

## Proteomic Analysis of *Vibrio cholerae* in Human Stool<sup>†</sup>

Regina C. LaRocque,<sup>1,2\*</sup> Bryan Krastins,<sup>3</sup> Jason B. Harris,<sup>1,4</sup> Lauren M. Lebrun,<sup>1</sup>  
Kenneth C. Parker,<sup>3</sup> Michael Chase,<sup>3</sup> Edward T. Ryan,<sup>1,2,5</sup> Firdausi Qadri,<sup>6</sup>  
David Sarracino,<sup>3</sup> and Stephen B. Calderwood<sup>1,2,7</sup>

Division of Infectious Diseases, Massachusetts General Hospital, Boston, Massachusetts<sup>1</sup>; Department of Medicine, Harvard Medical School, Boston, Massachusetts<sup>2</sup>; Harvard Medical School-Partners Healthcare Center for Genetics and Genomics, Cambridge, Massachusetts<sup>3</sup>; Department of Pediatrics, Harvard Medical School, Boston, Massachusetts<sup>4</sup>; Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts<sup>5</sup>; International Centre for Diarrhoeal Disease Research, Bangladesh, Dhaka, Bangladesh<sup>6</sup>; and Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, Massachusetts<sup>7</sup>

Received 13 May 2008/Returned for modification 24 May 2008/Accepted 20 June 2008

**An effective vaccine for *Vibrio cholerae* is not yet available for use in the developing world, where the burden of cholera disease is highest. Characterizing the proteins that are expressed by *V. cholerae* in the human host environment may provide insight into the pathogenesis of cholera and assist with the development of an improved vaccine. We analyzed the *V. cholerae* proteins present in the stools of 32 patients with clinical cholera. The *V. cholerae* outer membrane porin, OmpU, was identified in all of the human stool samples, and many *V. cholerae* proteins were repeatedly identified in separate patient samples. The majority of *V. cholerae* proteins identified in human stool are involved in protein synthesis and energy metabolism. A number of proteins involved in the pathogenesis of cholera, including the A and B subunits of cholera toxin and the toxin-coregulated pilus, were identified in human stool. In a subset of stool specimens, we also assessed which in vivo expressed *V. cholerae* proteins were recognized uniquely by convalescent-phase as opposed to acute-phase serum from cholera patients. We identified a number of these in vivo expressed proteins as immunogenic during human infection. To our knowledge, this is the first characterization of the proteome of a pathogenic bacteria recovered from a natural host.**

*Vibrio cholerae* is a gram-negative bacillus that exists within an aquatic reservoir and that can cause severe, dehydrating, and occasionally fatal diarrhea in humans (11). Strains of *V. cholerae* are differentiated serologically by the O side chain of lipopolysaccharide, and the majority of toxigenic strains belong to serogroup O1 or O139. *V. cholerae* O1 occurs in two biotypes, classical and El Tor, which differ in biochemical characteristics and phage susceptibility. Since 1817, there have been seven cholera pandemics during which disease has spread from the Indian subcontinent across Asia, Europe, Africa, and the Western Hemisphere. The disease is also endemic in flood-prone regions of South Asia, such as Bangladesh, where seasonal outbreaks typically occur in the spring and fall. The ongoing seventh pandemic of cholera is due to the O1 El Tor biotype of *V. cholerae*, and more than 100,000 diarrheal deaths per year are attributed to infection with this organism. Although a number of cholera vaccines have been developed over the past 100 years, an effective vaccine is not yet available for use in the developing world, where the burden of disease is highest (21).

Humans are the only known host and reservoir for *V. cholerae* outside its aquatic environment, and an important limita-

tion to the development and testing of cholera vaccines has been the lack of an optimal animal model of infection. It is consequently of great interest to understand which virulence factors are expressed by the organism directly in the human host environment, since in vivo expressed antigens of *V. cholerae* may represent targets of protective human immune responses. The capacity to perform such research has been facilitated by the sequencing of *V. cholerae* N16961, a clinical isolate of *V. cholerae* O1 El Tor (9, 18). The genome of *V. cholerae* is comprised of two circular chromosomes with approximately 3,900 predicted open reading frames, 40% of which are hypothetical or conserved hypothetical genes. Using gene microarrays based on the N16961 sequence, our group and others have evaluated the gene transcription profiles of vibrios immediately upon their being shed in stool or vomitus from the human host (1, 17, 20). These studies of gene expression within the human host environment indicate that the organism expresses clusters of selected virulence genes in the earliest phases of infection of the upper intestinal tract, as represented by organisms present in human vomitus, and experiences conditions of anaerobiosis, iron limitation, and nutrient deprivation in human stool.

Characterizing the proteins expressed by *V. cholerae* directly in the human host provides a useful complement to the findings of gene expression profiling experiments. Because of the short half-life of bacterial mRNA, gene expression profiling may miss certain in vivo induced transcripts, particularly those that are transiently expressed in a compartment that is difficult to sample, such as the human small intestine. In the case of *V.*

\* Corresponding author. Mailing address: Division of Infectious Diseases, GRJ 504, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114. Phone: (617) 726-3812. Fax: (617) 726-7416. E-mail: rclarocque@partners.org.

† Supplemental material for this article may be found at <http://iai.asm.org/>.

‡ Published ahead of print on 30 June 2008.

*cholerae*, such transcripts may relate to key steps in pathogenesis. The longer stability of the protein products of in vivo expressed transcripts may allow the detection of their presence in stool-shed organisms, even if transcription is no longer occurring; such proteins are among the direct mediators of pathogenesis and are also among the targets of human immune responses to infection.

Two previous studies have developed two-dimensional gel proteome reference maps of *V. cholerae* grown in vitro to stationary phase (5) or under various pH levels (10), but these conditions are unlikely to reflect the growth conditions experienced by the organism in the human host environment, and a complete characterization of the proteins present was not performed in these studies. In the present work, we provide a comprehensive description of the proteome of *V. cholerae* present in 32 separate human stool samples, compare these in vivo expressed proteins to previous results of gene expression profiling, and identify a number of these proteins as immunogenic during human infection.

#### MATERIALS AND METHODS

**Collection of patient stool samples.** Approximately 100 ml of "rice water" diarrheal stool was collected from each of 32 cholera patients upon presentation to the hospital of the International Centre for Diarrheal Disease Research in Dhaka, Bangladesh (ICDDR,B). All patients denied having received antibiotics prior to presentation, and the stool sample was collected prior to administration of antibiotics at the ICDDR,B. Quantitative culturing of the stool samples revealed a median of  $5.4 \times 10^6$  (range,  $7.0 \times 10^2$  to  $1.2 \times 10^8$ ) CFU of *V. cholerae* per milliliter of stool. Serogrouping of the *V. cholerae* isolate from each patient was performed, and all clinical samples were confirmed to contain *V. cholerae* O1 El Tor. Immediately after collection, the stool samples were centrifuged at  $500 \times g$  for 10 min at 4°C in order to remove particulate matter. Bacteria were then pelleted by centrifugation at  $13,500 \times g$  for 10 min at 4°C. The bacterial pellet was washed three times in phosphate-buffered saline (PBS) and subsequently lyophilized and stored at -80 degrees for transfer from Bangladesh to the Harvard-Partners Healthcare Center for Genetics and Genomics for evaluation by mass spectrometry (MS).

Human studies approval for collection of stool samples was obtained from the Massachusetts General Hospital and the ICDDR,B.

**MS.** Samples were prepared for analysis by MS as described previously (16). Briefly, 2 mg of lyophilized protein from each stool sample was rehydrated and incubated in a reducing solution of 8 M urea, 100 mM ammonium bicarbonate, 1% sodium dodecyl sulfate (SDS), and 10 mM dithiothreitol at 37°C for 1 h. The samples were alkylated by the addition of 500 mM iodoacetamide and then quenched with 2 M dithiothreitol. Following the addition of loading buffer, each sample was centrifuged at  $21,000 \times g$  for 10 min at room temperature, and 400  $\mu$ l of each sample was fractionated on a one-dimensional SDS-polyacrylamide gel electrophoresis Tris-glycine 8 to 16% gradient gel (Invitrogen) for 2.5 h at 125 V. Gels were shrunk for 16 h by the addition of 50% methanol and 7.5% acetic acid and then allowed to swell for 1 h by the addition of deionized water. Gels were stained with SimplyBlue Safe Stain (Invitrogen) for 4 to 6 h, imaged, and sliced horizontally into fragments of equal size based on the molecular weight markers. In-gel digestion was performed after the gel sections were destained with two washes of 50% methanol and 7.5% acetic acid, followed by three washes with 50 mM ammonium bicarbonate and 100% acetonitrile to remove the running buffer. Gel slices were dried for 10 min in a SpeedVac, and 200  $\mu$ l of sequencing grade trypsin (Promega) at a concentration of 6.6  $\mu$ g/ml in 50 mM ammonium bicarbonate was added to each gel slice. The gel slices were allowed to swell for 60 min on ice, after which the tubes were incubated at 37°C for 24 h. Peptides were extracted with two washes of 500  $\mu$ l of 50 mM ammonium bicarbonate and two washes of 500  $\mu$ l of 50% acetonitrile-0.1% formic acid. All extracts were frozen at -80 degrees, lyophilized to dryness, and redissolved in 60  $\mu$ l of 5% acetonitrile-0.1% formic acid. Samples were then loaded into a 96-well plate for MS analysis on an LTQ XL instrument (Thermo Fisher Scientific). For each run, 10  $\mu$ l of each reconstituted sample was injected with a Famos Autosampler; separation was done on a 75- $\mu$ m (diameter) by 20-cm-long column packed with C<sub>18</sub> medium running at a flow rate of 250 nl/min provided from a Surveyor MS pump with a flow splitter. A gradient of 5 to 28% acetonitrile and

0.35% formic acid was applied over the course of 80 min. Between each set of samples, standards from a mixture of peptides, 5 Angiotensin (Michrom Biore-sources), were run to ascertain column performance and observe any potential carryover that might have occurred. The LTQ XL was run in a top-eight configuration with one MS scan and eight tandem MS scans. Dynamic exclusion was set to 1 with a limit of 30 s.

An LTQ-Fourier transform mass spectrometer (Thermo Fisher Scientific) was used for the analysis of the eluates of the antibody column (see below); the injection volumes and sample gradients were as described above.

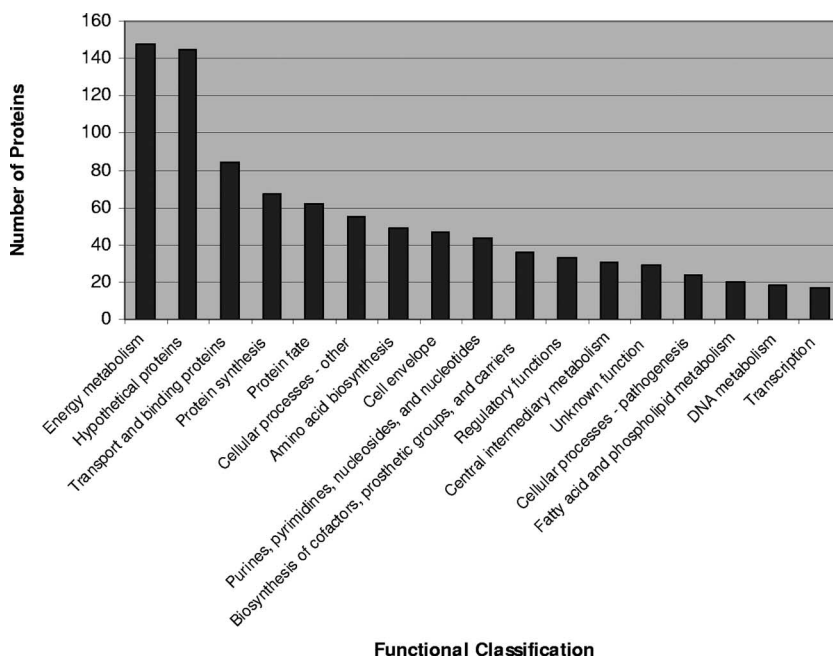
**Sera from patients with *V. cholerae*.** Acute-phase and convalescent-phase serum samples were obtained from five cholera patients enrolled in a previously described study of immune responses to *V. cholerae* infection (22). Briefly, these cholera patients were enrolled immediately upon presentation to the ICDDR,B hospital; serum samples were obtained at the time of enrollment and again 21 days later during convalescence. All patients were confirmed to be infected with *V. cholerae* O1 El Tor by stool culture, and all were noted to have more than a fourfold increase in vibriocidal antibody titer by day 21 compared to the acute stage of infection. Human studies approval for collection of serum samples was obtained from the Institutional Review Board of the Massachusetts General Hospital and the Ethical and Research Review Committees of the ICDDR,B. Serum samples were collected from different patients than those who provided the stool samples described above.

**Antibody capture columns.** Acute- and convalescent-phase serum samples from the five cholera patients were each pooled, and the acute- and convalescent-phase immunoglobulin G (IgG) fractions were purified using a HiTrap protein G high-performance column (Amersham Biosciences) according to the manufacturer's instructions. The purified acute- and convalescent-phase IgG fractions were each then separately coupled to a 1-ml HiTrap N-hydroxysuccinimide-activated high-performance column (Amersham Biosciences). Three stool protein samples (collected as described above) were used for this analysis. A total of 1.4 mg of each of the three lyophilized stool protein samples was rehydrated, diluted with 10 ml of PBS-0.2% octyl- $\beta$ -D-glucopyranoside-1 $\times$  protease inhibitor cocktail (Roche), and divided into two equal fractions for loading onto the acute- and convalescent-phase IgG columns. Each of the three stool protein samples was loaded separately over the two columns at a flow rate of 1 ml/min. The column was then rinsed with PBS-0.2% octyl- $\beta$ -D-glucopyranoside, and the IgG-bound proteins were eluted with 1 M acetic acid, pH 3.0. Between sample runs, an excess of 1 M acetic acid was run over the column, and the column was equilibrated with PBS-0.2% octyl- $\beta$ -D-glucopyranoside. The eluates of proteins bound to the IgG fractions of acute- and convalescent-phase sera on the two columns were each separately pooled, concentrated with a spin filter, reduced and alkylated, and loaded onto a one-dimensional SDS-polyacrylamide gel electrophoresis Tris-glycine 8 to 16% gradient gel (Invitrogen) for analysis by MS, as described above.

**Peptide identification.** Peptide identifications were made using SEQUEST (Thermo Fisher Scientific) through the Bioworks Browser, version 3.2. MS data were searched with a 2-Da window on the MS precursor with a 1.0-Da window on the fragment ions. A reverse database strategy was employed (7). The *V. cholerae* O1 El Tor N16961 and human sequences were reversed and concatenated with the forward sequences supplemented with common contaminants and filtered to obtain a false discovery rate less than or equal to 1%. The results presented here focus on the *V. cholerae* peptides that were identified in stool.

#### RESULTS

Thirty-two separate patient stool samples were included in this analysis of the *V. cholerae* proteome in human stool. The samples were obtained from adult patients with a mean age of  $29 \pm 11$  years; half of the patients were female. A total of 5,767 *V. cholerae* peptides, corresponding to 909 *V. cholerae* proteins, were identified in the stool samples. Table S1 in the supplemental material lists all of the peptides that were identified in the stool samples, including the corresponding *V. cholerae* proteins and the number of patient samples in which the peptides were encountered, as well as other details about the peptide identification. A total of 69% of proteins were identified by two or more unique peptides, and 68% of peptides were identified in two or more separate stool samples. In most cases when a peptide was encountered multiple times, there was a narrow



Functional Classification

FIG. 1. Functional classification of 909 *V. cholerae* proteins identified in human stool, according to The Institute for Genomic Research (www.tigr.org).

range of elution time, gel slice position, and charge state (see Table S1 in the supplemental material), suggesting that the chromatography was reproducible. Furthermore, the SEQUEST XCorr (cross-correlation) values were high (median, 4.1; range, 2.3 to 9.5), indicating the quality of the matches between the tandem mass spectra and the amino acid sequences in the database (8). Many of the proteins identified by only a single peptide are also likely to be correct for the following reasons: they were identified in an appropriate gel slice, given the protein’s molecular weight; they were identified in stool samples where other *V. cholerae* proteins were abundant; and they had favorable SEQUEST XCorr scores.

Figure 1 shows the functions of the *V. cholerae* proteins identified in human stool. The majority of *V. cholerae* proteins of known function identified in human stool are involved in protein synthesis and energy metabolism, including particularly anaerobic metabolism, fermentation, glycolysis, and gluconeogenesis. Proteins that have been proposed as hypothetical (54) or conserved hypothetical (91) proteins based on their DNA sequences comprise a large group. A number of these hypothetical proteins were previously identified in the construction of the two-dimensional gel proteome maps developed for in vitro grown *V. cholerae* (5, 10), including VC1101, VC1334, VCA0689, VCA0881, VCA1054, VCA0144, and VC2528. Twenty-six *V. cholerae* proteins involved in chemotaxis and motility, including a number of flagellar proteins, were also present in human stool.

As outlined in Table 1, 25 *V. cholerae* proteins related to pathogenesis were detected in the human samples. In particular, the B subunit of cholera toxin (CtxB) was found in 20 (63%) human stool samples, and the A subunit (CtxA) was found in 6 (19%). The more frequent identification of the B subunit may relate to the structural composition of cholera

TABLE 1. 25 pathogenesis proteins of *V. cholerae* identified in human stool samples

TIGR locus <sup>a</sup>	TIGR common name <sup>b</sup>	No. (%) of stool samples in which a corresponding peptide was identified	No. of unique peptide identifications
VC0831	Toxin coregulated pilus biosynthesis outer membrane protein C	25 (78)	47
VC0828	Toxin coregulated pilin (TcpA)	22 (69)	13
VC0409	MSHA pilin protein MshA	20 (63)	4
VC1456	Cholera enterotoxin, B subunit	20 (63)	5
VC0402	MSHA biogenesis protein MshL	12 (38)	12
VC0841	Accessory colonization factor AcfC	12 (38)	16
VC0824	TagD protein	11 (34)	15
VC0983	Regulatory protein ToxS	11 (34)	1
VC0829	Toxin coregulated pilus biosynthesis protein B	9 (28)	7
VC0833	Toxin coregulated pilus biosynthesis protein D	9 (28)	12
VC0837	Toxin coregulated pilus biosynthesis protein F	9 (28)	11
VC1621	Agglutination protein	9 (28)	16
VC0844	Accessory colonization factor AcfA	7 (22)	4
VCA0219	Hemolysin	7 (22)	9
VC0400	MSHA biogenesis protein MshJ	6 (19)	2
VC0830	Toxin coregulated pilus biosynthesis protein Q	6 (19)	3
VC1457	Cholera enterotoxin, A subunit	6 (19)	5
VC0825	Toxin coregulated pilus biosynthesis protein I	3 (9)	1
VC0827	Toxin coregulated pilus biosynthesis protein H	3 (9)	2
VC1130	DNA-binding protein VicH	3 (9)	2
VC0840	Accessory colonization factor AcfB	3 (9)	4
VC0959	Hemolysin, putative	3 (9)	2
VC0826	Toxin coregulated pilus biosynthesis protein P	2 (6)	1
VC0834	Toxin coregulated pilus biosynthesis protein S	2 (6)	3
VCA0446	Hemagglutinin	2 (6)	1

<sup>a</sup> TIGR, The Institute for Genomic Research.

<sup>b</sup> MSHA, mannose-sensitive hemagglutinin.

toxin, which is comprised of one enzymatically active A subunit coupled to a pentamer of B subunits. The major subunit, TcpA, of the toxin-coregulated pilus, which is essential for *V. cholerae* colonization of the human intestine, was also encountered in 22 (69%) human stool specimens. TcpA is encoded as part of a pathogenicity island (12), and a total of 13 proteins encoded by this island were present in human stool, including TcpF, a more recently recognized soluble colonization factor of *V. cholerae* (14, 15). Two transcription complexes, ToxRS and TcpPH, have been shown in vitro to act through a common downstream regulator, ToxT, to control expression of virulence factors in *V. cholerae* (3, 6). TcpP, TcpH, and ToxS were among the pathogenesis-related proteins identified in the human stool specimens.

Table S2 in the supplemental material specifies the number of peptides corresponding to each protein that were identified in each of the 32 patients. The outer membrane protein OmpU was the most frequently identified *V. cholerae* protein, and a peptide corresponding to OmpU was found in every patient sample. Of note, *ompU* was predicted to be among the top 20 most highly expressed genes of *V. cholerae* based on a theoretical study of codon usage frequencies (13). Other *V. cholerae* outer membrane proteins, including OmpV, OmpS, OmpA, OmpK, OmpC, and OmpW, were also repeatedly identified in multiple human stool samples.

A wide range of *V. cholerae* peptides was encountered in the different human stool samples. A median of 677 *V. cholerae* peptides was encountered in the 32 samples, with a range of 5 to 6,860 peptides in different patients. No correlation between the number of *V. cholerae* peptides identified and the quantity of vibrios recovered from the stool samples was observed ( $r = 0.04$ ). There was a nonsignificant trend for female patients to have more *V. cholerae* peptides identified in stool than male patients (1,552 versus 358 peptides;  $P = 0.07$  by a two-tailed  $t$  test). Since the proteomic content of human stool has not been characterized to date, we additionally searched our peptide data against the UniProtKB/Swiss-Prot protein knowledge base (<http://www.expasy.ch/sprot/>; release date 29 April 08) in order to ascertain if other microbial or food-derived proteins were competing with *V. cholerae* for identification. This database includes selected sequence data derived from >11,000 species, including *Homo sapiens*, *V. cholerae*, *Escherichia coli* (strain K12), and *Oryza sativa* (rice). A majority (57%) of the sequences in the knowledge base are of bacterial origin. The most abundantly detected proteins were those of *V. cholerae*, with some additional identification of *E. coli* proteins; there was little evidence of the presence of other bacterial species. Three rice-specific proteins were repeatedly identified in patient samples but only in small amounts.

In order to assess whether the proteome of *V. cholerae* identified in human stool correlated with the gene expression profile of the organism, we used data from a previous study in which we evaluated the transcriptional profile of *V. cholerae* present in five human stool samples different from the samples used in these experiments (17). In these gene profiling experiments, *V. cholerae* O1 El Tor N16961 genomic DNA was hybridized to each microarray as an internal standard, and hence a rank ordering of the most abundantly expressed genes with respect to genomic DNA was possible. Of the 750 most abundantly expressed *V. cholerae* genes in human stool, 139

TABLE 2. *V. cholerae* proteins present in human stool and recognized uniquely by the IgG fraction of convalescent-phase human serum

TIGR locus <sup>a</sup>	SWISS-PROT accession no.	TIGR common name	No. of unique peptides
VC0027	Q9KVV1	Threonine dehydratase	1
VC0083	Q9KVO6	Ubiquinone/menaquinone biosynthesis methyltransferase UbiE	1
VC0093	Q9KVP8	Glycerol-3-phosphate acyltransferase	1
VC0306	Q9KV51	Thioredoxin	1
VC0809	Q9KTS9	Hypothetical protein	1
VC1354	Q9KSA5	Conserved hypothetical protein	1
VC1456	P01556	Cholera enterotoxin, B chain precursor	3
VC1457	P01555	Cholera enterotoxin, A chain precursor	1
VC1604	Q9KRN2	Response regulator	1
VC1819	Q9KR28	Aldehyde dehydrogenase	1
VC2073	Q9KQC7	Conserved hypothetical protein	1
VCA0002	Q9KNG2	Hypothetical protein	1
VCA0455	Q9KMB8	Conserved hypothetical protein	1
VCA0809	Q9KLD7	Conserved hypothetical protein	1
VCA1028	Q56652	Maltoporin precursor (OmpS)	1

<sup>a</sup> TIGR, The Institute for Genomic Research.

(19%) had detectable protein products in our current study (see Table S3 in the supplemental material for a list). These results indicate that there may be a disparity between the abundance of transcripts by gene expression profiling and the abundance of at least some protein products by proteomic analysis.

We also assessed the immunogenicity of the *V. cholerae* proteins identified in three human stool samples using acute- and convalescent-phase human serum bound to an IgG antibody capture column. Table 2 lists the 15 *V. cholerae* proteins from human stool that were recognized uniquely by convalescent-phase but not acute-phase IgG. The most robustly identified protein was CtxB, a well-recognized immunogen in human cholera infection (4). Other known *V. cholerae* antigens identified by this approach were CtxA and VCA1028 (OmpS). Convalescent-phase serum also recognized a number of novel antigens uniquely, including six previously uncharacterized hypothetical or conserved hypothetical proteins. Of note, 8 of the 15 proteins recognized by convalescent-phase serum were not among the proteins identified in the human stool samples by MS, suggesting that the antibody column procedure may have identified immunogenic proteins that were of low abundance in stool.

## DISCUSSION

Humans are the only known host and reservoir for *V. cholerae* outside of the aquatic environment. Rabbit and mouse models of *V. cholerae* infection have been developed, but both have significant limitations and may not fully replicate the interaction between the organism and the human intestine. Elucidating the proteins that are expressed by *V. cholerae* directly in the human host is therefore useful for understanding the pathogenesis of cholera and for identifying protein targets

TABLE 3. 24 proteins identified in two strains of in vitro grown *V. cholerae* and in *V. cholerae* detected in human stool in the present study<sup>a</sup>

TIGR locus <sup>b</sup>	TIGR common name	Protein role	No. (%) of stool samples in which protein was identified
VC0633	Outer membrane protein OmpU	Cell envelope	32 (100)
VC0432	Malate dehydrogenase	Energy metabolism	29 (91)
VC2664	Chaperonin, 60-kDa subunit	Protein fate	29 (91)
VC2412	Pyruvate dehydrogenase	Energy metabolism	28 (88)
VC0731	Antioxidant, AhpC/Tsa family	Cellular processes	26 (81)
VC2213	Outer membrane protein OmpA	Cell envelope	26 (81)
VC0647	Polyribonucleotide nucleotidyltransferase	Transcription	25 (78)
VC0855	DnaK protein	Protein fate	25 (78)
VC2738	Phosphoenolpyruvate carboxykinase	Energy metabolism	25 (78)
VC2764	ATP synthase F1, beta subunit	Energy metabolism	24 (75)
VC0478	Fructose-bisphosphate aldolase, class II	Energy metabolism	23 (72)
VCA0744	Glycerol kinase	Energy metabolism	21 (66)
VC0430	Immunogenic protein	Cell envelope	19 (59)
VC1905	Alanine dehydrogenase	Energy metabolism	18 (56)
VC1915	Ribosomal protein S1	Protein synthesis	18 (56)
VC0171	Peptide ABC transporter	Transport and binding proteins	17 (53)
VC1141	Isocitrate dehydrogenase	Energy metabolism	17 (53)
VC1425	Spermidine/putrescine ABC transporter	Transport and binding proteins	15 (47)
VC2259	Elongation factor Ts	Protein synthesis	15 (47)
VC2746	Glutamate-ammonia ligase	Amino acid biosynthesis	15 (47)
VC0968	Cysteine synthase A	Amino acid biosynthesis	14 (44)
VC2342	Elongation factor G	Protein synthesis	14 (44)
VC0985	Heat shock protein HtpG	Protein fate	13 (41)
VC1097	Phosphate acetyltransferase	Energy metabolism	1 (3)

<sup>a</sup> In vitro grown *V. cholerae* of the classical strain (9) and El Tor strain (4).

<sup>b</sup> TIGR, The Institute for Genomic Research.

of protective immunity. We present here the first analysis of the proteome of *V. cholerae* organisms obtained directly from human stool; to our knowledge, this is also the first proteome of a pathogenic bacteria recovered from a natural host.

The majority of the *V. cholerae* proteins of known function encountered in human stool were proteins involved in energy metabolism or protein synthesis. This observation is consistent with previous work that has predicted such proteins to be highly expressed in fast-growing bacteria (13). As predicted and as noted in our results, ribosomal proteins, major transcription/translation processing factors, enzymes of essential energy metabolism pathways, and the principal genes of amino acid and nucleotide biosynthesis are among the most commonly encountered *V. cholerae* proteins in human stool. Many virulence factors known to be involved in the pathogenesis of cholera were also identified in human stool, lending validation to the proteomic results. Recent work has suggested a relationship between *V. cholerae* motility and chemotaxis and the pathogenesis of the organism, as well as the organism's transition to the aquatic environment (2); many proteins in this class were identified in stool. Of note, a significant proportion of the *V. cholerae* proteins identified in this analysis were previously categorized as hypothetical or conserved hypothetical proteins. The repeated identification of these proteins in multiple human samples can be considered confirmation of their existence.

Two previous studies have described the development of two-dimensional electrophoresis maps of in vitro grown *V. cholerae*, based on the identification of a limited number of "landmark" proteins that exhibit intense spots on a gel and that are presumably of high abundance. Hommais et al. created a

two-dimensional map based on 40 *V. cholerae* proteins that were identified by matrix-assisted laser desorption/ionization—time of flight MS and used this to identify pH-regulated proteins in a classical strain of *V. cholerae* O1 (10). Thirty-eight of the 40 landmark *V. cholerae* proteins selected for this study were identified in the cholera stool samples described here. Similarly, Coelho et al. identified 80 unique proteins of the *V. cholerae* O1 El Tor reference strain N16961 in their construction of a proteome reference map (5); 75 of these landmark *V. cholerae* proteins were present in cholera stool as described here. As outlined in Table 3, 24 proteins were selected as landmark proteins in both of these reference maps and were additionally identified in human stool. These repeatedly identified proteins are likely to represent proteins that are expressed in high abundance during in vitro growth as well as during infection of the human host and that may perform essential functions.

Recently, microarray expression profiling and in vivo expression technology have been used to elucidate the genes that are expressed by *V. cholerae* directly in the human host environment (1, 17, 19, 20). These studies have identified a number of candidate genes that are preferentially expressed during human infection. Elucidating the proteins of *V. cholerae* that are present in the human host provides complementary data to these studies of in vivo gene expression, as underscored by the fact that 57 (26%) of the 217 genes identified by in vivo expression technology were identified in the stool of cholera patients. Not surprisingly, our results indicate that the correlation between in vivo gene transcript and protein abundance is not linear since we found little overlap between the most abundantly transcribed *V. cholerae* genes in human stool and

the *V. cholerae* proteins that we detected in stool by MS. The discordance between gene expression and protein levels is also well illustrated by the case of *ctxAB*, the genes encoding cholera toxin. In previous transcriptional profiling studies, transcripts of *ctxAB* were detected at relatively low levels in stool-derived *V. cholerae* compared to optimal in vitro toxin-producing conditions (1). In our proteomic evaluation of human cholera stool, both the A and B subunits of cholera toxin were identified. This may indicate that only modest quantities of the *ctxAB* transcripts are associated with clinical infection; more likely, this finding supports the hypothesis that higher levels of *ctxAB* transcription occur transiently in organisms adherent to the intestinal mucosa, but levels of transcription are lower by the time the organisms are recovered from stool, despite persistence of the protein products.

Microbial genes that are uniquely expressed in the human host and that have long-lived protein products may be particularly relevant in the induction of immune responses. Following on this hypothesis, we used an MS-based approach to identify a subset of *V. cholerae* proteins present in human stool that were uniquely recognized by convalescent-phase but not acute-phase serum collected from cholera patients. This approach to identifying antigens worthy of further study was validated by our identification of the B subunit of cholera toxin, a well-characterized immunogen in *V. cholerae* infection. Further evaluation of the immunogenicity of the other antigens identified by this approach is warranted.

There was substantial variability in the number of *V. cholerae* proteins identified in the 32 stool samples. The reason for this is unclear. Surprisingly, we observed no correlation between the yield of *V. cholerae* peptides and the quantity of vibrios cultured from the stool specimen. All efforts were made to exclude patients who had previously received antibiotics; however, if patients' reporting of treatment at home was unreliable, prior use of antibiotics might have contributed to this variability. An abundance of protein material in stool, aside from human or vibrio-derived proteins, could also have differentially affected our ability to detect *V. cholerae* proteins in human stool. To our knowledge, no published data on the proteome of normal human stool is available to guide such an assessment, and hence it is difficult to exclude this as an explanation for the variability in our results. However, a search of our peptide data against the UniProtKB/Swiss-Prot protein knowledge base did not provide evidence for a large number of spectra other than those assigned to *V. cholerae*, *H. sapiens*, or, less commonly, *E. coli*.

The large number of patient stool samples evaluated in our study gives confidence to the proteins identified, but some limitations merit comment. First, since our MS approach is semiquantitative, we can detect only large differences in protein abundance within or between samples. Second, we cannot exclude the presence of additional, less abundant *V. cholerae* proteins that were not identified by MS or even abundant proteins that have few tryptic peptides in our analytical range. Lastly, our antibody column approach was limited to the identification of proteins recognized by IgG; proteins recognized by IgA or IgM may also have relevance to a mucosal infection such as cholera. Additionally, lipid and carbohydrate antigens are targets of human immune responses but would not have been evaluated by the techniques used here.

In summary, we present here a thorough proteome of *V. cholerae* recovered directly from the human host environment. Using an MS-based approach, we identified many of the known *V. cholerae* virulence factors in samples of cholera stool, lending credence to our findings. Our results demonstrate that microbial protein abundance does not directly relate to gene transcript levels as previously described in *V. cholerae* stool, and hence the proteomic approach described here offers an important complement to studies of *V. cholerae* genes that are expressed in vivo. Significantly, we have demonstrated that a high-throughput characterization of the presence and immunogenicity of microbial proteins in human samples is feasible. This may represent a novel approach to identifying vaccine targets that can be applied more broadly to other human pathogens or to microorganisms recovered from other host compartments.

#### ACKNOWLEDGMENTS

This research was supported by grants K01-TW07144 from the Fogarty International Center (R.C.L.), a Claflin Distinguished Scholar Award from the Massachusetts General Hospital (R.C.L.), K01-TW07409 from the Fogarty International Center (J.B.H.), AI40725 from the National Institute of Allergy and Infectious Diseases (E.T.R.), U01-AI058935 from the National Institute of Allergy and Infectious Diseases (S.B.C.), U54-AI057159 from the New England Regional Center of Excellence for Biodefense and Emerging Infectious Diseases (S.B.C., K.C.P., and D.S.), and by the ICDRR,B (F.Q.).

#### REFERENCES

- Bina, J., J. Zhu, M. Dziejman, S. Faruque, S. Calderwood, and J. Mekalanos. 2003. ToxR regulon of *Vibrio cholerae* and its expression in vibrios shed by cholera patients. *Proc. Natl. Acad. Sci. USA* **100**:2801–2806.
- Butler, S. M., and A. Camilli. 2005. Going against the grain: chemotaxis and infection in *Vibrio cholerae*. *Nat. Rev. Microbiol.* **3**:611–620.
- Carroll, P. A., K. T. Tashima, M. B. Rogers, V. J. DiRita, and S. B. Calderwood. 1997. Phase variation in *tcpH* modulates expression of the ToxR regulon in *Vibrio cholerae*. *Mol. Microbiol.* **25**:1099–1111.
- Cash, R. A., S. I. Music, J. P. Libonati, M. J. Snyder, R. P. Wenzel, and R. B. Hornick. 1974. Response of man to infection with *Vibrio cholerae*. I. Clinical, serologic, and bacteriologic responses to a known inoculum. *J. Infect. Dis.* **129**:45–52.
- Coelho, A., E. de Oliveira Santos, M. L. Faria, D. P. de Carvalho, M. R. Soares, W. M. von Kruger, and P. M. Bisch. 2004. A proteome reference map for *Vibrio cholerae* El Tor. *Proteomics* **4**:1491–1504.
- DiRita, V. J., C. Parsot, G. Jander, and J. J. Mekalanos. 1991. Regulatory cascade controls virulence in *Vibrio cholerae*. *Proc. Natl. Acad. Sci. USA* **88**:5403–5407.
- Elias, J. E., and S. P. Gygi. 2007. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4**:207–214.
- Eng, J. K., A. L. McCormack, and J. R. Yates. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**:976–979.
- Heidelberg, J. F., J. A. Eisen, W. C. Nelson, R. A. Clayton, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, L. Umayam, S. R. Gill, K. E. Nelson, T. D. Read, H. Tettelin, D. Richardson, M. D. Ermolaeva, J. Vamathevan, S. Bass, H. Qin, I. Dragoi, P. Sellers, L. McDonald, T. Utterback, R. D. Fleischmann, W. C. Nierman, and O. White. 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**:477–483.
- Hommais, F., C. Laurent-Winter, V. Labas, E. Krin, C. Tendeng, O. Soutourina, A. Danchin, and P. Bertin. 2002. Effect of mild acid pH on the functioning of bacterial membranes in *Vibrio cholerae*. *Proteomics* **2**:571–579.
- Kaper, J. B., J. G. Morris, Jr., and M. M. Levine. 1995. Cholera. *Clin. Microbiol. Rev.* **8**:48–86.
- Karaolis, D. K., S. Somara, D. R. Maneval Jr., J. A. Johnson, and J. B. Kaper. 1999. A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. *Nature* **399**:375–379.
- Karlin, S., J. Mrazek, A. Campbell, and D. Kaiser. 2001. Characterizations of highly expressed genes of four fast-growing bacteria. *J. Bacteriol.* **183**:5025–5040.
- Kirn, T. J., N. Bose, and R. K. Taylor. 2003. Secretion of a soluble coloni-

- zation factor by the TCP type 4 pilus biogenesis pathway in *Vibrio cholerae*. *Mol. Microbiol.* **49**:81–92.
15. **Kirn, T. J., and R. K. Taylor.** 2005. TcpF is a soluble colonization factor and protective antigen secreted by El Tor and classical O1 and O139 *Vibrio cholerae* serogroups. *Infect. Immun.* **73**:4461–4470.
  16. **Kudva, I. T., B. Krastins, H. Sheng, R. W. Griffin, D. A. Sarracino, P. I. Tarr, C. J. Hovde, S. B. Calderwood, and M. John.** 2006. Proteomics-based expression library screening (PELS): a novel method for rapidly defining microbial immunoproteomes. *Mol. Cell. Proteomics* **5**:1514–1519.
  17. **LaRocque, R. C., J. B. Harris, M. Dziejman, X. Li, A. I. Khan, A. S. Faruque, S. M. Faruque, G. B. Nair, E. T. Ryan, F. Qadri, J. J. Mekalanos, and S. B. Calderwood.** 2005. Transcriptional profiling of *Vibrio cholerae* recovered directly from patient specimens during early and late stages of human infection. *Infect. Immun.* **73**:4488–4493.
  18. **LaRocque, R. C., J. B. Harris, E. T. Ryan, F. Qadri, and S. B. Calderwood.** 2006. Postgenomic approaches to cholera vaccine development. *Expert Rev. Vaccines* **5**:337–346.
  19. **Lombardo, M. J., J. Michalski, H. Martinez-Wilson, C. Morin, T. Hilton, C. G. Osorio, J. P. Nataro, C. O. Tacket, A. Camilli, and J. B. Kaper.** 2007. An in vivo expression technology screen for *Vibrio cholerae* genes expressed in human volunteers. *Proc. Natl. Acad. Sci. USA* **104**:18229–18234.
  20. **Merrell, D. S., S. M. Butler, F. Qadri, N. A. Dolganov, A. Alam, M. B. Cohen, S. B. Calderwood, G. K. Schoolnik, and A. Camilli.** 2002. Host-induced epidemic spread of the cholera bacterium. *Nature* **417**:642–645.
  21. **Ryan, E. T., S. B. Calderwood, and F. Qadri.** 2006. Live attenuated oral cholera vaccines. *Expert Rev. Vaccines* **5**:483–494.
  22. **Saha, D., R. C. LaRocque, A. I. Khan, J. B. Harris, Y. A. Begum, S. M. Akramuzzaman, A. S. Faruque, E. T. Ryan, F. Qadri, and S. B. Calderwood.** 2004. Incomplete correlation of serum vibriocidal antibody titer with protection from *Vibrio cholerae* infection in urban Bangladesh. *J. Infect. Dis.* **189**:2318–2322.

---

*Editor:* W. A. Petri, Jr.