A proteome reference map for *Vibrio cholerae* El Tor

Ana Coelho[1, 2], Eidy de Oliveira Santos[1, 2], Mauro Luiz da Hora Faria[1, 2], Daniela Palermo de Carvalho[1, 3], Marcia Regina Soares[1, 3], Wanda Maria Almeida von Kruger[1, 3] and Paulo Mascarello Bisch[1, 3]

[1]Proteomics Network of Rio de Janeiro-FAPERJ
[2]Dep. Genética, I. Biologia
[3]I. Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

A proteome reference map has been constructed for *Vibrio cholerae* El Tor, in the p*I* range of 4.0 to 7.0. The map is based on two-dimensional gels (2-D) and the identification, by peptide mass fingerprint, of proteins in 94 spots, corresponding to 80 abundant proteins. Two strains are compared, strain N16961 and a Latin American El Tor strain C3294. The consensus map contains 340 spots consistently seen with both strains grown in Luria-Bertani broth (LB) or minimal M9 medium. The results were obtained from nine gels run with 18 cm immobilized pH gradient strips and precast gels. The 2-D gels were anchored to real N16961 proteins identified by mass spectrometry. Various energy metabolism components and periplasmic ATP-binding cassette (ABC) transporter proteins were identified among the abundant proteins. Two isoforms of OmpU were found. Five operons are proposed and seven hypothetical proteins were experimentally confirmed. Comparisons are made with protein 2-D gels for a classical strain and to microarray analysis available for the N16961 El Tor strain. New results were obtained from the proteome analysis, indicating an abundance of periplasmic ABC transporter proteins not found in microarray studies.

Keywords: Matrix-assisted laser desorption/ionization / Peptide mass fingerprint / Proteome / Two-dimensional gel electrophoresis / *Vibrio cholerae*

Received	30/8/03
Revised	2/11/03
Accepted	4/11/03

1 Introduction

Vibrio cholerae is an important human pathogen, causing the disease cholera, responsible for thousands of humans deaths every year. Strains of the El Tor biotype are responsible for the current seventh pandemic, which started in 1960. An El Tor strain was responsible for the Latin American cholera epidemic during the early 1990s. Many key discoveries in recent years have attracted renewed interest in this species [1, 2]. Among these are the appearance of a new epidemic strain O139, the finding that the toxin genes are carried by a phage [3], the formation of biofilms and quorum sensing regulation [4], and new studies with pathogenicity islands, toxins, colonization factors, and their regulation [5–8], including the study of the ToxR regulon *in vitro* and *in vivo* [9–11]. Whole-genome approaches are being used in the study of representative bacteria, and, in particular, in the case of *V. cholerae*. The complete genome of the El Tor strain N16961 was sequenced in 2000 [12], and the chromosomes of other strains or other species of the *Vibrio* genus, such as *V. vulnificus* and *V. parahaemolyticus* are being mapped or sequenced [13–16]. The microarray work of Merrell *et al.* [17] has laid the groundwork for complete mRNA studies applied to *V. cholerae*. These authors propose two different "expression stages" of the bacteria, with one of these occurring after human infection. Work on transcriptomes has progressed, including studies in a rabbit *in vivo* model [11, 18].

Abundant proteins from *V. cholerae* have been used in strain characterization in multilocus enzyme electrophoresis (MLEE) studies [19]. Gene regulation studies at the protein level started with the heat shock response [20] and the ToxR regulon. Convalescent human antisera have been recently employed to detect proteins

Correspondence: Dr. Ana Coelho, Department of Genetics, Instituto de Biologia, CCS, Universidade Federal do Rio de Janeiro, Cx Postal 68011, Ilha do Fundão, Rio de Janeiro, RJ, 21944-970, Brazil
E-mail: coelho@biologia.ufrj.br
Fax: +55-21-2562-6396

Abbreviations: ABC, ATP-binding cassette; **CMR**, comprehensive microbial resource; **LB**, Luria- Bertani broth; **MLEE**, multilocus enzyme electrophoresis; **TCA**, tricarboxylic acid

expressed during the disease [21]. The regulation of various genes by inorganic phosphate has been addressed with the use of 2-D gels [8]. Protein analysis has been described for a classical strain, comparing mild acid and neutral pH conditions [22]. Here, we analyzed proteins of El Tor strains in order to have a more complete picture of gene expression and to make comparisons with the available transcription data. This is a first step towards the description of the proteome of *V. cholerae* El Tor expressed under different conditions.

2 Materials and methods

2.1 Strains, media, and growth conditions

Two strains of *V. cholerae* El Tor were used: strain N16961, isolated in Bangladesh in 1975 [23] (origin Dr. J. Kaper, Univ. Maryland), and strain C3294, from the epidemic in Brazil, isolated in Rio de Janeiro in 1993 (Laboratorio Central Noel Nutels, RJ). N16961 was grown for 16 h either at 30°C in M9 minimal medium, with 0.4% glucose and 0.4% casamino acids or at 37°C in Luria-Bertani broth (LB)-rich medium (stationary phase). C3294 was grown at 37°C for 16 h in LB medium. Both were grown in a shaker at 125 rpm. The laboratory is certified for the use of Group II genetically modified organisms (CTNBio, certificate 0076/98, Instituto de Biologia, UFRJ).

2.2 Cell lysis

Aliquots of 1.2 mL were taken from cultures with an OD_{580} of 1.5. The cells were spun down, washed once with NaCl 0.9%, and resuspended in 500 μL lysis buffer (8 M urea, 4% CHAPS, 65 mM DTT, 1% Pharmalyte 3–10 (Amersham Biosciences, Uppsala, Sweden), 1 mM PMSF added just before use). Incubations were carried out for 2 h at room temperature (25°C), and cell lysis was confirmed by microscopy. Insoluble cellular debris were separated by centrifugation at 15K rpm, 10°C for 60 min [8], and the clear supernatants were kept at −80°C until use. Three sample preparations were made for strain N16961 (two in M9 and one in LB) and two for strain C3294 (LB). For preparative gels for protein identification, a sample cleanup/concentration step was done, using a CleanUp kit (Amersham Biosciences), and dissolving fivefold concentrated proteins in lysis buffer.

2.3 2-D PAGE

The first-dimensional isoelectric focusing and second-dimensional SDS-gel electrophoresis were performed according to the Amersham Biosciences manual [24]. Immobilized pH gradient strips in the p*I* range of 4–7 (linear, 18 cm) and an IPGphor instrument were used for isoelectric focusing. Strips were rehydrated at 60 V for 12 h, in the presence of the samples. The total volume for rehydration was 340 μL, of which 35 μL was a typical volume for the sample. The separations were carried out for 80 000 Vh. The strips were kept at −80°C until use. Two types of second-dimensional SDS-gels were used. Visualization gels were run with a Multiphor II equipment and precast ExcelGel XL 12–14% gradient SDS gels ($180 \times 250 \times 0.5$ mm). These gels were silver-stained using the EMBL method [25]. For preparative gels, an Ettan DALT*six* equipment was employed, with 26 cm × 20 cm × 1.5 mm, 12.5% (30% acrylamide: 0.8% bisacrylamide) gels, run with a Tris-glycine buffer system, for 30 min at 5 W/gel and 5 h at 20 W/gel. Preparative gels were Coomassie-stained (Coomassie Brilliant Blue R-250). 200 μg of protein were used for visualization gels, and 1 mg for preparative gels.

2.4 Gel analysis

The gels were scanned with an Image Scanner and analyzed with the Image Master 2D database software (Amersham Biosciences), followed by an additional visual analysis. Spots were considered real if detected in at least two gels for each strain. Gels from each of the two strains were analyzed and averaged separately and then compared. p*I* values were determined using a linear 4–7 distribution, and molecular weight (MW) determinations were based on a full-range Rainbow protein MW marker (Amersham Biosciences), using a logarithmic curve. The spots on the gels shown were marked with Adobe Photoshop over one gel from each strain.

2.5 Protein identification by peptide mass fingerprinting

Protein spots were cut from the gels and kept frozen at −80°C until used. The spots were processed as in [26], but reduction and alkylation of the samples were done before SDS-gel electrophoresis. 10–15 μL of 15 μg/mL porcine trypsin (Promega, Madison, WI, USA) were added to the dried gel fragments, and, after swelling, these fragments were covered with 15 μL of 25 mM ammonium bicarbonate. The samples were left overnight at 37°C and the peptides were then extracted with 50% ACN/5% TFA. A mixture of 1 μL of the sample and 1 μL of α-cyano-4-hydroxycinnamic acid matrix was spotted on a MALDI sample plate with a hydrophobic mask and allowed to crystallize at room temperature. A Voyager-DE Pro MALDI-TOF (Applied Biosystems, Foster City, CA, USA) equipment was used for the peptide mass finger-

print analysis in a reflector mode, with an accelerating voltage of 20 000 V and a laser intensity of 1700 V (grid voltage 76%, guide wire voltage 0.004%, delay time 150 ns). Peptides in the range of 800–4000 Da were analyzed. Spectra were acquired after calibration at a nearby spot with calibration mixture 1 or 2 (Sequazyme Peptide Mass Standard kit; PerSeptive Biosystems, Foster City, CA, USA) and with a minimum of 1500 shots for the actual samples. The peptide mass fingerprints were matched to the database with the Protein Prospector MS-Fit software. The NCBI database was usually searched for *Vibrio cholerae*, with a mass accuracy better than 50 ppm or 70 ppm. The criterion for identification was a first Mowse score above 10^4 and a 100-fold difference in the Mowse score from the first to the second possible identification within the species. Six or more peptides were matched in each case using this approach.

3 Results and discussion

3.1 2-D gels

A few preliminary gels were analyzed in the p*I* range of 3–10 (linear, data not shown). There was a distinct concentration of spots in the range of 4–7, which led us to use this range for most of the work. Nine 2-D gels (p*I* 4–7) were examined in detail in the construction of the map (filled ellipses in Figs. 1A and B). Four of these gels were run with extracts of the N16961 strain, which is widely used as a typical El Tor strain. The other five gels were run with the Latin American epidemic strain C3294. Two strains and different growth conditions were used in order to allow for more variation in protein expression. To construct a representative reference map of an El Tor strain, only the proteins present in all conditions and strains were selected. All of the visualization gels were electrophoresed under the same conditions, and there was a high degree of congruence between gels of the same strain grown in the same conditions. For each of the strains, the gels were analyzed and averaged. Visual inspection was also used to determine the presence of the spots. The N16961 average gel, with 539 spots (±27, standard deviation), was then compared with the C3294 average gel, with 516 spots (±53), and 340 spots were found in common between the gels of the two strains. Figure 1A shows one of the N16961 gels, and Fig. 1B one of the C3294 gels. The 340 spots constitute the reference map for *V. cholerae* El Tor, and represent ~65% of the spots found in each gel set.

In this type of gel the proteins detected are in the range of 10–110 kDa. A distribution of the spots by MW and p*I* is presented in Table 1, showing that the majority of the proteins detected are in the central region of the pH 4–7 gels. This is in disagreement with the predicted proteins from ORFs of the *V. cholerae* genome. A theoretical p*I* distribution of these proteins predicts a large number of 997 proteins in the p*I* range of 7–10 and MW 10–110 kDa, which was not detected in our experiments. This may reflect a difficulty in the solubilization or detection of basic proteins with our method, a problem also described by others [27]. It is not known how many of the proteins detected here are proteins processed by post-translational modifications, but the protein distribution found was very different from the theoretical map as presented in the TIGR database (www.tigr.org/tigr-scripts/CMR2/Pseudo2DGel.spl?db_data_id=108). From the 3820 ORFs predicted for N16961, the 539 spots detected correspond to 14.1%. There are 2005 N16961 ORFs with predicted products in the p*I* range of 4.0–7.0 and MW 10–110 kDa. The coverage for this range was 26.9%. Detailed studies with *Escherichia coli* and *Haemophilus influenzae* show that 2-D gels are capable of detecting up to 70% of the cell proteins, with gels covering a series of narrower p*I* ranges [28, 29].

Table 1. Experimental and theoretical number of proteins for *V. cholerae* El Tor, distributed according to p*I* and MW

MW (kDa)		p*I*			Total[a)]
		4–5	5–6	6–7	
50–110	N16961 exp.	19	59	14	92
	C3294 exp.	15	76	24	115
	Common spots exp.	10	40	9	59
	N16961 theor.	133	177	191	501
25–50	N16961 exp.	54	145	93	292
	C3294 exp.	61	145	78	284
	Common spots exp.	34	100	58	192
	N16961 theor.	229	308	379	916
10–25	N16961 exp.	52	73	30	155
	C3294 exp.	29	68	20	117
	Common spots exp.	25	47	17	89
	N16961 theor.	260	155	173	588
Total	N16961 exp.	125	277	137	539
	C3294 exp.	105	289	122	516
	Common spots exp.	69	187	84	340
	N16961 theor.	622	640	743	2005

a) The number of theoretical proteins for each range were obtained from the list of *V. cholerae* ORFs at http://www.tigr.org/tigr-scripts/CMR2/Pseudo2Dgel.spl?db_data_id=108. That table includes a total of 3820 ORFs, with 2005 in the p*I* range of 4–7 and MW 10–110 kDa.

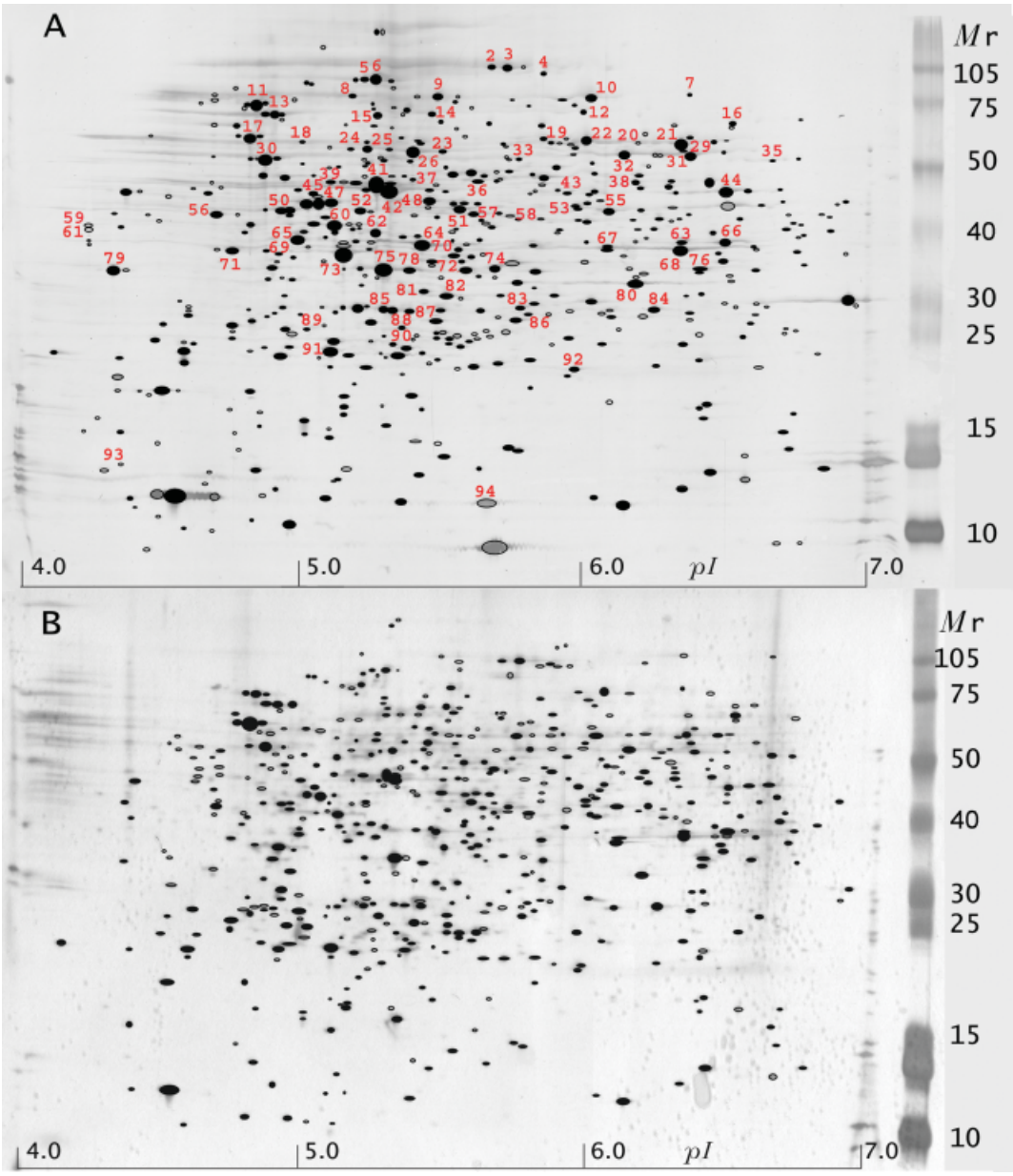

Figure 1. Proteome reference map for *V. cholerae* El Tor. One actual 2-D gel for each of the strains was used as a template to mark the 340 common spots that constitute the reference map (full ellipses) and spots found for only one of the strains or conditions (open ellipses). Red numbers mark the position of N16961 identified protein spots described in Table 2. The full-range Rainbow protein MW marker (Amersham Bioscieces) was used in each gel. (A) Strain N16961, grown at 30°C, in minimal M9 medium, with 0.4% glucose and 0.4% casamino acids. A total of 539 spots were found for the average gel of this strain. (B) Strain C3294, grown at 37°C in LB medium, with 516 spots present in the average gel.

3.2 Peptide mass fingerprint analysis

A total of 104 spots from gels of strain N16961 were cut from preparative gels and analyzed. The preparative gels were run with samples from cultures grown at 37°C in LB broth. A few additional spots were cut and identified from visualization gels. Ninety-four proteins were identified (Table 2). All of these were in the p*I* range of 4.0–7.0 and MW 12–110 kDa. Very abundant proteins identified previously in other bacterial species [27, 32] were also

Table 2. Identification of *V. cholerae* El Tor N16961 proteins by mass spectrometry

SSP[h)]	Genome ID[a)]	Description	Gene symbol	MW/pI ExPASy[b)]	MW/pI experim.[c)]	Mowse score[d)]	Score1/ Score2[e)]	Cover.[f)]	Number pep.[g)]	Function
1	VC0015	DNA gyrase, subunit B	gyrB	89.52/5.70	102/6.19	5.67E+07	2.48E+05	25%	17	DNA metabolism: DNA replication, recombination, and repair
2	VC0604	Aconitate hydratase 2*[i)]	acnB	93.59/5.39	98.0/5.69	1.50E+13	5.57E+08	47%	30	Energy metabolism: TCA cycle
3	VC0604	Aconitate hydratase 2*	acnB	93.59/5.39	98.0/5.75	1.94E+12	2.08E+10	36%	24	Energy metabolism: TCA cycle
4	VC0711	ClpB protein	clpB-1	95.83/5.47	94.0/5.86	1.28E+16	1.42E+12	55%	40	Protein fate: Degradation of proteins, peptides and glycopeptides
5	VC2342	Elongation factor G*	fusA-2	76.48/5.02	89.0/5.25	1.50E+17	4.42E+14	45%	33	Protein synthesis: translation factors
6	VC2342	Elongation factor G*	fusA-2	76.48/5.02	89.0/5.30	4.16E+13	2.31E+11	45%	28	Protein synthesis: translation factors
7	VC1560	Catalase/peroxidase	perA	80.65/5.79	82.0/6.35	3.78E+06	7.32E+02	26%	16	Cellular processes: detoxification
8	VC0647	Polyribonucleotide nucleotidyltransferase	pnp	76.58/5.05	80.2/5.21	6.31E+09	8.72E+05	33%	22	Transcription: degradation of RNA
9	VC1097	Phosphate acetyltransferase	pta	76.89/5.28	79.5/5.51	2.50E+14	8.11E+08	57%	31	Energy metabolism: fermentation
10	VC1141	Isocitrate dehydrogenase, NADP-dependent, monomeric type	icd	80.53/5.61	77.5/6.03	7.67E+14	3.33E+10	53%	31	Energy metabolism: TCA cycle
11	VC0855	DnaK protein	dnaK	68.76/4.80	73.5/4.89	4.21E+11	4.55E+08	50%	25	Protein fate: protein folding and stabilization
12	VC1872	Conserved hypothetical protein		74.11/5.56	71.8/6.00	1.98E+15	1.29E+12	67%	38	Hypothetical proteins: conserved
13	VC1915	Ribosomal protein S1	rpsA	61.05/4.86	70.0/4.96	8.33E+06	5.02E+04	34%	15	Protein synthesis: ribosomal proteins, synthesis, and modification
14	VC0985	Heat shock protein HtpG	htpG	72.20/5.26	68.3/5.49	2.10E+17	3.81E+13	65%	36	Protein fate: protein folding and stabilization
15	VC2562	2',3'-Cyclic-nucleotide 2'-phosphodiesterase	cpdB	75.03/5.35	66.2/5.29	1.21E+16	1.89E+07	52%	30	Purines, pyrimidines, nucleosides, and nucleotides: other
16	VC2089	Succinate dehydrogenase, flavoprotein subunit	sdhA	64.51/6.09	65.5/6.50	3.75E+10	2.60E+06	50%	22	Energy metabolism: TCA cycle
17	VC2664	Chaperonin, 60 kDa subunit	groEL-1	57.15/4.78	60.0/4.86	2.10E+08	1.91E+05	41%	12	Protein fate: protein folding and stabilization
18	VC2746	Glutamate-ammonia ligase	glnA	51.75/4.94	58.7/5.05	2.21E+06	2.18E+03	37%	12	Amino acid biosynthesis: glutamate family
19	VC2738	Phosphoenolpyruvate carboxykinase*	pckA	59.84/5.60	57.1/5.89	6.32E+15	2.53E+11	80%	30	Energy metabolism: glycolysis/gluconeogenesis
20	VC1091	Oligopeptide ABC transporter, peripl. oligo-bdg.ptn.*[j)]	oppA	61.14/6.02	57.0/6.17	3.00E+11	6.54E+06	48%	26	Transport and binding proteins: amino acids, peptides and amines
21	VC1091	Oligopeptide ABC transporter, peripl. oligo-bdg.ptn.*	oppA	61.14/6.02	57.0/6.31	5.92E+08	2.25E+05	38%	20	Transport and binding proteins: amino acids, peptides and amines
22	VC2738	Phosphoenolpyruvate carboxykinase*	pckA	59.84/5.60	56.2/5.99	2.41E+17	3.16E+13	73%	33	Energy metabolism: glycolysis/gluconeogenesis
23	VC0566	Protease DO*	htrA	48.37/5.74	55.6/5.50	3.72E+10	4.58E+06	63%	24	Protein fate: degradation of proteins, peptides and glycopeptides

Table 2. Continued

SSP[h)]	Genome ID[a)]	Description	Gene symbol	MW/p*I* ExPASy[b)]	MW/p*I* experim.[c)]	Mowse score[d)]	Score1/ Score2[e)]	Cover.[f)]	Number pep.[g)]	Function
24	VC0336	Phosphoglycerate mutase, 2,3-bisphospho-glycerate-independent	yibO	55.36/5.08	54.7/5.25	2.07E+05	1.93E+03	24%	11	Energy metabolism: glycolysis/gluconeogenesis
25	VC0171	Peptide ABC transporter, peripl. pept-bdg. ptn.*		60.61/5.76	54.5/5.34	1.72E+13	2.38E+09	58%	26	Transport and binding proteins: amino acids, peptides, and amines
26	VC0171	Peptide ABC transporter, peripl. pept-bdg. ptm*		60.61/5.76	54.5/5.42	1.94E+10	2.92E+06	38%	17	Transport and binding proteins: amino acids, peptides, and amines
27	VC1819	Aldehyde dehydrogenase*	aldA-2	55.31/5.75	54.0/6.13	1.14E+10	1.23E+07	39%	15	Energy metabolism: fermentation
28	VC1819	Aldehyde dehydrogenase*	aldA-2	55.31/5.75	53.2/6.25	2.43E+04	5.58E+02	19%	9	Energy metabolism: fermentation
29	VC0485	Pyruvate kinase I	pykF	50.44/5.79	53.2/6.30	2.42E+07	2.32E+04	52%	20	Energy metabolism: glycolysis/gluconeogenesis
30	VC2764	ATP synthase F1, beta-subunit	atpD	50.52/4.82	53.0/4.92	4.25E+12	3.08E+09	56%	24	Energy metabolism: ATP-proton motive force interconversion
31	VC2412	Pyruvate dehydrogenase, E3 component, lipoamide dehydrogenase	lpdA	50.99/5.79	52.2/6.35	7.72E+09	7.81E+07	49%	20	Energy metabolism: pyruvate dehydrogenase
32	VC2766	ATP synthase F1, alpha-subunit	atpA	56.78/5.54	52.1/6.13	1.51E+07	2.25E+04	39%	19	Energy metabolism: ATP-proton motive force interconversion
33	VCA0744	Glycerol kinase	glpK	55.65/5.47	52.0/5.78	1.76E+12	5.21E+08	53%	23	Energy metabolism: other
34	VC1923	Trigger factor	tig	47.95/5.00	51.5/5.15	1.13E+08	1.33E+04	38%	13	Protein fate: protein folding and stabilization
35	VC2295	NADH:ubiquinone oxidoreductase, Na translocating, alpha-subunit	nqrA	50.64/6.53	51.0/6.66	2.03E+08	1.14E+05	36%	16	Energy metabolism: electron transport
36	VC0566	Protease DO*	htrA	48.37/5.74	47.0/5.64	3.59E+06	7.71E+03	35%	16	Protein fate: degradation of proteins, peptides and glycopeptides
37	VC2447	Enolase*	eno	45.81/5.03	46.8/5.49	1.92E+09	7.26E+05	49%	16	Energy metabolism: glycolysis/gluconeogenesis
38	VC2602	Adenylosuccinate synthetase	purA	46.82/5.73	46.6/6.17	1.09E+07	7.88E+03	39%	14	Purines, pyrimidines, nucleosides, and nucleotides: purine ribonucleotide biosynthesis
39	VC2447	Enolase*	eno	45.81/5.03	45.3/5.14	1.72E+07	1.88E+02	42%	14	Energy metabolism: glycolysis/gluconeogenesis
40	VC2086	2-Oxoglutarate dehydrogenase, E2 component, dihydrolipoamide succinyltransferase	sucB	44.09/5.55	45.3/5.77	1.07E+10	1.32E+04	61%	25	Energy metabolism: TCA cycle

Table 2. Continued

SSP[h)]	Genome ID[a)]	Description	Gene symbol	MW/p*I* ExPASy[b)]	MW/p*I* experim.[c)]	Mowse score[d)]	Score1/ Score2[e)]	Cover.[f)]	Number pep.[g)]	Function
41	VC2447	Enolase*	eno	45.81/5.03	45.0/5.29	1.24E+15	4.30E+06	68%	26	Energy metabolism: glycolysis/gluconeogenesis
42	VC0362	Elongation factor TU	tufA	43.13/5.04	44.2/5.36	2.14E+08	3.17E+05	51%	14	Protein synthesis: translation factors
43	VC1098	Acetate kinase	ackA-1	42.84/5.58	43.8/5.93	9.03E+04	4.68E+03	31%	9	Energy metabolism: fermentation
44	VC1905	Alanine dehydrogenase	ald	39.84/5.88	43.7/6.47	8.77E+09	3.37E+06	56%	21	Energy metabolism: amino acids and amines
45	VC2436	Outer membrane protein TolC	tolC	47.75/5.10	42.8/5.05	3.76E+07	8.13E+04	36%	13	Protein fate: protein and peptide secretion and trafficking
46	VC1345	Oxireductase, putative		43.25/5.91	42.8/6.42	1.95E+12	2.98E+09	59%	26	Central intermediary metabolism: other
47	VC2187	Flagellin FlaC	flaC	39.91/4.93	42.7/5.13	1.90E+04	3.97E+02	25%	8	Cellular processes: chemotaxis and motility
48	VC0472	*S*-Adenosylmethionine synthase	metK	42.07/5.24	42.4/5.47	4.00E+05	6.20E+02	35%	13	Biosynthesis of cofactors, prosthetic groups, and carriers: other
49	VC0477	Phosphoglycerate kinase	pgk	41.56/4.96	42.1/5.08	3.18E+11	2.91E+08	58%	20	Energy metabolism: glycolysis/gluconeogenesis
50	VC2571	DNA-directed RNA polymerase, alpha-subunit	rpoA	36.42/4.83	41.8/4.97	5.28E+08	4.25E+05	71%	22	Transcription: DNA-dependent RNA polymerase
51	VC2085	Succinyl-CoA synthase, beta-subunit	sucC	41.40/5.35	41.8/5.59	3.50E+08	5.66E+06	55%	19	Energy metabolism: TCA cycle
52	VC2142	Flagellin FlaB	flaB	39.52/5.05	41.8/5.22	3.77E+08	9.00E+05	54%	15	Cellular processes: chemotaxis and motility
53	VCA0886	2-Amino-3-ketobutyrate coenzyme A ligase	kbl	43.27/5.51	41.8/5.97	5.22E+06	5.62E+02	38%	14	Energy metabolism: amino acids and amines
54	VC1344	4-Hydroxyphenylpyruvate dioxygenase*	hppD	41.72/5.53	41.7/5.77	3.58E+10	9.70E+07	46%	22	Energy metabolism: amino acids and amines
55	VC2618	Acetylornithine aminotransferase	argD	43.58/5.66	41.5/6.06	2.34E+06	1.28E+03	38%	13	Amino acid biosynthesis: glutamate family
56	VCA0945	Maltose ABC transporter peripl. maltose-bdg. ptn.	malE	43.81/4.89	41.4/4.77	5.30E+10	1.24E+05	54%	18	Transport and binding proteins: carbohydrates, organic alcohols, and acids
57	VC1344	4-Hydroxyphenylpyruvate dioxygenase*	hppD	41.72/5.53	41.3/5.65	1.11E+06	2.06E+03	45%	17	Energy metabolism: amino acids and amines
58	VC1344	4-Hydroxyphenylpyruvate dioxygenase*	hppD	41.72/5.53	41.0/5.77	8.06E+11	7.76E+08	56%	25	Energy metabolism: amino acids and amines
59	VC0633	Outer membrane protein OmpU*	ompU	37.66/4.51	40.2/4.37	1.08E+08	8.67E+04	62%	19	Cell envelope: other
60	VC0478	Fructose-bisphosphate aldolase, class II	fbaA	38.92/4.91	39.8/5.12	5.93E+05	4.49E+02	40%	10	Energy metabolism: glycolysis/gluconeogenesis
61	VC0633	Outer membrane protein OmpU*	ompU	37.66/4,51	39.6/4.37	5.72E+09	3.90E+06	64%	19	Cell envelope: other

Table 2. Continued

SSP[h]	Genome ID[a]	Description	Gene symbol	MW/pI ExPASy[b]	MW/pI experim.[c]	Mowse score[d]	Score1/ Score2[e]	Cover.[f]	Number pep.[g]	Function
62	VC0608	Iron(III) ABC transporter, peripl. iron-compound-bdg.ptn.		39.37/5.55	39.1/5.27	4.77E+06	1.15E+04	41%	13	Transport and binding proteins: cations
63	VC1091	Oligopeptide ABC transporter, peripl. oligo-bdg. ptn.*	oppA	61.14/6.02	39.1/6.31	9.21E+05	3.12E+02	27%	13	Transport and binding proteins: amino acids, peptides, and amines
64	VC0432	Malate dehydrogenase	mdh	36.96/6.85	38.7/5.44	5.91E+06	2.40E+04	48%	15	Energy metabolism: TCA cycle
65	VCA1113	Spermidine/putrescine ABC transporter, peripl. spermidine/ putrescine-bdg. ptn.		38.76/5.06	38.3/5.01	2.63E+11	5.88E+06	62%	25	Transport and binding proteins: amino acids, peptides, and amines
66	VC0968	Cysteine synthase A	cysK	34.24/5.93	38.2/6.48	3.62E+04	3.62E+02	39%	7	Amino acid biosynthesis: serine family
67	VC2000	Glyceraldehyde 3-phosphate dehydrogenase*	gapA-1	35.28/5.80	37.7/6.07	9.98E+07	5.01E+04	72%	17	Energy metabolism: glycolysis/gluco-neogenesis
68	VC2000	Glyceraldehyde 3-phosphate dehydrogenase*	gapA-1	35.28/5.80	37.7/6.30	5.67E+04	3.03E+02	29%	8	Energy metabolism: glycolysis/gluco-neogenesis
69	VC2213	Outer membrane protein OmpA	ompA	34.29/5.07	37.6/4.97	1.87E+07	1.42E+06	59%	11	Cell envelope: other
70	VCA0685	Iron(III) ABC transporter, peripl. iron-compound-bdg. ptn.		37.74/5.91	37.3/5.58	3.58E+05	1.84E+02	32%	9	Transport and binding proteins: cations
71	VC1425	Spermidine/putrescine ABC transporter, peripl. spermidine/ putrescine-bdg. ptn.	potD-2	40.78/4.95	37.0/4.80	9.51E+09	3.56E+06	42%	21	Transport and binding proteins: amino acids, peptides, and amines
72	VC1929	C4-dicarboxylate-binding periplasmic protein	dctP-2	37.10/5.46	36.7/5.55	1.30E+10	6.74E+07	53%	20	Transport and binding proteins: carbo-hydrates, organic alcohols, and acids
73	VC1362	Amino acid ABC transporter, peripl. aa-bdg. ptn.		36.85/5.19	36.5/5.15	5.24E+11	4.46E+08	69%	21	Transport and binding proteins: amino acids, peptides, and amines
74	VC1334	Conserved hypothetical protein		34.84/5.77	35.5/5.69	1.82E+05	2.93E+02	37%	13	Hypothetical proteins: conserved
75	VC2259	Elongation factor Ts	tsf	29.85/5.16	35.1/5.31	6.13E+07	1.25E+06	65%	14	Protein synthesis: translation factors
76	VC0430	Immunogenic protein		35.26/6.61	35.1/6.38	1.93E+04	8.60E+02	25%	8	Cell envelope: other
77	VCA0877	Hydrolase, putative		38.94/7.02	35.0/5.74	4.35E+05	3.53E+02	37%	9	Central intermediary metabolism: other
78	VC1101	Conserved hypothetical protein		33.28/5.54	34.1/5.40	1.16E+06	5.34E+03	40%	11	Hypothetical proteins: conserved
79	VC1325	Galactoside ABC transporter, peripl. D-galactose/D-glucose-bdg. ptn.	mglB	34.78/4.68	33.1/4.48	5.42E+11	2.85E+08	71%	20	Transport and binding proteins: carbo-hydrates, organic alcohols, and acids
80	VC2084	Succinyl-CoA synthase, alpha-subunit	sucD	29.94/5.79	32.3/6.15	3.85E+05	5.51E+04	70%	13	Energy metabolism: TCA cycle
81	VC2350	Deoxyribose-phosphate aldolase	deoC	27.96/5.23	31.3/5.45	6.11E+06	3.70E+05	71%	14	Energy metabolism: other
82	VC2615	Conserved hypothetical protein		29.47/5.29	30.7/5.52	3.72E+07	2.44E+04	59%	14	Hypothetical proteins: conserved

Table 2. Continued

SSP[h]	Genome ID[a]	Description	Gene symbol	MW/pI ExPASy[b]	MW/pI experim.[c]	Mowse score[d]	Score1/ Score2[e]	Cover.[f]	Number pep.[g]	Function
83	VCA1039	Amino acid ABC transporter, peripl. aa-bdg. ptn.		30.85/6.11	29.3/5.77	5.52E+04	7.96E+02	32%	11	Transport and binding proteins: amino acids, peptides and amines
84	VC1034	Uridine phosphorylase	udp-1	27.66/5.71	28.7/6.21	3.40E+08	5.23E+05	73%	13	Purines, pyrimidines, nucleosides and nucleotides: other
85	VC1863	Amino acid ABC transporter, peripl. aa-bdg. ptn.		28.71/5.33	27.7/5.33	2.53E+06	2.99E+03	65%	13	Transport and binding proteins: amino acids, peptides and amines
86	VC2088	Succinate dehydrogenase, iron-sulfur protein	sdhB	26.39/5.44	27.2/5.78	1.48E+08	4.90E+05	62%	17	Energy metabolism: TCA cycle
87	VC0986	Adenylate kinase	adk	23.28/5.20	27.1/5.48	6.32E+05	1.47E+03	53%	11	Purines, pyrimidines, nucleosides, and nucleotides: nucleotide and nucleoside interconversions
88	VC2347	Purine nucleoside phosphorylase	deoD-1	26.12/5.20	26.3/5.36	2.82E+05	2.79E+02	56%	12	Purines, pyrimidines, nucleosides, and nucleotides: other
89	VC0010	Amino acid ABC transporter, peripl. aa-bdg. ptn.		27.30/4.98	24.0/5.02	1.01E+05	8.34E+02	50%	10	Transport and binding proteins: amino acids, peptides, and amines
90	VCA1054	Conserved hypothetical protein		22.35/5.15	23.1/5.37	1.19E+08	8.35E+05	85%	14	Hypothetical proteins: conserved
91	VC0731	Antioxidant, AhpC/Tsa family		22.86/5.37	22.0/5.10	1.81E+05	1.66E+04	46%	9	Cellular processes: detoxification
92	VC2767	ATP synthase F1, delta-subunit	atpH	19.56/5.57	20.7/5.95	1.09E+07	1.12E+02	66%	10	Energy metabolism: ATP-proton motive force interconversion
93	VCA0881	Hypothetical protein		14.39/4.45	13.0/4.57	1.60E+04	1.51E+03	52%	6	Hypothetical proteins
94	VCA0689	Conserved hypothetical protein		13.13/5.50	12.7/5.62	3.06E+04	3.29E+02	64%	9	Hypothetical proteins: conserved

a) Genome ID corresponds to the gene identification in the TIGR *V. cholerae* genome homepage (http://www.tigr.org/tigr-scripts/CMR2/GenomePage3.spl?database=gvc) [12]
b) Theorethical molecular weight and p*I*, as calculated using the ExPASy tool Compute p*I*/Mw, http://www.expasy.org/tools/pi_tool.html [30]
c) Manual determination, from curves constructed from the El Tor N16961 gels
d) Mowse ranking score obtained with the use of the MS-Fit software at http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm [31]. A value >10e4 was required for identification
e) Mowse score for the first putative identification/Mowse score for the second putative identification, both for *V. cholerae.* A value > 100 was required for identification
f) Coverage
g) Number of experimental peptides matched to protein predicted peptides
h) Proteins were listed according to decreasing experimental MW. Numbering of the spots is from left to right and from top to bottom on the gel
i) Proteins with repeated identification from different spots are marked with asterisks
j) Peripl. substrate-bdg. ptn., periplasmic substrate-binding protein; aa, amino acid; oligo, oligopeptide; pept, peptid

detected in *V. cholerae* El Tor, such as elongation factor TU, enolase, DnaK protein, chaperonin 60 kDa subunit, glyceraldehyde 3-phosphate dehydrogenase, and malate dehydrogenase. Aconitate hydratase, isocitrate dehydrogenase, malate dehydrogenase, alanine dehydrogenase, and adenylate kinase activities, among others, have been previously detected by MLEE in *V. cholerae* [19], relying on the high level of these enzymes in the cells. A theoretical work [33] has predicted various highly expressed genes for four fast-growing bacterial species, including *V. cholerae*, based on codon usage difference between highly expressed genes and others. Thirty-five of the proteins found here correspond to genes included in their list of 172 predicted highly expressed genes of *V. cholerae.*

The genes corresponding to the identified proteins were located mainly in chromosome 1, with only 12.5% of the genes located in chromosome 2. Considering the size of the genomes, this is significantly lower than expected from a random proportion of 28.7% of proteins from chromosome 2. Within chromosome 1 itself there was a concentration of proteins from genes located in the left upper quarter of the chromosome map, the same containing various rRNA operons. The reliability of the identification was checked in a blind test. A peptide mass fingerprint analysis of 20 spots from replicate gels, carried out by different researchers, resulted in a complete agreement in the identification. There was also a good agreement between predicted and experimental MW and p*I*s, with expected exceptions (see below).

3.2.1 Abundant protein functions: energy metabolism and transport

The proteins identified are listed in Table 3 according to their cellular function. More than one-third of these are involved in energy metabolism. Eight proteins of the tricarboxylic acid (TCA) cycle and seven from the glycolysis/gluconeogenesis pathway were identified. Another large proportion of the identified proteins belongs to the group of periplasmic substrate-binding subunits of ATP-binding cassette (ABC) transporters. Fifteen spots, corresponding to twelve proteins, were identified in this group, representing 15% of the total independent proteins. Regulatory function and signal transduction proteins were not found, and these are normally present in small amounts in the cells [32].

3.2.2 Hypothetical and conserved hypothetical proteins identified as abundant proteins

Seven proteins, previously proposed as hypothetical proteins based on DNA sequence, were verified experimentally by mass spectrometry, and should be considered now as real proteins, corresponding to genes VC1101, VC1334, VC1872, VC2615, VCA0689, VCA0881, and VCA1054. Their sizes and p*I* were in good agreement with the theoretical predictions.

Table 3. Cellular role categories

Categories[a)]	Number of spots with identified proteins	Number of independent proteins[b)]
Amino acid biosynthesis	3	3
Biosynthesis of cofactors, prosthetic groups, and carriers	1	1
Cell envelope	4	3
Cellular processes	4	4
Central intermediary metabolism	2	2
DNA metabolism	1	1
Energy metabolism, total	36	28
(TCA cycle)	(9)	(8)
(glycolysis/gluconeogenesis)	(11)	(7)
Fatty acid and phospholipid metabolism	0	0
Other categories	0	0
Protein fate	8	7
Protein synthesis	5	4
Purines, pyrimidines, nucleosides, and nucleotides	5	5
Regulatory functions	0	0
Signal transduction	0	0
Transcription	2	2
Transport and binding proteins	16	13
(ABC transporters)	(15)	(12)
(other transporters)	(1)	(1)
Conserved hypothetical	6	6
Other hypothetical	1	1
Unknown function	0	0
Total	94	80

a) Categories were taken from the TIGR-CMR (http://www.tigr.org/tigr-scripts/CMR2/gene_table.spl?db=CMR).
b) Independent proteins are encoded in different genes.

3.2.3 Operons

Some pairs of proteins encoded in adjacent genes were identified. This strongly suggests the presence of operons, because these proteins were abundant enough to

be found within the limited number of proteins analyzed. In particular, we identified five subunits of proteins involved in the TCA cycle from one single region of the chromosome. Table 4 shows some neighbor genes of N16961 with identified proteins, including an analysis of their presence in the same organization and orientation in other genomes. This analysis was carried out in two different ways. First, the cluster of orthologous groups of proteins (COG) of each protein was determined, and its genome context examined with the tool provided on the COG web page, http://www.ncbi.nlm.nih.gov/COG/xindex.html [34]. Second, the confidence probability that a particular pair is an operon was estimated with the comprehensive microbial resource (CMR) operon prediction software [35]. Five of the six gene groups analyzed were found in the same order and orientation in other bacterial genomes; thus we propose that they belong to operons. In the case of VC1344 (*hppD*) and VC1345 (oxireductase), the available data is not sufficient to allow prediction of an operon, since *hppD* is found close to different partners within various genomes. Even so, the two genes appear in the same arrangement in the three *Vibrio* species with a sequenced genome and in *Caulobacter crescentus.*

3.2.4 Processing

Fourteen spots contained repeats of proteins identified from other spots. The total number of independent proteins, encoded in different genes, was 80. Table 2 shows the proteins with repeats marked with asterisks. Such observations have been made for other species and the migration variants generally result from post-translational modifications of the proteins. Proteins from three different spots were identified as the product of VC1091, an oligopeptide ABC transporter, the periplasmic oligo-binding

Table 4. Adjacent genes coding for abundant identified proteins

Genome ID[a)]	Gene symbol or description	COG[b)]	Function	CMR operon prediction confidence levels[c)]
VC0477	pgk	0126	Glycolysis/gluconeogenesis	93.43, *n* = 12
VC0478	fbaA	0191	Glycolysis/gluconeogenesis	
VC0985	htpG	0326	Protein folding and stabilization	65.92, *n* = 1
VC0986	adk	0563	Nucleotide and nucleoside interconversions	
VC1098	ackA-1	0282	Energy metabolism: fermentation	100, *n* = 20
VC1097	pta-1	0857	Energy metabolism: fermentation	
	pta-2	0280		
VC1344	hppD	3185	Energy metabolism: amino acids and amines	Genes found together in same orientation only in species
VC1345	Oxireductase	3508	Central intermediary metabolism	(*V. vulnificus, V. parahaemolyticus*, and *C. crescentus*)
VC2089	sdhA	1053	Energy metabolism: TCA cycle	100, *n* = 35 (VC2088x VC2089)
VC2088	sdhB	0479	Energy metabolism: TCA cycle	
VC2086	sucB	0508	Energy metabolism: TCA cycle	100, *n* = 22 (VC2086x VC2087)
VC2085	sucC	0045	Energy metabolism: TCA cycle	100, *n* = 28 (VC2084xVC2085)
VC2084	sucD	0074	Energy metabolism: TCA cycle	
VC2767	atpH	0712	ATP-proton motive force interconversion	
VC2766	atpA	0056	ATP-proton motive force interconversion	85.09, *n* = 5 (VC2764x VC2766)
VC2764	atpD	0055	ATP-proton motive force interconversion	

a) In all cases transcription of the adjacent genes is in the same direction. Genome IDs are gene names from the TIGR *V. cholerae* El Tor N16961 genome.
b) Cluster of orthologous groups of proteins [34]. The genome context of each COG was inspected for the presence of the other gene nearby (http://www.ncbi.nlm.nih.gov/COG/xindex.html).
c) Operons prediction in the CMR (http://www.tigr.org/tigr-scripts/operons/operons.cgi) [35]
Confidence is an estimation of the lower boundary of the probability that the two corresponding genes are located in the same operon, *n* is a number of other genomes (for 73 bacterial and archaeal genomes) that have the same pair of genes located in the same direction. Confidence values given are underestimated values, as the software disregards a valid correct organization from a phylogenetically close species.

protein OppA. Two of these, from spots 20 and 21, have approximately the same MW, 61.14 kDa, only having a different p*I*, while the other one, from spot 63, has a much smaller size, with an MW of 39.1 kDa. The analysis of the identified peptides shows that the protein from spot 63 carries the amino first half of the protein. No peptides were identified from amino acid 325 onwards, from a total of 543 amino acids, suggesting that OppA is processed, even though a nonspecific degradation was not ruled out.

The majority of the identified proteins had an experimental MW within 10% of the expected value (Table 2). Some proteins presented a lower MW than predicted: peptide ABC transporter (spots 25 and 26), 2', 3'-cyclic-nucleotide 2'-phosphodiesterase (spot 15), and outer membrane protein TolC (spot 45). Other proteins were found with a higher experimental mass than expected: elongation factor G (spots 5 and 6), ribosomal protein S1 (spot 13), protease DO (spot 23), glutamate-ammonia ligase (spot 18), elongation factor Ts (spot 75), DNA-directed RNA polymerase, α-subunit (spot 50), deoxyribose-phosphate aldolase (spot 81), and adenylate kinase (spot 87). This represents about 15% of the proteins, and the mass difference in these cases was in the range of 10–20%. Further sequencing studies will be necessary to confirm or reassign the start sites in these cases.

3.2.5 OmpU and HtpG

The adjacent gene pair *htp*G and *adk* (Table 4), coding for the heat shock protein HtpG and adenylate kinase, is particularly interesting. The important regulatory gene *toxR* is located adjacent to *htpG* in a divergent orientation and there is experimental evidence for an opposite temperature regulation of these two genes [36]. Growth of El Tor strains in LB at 37°C, as used in preparing the extracts for protein identification, leads to a low level of ToxR production, and an increased level of HtpG in comparison to growth at 22°C. Here, we find high levels of production of both HtpG and Adk, suggesting an increased production of Adk directly in response to an increase in temperature. Another ToxR-related protein was identified, OmpU, the major outer membrane protein of *V. cholerae*, induced by ToxR [37]. In agreement with the expected low amounts of ToxR in the conditions used, OmpU was detected, but not as a major spot. An important finding of this work was the identification of OmpU from two nearby spots, one above the other, suggesting at least two forms of this protein.

3.2.6 Comparison to classical *V. cholerae*

Recently, a protein map for the strain O395 of *V. cholerae*, a strain from the classical biotype, was presented, and 40 reference proteins were defined [22]. Twenty-four of these were also identified as abundant proteins in our work. There is a general difference in the attribution of p*I* values between the two works. They estimate a more basic p*I* for the majority of the reference proteins, in some cases a difference of 0.5 in the p*I* scale compared to our values, particularly for the proteins with p*I* $>$ 5. Although the gels presented in that paper cover a larger range of MW, they span p*I* from 3.5 to 10, while our gels expand the p*I* range from 4 to 7, in which the majority of proteins is concentrated, thus allowing for more accurate p*I* determinations.

The relative positions of the reference proteins on the maps of the two studies were compared. Twenty-two of the twenty-four common reference proteins are in equivalent positions in both maps. However, DnaK (spot 11) on our map is at the right of GroEL-1 (spot 17), whereas in their work it is more acidic, to the left of GroEL-1. On our map, GplK (spot 33) appears at p*I* 5.78, to the left of PckA (spot 19), while on their map GplK is at p*I* 6.61, to the right of PckA. In addition to the reference proteins they present 17 low-pH negatively regulated proteins. Three of these were independently identified here, RpoA, ClpB-1, and the VC1101 product, but their positions on our map were different from theirs. The different position of some proteins could be due to the fact that migration variants exist for some proteins, and a limited number of spots was identified in both works. An alternative explanation is that these are biotype differences. New studies focused at these spots will clarify the issue. El Tor and classical biotypes are quite different in genome organization and even in genome content [15, 38, 39], thus it is reasonable to expect that their proteomes will be different. The common abundant proteins detected are just the beginning of the construction of a complete proteome map of the strains.

3.2.7 Comparison to mRNA microarray analysis

A comparison was also made with the results of gene expression obtained by microarray analysis of strain N16961 [18]. Considering that the 2-D gels in the MW and p*I* range used here contain most of the proteins in the cells, in broad terms, the 500 more abundant proteins are being detected. A parallel was made with the top 500 mRNAs identified in that paper (growth in LB medium). From the subgroup of 80 identified proteins, the mRNA for 38 of these were present among those 500 mRNAs. This is a proportion of 38/80, predicting a general agreement of 48% between the two datasets of the 500 most-expressed genes. The growth phase was not the same, exponential in their case and stationary in our experiments, thus this comparison has to be looked at with some care.

Two major trends were noticed: few ribosomal proteins were identified in our experiments. Xu *et al.* [18] detected 54 ribosomal protein mRNAs within the top 500 mRNAs. Many ribosomal proteins were also predicted as highly expressed proteins [33]. Nevertheless, only one ribosomal protein was identified in the 80 proteins described here. This is largely explained by the small size of ribosomal proteins. The MW of 41 of these 54 ribosomal proteins (76%) is less than 17 kDa, which was the lower limit of the preparative gels used here. Another possible explanation could be the difficulty in their extraction/solubilization from the ribosomes in the conditions used or the fact that these are basic proteins, more difficult to be detected on 2-D gels [27]. However, there was a better representation of various ABC transporter proteins in our experiments. Twenty-one genes having very low expression by their LB mRNA data, included in the 1500 genes with lowest expression, were found to be well-expressed with the identification of the proteins (top 500 proteins) (Table 5). A large proportion of these (one-third) corresponds to periplasmic substrate-binding subunits of ABC transporters. The mRNA of only two of the twelve periplasmic substrate-binding subunits of ABC transporters identified here were in the top 500 LB mRNA rank [18]. Possible explanations for these differences are not immediate, although the different growth phase could be an explanation. The mRNAs corresponding to these periplasmic proteins may have a particular property leading to a difficult extraction, or the proteins may be more stable than average. These substrate-binding subunits of ABC transporters were not predicted as highly expressed genes by their codon usage [33].

Table 5. Proteins identified in this work (top 500 proteins) and ranked in the 1500 lowest expression range by mRNA data [18]

Description	Genome ID
ABC transporters, periplasmic solute-binding proteins	
Peptide ABC transporter, periplasmic peptide-binding protein	VC0171
Galactoside ABC transporter, periplasmic D-galactose/D-glucose-binding protein, mglB	VC1325
Amino acid ABC transporter, periplasmic amino acid-binding protein	VC1362
Amino acid ABC transporter, periplasmic amino acid-binding protein	VC1863
Iron(III) ABC transporter, periplasmic iron-compound-binding protein	VCA0685
Amino acid ABC transporter, periplasmic amino acid-binding protein	VCA1039
Spermidine/putrescine ABC transporter, periplasmic spermidine/putrescine-binding protein	VCA1113
Other proteins	
Conserved hypothetical protein	VC1101
4-Hydroxyphenylpyruvate dioxygenase, hppD	VC1344
Oxireductase, putative	VC1345
Catalase/peroxidase, perA	VC1560
Aldehyde dehydrogenase, aldA-2	VC1819
Conserved hypothetical protein	VC1872
C4-dicarboxylate-binding periplasmic protein, dctP-2	VC1929
Conserved hypothetical protein	VC2615
Acetylornithine aminotransferase, argD	VC2618
Conserved hypothetical protein	VCA0689
Glycerol kinase, glpK	VCA0744
Hydrolase, putative	VCA0877
Hypothetical protein	VCA0881
Conserved hypothetical protein	VCA1054
Total	21

4 Concluding remarks

This is a first presentation of a map of the proteins of the major human pathogen *V. cholerae* El Tor. The use of different strains and growth conditions gives a broader spectrum of protein expression under laboratory conditions. For each strain and condition used ~525 spots were detected, of which 340 were common. Further studies with narrower p*I* ranges will be done, in order to expand the proteome map of *V. cholerae*. Eighty independent proteins were identified by MALDI-TOF-MS, corresponding to 94 2-D gel spots. Five operons are proposed. The protein map presented here for *V. cholerae* El Tor complements the data for a classical strain [22]. A database with the present and future data will be organized within the Federated 2-D database network [40]. It will be useful for comparative studies between different strains, and global regulation studies for mutants of interest. The fact that *V. cholerae* is being approached both by microarray and proteome analysis will probably lead to cross-fertilization of the studies and a better understanding of its gene expression, protein stability, post-translational modifications, and physiology.

This work was developed within the framework of the Proteomics Network of Rio de Janeiro, supported by FAPERJ-RJ. The network involves many more scientists

who contributed with ideas, equipment, and helpful discussions. We acknowledge important contributions from FUJB (Fundação Universitária José Bonifácio). M.R.S. and D.P. de C. are recipients of post-doctoral scholarships from CNPq (Brazil). E.O.S. and M.L.H.F. receive M.Sc. scholarships (CAPES and CNPq, respectively).

5 References

[1] Kaper, J. B., Glenn Morris Jr., J., Levine, M. M., *Clin. Microbiol. Rev.* 1995, *8*, 48–86.

[2] Reidl, J., Klose, K. E., *FEMS Microbiol. Rev.* 2002, *26*, 125–139.

[3] Waldor, M. K., Mekalanos, J. J., *Science* 1996, *272*, 1910–1914.

[4] Zhu, J., Miller, M. B., Vance, R. E., Dziejman, M. *et al.*, *Proc. Natl. Acad. Sci. USA* 2002, *99*, 3129–3134.

[5] Karaolis, D. K., Johnson, J. A., Bailey, C. C., Boedeker, E. C. *et al.*, *Proc. Natl. Acad. Sci. USA* 1998, *95*, 3134–3139.

[6] Coelho, A., Andrade, J. R. C., Vicente, A. C. P., DiRita, V. J., *Infect. Immun.* 2000, *68*, 1700–1705.

[7] Lin, W., Fullner, K. J., Clayton, R., Sexton, J. A. *et al.*, *Proc. Natl. Acad. Sci. USA* 1999, *96*, 1071–1076.

[8] von Kruger, W. M., Humphreys, S., Ketley, J. M., *Microbiology* 1999, *145*, 2463–2475.

[9] Skorupski, K., Taylor, R. K., *Mol. Microbiol.* 1997, *25*, 1003–1009.

[10] Krukonis, E. S., DiRita, V. J., *Mol. Cell* 2003, *12*, 157–615.

[11] Bina, J., Zhu, J., Dziejman, M., Faruque, S. *et al.*, *Proc. Natl. Acad. Sci. USA* 2003, *100*, 2801–2806.

[12] Heidelberg, J. F., Eisen, J. A., Nelson, W. C., Clayton, R. A. *et al.*, *Nature* 2000, *406*, 477–483.

[13] Kim, Y. R., Lee, S. E., Kim, C. M., Kim, S. Y. *et al.*, *Infect. Immun.* 2003, *71*, 5461–5471.

[14] Makino, K., Oshima, K., Kurokawa, K., Yokoyama, K. *et al.*, *Lancet* 2003, *361*, 743–749.

[15] Trucksis, M., Michalski, J., Deng, Y. K., Kaper, J. B., *Proc. Natl. Acad. Sci. USA* 1998, *95*, 14464–14469.

[16] Tagomori, K., Iida, T., Honda, T., *J. Bacteriol.* 2002, *184*, 4351–4358.

[17] Merrell, D. S., Butler, S. M., Qadri, F., Dolganov, N. A. *et al.*, *Nature* 2002, *417*, 642–645.

[18] Xu, Q., Dziejman, M., Mekalanos, J. J., *Proc. Natl. Acad. Sci. USA* 2003, *100*, 1286–1291.

[19] Salles, C. A., Momen, H., *Trans. Royal Soc. Trop. Med. Hyg.* 1991, *85*, 544–547.

[20] Sahu, G. K., Chowdhury, R., Das, J., *Infect. Immun.* 1994, *62*, 5624–5631.

[21] Hang, L., John, M., Asaduzzaman, M., Bridges, E. A. *et al.*, *Proc. Natl. Acad. Sci. USA* 2003, *100*, 8508–8513.

[22] Hommais, F., Laurent-Winter, C., Labas, V., Krin, E. *et al.*, *Proteomics* 2002, *2*, 571–579.

[23] Levine, M. M., Black, R. E., Clements, M. L., Nalin, D. R. *et al.*, in: Holme, T., Holmgreen, J., Merson, M. H., Möllby, R. (Eds.), *Acute Enteric Infections in Children: New Prospects for Treatment and Prevention*, Elsevier-North-Holland, Amsterdam 1981, pp. 443–459.

[24] Berkelman, T., Stenstedt, T., *2-D Electrophoresis*, Amersham Biosciences, Uppsala, Sweden 1998.

[25] Shevchenko, A., Wilm, M., Vorm, O., Mann, M., *Anal. Chem.* 1996, *68*, 850–858.

[26] Jensen, O. N., Wilm, M., Schevchenko, A., Mann, M., *Methods Mol. Biol.* 1998, *112*, 513–530.

[27] Link, A. J., Robison, K., Church, G. M., *Electrophoresis* 1997, *18*, 1259–1313.

[28] Langen, H., Takacs, B., Evers, S., Berndt, P. *et al.*, *Electrophoresis* 2000, *21*, 411–429.

[29] Tonella, L., Hoogland, C., Binz, P. A., Appel, R. D. *et al.*, *Proteomics* 2001, *1*, 409–423.

[30] Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J.-C. *et al.*, in: Link, A. J., (Ed.), *2-D Proteome Analysis Protocols*, Humana Press, Totowa, NJ 1998, 531–552.

[31] Clauser, K. R., Baker, P. R., Burlingame, A. L., *Anal. Chem.* 1999, *71*, 2871–2882.

[32] VanBogelen, R. A., Schiller, E. E., Thomas, J. D., Neidhardt, F. C., *Electrophoresis* 1999, *20*, 2149–2159.

[33] Karlin, S., Mrázek, J., Campbell, A., Kaiser, D., *J. Bacteriol.* 2001, *183*, 5025–5040.

[34] Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A. *et al.*, *Nucleic Acids Res.* 2001, *29*, 22–28.

[35] Ermolaeva, M. D., White, O., Salzberg, S. L., *Nucleic Acids Res.* 2001, *29*, 1216–1221.

[36] Parsot, C., Mekalanos, J. J., *Proc. Natl. Acad. Sci. USA* 1990, *87*, 9898–9902.

[37] Miller, V. L., Mekalanos, J. J., *J. Bacteriol.* 1988, *170*, 2575–2583.

[38] Calia, K. E., Waldor, M. K., Calderwood, S. B., *Infect. Immun.* 1998, *66*, 849–852.

[39] Dziejman, M., Balon, E., Boyd, D., Fraser, C. *et al.*, *Proc. Natl. Acad. Sci. USA* 2002, *99*, 1556–1561.

[40] Appel, R. D., Bairoch, A., Sanchez, J. C., Vargas, J. R. *et al.*, *Electrophoresis* 1996, *17*, 540–546.

