ELSEVIER

Journal of Chromatography B, 782 (2002) 227–243

JOURNAL OF CHROMATOGRAPHY B

www.elsevier.com/locate/chromb

Global analysis of a ‘‘simple’’ proteome: *Methanococcus jannaschii*

Carol S. Giometti[a,*], Claudia Reich[b], Sandra Tollaksen[a], Gyorgy Babnigg[a], Hanjo Lim[c], Wenhong Zhu[c], John Yates III[c], Gary Olsen[b]

[a]*Biosciences Division, Argonne National Laboratory, 9700 South Cass Avenue, Building 202, Room B117, Argonne, IL 60439, USA*
[b]*Biochemistry Department, University of Illinois, Urbana, IL, USA*
[c]*Mass Spectrometry Group, The Scripps Institute, La Jolla, CA, USA*

Abstract

The completed genome of *Methanococcus jannaschii*, including the main chromosome and two extra-chromosomal elements, predicts a proteome comprised of 1783 proteins. How many of those proteins are expressed at any given time and the relative abundance of the expressed proteins, however, cannot be predicted solely from the genome sequence. Two-dimensional gel electrophoresis coupled with peptide mass spectrometry is being used to identify the proteins expressed by *M. jannaschii* cells grown under different conditions as part of an effort to correlate protein expression with regulatory mechanisms. Here we describe the identification of 170 of the most abundant proteins found in total lysates of *M. jannaschii* grown under optimal fermentation conditions. To optimize the number of proteins detected, two different protein specific stains (Coomassie Blue R250 or silver nitrate) and two different first dimension separation methods (isoelectric focusing or nonequilibrium pH gradient electrophoresis) were used. Thirty-two percent of the proteins identified are annotated as hypothetical (21% conserved hypothetical and 11% hypothetical), 21% are enzymes involved in energy metabolism, 12% are proteins required for protein synthesis, and the remainder include proteins necessary for intermediary metabolism, cell division, and cell structure. Evidence of post-translational modification of numerous *M. jannaschii* proteins has been found, as well as indications of incomplete dissociation of protein–protein complexes. These results demonstrate the complexity of proteome analysis even when dealing with a relatively simple genome.

Keywords: Proteomics; *Methanococcus jannaschii*; Two-dimensional gel electrophoresis

1. Introduction

M. jannaschii, an extreme thermophile (optimal growth temperature is 83 °C), was isolated from a deep sea hydrothermal vent. An obligate anaerobe, the autotrophic *M. jannaschii* utilizes H_2 and CO_2 for energy production, producing methane (CH_4). Methanotrophic bacteria can oxidize the released methane or the methane can escape into the atmosphere as a significant greenhouse gas. Due to its infrared spectrum, on a molar basis, CH_4 is a greater contributor to the greenhouse effect than CO_2. In certain contexts, such as anaerobic sludge digesters and landfills, methane gas produced is sometimes captured and used as fuel, thereby generating useful energy and protecting the atmosphere. Of all hydrocarbon fuels, CH_4 yields the most energy per unit CO_2 produced, and burning CH_4 is less harmful than releasing it. The biochemical mechanisms used by methanogenic microorganisms in general, and more

*Corresponding author. Tel.: +1-630-252-3839; fax: +1-630-252-5517.
E-mail address: csgiometti@anl.gov (C.S. Giometti).

1570-0232/02/$ – see front matter
PII: S1570-0232(02)00568-8

recently those that thrive in extreme environments, to convert CO_2 to methane is, therefore, of considerable interest in the context of both carbon management and energy production.

In 1996, the complete genome sequence of *M. jannaschii* was published [1]. Based on genetic sequence analysis, ~1783 proteins are encoded in this genome. Of these, ~70% have been assigned tentative functions based on DNA sequence similarity with known archaeal, bacterial, or eukaryotic proteins [1]. As yet, however, few *M. jannaschii* proteins have been purified or their activities studied. Having the complete genome sequence greatly facilitates protein identification, permits correlation of gene expression with genome organization, and provides a closed world in which all functions must have a corresponding gene among those in the genome. Thus, *M. jannaschii* is a prime candidate for exhaustive proteome analysis. Each protein resolved by 2DE is a verification of a gene sequence prediction. By manipulating the growth environment and then examining the cells for changes in protein composition, the co-regulation of proteins can be determined and regulatory mechanisms deduced.

Two-dimensional gel electrophoresis (2DE), with charge separation by isoelectric focusing in the first dimension followed by polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate (SDS–PAGE) in the second dimension, provides high-resolution separation and detection of a majority of the proteins produced by a biological system. The advent of peptide mass spectrometry methods with detection sensitivities in the femtomole range provides the capability to determine the masses of proteolytic fragments of proteins eluted from 2DE gels. Protein identifications can then be based on mass identity with the hypothetical proteolytic fragments generated from complete genome sequence open reading frame (ORF) databases.

Using the combined high-resolution separation of 2DE with the precision of peptide mass spectrometry, the most abundant *M. jannaschii* proteins have been identified based on comparison with the *M. jannaschii* ORF database. Differences in the staining specificity of Coomassie Blue R250 and silver nitrate, generally considered to be nonspecific protein dyes, necessitated the comparison of 2DE patterns of *M. jannaschii* proteins stained with one or the other stain in order to optimize the number of proteins detected. To further optimize the number of proteins detected, *M. jannaschii* protein samples were separated by isoelectric focusing (IEF) and by nonequilibrium pH gradient electrophoresis (NEPHGE) to allow for the detection of proteins with a broad range of isoelectric points. The results discussed show that even the smallest of genomes gives rise to proteome complexity.

2. Materials and methods

2.1. Cell growth

M. jannaschii cells were grown in a fermentor at 83 °C in a minimal mineral salts medium containing NaCl, $MgCl_2$, NH_4Cl, KCl, Na_2S, KH_2PO_4, Na_2CO_3, K_2HPO_4, $CoCl_2$, $FeCl_3$, $ZnCl_2$, $MnCl_2$, $CaCl_2$, $NiCl_2$, $CuSO_4$, Na_2MoO_4, Na_2SeO_4, Na_2WO_4, HCl, nitrilotriacetic acid, and resazurin. Cells were collected while in the exponential growth phase, washed twice with 10 m*M* Tris–HCl, pH 7.5, containing 0.15 *M* NaCl (Tris–saline), collected by centrifugation, and rapidly frozen.

2.2. Sample preparation

Frozen *M. jannaschii* cells were thawed and immediately mixed with an equal volume of a solution containing protease inhibitors (Boehringer Mannheim complete mini protease inhibitor cocktail), 9 *M* urea, 4% (v/v) Nonidet P40 (a nonionic detergent), 2% ampholytes (BioRad pH 8–10), and 2% (v/v) 2-mercaptoethanol. The cell lysate was left at room temperature for 10 min and then centrifuged at 400,000 *g* for 10 min in a Beckman TL100 tabletop ultracentrifuge. Protein concentrations of the resulting clarified lysates were determined by using the Ramagli modification of the Bradford protein assay [2].

2.3. Electrophoresis

Isoelectric focusing (IEF) was done using a mixture (50:50) of pH 3–10 and pH 5–7 carrier ampholytes (BioRad) in 10-inch tube gels with a diameter of 1.5 mm for 14,000 V-h essentially as

described by Anderson and Anderson [3]. Nonequilibrium pH gradient electrophoresis (NEPHGE) was done using pH 3–10 ampholytes (BioRad) for 6000 V-h as described by O'Farrell et al. [4]. Aliquots of *M. jannaschii* whole cell lysates containing 20 μg of protein (for silver staining) or 300 μg of protein (for Coomassie Blue staining) were loaded onto each IEF gel. After IEF or NEPHGE, gels were equilibrated in a buffer containing sodium dodecyl sulfate (SDS) as originally described by O'Farrell et al. [5] and then loaded onto 10–17% polyacrylamide gradient gels cast as slabs [6]. The second-dimension separation was done using the Laemmli buffer system as previously described [5]. Proteins were then fixed and stained in the gels using Coomassie Blue R250 in 2.5% (v/v) phosphoric acid and 50% (v/v) ethanol [7] or fixed in 50% (v/v) ethanol with 0.1% (v/v) formaldehyde and 1% (v/v) acetic acid for subsequent staining with silver nitrate [8].

2.4. Digitization and analysis of 2DE patterns

Digitization of the 2DE protein patterns was done using an Eikonix 1412 CCD scanner interfaced with a VAXstation 9000-40. The images were converted to parameter lists using the TYCHO 2DE data analysis system to determine the number of spots present, their relative integrated densities, and to generate reference spot numbers for tracking protein identifications [9]. The reference spot numbers (master spot numbers) are linked through an Oracle database table to the x and y coordinates of each protein spot detected and, hence, can be used to relate protein identifications from mass spectrometry analysis back to the appropriate protein spot locations within the 2DE image patterns (see http://ProteomeWeb.anl.gov).

2.5. Protein identification by peptide mass spectrometry

Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry peptide mass mapping and capillary liquid chromatography–electrospray tandem mass spectrometry (μ-LC–ESI-MS/MS) were used to identify *M. jannaschii* proteins. Approximately 300 μg of protein were loaded onto replicate IEF gels. After second dimension separation, the proteins were detected by staining with Coomassie Blue [7] for 24 h followed by five cycles of destaining in 20% (v/v) ethanol. Protein spots to be identified were cut from 2–5 replicate gels (number of spots required varied with abundance of individual proteins), gel pieces containing replicate proteins were combined, and the proteins were digested in-gel with trypsin (Promega sequence grade trypsin, 12.5 μg/μl). The peptides were eluted from the gel pieces by extracting three times, first with equal parts of 25 m*M* ammonium bicarbonate and acetonitrile and then twice with equal parts of 5% (v/v) formic acid and acetonitrile. Extracted tryptic peptides from the gel spots were desalted and concentrated with a commercial ZipTip C_{18} pipette tip (Millipore, Bedford, MA) prior to mass spectrometric analysis.

2.5.1. MALDI-TOF peptide mass mapping

Tryptic digest samples, mixed with α-cyano-4-hydroxycinnamic acid, were spotted onto a MALDI target plate and then transported into a Voyager DE-STR (PE Biosystem, Framingham, MA) MALDI-TOF mass spectrometer equipped with delayed extraction and reflection. MALDI-TOF mass spectra for each sample spot were generated by averaging 64–126 N_2 laser shots. Proteins were then identified using PROQUEST, a peptide mass mapping database search algorithm developed and used in the Yates laboratory, that compares experimentally obtained mass-to-charge (m/z) values with theoretically calculated m/z values of tryptic peptides of proteins from an *M. jannaschii* open reading frame database.

2.5.2. μ-LC–ESI tandem mass spectrometry

Tryptic digests were loaded onto a 365×100 μm fused-silica capillary (FSC) column packed with 10 μm POROS 10 R2 packing material (PE Biosystem, Framingham, MA) at a length of 10–15 cm. The tryptic peptides were separated with a 30-min linear gradient of 0–60% solvent B (A: 0.5% acetic acid, B: 80% acetonitrile–0.5% acetic acid), and then entered an LCQ ion trap mass spectrometer (Finnigan MAT, San Jose, CA). Tandem mass spectra were automatically collected under computer control during the 30-min LC–MS runs. Obtained MS–MS spectra were then directly subjected to SEQUEST database searches [10] without need of manual

interpretation. SEQUEST identified proteins in a spot by correlating experimental MS–MS spectra to protein sequences in the *M. jannaschii* database.

3. Results

3.1. Differential staining of M. jannaschii proteins

The 2DE protein pattern (first dimension pH range of 4–7) of *M. jannaschii* whole cell lysate from cells grown under optimal culture conditions contained ~1300 protein spots when Coomassie Blue R250 was used for protein detection (Fig. 1A) and ~1200 protein spots when silver nitrate was used for protein detection (Fig. 1B). Despite the 15-fold difference in protein amount loaded onto the first dimension IEF gel for the two different staining methods, the overall staining intensity of these patterns is comparable, verifying the greater detection sensitivity of the silver stain [8]. Differential staining of individual proteins by Coomassie Blue R250 and silver was observed, with some intensely stained proteins in the Coomassie-stained patterns appearing as more minor components of the silver-stained patterns while other proteins more dominant in the silver-stained patterns were under-represented in the Coomassie-stained patterns. Although both Coomassie Blue R250 and silver share the common feature of reacting most strongly with basic amino acids [11], other as yet undefined chemical interactions between proteins and these two dye reagents produce differential staining intensities for some proteins.

3.2. Identification of M. jannaschii proteins

Proteins for tryptic peptide analysis were selected from the Coomassie-stained pattern on the basis of their relatively high abundance and, in some cases, on the basis of observed changes in abundance in response to different growth conditions. Since the proteins selected were relatively abundant, only 2–3 replicate spots were needed in most cases to provide sufficient material for successful MALDI-TOF or LC–MS/MS analysis. From 166 unique protein spots for which tryptic peptide masses were obtained

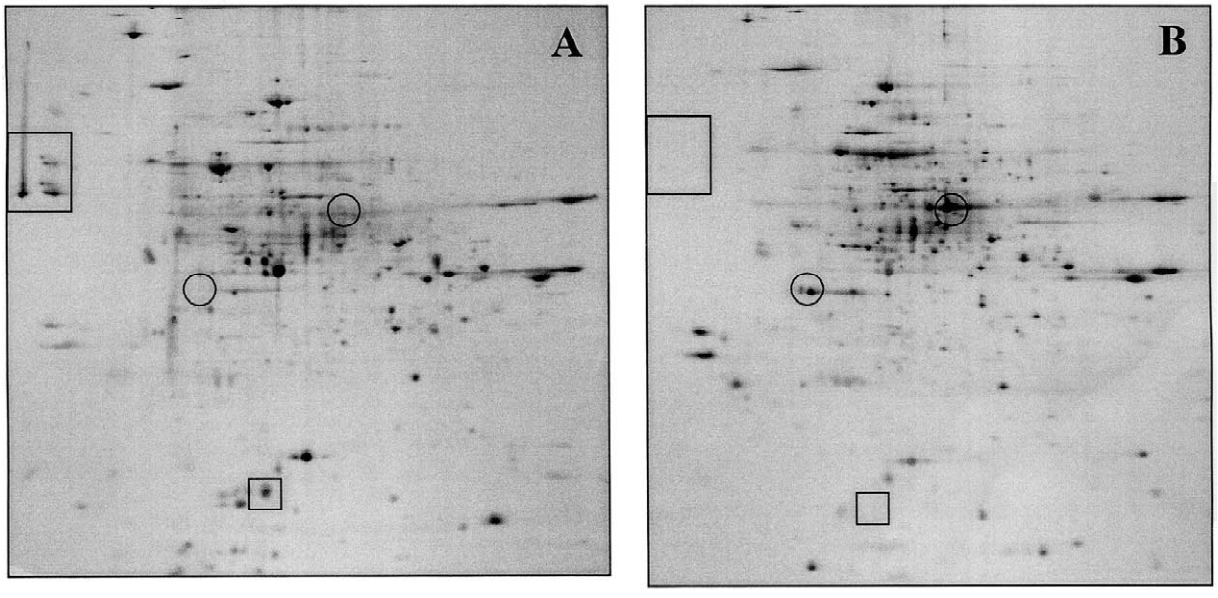


Fig. 1. Two-dimensional electrophoresis patterns of *Methanococcus jannaschii* proteins from whole cell lysates. Proteins were separated in the first dimension using isoelectric focusing with carrier ampholytes (50% pH 3–10 and 50% pH 5–7). (A) Coomassie Blue-stained proteins; (B) silver-stained proteins. Patterns are oriented with the acidic proteins to the left and basic proteins to the right, high molecular mass proteins toward the top and low molecular mass proteins toward the bottom. Rectangles indicate proteins detected with Coomassie but not silver, while circles indicate proteins detected with silver but not Coomassie.

(Fig. 1 shows a subset of those sampled; for the complete set, see http://ProteomeWeb.anl.gov), 170 different *M. jannaschii* ORFs were identified (Table 1). These identifications were based on the observation of peptides unique to a single ORF sequence. Thus, in the numerous cases were multiple ORFs were identified within a single protein spot (e.g. peptides matching Mj0083, Mj00842, Mj0845, and Mj0846 were all detected within a single protein spot cut from 2DE gels; single arrow in Fig. 2), the identification of each ORF was determined on the basis of unique peptide masses and amino acid sequences present within the sample analyzed. In some cases, the same ORF has been identified for multiple spots (e.g. peptides matching Mj0324; dashed arrow in Fig. 2; peptides matching Mj0822; rectangle in Fig. 2).

The highly reproducible 2DE patterns generated by using the methods described by Anderson and Anderson [3,6] ensured that replicate spots were cut precisely from multiple gels. Independent analysis of protein spots cut from different sets of gels confirmed that the detection of multiple proteins within some protein spots was reproducible (data not shown). Although a detailed analysis of each case of multiple proteins within an apparently single protein spot will be required to define the source of multiplicity, in some cases deductions can be made on the basis of the protein identifications. For example, methyl coenzyme M reductase I alpha subunit (Mj0846; M_w ~61 kDa), methyl coenzyme M reductase I beta subunit (Mj00842; M_w ~48 kDa), methyl coenzyme M reductase I gamma subunit (Mj0845; M_w ~30 kDa), and methyl coenzyme M reductase II alpha subunit (Mj0083; M_w ~61 kDa) are all involved in the final step of methanogenesis and are all found within the same protein spot after 2DE (solid arrow in Fig. 2). Methyl coenzyme M reductase I is a hexamer consisting of two identical trimers made up of an alpha, a beta and a gamma subunit. The presence of peptides representative of all three of these subunits in the same protein spot after 2DE separation suggests the multiprotein complex was not completely disrupted by the sample preparation conditions used. However, an intact hexamer would be ~139 kDa, i.e. larger than the mass observed. The lower molecular mass observed (~108 kDa) could result from anomalous binding of sodium dodecyl sulfate to intact complexes or could indicate a mixture of partially dissociated complexes (e.g. a combination of alpha-beta and beta-gamma subunits). Methyl coenzyme M reductase II, identified on the basis of a unique tryptic peptide sequence obtained from the same 2DE protein spot containing the three methyl coenzyme M reductase I subunits, is an isofunctional enzyme of methyl coenzyme M reductase I that is co-expressed with reductase I in pure cultures of methanogens [12]. The detection of the reductase II alpha subunit in combination with reductase I subunits is consistent with the retention of protein complex structure. Incomplete dissociation of archaeal protein complexes is a feasible possibility since many proteins synthesized by hyperthermophilic organisms such as *M. jannaschii* are believed to be resistant to conditions known to dissociate and denature proteins synthesized by mesophilic organisms.

The appearance of an ORF protein product in multiple locations within the 2DE patterns of *M. jannaschii* suggests post-translational modification of the actual ORF product. For example, Mj0324, annotated as elongation factor, is found in multiple locations varying in either isoelectric point or molecular mass (dotted arrows in Fig. 2). In eukaryotes, elongation factor is phosphorylated, suggesting that the different isoelectric forms observed in the *M. jannaschii* protein pattern represent different degrees of protein phosphorylation. The different molecular mass forms of the *M. jannaschii* elongation factor could indicate limited proteolysis of the protein. Mj0822, another ORF found in multiple locations, although in a more restricted region of the 2DE pattern than the products of ORF0324, is annotated as S-layer protein (boxed region in Fig. 2). S-layer proteins in Archaea are glycosylated, generating both molecular mass and isoelectric point heterogeneity [13]. The S-layer protein spots are among the proteins stained strongly with Coomassie Blue R250, but undetectable by silver stain, suggesting that the glycosylation of the protein interferes with the detection by silver. These results emphasize the level of complexity represented by a proteome, including not only variations in the relative abundance of protein products, but also in their structure (i.e. post-translational modifications) and their associations with other proteins.

Table 1
Methanococcus jannaschii protein identifications

MSN	MALDI peptides	MS–MS peptides	M_w	p*I*	ORF	Annotation
MJIEF4	12	9	61227	5.12	MJ0846	Methyl coenzyme M reductase I, subunit alpha (mcrA)
MJIEF4		4	61241	5.13	MJ0083	Methyl coenzyme M reductase II, subunit alpha (mtrA)
MJIEF4	5	10	47789	5.61	MJ0842	Methyl coenzyme M reductase I, subunit beta (mcrB)
MJIEF4	7	10	30155	5.27	MJ0845	Methyl coenzyme M reductase I, subunit gamma (mcrG)
MJIEF3	7	10	47349	6.04	MJ0748	Flavoprotein (fpaA)
MJIEF6	31	18	100341	5.4	MJ1156	AAA family ATPase
MJIEF70	11	4	49067	5.26	MJ0937	Phosphoribosylamine-glycine ligase (purD)
MJIEF40	12	17	49067	5.26	MJ0937	Phosphoribosylamine-glycine ligase (purD)
MJIEF40	5	8	51356	5.35	MJ0216	H^+-transporting ATP synthase, subunit B (atpB)
MJIEF13	5	7	34877	5.5	MJ1636	N^5,N^{10}-methenyl-tetrahydromethanopterin cyclohydrolase (mch)
MJIEF62	7	8	47250	7.16	MJ0324	Translation elongation factor EF-1, subunit alpha
MJIEF144	14	3	61534	5.56	MJ0765	[6Fe-6S] prismane-containing protein
MJIEF144		3	49067	5.26	MJ0937	Phosphoribosylamine-glycine ligase (purD)
MJIEF50		5	47789	5.61	MJ0842	Methyl coenzyme M reductase I, subunit beta (mcrB)
MJIEF50		3	42984	5.48	MJ0007	2-Hydroxyglutaryl-CoA dehydratase, putative
MJIEF50		6	38689	5.42	MJ0784	H_2-forming N^5,N^{10}-methylene-tetrahydromethanopterin dehydrogenase
MJIEF69	9	17	43477	5.76	MJ0622	Cell division protein (ftsZ)
MJIEF69		3	40907	5.4	MJ0210	Succinyl-CoA synthetase, beta subunit
MJIEF69		3	45555	5.43	MJ0732	Flavoprotein (fprA)
MJIEF69		7	38689	5.42	MJ0784	H_2-forming N^5,N^{10}-methylene-tetrahydromethanopterin dehydrogenase
MJIEF107	7		71962	8.17	MJ1610	Glucoamylase, putative
MJIEF107		7	43505	5.76	MJ0622	Cell division protein (ftsZ)
MJIEF107		6	45555	5.43	MJ0732	Flavoprotein (fprA)
MJIEF174	14	11	42386	5.87	MJ0959	Aspartate aminotransferase (aspC)
MJIEF75	10	7	39926	5.91	MJ0895	Flagella-related protein D, putative
MJIEF220	13	14	30155	5.27	MJ0845	Methyl coenzyme M reductase I, subunit gamma (mcrG)
MJIEF117	10	21	30155	5.27	MJ0845	Methyl coenzyme M reductase I, subunit gamma (mcrG)
MJIEF88	12	19	36830	5.69	MJ0715	H(2)-dependent methylene-tetrahydromethanopterin dehydrogenase
MJIEF88		4	30155	5.27	MJ0845	Methyl coenzyme M reductase I, subunit gamma (mcrG)
MJIEF88		3	47250	7.16	MJ0324	Translation elongation factor EF-1, subunit alpha
MJIEF115	9	3	32424	6.06	MJ0318	Formylmethanofuran:tetrahydromethanopterin formyltransferase (ftr)
MJIEF278	10	15	35331	6.78	MJ1534	N^5,N^{10}-methylene-tetrahydromethanopterin reductase (mer)
MJIEF581	9	8	38188	7.1	MJ1338	H(2)-dependent methylene-tetrahydromethanopterin dehydrogenase
MJIEF581		5	35331	6.78	MJ1534	N^5,N^{10}-methylene-tetrahydromethanopterin reductase (mer)
MJIEF581		5	30155	5.27	MJ0845	Methyl coenzyme M reductase I, subunit gamma (mcrG)
MJIEF148	10	4	35331	6.78	MJ1534	N^5,N^{10}-methylene-tetrahydromethanopterin reductase (mer)
MJIEF128	15	11	47250	7.16	MJ0324	Translation elongation factor EF-1, subunit alpha
MJIEF128		4	30155	5.27	MJ0845	Methyl coenzyme M reductase I, subunit gamma (mcrG)
MJIEF128		7	35331	6.78	MJ1534	N^5,N^{10}-methylene-tetrahydromethanopterin reductase (mer)
MJIEF65	10	5	40234	5.09	MJ1249	Conserved hypothetical protein
MJIEF65		3	47250	7.16	MJ0324	Translation elongation factor EF-1, subunit alpha
MJIEF65		3	35331	6.78	MJ1534	N^5,N^{10}-methylene-tetrahydromethanopterin reductase (mer)
MJIEF1263	9	5	26227	6.03	MJ0308	Conserved hypothetical protein
MJIEF1263		5	35331	6.78	MJ1534	N^5,N^{10}-methylene-tetrahydromethanopterin reductase (mer)
MJIEF1263		3	30155	5.27	MJ0845	Methyl coenzyme M reductase I, subunit gamma (mcrG)
MJIEF194	8	13	30286	5.86	MJ1035	N^5,N^{10}-methylene-tetrahydromethanopterin dehydrogenase (mtd)
MJIEF169	9	5	25931	6.69	MJ0825	Peptidyl-prolyl *cis–trans* isomerase, FKBP-type rotamase
MJIEF195	9	7	22152	5.78	MJ0479	Adenylate kinase (adk)
MJIEF195	7	8	23306	6.04	MJ1646	Orotate phosphoribosyl transferase
MJIEF443		3	22700	4.91	MJ0891	Flagellin B1 (flaB1)

Table 1. Continued

MSN	MALDI peptides	MS–MS peptides	M_w	p*I*	ORF	Annotation
MJIEF170	8	3	18777	5.4	MJ1053	Hypothetical protein
MJIEF469		7	15905	5.2	MJ0746	Conserved hypothetical protein
MJIEF469		4	16296	5.46	MJ1199	Invalid gene
MJIEF759	6	6	15882	5.02	MJ0896	Flagella-related protein E, putative
MJIEF252	5	3	188878	6.86	MJ0531	Conserved hypothetical protein
MJIEF364	6	11	14094	6.51	MJ0733	Hypothetical protein
MJIEF322	6	10	12443	5.76	MJ1539	Hypothetical protein
MJIEF1051	5	14	60510	4.27	MJ0822	S-layer structural protein
MJIEF1250	4	25	60529	4.27	MJ0822	S-layer structural protein
MJIEF1244	7	22	60529	4.27	MJ0822	S-layer structural protein
MJIEF1051	7	24	60529	4.27	MJ0822	S-layer structural protein
MJIEF17	13	24	66426	5.06	MJ0217	H^+ transporting ATP synthease, subunit A (atpA)
MJIEF25	13	18	51356	5.35	MJ0216	H^+ transporting ATP synthease, subunit B (atpB)
MJIEF130	9	21	51356	5.35	MJ0216	H^+ transporting ATP synthease, subunit B (atpB)
MJIEF27	10	21	51356	5.35	MJ0216	H^+ transporting ATP synthease, subunit B (atpB)
MJIEF1263	10	19	44810	4.85	MJ0216	H^+ transporting ATP synthease, subunit C (atpC)
MJIEF194	5	26	39932	5.91	MJ0895	Flagella-related protein D, putative
MJIEF95	11	14	36452	4.92	MJ0854	N^5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit H
MJIEF1223	3	8	22559	5.4	MJ0892	Flagellin B2 (flaB2)
MJIEF443	3	8	22682	4.91	MJ0891	Flagellin B1 (flaB1)
MJIEF1257		4	10344	4.41	MJ0508	LSU ribosomal protein L12A (P1)
MJIEF215	3	9	15873	5.02	MJ0896	Flagella-related protein E, putative
MJIEF252	3	8	18872	6.86	MJ0531	Conserved hypothetical protein
MJIEF498	3	7	12667	5.53	MJ1203	LSU ribosomal protein L7AE
MJIEF511	2	9	11823	5.8	MJ0394	SSU ribosomal protein S24E
MJIEF322	6	10	12432	5.76	MJ1539	Hypothetical protein
MJIEF33	13	14	80866	4.92	MJ1264	Phosphoribosylformylglycinamidine synthase II (purL)
MJIEF33	6	4	25417	5.99	MJ1648	Phosphoribosylformylglycinamidine synthase I (purQ)
MJIEF92	10	10	80866	4.92	MJ1264	Phosphoribosylformylglycinamidine synthase II (purL)
MJIEF92	9	11	32249	5.43	MJ1366	Ribose-phosphate pyrophosphokinase, (prsA)
MJIEF148	10	10	35353	6.78	MJ1534	N^5,N^{10}-methylene-tetrahydromethanopterin reductase (mer)
MJIEF57	15	9	30174	5.27	MJ0845	Methyl coenzyme M reductase I, subunit gamma (mcrG)
MJIEF65	10	18	40260	5.09	MJ1249	Conserved hypothetical protein
MJIEF65	3	3	36471	4.92	MJ0854	N^5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit H (mtrH)
MJIEF39	3	16	61573	5.56	MJ0765	[6Fe-6S] prismane-containing protein
MJIEF88	8	9	32090	5.46	MJ1334	UDP-glucose pyrophosphorylase
MJIEF290	12	12	32096	5.35	MJ1008	Branched-chain amino acid aminotransferase (ilvE)
MJIEF139	7	8	14411	5.33	MJ0763	Hypothetical protein
MJIEF139		8	47789	5.61	MJ0842	Methyl coenzyme M reductase I, subunit beta (mcrB)
MJIEF32	6	8	30306	5.86	MJ1035	N^5,N^{10}-methylene-tetrahydromethanopterin dehydrogenase (mtd)
MJIEF626	6	2	30306	5.86	MJ1035	N^5,N^{10}-methylene-tetrahydromethanopterin dehydrogenase
MJIEF14	19	16	58772	5.04	MJ0999	Thermosome subunit (ths)
MJIEF14		3	60548	4.27	MJ0822	S-layer structural protein
MJIEF18	15	12	65700	5.65	MJ1355	Conserved hypothetical protein
MJIEF1250	8	17	60548	4.27	MJ0822	S-layer structural protein
MJIEF142	14	10	35009	5.5	MJ1455	Hypothetical protein
MJIEF142		3	47380	6.04	MJ0748	Flavoprotein (fpaA)
MJIEF115	9	11	36854	5.69	MJ0715	H(2)-dependent methylenetetrahydromethanopterin dehydrogenase
MJIEF278	12	9	35353	6.78	MJ1534	N^5,N^{10}-methylene-tetrahydromethanopterin reductase (mer)
MJIEF268	10	5	35866	6.35	MJ0094	Conserved hypothetical protein

Table 1. Continued

MSN	MALDI peptides	MS–MS peptides	M_w	p*I*	ORF	Annotation
MJIEF128	12	11	47280	7.16	MJ0324	Translation elongation factor EF-1, subunit alpha
MJIEF1221	17	9	31719	6.26	MJ1225	Conserved hypothetical protein
MJIEF333	12	12	26244	6.03	MJ0308	Conserved hypothetical protein
MJIEF1223	4	5	22577	5.4	MJ0892	Flagellin B2 (flaB2)
MJIEF1267	12	16	29634	5.45	MJ1615	Hypothetical protein
MJIEF666	7	13	29130	5.46	MJ0924	Capsular polysaccharide biosynthesis protein
MJIEF1264	12	12	25559	5.22	MJ0736	Alkyl hydroperoxide reductase
MJIEF188	7	5	25611	5.28	MJ1259	Uridylate kinase (pyrH)
MJIEF313	7	7	22166	5.78	MJ0479	Adenylate kinase (adk)
MJIEF195	11	10	23306	6.04	MJ1646	Orotate phosphoribosyl transferase
MJIEF195		3	22166	5.78	MJ0479	Adenylate kinase (adk)
MJIEF467	4	5	22367	7.87	MJ0312	Hypothetical protein
MJIEF410	4	5	22367	7.87	MJ0312	Hypothetical protein
MJIEF170	11	13	18770	5.4	MJ1053	Hypothetical protein
MJIEF358	13	10	18770	5.4	MJ1053	Hypothetical protein
MJIEF492	7	7	17669	4.44	MJ0278	Peptidyl-prolyl isomerase, FKBP-rotamase
MJIEF182	7	11	17707	5.65	MJ1509	Hypothetical protein
MJIEF469	8	5	15895	5.2	MJ0746	Conserved hypothetical protein
MJIEF469		4	16296	5.46	MJ1199	Invalid gene
MJIEF216	6	9	14215	5.86	MJ0405	Conserved hypothetical protein
MJIEF252	8	10	18890	6.86	MJ0531	Conserved hypothetical protein
MJIEF252	5	6	15459	6.13	MJ0922	Conserved hypothetical protein
MJIEF252		3	15359	8.08	MJ1158	Hypothetical protein
MJIEF364	7	6	14084	6.51	MJ0733	Hypothetical protein
MJIEF364		6	63908	5.35	MJ1231	Oxaloacetate decarboxylase alpha chain (oadA)
MJIEF364	3	2	13698	6.41	MJ0987	Conserved hypothetical protein
MJIEF322	9	10	12433	5.76	MJ1539	Hypothetical protein
MJIEF133	6	4	8465	5.17	MJ0990	Conserved hypothetical protein
MJIEF178	10	2	38178	5.35	MJ0203	Phosphoribosylformylglycinamidine cyclo-ligase (purM)
MJIEF178	7	3	30174	5.27	MJ0845	Methyl coenzyme M reductase I, subunit gamma (mcrG)
MJIEF190	14	12	30155	5.27	MJ0845	Methyl coenzyme M reductase I, subunit gamma (mcrG)
MJIEF190		11	36630	5.36	MJ0234	Anthranilate synthase component II (trpD)
MJIEF220	10	5	40798	5.33	MJ0136	Conserved hypothetical protein
MJIEF220	7		38924	5.14	MJ0370	Cell division protein (ftsZ)
MJIEF220	4		37977	5.19	MJ1065	Polysaccharide biosynthesis protein
MJIEF50		4	43505	5.76	MJ0622	Cell division protein (ftsZ)
MJIEF94		3	46249	5.88	MJ1266	High-affinity branched-chain amino acid ABC transporter (braC)
MJIEF99		7	47610	7.17	MJ1404	Hypothetical protein
MJIEF51		16	44829	4.85	MJ0219	H^+-transporting ATP synthase, subunit C (atpC)
MJIEF64		9	26328	4.65	MJ0851	N^5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase,subunit A (mtrA)
MJIEF65		20	36471	4.92	MJ0854	N^5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase, subunit H (mtrH)
MJIEF65		19	40260	5.09	MJ1249	Conserved hypothetical protein
MJIEF65		3	30174	5.27	MJ0845	Methyl coenzyme M reductase I, subunit gamma (mcrG)
MJIEF65		3	29461	5.23	MJ1171	Formylmethanofuran dehydrogenase, subunit C (tungsten) (fwdC)
MJIEF81		13	22577	5.4	MJ0892	Flagellin B2 (flaB2)
MJIEF86		23	54722	5.68	MJ0667	Thymidine phosphorylase
MJIEF106		22	38212	7.1	MJ1338	H(2)-dependent methylene-tetrahydromethanopterin dehydrogenase
MJIEF106		4	47280	7.16	MJ0324	Translation elongation factor EF-1, subunit alpha

Table 1. Continued

MSN	MALDI peptides	MS–MS peptides	M_w	p*I*	ORF	Annotation
MJIEF119		9	22700	4.91	MJ0891	Flagellin B1 (flaB1)
MJIEF125		13	32249	5.43	MJ1336	Ribose-phosphate pyrophosphokinase, (prsA)
MJIEF125		3	30174	5.27	MJ0845	Methyl coenzyme M reductase I, subunit gamma (mcrG)
MJIEF166		7	22700	4.91	MJ0891	Flagellin B1 (flaB1)
MJIEF166		3	40260	5.09	MJ1249	Conserved hypothetical protein
MJIEF307		6	22700	4.91	MJ0891	Flagellin B1 (flaB1)
MJIEF1224		3	22700	4.91	MJ0891	Flagellin B1 (flaB1)
MJIEF1225		5	22700	4.91	MJ0891	Flagellin B1 (flaB1)
MJIEF1227		17	47280	7.16	MJ0324	Translation elongation factor EF-1, subunit alpha
MJBASO1		11	58772	5.04	MJ0999	Thermosome subunit (ths)
MJBASO2		20	47280	7.16	MJ0324	Translation elongation factor EF-1, subunit alpha
MJBASO2		3	47609	7.17	MJ1404	Hypothetical protein
MJBASO3		27	100403	5.4	MJ1156	AAA family ATPase
MJBASO4		16	58772	5.04	MJ0999	Thermosome subunit (ths)
MJBASO4		9	65700	5.65	MJ1355	Conserved hypothetical protein
MJBASO5		18	47280	7.16	MJ0324	Translation elongation factor EF-1, subunit alpha
MJBASO6		14	38212	7.1	MJ1338	H(2)-dependent methylenetetrahydromethanopterin dehydrogenase
MJBASO7		14	35353	6.78	MJ1534	N^5,N^{10}-methylene-tetrahydromethanopterin reductase (mer)
MJBASO8		7	35353	6.78	MJ1534	N^5,N^{10}-methylene-tetrahydromethanopterin reductase (mer)
MJBASO9		4	27411	7.79	MJ0592	Conserved hypothetical protein
MJBASO10		12	14180	5.88	MJ1228	Translation initiation factor aIF-5A
MJBASO11		5	12451	5.76	MJ1539	Hypothetical protein
MJBASO12		8	11842	5.8	MJ0394	SSU ribosomal protein S24E
MJBASO13		11	12490	7.78	MJ0059	Nitrogen regulatory protein P-II (glnB)
MJBASO14		6	10941	6.6	MJ0218	H^+-transporting ATP synthase, subunit F (atpF)
MJBASO15		3	10034	6.74	MJ1118	Conserved hypothetical protein
MJBASO16		7	11703	8.62	MJ0742	Conserved hypothetical protein
MJBASO16		3	9856	8.8	MJ0178	LSU ribosomal protein L23P (rplW)
MJBASO17		4	10292	8.89	MJ1325	Transcriptional regulator
MJBASO17		3	8739	8.5	MJ0657	LSU ribosomal protein L14E
MJBASO18		4	12233	9.66	MJ0322	SSU ribosomal protein S10P (rpsJ)
MJBASO19		11	23325	9.33	MJ0461	SSU ribosomal protein S3P (rpsC)
MJBASO19		4	33098	9.14	MJ0266	Pyruvate ferredoxin oxidoreductase, subunit beta (porB)
MJBASO20		13	23839	9.6	MJ0475	SSU ribosomal protein S5P (rpsE)
MJBASO20		4	24285	8.9	MJ1237	Proteasome, subunit beta (psmB)
MJBASO21		6	21760	9.66	MJ1047	SSU ribosomal protein S7P (rpsG)
MJBASO22		8	23325	9.33	MJ0461	SSU ribosomal protein S3P (rpsC)
MJBASO23		7	23325	9.33	MJ0461	SSU ribosomal protein S3P (rpsC)
MJBASO23		5	25708	9.62	MJ0982	SSU ribosomal protein S2P
MJBASO24		5	24951	9.25	MJ0651	H^+-transporting ATP synthase, subunit D (atpD)
MJBASO24		5	21760	9.66	MJ1047	SSU ribosomal protein S7P (rpsG)
MJBASO25		32	46652	5.67	MJ1192	Methylviologen-reducing hydrogenase, alpha chain (vhuA)
MJBASO25		4	46107	5.58	MJ0499	3-Isopropylmalate dehydratase large subunit/aconitase-related protein
MJBASO26		17	39951	5.91	MJ0895	Flagella-related protein D, putative
MJBASO26		7	36751	5.68	MJ0509	LSU ribosomal protein L10E
MJBASO26		4	42413	5.87	MJ0959	Aspartate aminotransferase (aspC)
MJIEF1231		3	22598	5.6	MJ0892	Flagellin B2 (flaB2)
MJIEF1233		5	22598	5.6	MJ0892	Flagellin B2 (flaB2)
MJIEF272		3	22598	5.6	MJ0892	Flagellin B2 (flaB2)
MJIEF81		4	21758	4.7	MJ0653	Conserved hypothetical protein

Table 1. Continued

MSN	MALDI peptides	MS–MS peptides	M_w	p*I*	ORF	Annotation
MJIEF22		16	134372	6.6	MJ0542	Phosphoenolpyruvate synthase
MJIEF1027		5	81046	6.1	MJ1048	Translation elongation factor EF-2
MJIEF15		12	77197	5.4	MJ0222	H^+-transporting ATP synthase, subunit I (atpI)
MJIEF15		3	61208	6	MJ1174	CTP synthase (pyrG)
MJIEF7		54	58738	5.1	MJ0999	Thermosome subunit (ths)
MJIEF38		18	56956	6	MJ1018	Phosphoglycerate dehydrogenase (serA)
MJIEF38		5	61530	5.3	MJ0846	Methyl coenzyme M reductase I, subunit alpha (mcrA)
MJIEF86		22	54809	5.9	MJ0667	Thymidine phosphorylase
MJIEF86		5	47898	5.9	MJ0842	Methyl coenzyme M reductase I, subunit beta (mcrB)
MJIEF86		5	56651	6	MJ1040	DNA-directed RNA polymerase, subunit B″ (rpoB2)
MJIEF86		4	56847	6.2	MJ1429	Conserved hypothetical protein
MJIEF49		43	47898	5.9	MJ0842	Methyl coenzyme M reductase I, subunit beta (mcrB)
MJIEF49		4	47632	6.5	MJ0748	Flavoprotein (fpaA)
MJIEF76		24	47898	5.9	MJ0842	Methyl coenzyme M reductase I, subunit beta (mcrB)
MJIEF76		8	47632	6.5	MJ0748	Flavoprotein (fpaA)
MJIEF76		4	51517	5.7	MJ1346	Glutamine synthetase (glnA)
MJIEF55		15	47632	6.5	MJ0748	Flavoprotein (fpaA)
MJIEF55		12	49841	6.3	MJ0603	Glutamate-1-semialdehyde aminotransferase (hemL)
MJIEF55		10	43098	5.6	MJ0007	2-Hydroxyglutaryl-CoA dehydratase, putative
MJIEF55		4	47898	5.9	MJ0842	Methyl coenzyme M reductase I, subunit beta (mcrB)
MJIEF55		4	43260	5.7	MJ0532	Geranylgeranyl hydrogenase
MJIEF135		17	39997	5.6	MJ1142	Arsenical pump-driving ATPase, putative
MJIEF135		11	36768	5.8	MJ0509	LSU ribosomal protein L10E
MJIEF135		8	47632	6.5	MJ0748	Flavoprotein (fpaA)
MJIEF135		7	43284	6.1	MJ0812	Conserved hypothetical protein
MJIEF135		5	47898	5.9	MJ0842	Methyl coenzyme M reductase I, subunit beta (mcrB)
MJIEF135		5	38730	5.6	MJ0784	H_2-forming N^5,N^{10}-methylene-tetrahydromethanopterin dehydrogenase
MJIEF135		4	43098	5.6	MJ0007	2-Hydroxyglutaryl-CoA dehydratase, putative
MJIEF135		3	35067	5.9	MJ0034	Conserved hypothetical protein
MJIEF75		23	39977	6	MJ0895	Flagella-related protein D, putative
MJIEF75		14	42528	6.2	MJ0959	Aspartate aminotransferase (aspC)
MJIEF75		5	46425	6.1	MJ1266	High-affinity branched-chain amino acid ABC transporter + G135 (braC)
MJIEF75		3	36768	5.8	MJ0509	LSU ribosomal protein L10E
MJIEF52		28	38874	5.2	MJ0370	Cell division protein (ftsZ)
MJIEF52		10	45005	4.9	MJ0219	H^+-transporting ATP synthase, subunit C (atpC)
MJIEF52		5	40941	5.4	MJ0136	Conserved hypothetical protein
MJIEF52		3	39480	5.4	MJ0205	Aspartate-semialdehyde dehydrogenase (asd)
MJIEF52		3	37928	5.7	MJ0561	Adenylosuccinate synthetase (purA)
MJIEF51		18	38874	5.2	MJ0370	Cell division protein (ftsZ)
MJIEF51		11	45005	4.9	MJ0219	H^+-transporting ATP synthase, subunit C (atpC)
MJIEF51		8	35914	5.2	MJ0677	Conserved hypothetical protein
MJIEF51		4	40941	5.4	MJ0136	Conserved hypothetical protein
MJIEF51		3	39954	5.2	MJ0862	Conserved hypothetical protein
MJIEF125		14	32337	5.6	MJ1366	Ribose-phosphate pyrophosphokinase, (prsA)
MJIEF217		7	22842	5.1	MJ0827	Conserved hypothetical protein
MJIEF217		3	28616	5.7	MJ0837	Hypothetical protein
MJIEF151		27	29208	5.6	MJ0924	Capsular polysaccharide biosynthesis protein, putative
MJIEF151		4	28942	5.8	MJ0761	Conserved hypothetical protein

Table 1. Continued

MSN	MALDI peptides	MS–MS peptides	M_w	p*I*	ORF	Annotation
MJIEF1254		5	39977	6	MJ0895	Flagella-related protein D, putative
MJIEF1254		3	15907	5.1	MJ0896	Flagella-related protein E, putative
MJIEF1254		4	29208	5.6	MJ0924	Capsular polysaccharide biosynthesis protein
MJIEF216		9	14863	5.6	MJ0614	Hypothetical protein
MJIEF233		10	29688	6.4	MJ0400	Conserved hypothetical protein
MJIEF172		6	29688	6.4	MJ0400	Conserved hypothetical protein
MJIEF169		21	26326	6.4	MJ0308	Conserved hypothetical protein
MJIEF169		3	29688	6.4	MJ0400	Conserved hypothetical protein
MJIEF278		18	38267	7.6	MJ1338	H(2)-dependent methylenetetrahydromethanopterin dehydrogenase
MJIEF278		11	47324	7.6	MJ0324	Translation elongation factor EF-1, subunit alpha
MJIEF88		30	36936	6	MJ0715	H(2)-dependent methylenetetrahydromethanopterin dehydrogenase
MJIEF88		9	30278	5.5	MJ0845	Methyl coenzyme M reductase I, subunit gamma (mcrG)
MJIEF88		6	47632	6.5	MJ0748	Flavoprotein (fpaA)
MJIEF88		6	47324	7.6	MJ0324	Translation elongation factor EF-1, subunit alpha
MJIEF88		3	30934	6.1	MJ1246	Succinyl-CoA synthetase, alpha subunit (sucD)
MJIEF281		18	30934	6.1	MJ1246	Succinyl-CoA synthetase, alpha subunit (sucD)
MJIEF281		11	35424	7.2	MJ1534	N^5,N^{10}-methylene-tetrahydromethanopterin reductase (mer)
MJIEF281		5	30278	5.5	MJ0845	Methyl coenzyme M reductase I, subunit gamma (mcrG)
MJIEF281		6	36936	6	MJ0715	H(2)-dependent methylenetetrahydromethanopterin dehydrogenase
MJIEF281		6	47324	7.6	MJ0324	Translation elongation factor EF-1, subunit alpha
MJIEF281		4	35237	5.6	MJ1455	Hypothetical protein
MJIEF1280		8	39977	6	MJ0895	Flagella-related protein D
MJIEF1280		6	36768	5.8	MJ0509	LSU ribosomal protein L10E
MJIEF1280		7	42528	6.2	MJ0959	Aspartate aminotransferase (aspC)
MJIEF1280		6	46425	6.1	MJ1266	High-affinity branched-chain amino acid ABC transporter (braC)
MJIEF1280		4	36519	6.2	MJ1602	Homoserine dehydrogenase (hom)
MJIEF1281		19	47632	6.5	MJ0748	Flavoprotein (fpaA)
MJIEF1281		15	39977	6	MJ0895	Flagella-related protein D, putative
MJIEF1281		10	42528	6.2	MJ0959	Aspartate aminotransferase (aspC)
MJIEF1281		5	38730	5.6	MJ0784	H_2-forming N^5,N^{10}-methylene-tetrahydromethanopterin dehydrogenase
MJIEF1281		7	47898	5.9	MJ0842	Methyl coenzyme M reductase I, subunit beta (mcrB)
MJIEF1281		4	43098	5.6	MJ0007	2-Hydroxyglutaryl-CoA dehydratase, putative
MJIEF1281		5	11710	8.6	MJ0742	Conserved hypothetical protein
MJIEF1281		4	46425	6.1	MJ1266	High-affinity branched-chain amino acid ABC transporter (braC)
MJIEF1281		5	36519	6.2	MJ1602	Homoserine dehydrogenase (hom)
MJIEF1271		32	81046	6.1	MJ1048	Translation elongation factor EF-2
MJIEF1271		5	47632	6.5	MJ0748	Flavoprotein (fpaA)
MJIEF1271		5	16226	7.4	MJ1265	Nucleoside diphosphate kinase (ndk)
MJIEF1271		3	61530	5.3	MJ0846	Methyl coenzyme M reductase I, subunit alpha (mcrA)
MJIEF1271		3	47898	5.9	MJ0842	Methyl coenzyme M reductase I, subunit beta (mcrB)
MJIEF77		55	81046	6.1	MJ1048	Translation elongation factor EF-2
MJIEF77		7	61530	5.3	MJ0846	Methyl coenzyme M reductase I, subunit alpha (mcrA)
MJIEF77		5	78414	8.3	MJ0590	Conserved hypothetical protein
MJIEF77		4	61517	5.3	MJ0083	Methyl coenzyme M reductase II, subunit alpha (mtrA)
MJIEF77		4	68707	5.5	MJ1378-C	Carbamoyl-phosphate synthase, large subunit
MJIEF77		4	77197	5.4	MJ0222	H^+-transporting ATP synthase, subunit I (atpI)
MJIEF77		3	26326	6.4	MJ0308	Conserved hypothetical protein
MJIEF77		5	18325	7.5	MJ0577	Conserved hypothetical protein
MJIEF167		11	12596	5.2	MJ0989	Conserved hypothetical protein

Table 1. Continued

MSN	MALDI peptides	MS–MS peptides	M_w	p*I*	ORF	Annotation
MJIEF227		17	9459	5.5	MJ0853	N^5-methyl-tetrahydromethanopterin:coenzyme M methyltransferase,subunit G(mtrG)
MJIEF1273		17	35049	5.6	MJ1636	N^5,N^{10}-methenyl-tetrahydromethanopterin cyclohydrolase (mch)
MJIEF1273		10	47909	5.5	MJ1391	Putative aminotransferase
MJIEF1273		5	61530	5.3	MJ0846	Methyl coenzyme M reductase I, subunit alpha (mcrA)
MJIEF367		4	23105	8	MJ0220	H^+-transporting ATP synthase, subunit E (atpE)
MJIEF201		4	25602	5.3	MJ0736	Alkyl hydroperoxide reductase, putative
MJIEF149		26	25602	5.3	MJ0736	Alkyl hydroperoxide reductase, putative
MJIEF149		7	20584	5.8	MJ0536	2-Ketoglutarate ferredoxin oxidoreductase, subunit gamma (korG)
MJIEF187		25	27666	4.8	MJ0247	Proliferating-cell nuclear antigen (pcnA)
MJIEF187		5	24452	4.6	MJ0048	Conserved hypothetical protein
MJIEF187		3	14766	6.3	MJ0866	HIT family protein
MJIEF187		3	18931	4.7	MJ1049	Hypothetical protein
MJIEF187		3	22410	5.1	MJ1356	Hypothetical protein
MJIEF241		6	27666	4.8	MJ0247	Proliferating-cell nuclear antigen (pcnA) G31
MJIEF165		3	22410	5.1	MJ1356	Hypothetical protein

Proteins were selected for identification based on their abundance and on their quantitative changes in response to different growth conditions. MSN, master spot number of the protein spot analyzed (provides an identifier for Oracle database management system (each MALDI or MS–MS analysis was done with peptides extracted from 2–5 replicate gel pieces containing the protein spot of interest); IEF refers to proteins in 2DE patterns for which IEF was used for the first dimension separation, while BASO refers to proteins in 2DE patterns for which NEPHGE was used for the first dimension separation); MALDI peptides, the number of observed peptides with masses obtained by MALDI-TOF that produced matches with the *M. jannaschii* ORF database as described in Section 2; MS–MS peptides, the number of observed peptides with masses obtained by LC–MS/MS that produced matches with the *M. jannaschii* ORF database as described in Section 2; M_w, predicted molecular mass based on sequence of ORF; p*I*, predicted isoelectric point based on sequence of ORF; ORF, *M. jannaschii* open reading frame predicting tryptic peptides with masses most comparable to the observed peptide masses; annotation, annotation for the ORF in the TIGR Microbial Genome sequence database (http://www.tigr.org/tdb/mdb/mdbcomplete.html).

3.3. Separation and identification of alkaline M. jannaschii proteins

The genome of *M. jannaschii* predicts a substantial number of proteins with isoelectric points above pH 7 which are not resolved by isoelectric focusing with carrier ampholytes due to cathodic drift. Therefore, NEPHGE was used for the first-dimension separation of *M. jannaschii* proteins to detect these more alkaline proteins. Under the conditions used for the NEPHGE separation (i.e. 6000 V-h), proteins more basic than those detected by IEF were revealed (Fig. 3). Twenty-six proteins (Table 1, master spot numbers beginning with MJBASO) were selected from the NEGPHE 2DE pattern stained with Coomassie Blue R250 for identification by tryptic peptide mass analysis, and 29 ORFs were matched. Since ampholytes and proteins do not focus under the nonequilibrium conditions used for NEPHGE separation, the actual pH gradient of the NEPHGE gels could not be determined. However, the predicted isoelectric points for the proteins identified ranged from pH 5.5 up to 9.7. These results clearly demonstrate the usefulness of the NEPHGE method for the detection of the very alkaline *M. jannaschii* proteins.

3.4. Functional categories of the identified M. jannaschii proteins

The functional categories represented by the ORF products identified in this initial phase of the *M. jannaschii* proteome project were determined (Fig. 4). The greatest percentage of the identified proteins fell into the hypothetical and conserved hypothetical categories, with the next largest percentage in the energy metabolism category. Proteins involved in protein synthesis, nucleotide synthesis, and amino acid synthesis comprised the next largest segment of the proteins identified. The remainder of the identified proteins was distributed among protein catabolism, transport and binding, regulatory processes, transcription, cell structure (e.g. cell envelope),

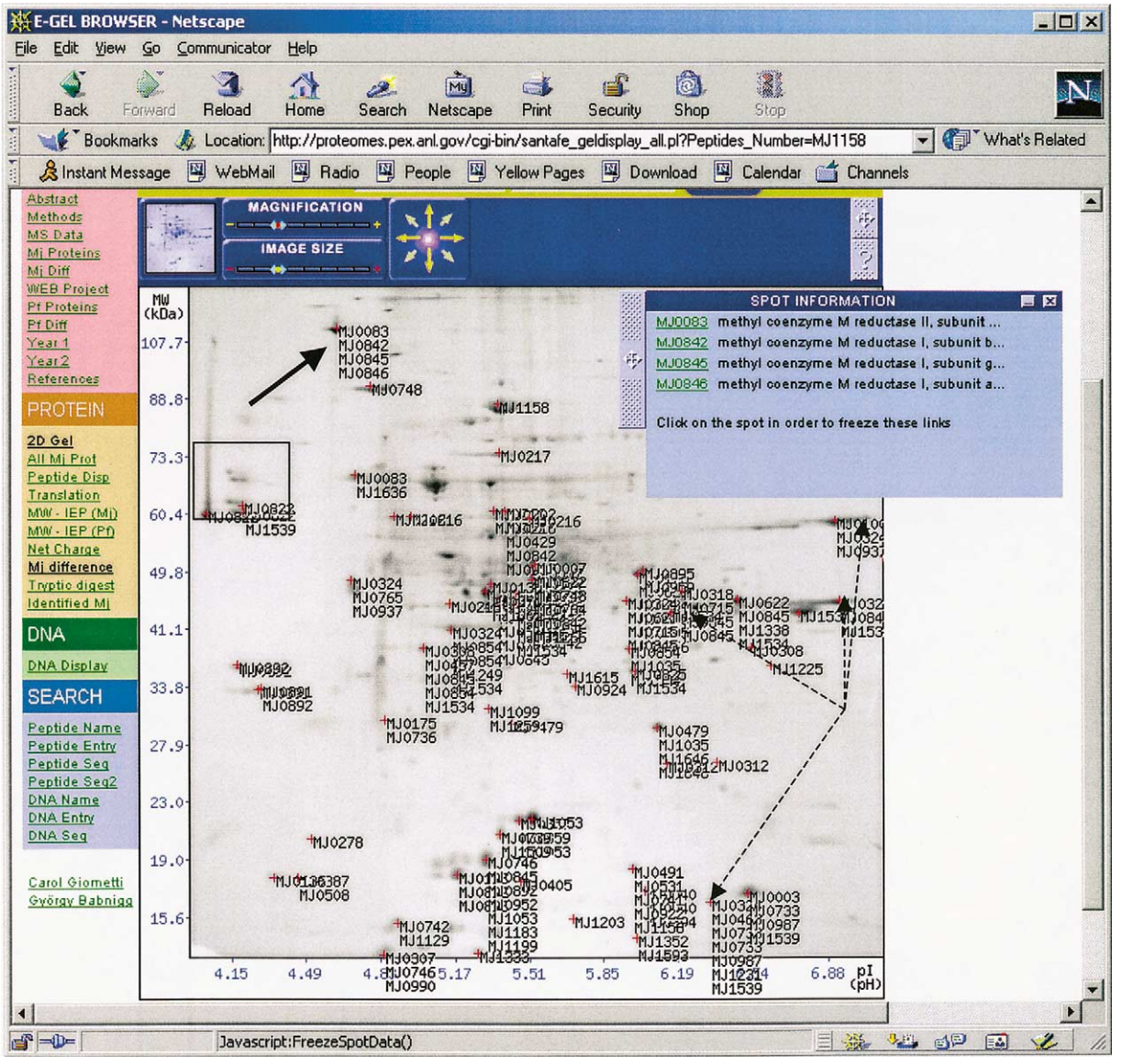


Fig. 2. Identified *Methanococcus jannaschii* proteins. The open reading frames in the *M. jannaschii* gene sequence that predict proteins with tryptic peptide masses most similar to those obtained from proteins extracted from 2DE gels are indicated. The correlation of observed peptide masses and sequences with specific ORFs was based on the detection of peptides unique to a single ORF. The occurrence of multiple ORF products in a single protein spot cut from the 2DE gels complicates the annotation, so a world wide web interface has been developed to simplify browsing the identifications (see http://ProteomeWeb.anl.gov). A complete list of the proteins identified is shown in Table 1. The rectangle indicates the location of S-layer protein; solid arrow indicates location of a protein spot containing the products of multiple ORFs (described in the text); the dashed arrows indicate the multiple locations of products predicted by the same ORF (described in the text).

cofactor biosynthesis, central intermediary metabolism, DNA metabolism, and fatty acid and phospholipid metabolism, as well as a category annotated as ''unknown function''. Since a majority of the proteins identified were those expressed in the highest relative abundance, these results suggest that

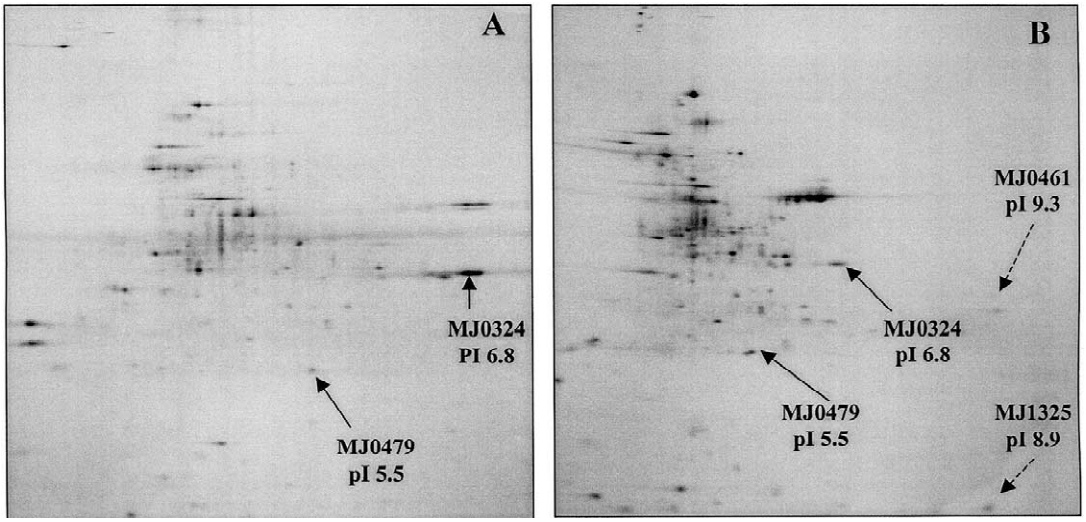


Fig. 3. Proteins in *M. jannaschii* whole cell lysates separated by different first dimension methods. *M. jannaschii* whole cell lysate proteins were separated (A) by isoelectric focusing (IEF) using the same carrier ampholytes described for Fig. 1 or (B) by nonequilibrium pH gradient electrophoresis (NEPHGE) using pH 3–10 carrier ampholytes, followed by SDS–PAGE in the second dimension and silver stained for protein detection. In contrast to IEF, a stable pH gradient is not established using the NEPHGE method, so a pH scale cannot be accurately provided. The isoelectric points of some identified proteins are provided to demonstrate the pH range generated for the NEPHGE pattern shown. Solid arrows indicate the location of proteins in common in both panels A and B, while dashed arrows indicate proteins with very alkaline isoelectric points that cannot be detected with isoelectric focusing using carrier ampholytes. MJ numbers correspond to ORF designations in Table 1.

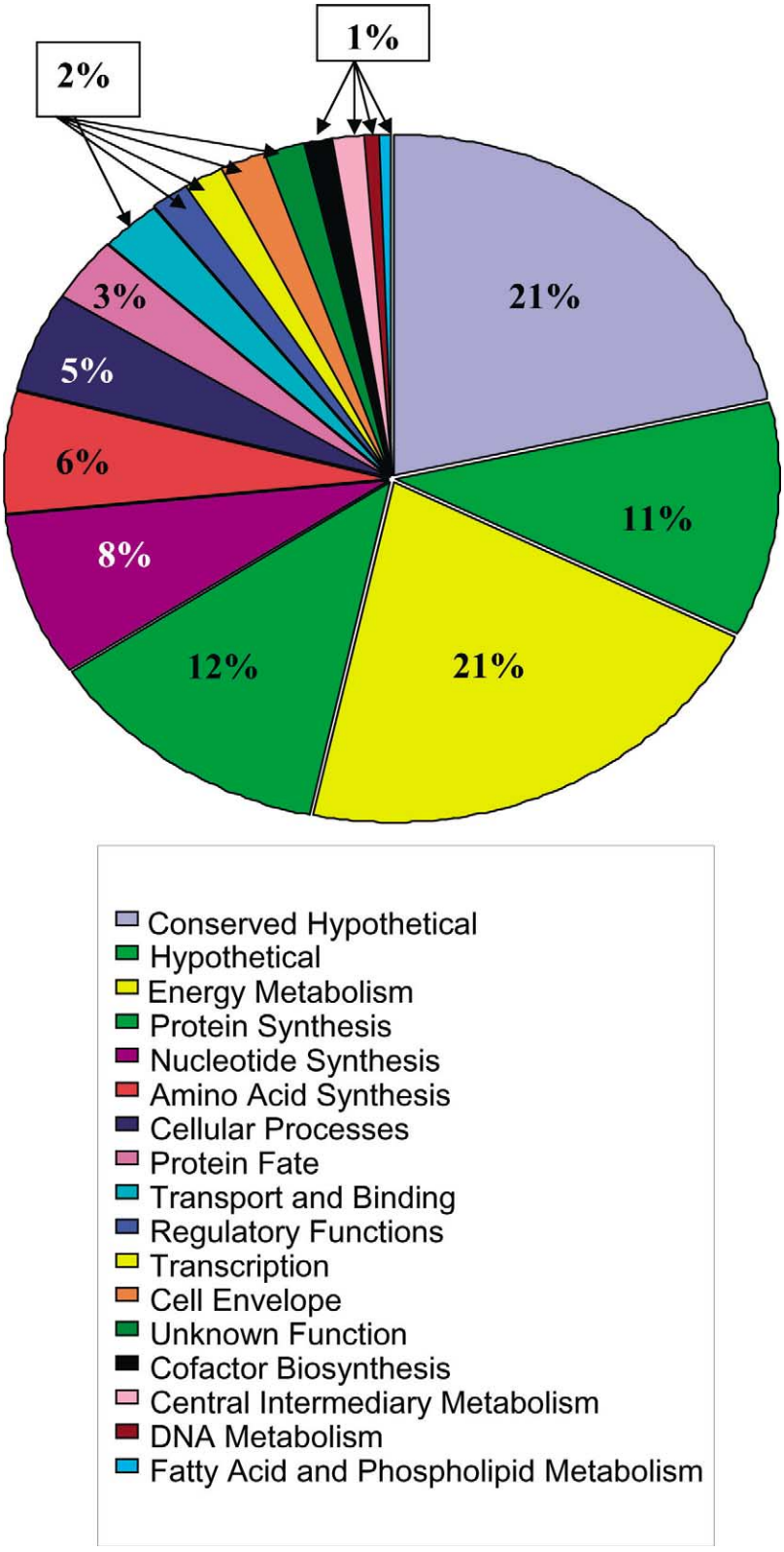
M. jannaschii, when grown under optimal conditions, predominantly produces proteins of as yet unknown function and proteins required for energy generation.

4. Discussion

In contrast to the analysis of a genome sequence, which is a static entity, proteome analysis is extremely complex by virtue of the dynamic nature of protein expression. In addition to the lack of one-to-one correspondence between an open reading frame of genome sequence and the expression of the corresponding protein due to post-translational modifications, proteins are expressed in varying relative abundance in response to cell growth conditions. The high resolution capability of 2DE together with the availability of a range of protein-specific stains that can be used to obtain quantitative measurements of protein abundance provide a seemingly compatible analytical approach to the global analysis of proteomes. Use of 2DE for proteome analysis, however, must be done judiciously with an awareness of its strengths and its limitations. The heterogeneous constitution of proteins necessitates the use of multiple 2DE methods even for an organism as apparently simple as *M. jannaschii*.

As shown in this study, the use of a single approach to protein detection in 2DE could doom a proteome study to missing even abundant proteins. The S-layer protein (Mj0822), for example, is a major component of the *M. jannaschii* protein repertoire, but silver stain failed to reveal this protein. To

Fig. 4. Functional distribution of *M. jannaschii* proteins identified from 2DE patterns. This pie chart shows the relative distribution of the *M. jannaschii* proteins identified (Table 1) within the annotated functional groups. The identification process thus far has focused on the most abundant proteins detected by Coomassie Blue R250 (see Fig. 2). In addition to the designations of ''conserved hypothetical'' and ''hypothetical'' proteins which refer to ORF sequences indicative of protein coding regions but with no sequence homology with proteins having known functions, the designation ''unknown function'' refers to ORFs encoding known proteins (i.e. proteins that are known to be expressed) with as yet characterized function.


1%
2%
21%
11%
21%
12%
8%
6%
5%
3%
Conserved Hypothetical
Hypothetical
Energy Metabolism
Protein Synthesis
Nucleotide Synthesis
Amino Acid Synthesis
Cellular Processes
Protein Fate
Transport and Binding
Regulatory Functions
Transcription
Cell Envelope
Unknown Function
Cofactor Biosynthesis
Central Intermediary Metabolism
DNA Metabolism
Fatty Acid and Phospholipid Metabolism

optimize the number of expressed proteins detected, therefore, a comparison of the proteins detected by different staining methods is advisable as part of any proteome analysis using electrophoretic methods. In addition, analysis of subcellular fractions of the cell provides the opportunity to detect proteins underrepresented in whole cell lysates. Analysis of cytosol and crude membrane fractions from *M. jannaschii*, for example, are in progress in order to optimize the chance of detecting minor proteins and to obtain sufficient amounts of the more minor proteins seen in the whole lysate patterns for peptide mass analysis.

The pH gradient limitations of 2DE must also be taken into account in order to optimize the number of expressed proteins detected. Cathodic drift causes poor resolution of the most alkaline proteins in a protein sample when carrier ampholytes are used for isoelectric focusing. The availability of immobilized pH gradient electrophoresis (IPG) strips provides an alternative to the use of carrier ampholytes for the IEF separation that improves the resolution of the proteins with alkaline isoelectric points, although sometimes compromising the resolution of the neutral region. However, the electrophoresis of some proteins out of the IPG strips into the second dimension gels is inefficient (data not shown), so quantitative reproducibility is decreased. In contrast, the use of the NEPHGE method provides good quantitative reproducibility for the alkaline proteins, although this approach does necessitate the analysis of two protein patterns (one using IEF in the first dimension and one using NEPHGE in the first dimension) for each sample rather than one.

M. jannaschii is an autotrophic microbe that requires no organic molecules for survival. The simplicity of the *M. jannaschii* genome suggested a proteome that was likely to present essentially a one-to-one correspondence between the predicted proteins and the expressed proteins. With the methods used thus far to analyse the proteome of *M. jannaschii* by 2DE, patterns containing over 1200 protein components have been generated and identifications based on tryptic peptide masses have been obtained for ~170 of those proteins. The complexity of the protein identifications, that is the observation of multiple proteins in an apparently single protein spot and the identification of the same protein in multiple positions in the 2DE pattern, indicates the magnitude of the process of proteome analysis. Instead of a one-to-one correspondence, these results show that a single ORF can give rise to multiple forms of proteins that vary not only in isoelectric point, but also in molecular mass. The actual position of a protein in the 2DE pattern relative to its predicted position suggests the addition of chemical groups such as phosphates or sugars, in the case of multiple forms with different isoelectric points. A molecular mass that is higher than predicted from the ORF indicates an association with other proteins or aggregates of the same protein, while a lower than predicted molecular mass indicates proteolytic cleavage. While some of the molecular mass heterogeneity may result from sample preparation or anomalous behavior in the electrophoresis system, a majority of the observed differences are thought to be a result of cellular processes. Thus, the observed complexity of protein associations and protein structural changes revealed by 2DE provide the starting point for study of protein interactions and protein processing in the context of *M. jannaschii* cellular regulation.

The functional categories of the expressed *M. jannaschii* proteins identified thus far provide an interesting commentary on this little understood microbe. The results summarized in this paper were from an analysis that focused primarily on the most abundant proteins observed in the 2DE patterns of *M. jannaschii* grown under optimal laboratory conditions. Since ~34% of the ORFs in the *M. jannaschii* genome are annotated as hypothetical or conserved hypothetical proteins (i.e. DNA sequences predicted to code for proteins with as yet undetermined functions) or proteins with unknown function (i.e. DNA sequences coding for proteins known to be expressed but for which no function has yet been characterized), the finding that 32% of the proteins identified thus far correspond to hypothetical or conserved hypothetical ORFs is not surprising. The observation that so many proteins with undetermined function are among the most abundant proteins synthesized by this microbe when grown under optimal conditions is, however, significant. This would suggest that these proteins serve some as yet unidentified ''housekeeping'' functions unique to the hyperthermophilic way of life. The next most prevalent group of abundant proteins identified was the enzymes associated with methanogenesis, appar-

ently the sole source of energy generation for this microbe. The abundance of these proteins has been observed to change only when cells are grown with depleted hydrogen [14,15], indicating how important the maintenance of a sufficient abundance of these key proteins is to cell survival.

As this report shows, analysis of the *M. jannaschii* proteome by using 2DE is revealing more than simply the identification of which ORFs are actually expressed under particular growth conditions. Further studies are needed to characterize the post-translational modifications indicated by the observation of the same protein in multiple migration positions and to verify the in vivo association of the proteins found to co-migrate on the 2DE gels. A powerful tool for global analysis of the proteins expressed by a biological system, these results demonstrate that 2DE is often only the starting point in proteome analysis. More than a list of parts, a proteome is the sum of how those parts interact and change in response to stimuli. For even a simple genome like that of *M. jannaschii*, the corresponding proteome is dynamic and complex.

Acknowledgements

This work was supported by the US Department of Energy, Office of Biological and Environmental Research, through the Microbial Genome Program under contract no. W-31-109-ENG-38. The submitted manuscript has been created by the University of Chicago as Operator of Argonne National Laboratory (''Argonne'') under Contract No. W-31-109-ENG-38 with the US Department of Energy. The US Government retains for itself, and others acting on its behalf, a paid-up, nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government.

References

[1] C.J. Bult, O. White, G.J. Olsen, L. Zhou, R.D. Fleischmann, G.G. Sutton, J.A. Blake, L.M. Fitzgerald, R.A. Clayton, J.D. Gocayne, A.R. Kerlavage, B.A. Dougherty, J.-F. Tomb, M.D. Adams, C.I. Reich, R. Overbeek, E.F. Kirkness, K.G. Weinstock, J.M. Merrick, A. Glodek, J.L. Scott, N.S.M. Geoghagen, J.F. Wiedman, J.L. Fuhrmann, D. Nguyen, T.R. Utterback, J.M. Kelley, J.D. Peterson, P.W. Sadow, M.C. Hanna, M.D. Cotton, K.M. Roberts, M.A. Hurst, B.P. Kaine, M. Borodovsky, H.-P. Klenk, C.M. Fraser, H.O. Smith, C.R. Woese, J.C. Venter, Science 273 (1996) 1058.
[2] L.S. Ramagli, L.V. Rodriguez, Electrophoresis 6 (1985) 559.
[3] N.G. Anderson, N.L. Anderson, Anal. Biochem. 85 (1978) 331.
[4] P.Z. O'Farrell, H.M. Goodman, P.H. O'Farrell, Cell 12 (1977) 1133.
[5] P.H. O'Farrell, J. Biol. Chem. 250 (1975) 4007.
[6] N.L. Anderson, N.G. Anderson, Anal. Biochem. 85 (1978) 341.
[7] N.L. Anderson, S.L. Nance, S.L. Tollaksen, F.A. Giere, N.G. Anderson, Electrophoresis 6 (1985) 592.
[8] C.S. Giometti, M.A. Gemmell, S.L. Tollaksen, J. Taylor, Electrophoresis 12 (1991) 536.
[9] N.L. Anderson, J. Taylor, A.E. Scandora, B.P. Coulter, N.G. Anderson, Clin. Chem. 27 (1982) 1807.
[10] J.K. Eng, A.L. McCormack, J.R. Yates III, J. Am. Soc. Mass Spectrom. 5 (1994) 976.
[11] C.R. Merril, Methods Enzymol. 182 (1990) 447.
[12] H.W. Luo, H. Zhang, T. Suzuki, S. Hattori, Y. Kamagata, Appl. Environ. Microbiol. 68 (2002) 1173.
[13] C. Schaffer, P. Messner, Biochimie 83 (2001) 591.
[14] C.S. Giometti, C.I. Reich, S.L. Tollaksen, G. Babnigg, H. Lim, J.R. Yates III, G.J. Olsen, Eur. J. Mass Spectrom. 7 (2001) 195.
[15] B. Mukhopadhyay, E.F. Johnson, R.S. Wolfe, Proc. Natl. Acad. Sci. USA 97 (2000) 11522.