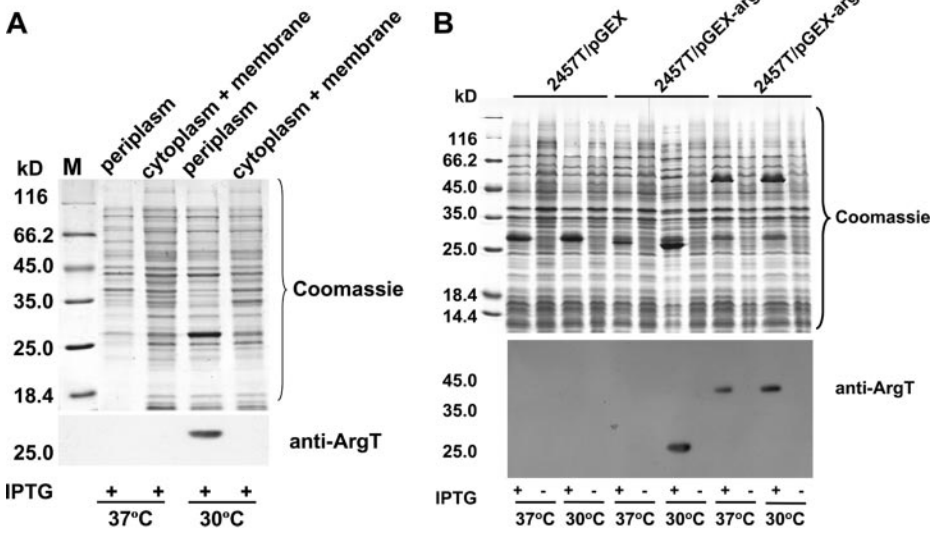

FIG. 5. **ArgT undergoes proteolysis in periplasmic space of *S. flexneri*.** *A*, the periplasmic proteins (*periplasm*) and the remaining whole-cell proteins (*cytoplasm + membrane*) of induced 2457T/pProEX-argT cells were extracted from cultures grown at 37 and 30 °C. ArgT expression products were only detected in the periplasm of 2457T grown at 30 °C. *B*, 2457T/pGEX-argT (with signal peptide) and 2457T/pGEX-argT2 (without signal peptide) were induced by IPTG at 30 °C and then incubated at 37 and 30 °C for 4 h. Due to the cleavage of the signal peptide, the ArgT detected in 2457T/pGEX-argT is smaller than that in 2457T/pGEX-argT2. *M*, protein marker (Fermentas).

(2457TΔ*htrA*) was constructed using the lambda *red* recombination strategy. Then, the ArgT-overexpressing plasmid was transformed into the 2457TΔ*htrA* strain. 2457TΔ*htrA*/pProEX-argT was induced to express ArgT by IPTG at 37 and at 30 °C. The results (Fig. 7*A*) showed that ArgT could not be degraded in the *htrA* deletion mutant. Thus, we concluded that HtrA is related to ArgT proteolysis *in vivo*.

Skórko-Glonek *et al.* (11) reported that mutation of the Ser residue to Ala at position 210 of the mature HtrA could totally eliminate the proteolytic activity of HtrA without changing its structure. To test whether HtrA protease degraded ArgT directly without the participation of any other proteins, GST-ArgT, GST-HtrA, and GST-$HtrA_{S210A}$ fusion proteins were purified, and proteolysis assays were carried out *in vitro*. As shown in Fig. 7*B*, ArgT was degraded by HtrA *in vitro* at 37 °C but not at 30 °C. Moreover, ArgT was not degraded by the inactivated $HtrA_{S210A}$. Thus, we concluded that HtrA is the protease responsible for the temperature-dependent degradation of ArgT. Periplasmic protease HtrA, a known virulence factor (34), is responsible for the degradation of ArgT. It has been reported that HtrA facilitates IcsA surface expression and is required for efficient intercellular spread of *S. flexneri* (35). HtrA is also found to be a virulence factor in other bacteria, such as *Salmonella typhimurium* (36) and *Listeria monocytogenes* (37).

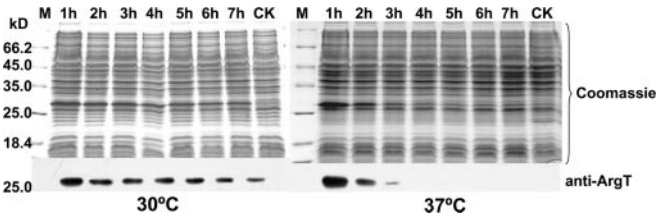

FIG. 6. **ArgT, overexpressed at 30 °C, was degraded at 37 °C without new protein synthesis.** The cultures of 2457T/pProEX-argT were induced to express ArgT at 30 °C for 4 h. Then, chloramphenicol was added to prevent new protein synthesis. After that, the cells were incubated at 30 (*left*) and 37 °C (*right*) for another 6 h. *CK*, the cultures of 2457T/pProEX-argT induced to express at 37 °C. *M*, protein marker (Fermentas).

Identification of Cleavage Site in ArgT

To identify the HtrA cleavage sites of ArgT, the fragments of ArgT degraded by HtrA were separated by 15% SDS-PAGE (Fig. 8*A*). ArgT was cleaved into two fragments (Fig. 8*A*, frag1 and frag2). The fragments were cut out and digested by trypsin and Glu-C, respectively (trypsin cleaves peptide on the carboxylic side of lysine and arginine; Glu-C cleaves peptide on the carboxylic side of Glu in ammonium bicarbonate at pH 7.8). The resolved digestion products were analyzed by tandem mass spectrometry (MALDI-TOF/TOF). The PMF result of the larger fragment (frag1; about 32 kDa) matched some peptides of GST (*m/z* 1032.57, 1094.56, 1182.68, 1516.81, 1801.94, 2229.08, 2269.12, 2326.12, and 2357.18), suggesting this fragment was the N terminus of ArgT, and the MS/MS results showed that one of the peptides (molecular weight, 1141.55) of the smaller fragment (frag2; about 14 kDa) matched the C-terminal region of ArgT. Furthermore, the sum of the molecular weights of these two fragments was similar to the molecular weight of intact ArgT. All these observations suggested that these two fragments might be produced by one cleavage in ArgT.

If we can identify the N-terminal or C-terminal sequence of the cleavage fragments, we can determine the cleavage site. That is, our targets are those ion peaks whose experimental ion masses did not match the theoretical values calculated from protein digestion of ArgT. One such peptide digested by trypsin (molecular weight, 1726.842) and one peptide digested by Glu-C (molecular weight, 2583.301) were identified. They represented the same cleavage site according to their amino acid sequences (Fig. 8, *B*, *C*, and *D*). Thus, we concluded that the cleavage site in ArgT was between position 160 (Val) and position 161 (Ala). This is in agreement with a report suggesting a preference for Val and Ile as the residue preceding the HtrA cleavage site (34).

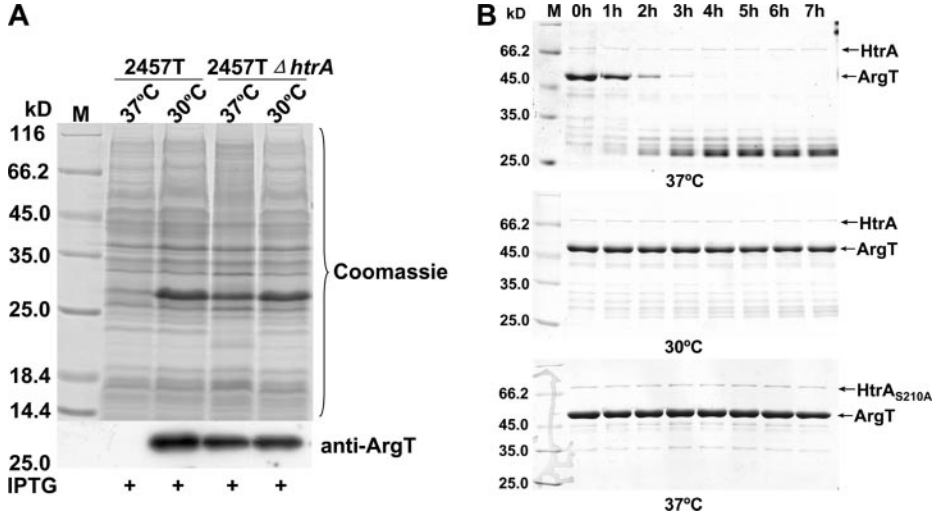

FIG. 7. **ArgT was degraded by HtrA protease *in vivo* and *in vitro*.** *A*, ArgT could not be degraded in the *htrA* deletion mutant. Plasmid pProEX-HTb-argT was transformed into 2457T and 2457TΔ*htrA*, and then strains carrying the recombinant plasmids were induced by IPTG at 37 and 30 °C, respectively. *B*, 50 μg of GST-ArgT and 0.5 μg of GST-HtrA/HtrAS210A fusion proteins were incubated in a final volume of 100 μl of 50 mM Tris-Cl, pH 8.0 buffer at 37 and 30 °C. 15 μl of the incubation was removed every hour and loaded for SDS-PAGE. *M*, protein marker (Fermentas).

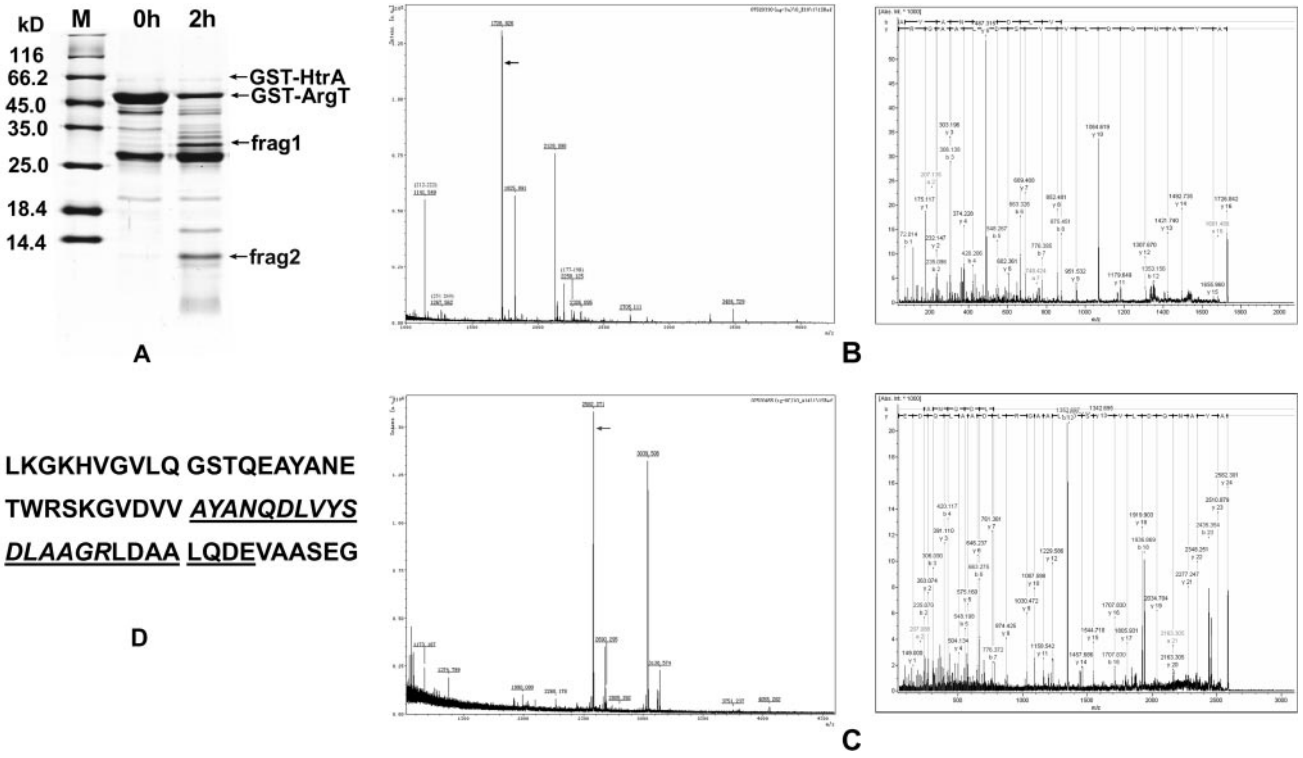

FIG. 8. **Identification of cleavage site in ArgT.** *A*, GST-ArgT and GST-HtrA were incubated at 37 °C for 2 h, and then the degradation products were separated by 15% SDS-PAGE. The two cleavage fragments (frag1 and frag2) were cut out and digested by trypsin and Glu-C. The resolved digestions were analyzed by MALDI-TOF/TOF. *B*, the PMF result of frag2 digested by trypsin and the MS/MS result of the target peptide (theoretical *m/z* 1725.82; sequence, AYANQDLVYSDLAAGR). *C*, the PMF result of frag2 digested by Glu-C and the MS/MS result of the target peptide (theoretical *m/z* 2581.22; sequence, AYANQDLVYSDLAAGRLDAALQDE). These two peptides suggest the same cleavage site between position 160 and position 161 (VV ↓ AY) of ArgT. *D*, part of the sequence of ArgT. The amino acid sequences of the above two peptides are *italicized* and *underlined*, respectively. *M*, protein marker (Fermentas).

Conclusion

In this study, the proteome reference maps of the log phase and stationary phase *S. flexneri* 2a 2457T cells grown at 30 and 37 °C were thoroughly analyzed by the use of multiple overlapping narrow pH range (pH 4.0–5.0, pH 4.5–5.5, pH 5.0–6.0, pH 5.5–6.7, and pH 6–11) two-dimensional gel electrophoresis. Altogether, 755 spots representing 588 different protein entries were identified by MALDI-TOF/TOF MS. The majority of the known key virulence factors of *S. flexneri* were located in the 2-DE maps for the first time. We also analyzed the abundance of identified proteins at different stages and confirmed the expression of 70 hypothetical proteins. Moreover, six misannotated proteins were experimentally identified.

The comparative proteomics analysis also revealed that protein ArgT of *S. flexneri* was not detected at 37 °C but was present at 30 °C, contrary to the expression of virulence genes. We then showed that ArgT was degraded by the periplasmic protease HtrA, a known virulence-related protein. All of these findings suggested that the temperature-dependent proteolysis of ArgT at 37 °C is possibly related to the virulence of *S. flexneri*. Thus, our future work will focus on the mutation of *argT* to construct a mutant strain that can express functional ArgT proteins at 37 °C. We will then be able to evaluate the virulence of this mutant by competitive invasion assays in HeLa cells and BALB/c mice.

Acknowledgments—We are grateful to Dr. Jinbin Sun and Ping Zheng for advice on compiling this paper and Dr. Feng Shao and Yongqiang Jiang for technical assistance and helpful discussions.

* This work was supported by the National Key Basic Research Program of China (973 Program, Grant 2005CB522904), Megaprojects of Science Research for the 11th Five-year Plan (Grant 2009ZX10004-103), and the National Natural Science Foundation of China (Grants 30470101 and 30700035).

S This article contains supplemental Figs. 1–4 and Tables 1 and 2.

§ Both authors contributed equally to this work.

§§ To whom correspondence may be addressed: National Inst. for the Control of Pharmaceutical and Biological Products, 2 Tiantanxili, 100050 Beijing, China. Tel.: 86-10-67058416; Fax: 86-10-67058402; E-mail: zengming@nicpbp.org.cn.

¶¶ To whom correspondence may be addressed: Beijing Inst. of Biotechnology, State Key Laboratory of Pathogen and Biosecurity, 20 Dongdajie St., Fengtai District, 100071 Beijing, China. Tel.: 86-10-63802181; Fax: 86-10-63833521; E-mail: wanghl@nic.bmi.ac.cn.

REFERENCES

1. Wei, J., Goldberg, M. B., Burland, V., Venkatesan, M. M., Deng, W., Fournier, G., Mayhew, G. F., Plunkett, G., 3rd, Rose, D. J., Darling, A., Mau, B., Perna, N. T., Payne, S. M., Runyen-Janecky, L. J., Zhou, S., Schwartz, D. C., and Blattner, F. R. (2003) Complete genome sequence

and comparative genomics of Shigella flexneri serotype 2a strain 2457T. *Infect. Immun.* **71,** 2775–2786
2. Liao, X., Ying, T., Wang, H., Wang, J., Shi, Z., Feng, E., Wei, K., Wang, Y., Zhang, X., Huang, L., Su, G., and Huang, P. (2003) A two-dimensional proteome map of Shigella flexneri. *Electrophoresis* **24,** 2864–2882
3. Buchrieser, C., Glaser, P., Rusniok, C., Nedjari, H., D'Hauteville, H., Kunst, F., Sansonetti, P., and Parsot, C. (2000) The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of Shigella flexneri. *Mol. Microbiol.* **38,** 760–771
4. Le Gall, T., Mavris, M., Martino, M. C., Bernardini, M. L., Denamur, E., and Parsot, C. (2005) Analysis of virulence plasmid gene expression defines three classes of effectors in the type III secretion system of Shigella flexneri. *Microbiology* **151,** 951–962
5. Maurelli, A. T., Blackmon, B., and Curtiss, R., 3rd (1984) Temperature-dependent expression of virulence genes in Shigella species. *Infect. Immun.* **43,** 195–201
6. Konkel, M. E., and Tilly, K. (2000) Temperature-regulated expression of bacterial virulence genes. *Microbes Infect.* **2,** 157–166
7. Görg, A., Weiss, W., and Dunn, M. J. (2004) Current two-dimensional electrophoresis technology for proteomics. *Proteomics* **4,** 3665–3685
8. Desjardin, L. E., Hayes, L. G., Sohaskey, C. D., Wayne, L. G., and Eisenach, K. D. (2001) Microaerophilic induction of the alpha-crystallin chaperone protein homologue (hspX) mRNA of Mycobacterium tuberculosis. *J. Bacteriol.* **183,** 5311–5316
9. Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.* **97,** 6640–6645
10. Berducci, G., Mazzetti, A. P., Rotilio, G., and Battistoni, A. (2004) Periplasmic competition for zinc uptake between the metallochaperone ZnuA and Cu,Zn superoxide dismutase. *FEBS Lett.* **569,** 289–292
11. Skórko-Glonek, J., Krzewski, K., Lipińska, B., Bertoli, E., and Tanfani, F. (1995) Comparison of the structure of wild-type HtrA heat shock protease and mutant HtrA proteins. A Fourier transform infrared spectroscopic study. *J. Biol. Chem.* **270,** 11140–11146
12. Chen, H., Gill, A., Dove, B. K., Emmett, S. R., Kemp, C. F., Ritchie, M. A., Dee, M., and Hiscox, J. A. (2005) Mass spectroscopic characterization of the coronavirus infectious bronchitis virus nucleoprotein and elucidation of the role of phosphorylation in RNA binding by using surface plasmon resonance. *J. Virol.* **79,** 1164–1179
13. O'Connor, C. D., Clarke, I. N., and Skipp, P. (2006) Quest for complete proteome coverage, in *Microbial Proteomics: Functional Biology of Whole Organisms* (Humphery-Smith, I., and Hecker, M., eds) 1st Ed., pp. 27–38, John Wiley & Sons, New York
14. Drews, O., Reil, G., Parlar, H., and Görg, A. (2004) Setting up standards and a reference map for the alkaline proteome of the Gram-positive bacterium Lactococcus lactis. *Proteomics* **4,** 1293–1304
15. Comeron, J. M., and Aguadé, M. (1998) An evaluation of measures of synonymous codon usage bias. *J. Mol. Evol.* **47,** 268–274
16. Link, A. J., Robison, K., and Church, G. M. (1997) Comparing the predicted and observed properties of proteins encoded in the genome of Escherichia coli K-12. *Electrophoresis* **18,** 1259–1313
17. Guillot, A., Gitton, C., Anglade, P., and Mistou, M. Y. (2003) Proteomic analysis of Lactococcus lactis, a lactic acid bacterium. *Proteomics* **3,** 337–354
18. Büttner, K., Bernhardt, J., Scharf, C., Schmid, R., Mäder, U., Eymann, C., Antelmann, H., Völker, A., Völker, U., and Hecker, M. (2001) A comprehensive two-dimensional map of cytosolic proteins of Bacillus subtilis. *Electrophoresis* **22,** 2908–2935
19. Yuan, J., Zhu, L., Liu, X., Li, T., Zhang, Y., Ying, T., Wang, B., Wang, J., Dong, H., Feng, E., Li, Q., Wang, J., Wang, H., Wei, K., Zhang, X., Huang, C., Huang, P., Huang, L., Zeng, M., and Wang, H. (2006) A proteome reference map and proteomic analysis of Bifidobacterium longum NCC2705. *Mol. Cell. Proteomics* **5,** 1105–1118
20. Jing, H. B., Yuan, J., Wang, J., Yuan, Y., Zhu, L., Liu, X. K., Zheng, Y. L., Wei, K. H., Zhang, X. M., Geng, H. R., Duan, Q., Feng, S. Z., Yang, R. F., Cao, W. C., Wang, H. L., and Jiang, Y. Q. (2008) Proteome analysis of Streptococcus suis serotype 2. *Proteomics* **8,** 333–349
21. Jin, Q., Yuan, Z., Xu, J., Wang, Y., Shen, Y., Lu, W., Wang, J., Liu, H., Yang, J., Yang, F., Zhang, X., Zhang, J., Yang, G., Wu, H., Qu, D., Dong, J., Sun, L., Xue, Y., Zhao, A., Gao, Y., Zhu, J., Kan, B., Ding, K., Chen, S., Cheng, H., Yao, Z., He, B., Chen, R., Ma, D., Qiang, B., Wen, Y., Hou, Y., and Yu, J. (2002) Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli K12 and O157. *Nucleic Acids Res.* **30,** 4432–4441
22. Gupta, N., Tanner, S., Jaitly, N., Adkins, J. N., Lipton, M., Edwards, R., Romine, M., Osterman, A., Bafna, V., Smith, R. D., and Pevzner, P. A. (2007) Whole proteome analysis of post-translational modifications: applications of mass-spectrometry for proteogenomic annotation. *Genome Res.* **17,** 1362–1377
23. Zhu, L., Liu, X. K., Zhao, G., Zhi, Y. D., Bu, X., Ying, T. Y., Feng, E. L., Wang, J., Zhang, X. M., Huang, P. T., and Wang, H. L. (2007) Dynamic proteome changes of Shigella flexneri 2a during transition from exponential growth to stationary phase. *Genomics Proteomics Bioinformatics* **5,** 111–120
24. Chang, D. E., Smalley, D. J., and Conway, T. (2002) Gene expression profiling of Escherichia coli growth transitions: an expanded stringent response model. *Mol. Microbiol.* **45,** 289–306
25. Nyström, T. (2004) Stationary-phase physiology. *Annu. Rev. Microbiol.* **58,** 161–181
26. Benjelloun-Touimi, Z., Sansonetti, P. J., and Parsot, C. (1995) SepA, the major extracellular protein of Shigella flexneri: autonomous secretion and involvement in tissue invasion. *Mol. Microbiol.* **17,** 123–135
27. Mavris, M., Page, A. L., Tournebize, R., Demers, B., Sansonetti, P., and Parsot, C. (2002) Regulation of transcription by the activity of the Shigella flexneri type III secretion apparatus. *Mol. Microbiol.* **43,** 1543–1553
28. Meng, L. M., Kilstrup, M., and Nygaard, P. (1990) Autoregulation of PurR repressor synthesis and involvement of purR in the regulation of purB, purC, purL, purMN and guaBA expression in Escherichia coli. *Eur. J. Biochem.* **187,** 373–379
29. Celis, T. F., Rosenfeld, H. J., and Maas, W. K. (1973) Mutant of Escherichia coli K-12 defective in the transport of basic amino acids. *J. Bacteriol.* **116,** 619–626
30. Weichart, D., Querfurth, N., Dreger, M., and Hengge-Aronis, R. (2003) Global role for ClpP-containing proteases in stationary-phase adaptation of Escherichia coli. *J. Bacteriol.* **185,** 115–125
31. Wick, L. M., Quadroni, M., and Egli, T. (2001) Short- and long-term changes in proteome composition and kinetic properties in a culture of Escherichia coli during transition from glucose-excess to glucose-limited growth conditions in continuous culture and vice versa. *Environ. Microbiol.* **3,** 588–599
32. Kim, D. Y., and Kim, K. K. (2005) Structure and function of HtrA family proteins, the key players in protein quality control. *J. Biochem. Mol. Biol.* **38,** 266–274
33. Kim, K. I., Park, S. C., Kang, S. H., Cheong, G. W., and Chung, C. H. (1999) Selective degradation of unfolded proteins by the self-compartmentalizing HtrA protease, a periplasmic heat shock protein in Escherichia coli. *J. Mol. Biol.* **294,** 1363–1374
34. Pallen, M. J., and Wren, B. W. (1997) The HtrA family of serine proteases. *Mol. Microbiol.* **26,** 209–221
35. Purdy, G. E., Hong, M., and Payne, S. M. (2002) Shigella flexneri DegP facilitates IcsA surface expression and is required for efficient intercellular spread. *Infect. Immun.* **70,** 6355–6364
36. Bäumler, A. J., Kusters, J. G., Stojiljkovic, I., and Heffron, F. (1994) Salmonella typhimurium loci involved in survival within macrophages. *Infect. Immun.* **62,** 1623–1630
37. Stack, H. M., Sleator, R. D., Bowers, M., Hill, C., and Gahan, C. G. (2005) Role for HtrA in stress induction and virulence potential in Listeria monocytogenes. *Appl. Environ. Microbiol.* **71,** 4241–4247