6 DISCUSSION

6.1 STRATEGY FOR PROTEOMICS STUDY OF MODERATELY HALOPHILIC BACTERIA

Genomic and proteomic studies of microorganisms provide us huge amount of data, and bioinformatic studies related to these data lead us to the knowlege about these microorganisms. There are some model microorganisms which have been studied over years giving very important results and knowledge about their genetic material and functional properties. It is in common between these model organisms that their genome is completely known which provides a clear base for further functional studies and mainly for proteomic studies.

Although various techniques are available to study microorganisms which have known genomes, these methods need to be overviewed and modified when only limited genome data are known for a microorganism to be studied. Besides, there are several microorganisms which are isolated from a variety of natural sources. These native, isolated microorganisms are important in many different ways such as providing good sources for industrial enzymes, for studying biological diversity, their use for important pharmaceutical applications, or as a model to understand how these organisms survive in non-conventional environments. For such unknown organisms no genomic data are known, and studies must be carefully planned to gain trustable informations at the end.

Functional studies of bacteria using proteomic tools has become an important way to collect data about the bacterial world in recent years. At the Marmara University in Istanbul 11 new isolates from Çamaltı Salten Area were studied in order to classify these strains, and preliminary data about these microorganisms were obtained [108]. By the information gained from this work, it was decided to select the study with Isolate No 6 and to compare this strain to *Halomonas salina* since they showed 98 % similarty on the bases of 16S rRNA and showed high similarity on their biochemical characteristics and salt dependence. However, DNA-DNA hybridization gave no

similarities among them and therefore, Isolate No 6 was indicated as a new strain from the *Halomonas* family.

From then on, in this thesis it was planned to achive some information in two different aspects. The first was to find out a satisfying working programme for such microorganisms where limited or no genome data are available, and the second was to gain some information about the moderate halophiles by proteomics studies. This approach allows a rather fast and reliable comparison of their proteomes for classification and might allow to use in part genetic data available for homologous microorganisms.

After gaining some information about the new isolate, the culturing conditions have been optimised. The sample preparation steps were determined according to which type of sample preparation would be best to study. By this step the decision was made whether the whole cell or a particular compartment of the cells were of interest, and whether an additional fractionation was needed and which lysis buffer and which method of cell distruption would be applied.

The high resolution 2-DE separation (30 x 23 x 0.1 cm) of the complex proteome of these microorganisms and staining with silver, made it possible to visualize whole cell proteins of these microorganisms at given conditions. As an important step, it had to be decided which pI scale for the first dimension should be applied, the gel size optimization and gel staining methods had to be optimized. For the pI scale decision, the biochemical characteristics of the moderatey halophilic bacteria were taken into consideration. It was known that their proteins show slightly acidic character and in order to see the maximum protein expression profile on the gel, the pI scale was optimized in the range of pH 3 to pH 7,5 after several trials. As it can be seen from the gel images in the previous section, the 2-DE images of the moderately halophilic bacteria were sucessfully resolved with the NEPGHE technique. In our study, one of the microorganisms studied was *Halomonas salina* (DSMZ 5928) which is a Gram negative aerobic moderately halophilic bacteria. In EMBL /GeneBank databases, 20 entries have been found about *Halomonas salina*. The second organism studied was the new isolate which is indicated as Isolate No 6. In the case of these two organisms, there was no entire genome knowledge and therefore no database available.

Selected spots of the high resolution 2 DE-gels of analysis combined with bioinformatic tools was the key element for identification of the proteins. In order to obtain reliable results for our microorganisms, different MS techniques were investigated. Firstly, MALDI-TOF MS was used, secondly ESI MS was applied.

Usually, peptide mass fingerprinting (PMF) can be applied easily where tryptic peptide masses are acquired and compared with theoretical peptide masses calculated for all the proteins in a given protein sequence database. This approach was thought to give us information about the homologous proteins from various similar microorganisms. But unfortunately, from all 20 selected spots investigated (10 from each halophile), only 5 meaningful results were acquired. From 20 spots investigated, 9 protein spots gave very weak spectral signals which did not result in an unambiguous protein identification. 11 protein spots gave good spectral results these organisms were manually cut and tryptically digested and prepared for further analysis. In order to clean up the digested samples, the usual ZipTip procedure was applied. It was seen that the ZipTip cleaning step enriched the quality of the final MS and MS/MS spectra. Mass but only 5 of them gave acceptable scores for matched peptides. These results were from spots 2816, 2822 and 2839 which were proteins of *Halomonas salina*, and spots A4 and 629 of Isolate No 6 (given in table 5.3 and 5.4). It was remarkable that spot A4 which was a common spot of both halophiles, gave the same homologous protein result with different MS methods applied to two different microorganisms (given in table 5.3 and 5.4) Spot A4 which was studied from *Halomonas salina* was identified using NanoLC ESI QTOF MS whereas the same spot of Isolate No 6 was studied using MALDI TOF. Both gave the homologuos protein Acetyl-CoA acetyltransferase.

For bacteria for which the fully sequenced genomes are available the annotated PMF spectra can be the simplest and most reliable method for protein identification. But in our case, with this method only 35% of the selected spots could be identified regardless whether MALDI-MS or NanoLC-MS techniques were applied.

It was clearly seen in this thesis that for the microorganisms whose genomes are not fully sequenced and annotated, there was a need for additional information in order to arrive at an unambiguous protein identity which is an essential step for the comparison of the proteomes of different microorganisms. In this case, further fragmentation of selected peptides of a selected protein by MS/MS thought to give an increase in number of identified proteins.

Finally, NanoLC ESI QTOF MS/MS and N-terminal sequencing methods were used to obtain better results for the protein identification. With the MS/MS method, ionized peptides were further fragmented yielding partial or complete amino acid sequences. Matched peptides were listed and ion scores were obtained individually. The results from NanoLC ESI MS/MS were evaluated automatically by ‘matrix science’ and the best hits were listed. From 30 selected protein spots (15 protein

spots for each microorganism) 5 spots produced very weak spectra, and from the other 25 spots, 20 yielded very high scored identifications. 8 of these 18 spots were proteins from Isolate No 6 which are A2, A3, A5, A6, 603, 623, 624, 626 and the other 12 spots were from *Halomonas salina* which are spots A2 (Group 1,2), A3, A4, A5, A6, 285, 292, 2819, 2821 (Group 1,2), and 2823. The results are summerized in table 5.3 and 5.4. The accomplishment of this identification method used in our study was 80%.

N-terminal sequencing by the automated Edman sequencing method was only applied to samples from Isolate No 6. N-terminal sequencing was run from PVDF blotted samples of proteins separated by the 2-DE unstained gels, and 9-12 amino acids each from the N-terminal sequence of the proteins have been sequenced. Totally 8 spots were studied and 7 of them gave very reliable sequencing results and 6 sequences showed homology to various proteins with the 'blast' search (given in tables 5.3, 5.4). These spots are A1, 62, 617, 60011, 60013, and 60021. This identification method gave 85 % success which was the best of all. In table 6.1, the success of the identification methods has been summarized.

Table 6.1 Summarized overview for a number of studied protein spots

	Total # spots	**Weak spectra**	**# Identified spots**	**% success**
MALDI TOF	20	9	5	35
NanoLC ESI QTOF MS/MS	30	5	20	80
N-Terminal Sequencing	8	1	6	85

The key points of matching homologous proteins using MS/MS and N-terminal sequencing data were the blast searches performed and the comparison of both the theoretical and practical MW and pI values obtained. When the significant hit/hits of a protein spot was acquired from NCBInr automatically, firstly the matched peptides were examined. The individual ion scores of each matched peptide are controlled, and the ones above the given minimum limit have been accepted. These individual peptides which were further fragmented and their amino acid sequences supplied, have been blast searched on the NCBI and ExPASy sites. Conserved domains of the matched proteins have been detected between different species. The best hit homologous protein was also checked by controlling the pI and Mw values as obtained from the 2-DE gels in comparison to the theoretical values. This control, decreased the false matches, in considereable amount. Further, from N-terminal sequencing, 9 to 12 amino acids have been determined and homology searches have

been performed on the ExPASy site. Conserved sequences enabled homology matches between species.

Another important point in our study, was to restrict taxonomy at the "mascot search" level. Both archaeal and bacterial data have been used for database searches. This helped us to minimize uncertainties caused by the large number of sequences in the public databases.

Out of 58, 14 protein spots gave low MS spectra quality probably because of their low expression levels causing faint protein spots on 2-DE gels. The quality of spectra could have been affected from the nature of proteins as well. As shown in 2-DE gel images in the previous section, most proteins have a slightly acidic character and without many basic residues trypsin can not yield good fragments to give the appropriate ions to produce many intensive peaks. Also if we take contaminations into account, we might have probably lowered the quality of spectra especially when the protein amount was very low.

Out of 58 proteins investigated, 13 protein spots did not give any reliable identification and comparison to a homologous protein from other microorganisms although good spectra have been obtained. As expected, when the proteins or their homologous are not present in the databases, identification can not be supplied even though intensive ionic peaks have been achieved. Further, in this study, for some of the protein spots, more than one homologous protein have been matched with sufficient scores. In such cases, these results were expected to be false positives and were not accepted.

According to the results and comments, a flow chart of this study is constructed showing the experimental steps followed and the interaction of the techniques employed. This is shown in figure 6.1 below. This chart represents the organizations and control steps of this work. Also, additional fractionation steps can be applied in order to simplify the sample to be analyzed. On each step represented on the flow chart, several questions have been observed and answered.

As mentioned above it was important to have preliminary data about the interested microorganisms. It could be proven by the proteomics work presented here that the new isolate No 6 can be classified to the moderate halophiles. Furthermore, a satisfying working programme for any classification of microorganisms of limited or no genomic data could be achieved.

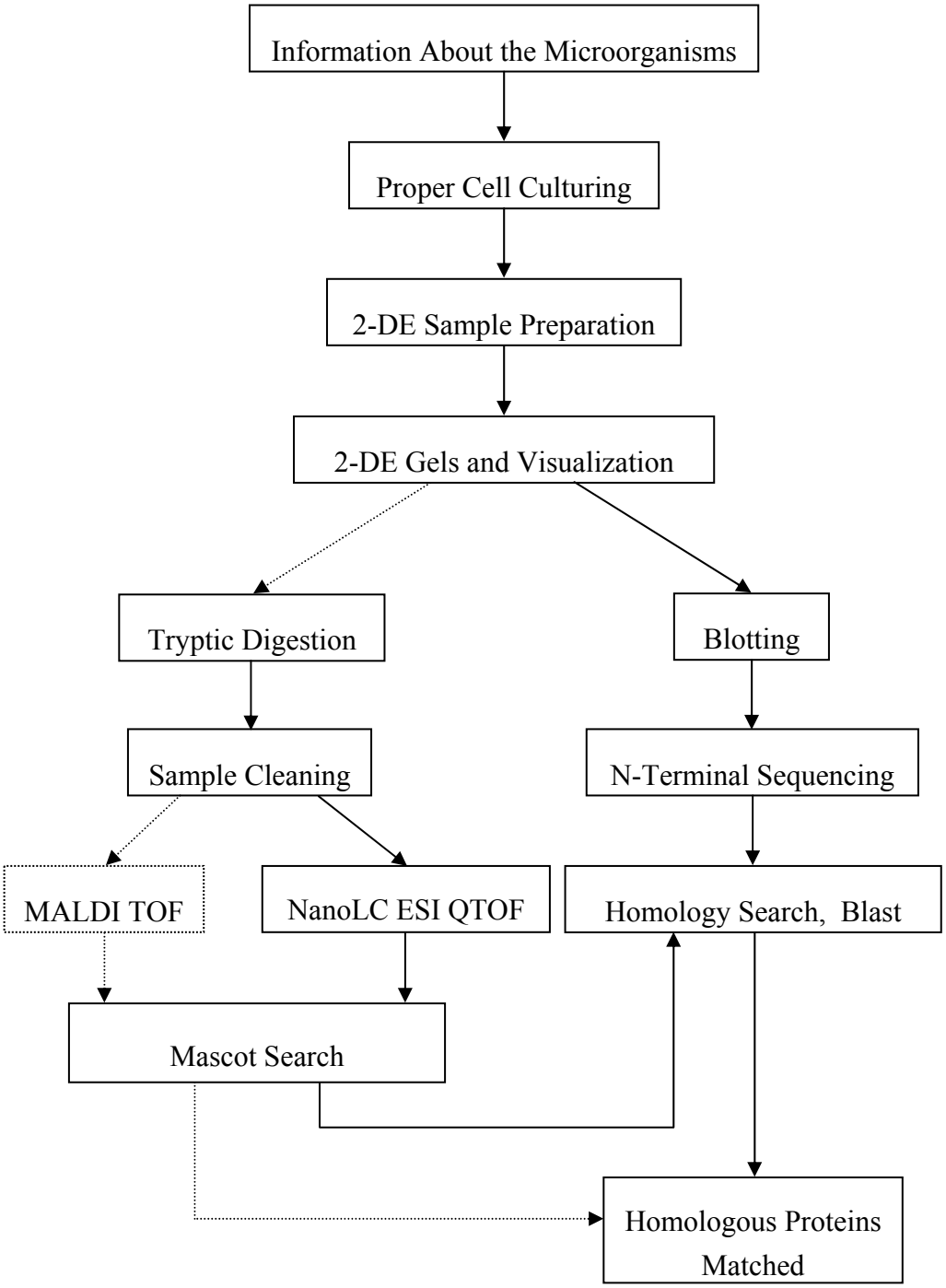

Figure 6.1 The schematic representation of the important steps followed. On this scheme, a dashed step represents the tried but unpreferred method of protein identification according to the results obtained.

The gel size was chosen as the large gel size in order to have a good resolution. Silver staining was chosen for analytical gels since this staining method was very sensitive. In our experience, the silver staining method used in this work gave very clear background with bright protein spots on gels. For MS analysis, the gels were stained with Colloidal Coomassie G-250 since destaining procedures were faster and easier. It was noted that when the spots were faint and / or spot pieces were contaminated, the spectra gained from mass sperctrometry gave low densities with high noise signals. In order to prevent this only the very dense parts (the central parts) of the spots were cut out of the gels and studied.

On the digestion step, the enzyme to be used must be carefully selected. The most commonly used enzyme is trypsin and in this study this enzyme was also applied because of moderate acidic character of the proteins. Also in this step, it should be decided if an extra sample cleaning would be needed or not. In our case, we have performed ZipTip cleaning to the digested samples, since salt and other contaminations lowered our previous trial of MS analysis.

On the step of Mascot Search, it should be kept in mind that there are two sides of restricting the taxonomy in the search. Firstly, it is important to see whether a keratine contamination is the case or not. Clearly, useful information must be extracted from the complete dataset but would be huge if contaminations would occur. So the solution found was first to perform the searches without limitations, and if the results came out without a keratine contamination, then the search was run again with restricted taxonomy. Matched homologous proteins were searched with several database sets and information about homologous, similar strains or conserved sequences between strains have been supplied. This data searches provide us more external information about halophiles, their positions in the diverse environment using proteomic data.

6.2 PROTEIN IDENTIFICATION OF HALOMONAS SPECIES

Halophiles were grown in 4 different physicochemical conditions altering temperature and salt content, and for each conditions 2-DE profiles were prepared.

Totally 15 protein spots have been identified with high scoring from *Halomonas salina* by MALDI TOF MS, NanoLC ESI QTOF MS/MS and N-terminal sequencing through cross-species homologous protein searching in public databases. These proteins were mainly involved in energy metabolism, amino acid metabolism, nucleotide metabolism and transportation, DNA metabolism and translation.

Identified proteins of *H. salina* show high sequence similarities to the sequences of homologous proteins from *C. salexigens*, *P. bermudensis*, *M. acetivorans*, *E. carotovora*, *S. solfataricus*, *C. violaceum* and *Halomonas sp* as given in table 6.3.

Two proteins, nucleoside diphosphate kinase and NAD+ dependent aldehyde dehydrogenase were identified from *Halomonas* genus since the genetic data were reported and annotated in databases [120, 121].

Aconitate hydratase was studied from both Group1 and Group 2 gels of *Halomonas salina* and it was down regulated in group 2 conditions compared to group 1, as seen in figures 5.1 and 5.2. By increasing the salt concentration in the growth media from 5 % to 20 % which is an obvious stress for the microorganism, the expression of aconitate hydratase is suppressed leading to a decrease in energy production. This protein was also identified from Isolate No 6.

NAD+ dependent aldehyde dehydrogenase was identified both from *H. salina* and Isolate No 6 as spot No A3. As seen in figure 5.12 and 5.15, for Isolate No 6 and *H.salina* respectively, probable NAD+ dependent aldehyde dehydrogenase is stronger expressed in low salt medium (group1 and group 3) which points to that low salt concentrations are optimal for NAD+ dependent aldehyde dehydrogenase synthesis, supporting the optimal function conditions being low in salt content [122].

From Isolate No 6 gels, two identified proteins, namely protein 60011 as L-ectoine synthase (fragment) and 626 as aspartate semialdehyde deyhdrogenase take role in osmoprotectant synthesis. This result gives us an idea how Isolate No 6 adapts hypersaline environment and resists any salt stress. This microorganism like many other halophiles accumulate ectoine instead of accumulating intracellular salt.

Spots 60013, superoxide dismutase, A1 major outer membrane protein from Isolate No 6, spot A6 periplasmic phosphate binding protein both from Isolate No 6 and *H.salina*, and spot 2821, a hypothetical protein from *H.salina,* are proteins involved in solute diffusion, and ion transportation. Spot 2821 was studied from group 1 and group 2 gels of *H. salina* and found to be down regulated in group 2 where the salt concentration was higher.

Spot A5, malate dehydrogenase, was studied from both, Isolate No 6 and *H. salina,* and probably regulation can be seen in figure 5.12 and 5.15. As seen in the figures, increasing salt concentration and increases in temperature decreases protein

expression. This result also supports that malate dehydrogenase is inhibited by high salt content [123].

The tables 6.2 and 6.3 are given below to summarize the identified proteins and indicate the important functions of these proteins.

Isolate No 6 and *H. salina* are from the bacterial family Halomonadaceae like *C. salexigens*, but differ at the genus level being *Halomonas* and *Chromohalobacter*. Although the genome of *C. salexigens* has been fully studied, the genome knowledge of *H. salina* and new isolate Isolate No 6 are too limited or even not available at all, respectively. It can be seen that most of the identified proteins of *H. salina* and Isolate No 6 show high sequence similarities to the sequences of homologous proteins from *Chromohalobacter salexigens*. It can be stated that there is a sequence conservation of the identified proteins between these microorganisms.

Table 6.2 An overview of Isolate No 6 proteins identified using MALDI TOF MS, ESI-QTOF MS/MS and N-terminal analysis

Spot	Origin Organism	Homologous Proteins	Family	Rxn(s)	Role in
A1	B.S.*	Major outer membrane protein precursor	Membrane protein		Structural rigidity Porin formation Solute diffusion
A2	*C. salexigens* DSM 3043	Aconitate hydratase	Lyases (EC. 4.2.1.3)	citrate ↔isocitrate citrate ↔cis-Aconitate + H_2O isocitrate↔cis-Aconitate + H_2O	TCA cycle
A3	*C. salexigens* DSM 3043	Succinic semialdehyde dehydrogenase	Oxidoreductase (EC. 1.2.1.16)	succinate semialdehyde + NAD(P)+ + H2O ↔ succinate + NAD(P)H + 2 H+	Glutamate metabolism Tyrosine metabolism Butanoate metabolism
A4	*Cellulophage sp.MED134*	Acetyl-CoA-acetyltransferase	Transferase (EC. 2.3.1.9)	2 acetyl-CoA ↔ CoA + acetoacetyl-CoA Acetyl-CoA + Butanoyl-CoA ↔ CoA + 3-Oxohexanoyl-CoA	Fatty acid metabolism Keton bodies metabolism Amino acid degradation Tryptophan metabolism Pyruvate metabolism Benzoate degrad. Propanoate metabolism Butanoate metabolism Two-component system
A5	*A. ehrlichhei MLHE-1*	Malate dehydrogenase	Oxidoreductase (EC. 1.1.1.37)	(S)-malate + NADP+ ↔ pyruvate + CO_2 + NADPH (S)-Malate + NAD+ ↔ pyruvate + CO_2 + NADH	Carbohydrate metabolism Energy metabolism

Spot	Origin Organism	Homologous Proteins	Family	Rxn(s)	Role in
A6	*C. salexigens* DSM 3043	Periplasmic phosphate binding protein	Periplasmic component	Inorganic ion transport	Transportation
62	B.S.*	Alkyl hydroperoxide reductase	Oxidoreductase (EC 1.11.1.15)	**2** R'-SH + ROOH = R'-S-S-R' + H_2O + ROH	Genetic information processing Folding, sorting, degradation Reducing hyperoxides
603	*C. salexigens* DSM 3043	Acetate--CoA ligase	Ligase (EC 6.2.1.1)	ATP + acetate + CoA ↔ AMP + diphosphate + acetyl-CoA Acetyl adenylate + CoA ↔ AMP + Acetyl-CoA ATP + Acetate ↔ Pyrophosphate + Acetyl adenylate Propinol adenylate + CoA ↔ AMP + Propanoyl-CoA ATP + Propanoate ↔ Pyrophosphate + Propinol adenylate	Glycolysis Gluconeogenesis Pyruvate metabolism Propanoate metabolism Reductive carboxylate cycle (CO2 fixation)
617	B.S.*	Universal Stress Protein A	USP		Defensive against stress
623	*C. salexigens* DSM 3043	Dihydrolipoamide dehydrogenase	Oxidoreductase (EC 1.8.1.4)	Dihydrolipoylprotein + NAD+ ↔ Lipoylprotein + NADH + H+ Enzyme N6-(dihydrolipoyl)lysine + NAD+ ↔ Enzyme N6-(lipoyl)lysine+ NADH + H+	Glycolysis Gluconeogenesis TCA cycle Pyruvate metabolism Amino acid degradation Amino acid metabolism

Spot	Origin Organism	Homologous Proteins	Family	Rxn(s)	Role in
624	*C. salexigens* DSM 3043	Translation elongation factor Tu	G- protein		Protein biosynthesis Translocation of synthesized protein
626	*C. salexigens* DSM 3043	Aspartate semialdehyde dehydrogenase	Oxidoreductase (1.2.1.11)	L-Aspartate 4-semialdehyde + ortophosphate + NADP+ ↔ 4-Phospho-L-aspartate + NADPH + H+	Amino acid metabolism Lysine biosynthesis Ectoine synthesis
629	*Marinobacter sp.* ELB17	Type I restriction - modification system M-subunit		Protection against invasion of foreign DNA by endonucleolytic cleavage of DNA	Bacterial cell protection
60011	B.S.*	L-ectoine synthase (fragment)	Lyases (EC 4.2.1.108)	N4-acetyl-L-2,4-diaminobutanoate ↔ L-ectoine + H_2O	Amino acid metabolism Osmoprotectant synthesis
60013	B.S.*	Superoxide dismutase	Oxidoreductase (EC 1.15.1.1)	2 O_2.- + 2 H+ ↔ H_2O_2 + Oxygen	Inorganic ion transport and metabolism Cell processing Cell signaling
60021	B.S.*	Succinyl-CoA ligase [ADP-forming] subunit alpha	Ligase (6.2.1.5)	ATP + Succinate + CoA ↔ ADP + Orthophosphate + Succinyl-CoA ATP + Itaconate + CoA ↔ ADP + Orthophosphate + Itaconyl-CoA	TCA cycle Propanoate metabolism Reductive carboxylate cycle (CO_2 fixation)

* states that the identification was performed by blast search.

Table 6.3 An overview of *H. salina* proteins identified using MALDI TOF MS, ESI-QTOF MS/MS and N-terminal analysis

Spot	Origin Organism	Homologous Proteins	Family	Rxn(s)	Role in
A2	*C. salexigens* DSM 3043	Aconitate hydratase	Lyases (EC. 4.2.1.3)	citrate ↔isocitrate citrate ↔cis-Aconitate + H_2O isocitrate↔cis-Aconitate + H_2O	TCA cycle Stabilize mtDNA
A3	*Halomonas salina*	NAD+ dependent Aldehyde dehydrogenase	Oxidoreductase (EC. 1.2.1.16)	succinate semialdehyde + NAD(P)+ + H2O ↔ succinate + NAD(P)H + 2 H+	Glutamate metabolism Tyrosine metabolism Butanoate metabolism
A4	*C. salexigens* DSM 3043	Acetyl-CoA-acetyltransferase	Transferase (EC. 2.3.1.9)	2 acetyl-CoA ↔ CoA + acetoacetyl-CoA Acetyl-CoA + Butanoyl-CoA ↔ CoA + 3-Oxohexanoyl-CoA	Fatty acid metabolism Keton bodies metabolism Amino acid degradation Tyrptophan metabolism Pyruvate metabolism Benzoate degrad. Propanoate metabolism Butanoate metabolism Two-component system
A5	*P. Bermudensis*	Malate dehydrogenase	Oxidoreductase (EC. 1.1.1.37)	(S)-malate + NADP+ ↔ pyruvate + CO_2 + NADPH (S)-Malate + NAD+ ↔ pyruvate + CO_2 + NADH	Carbohydrate metabolism Energy metabolism
A6	*C. salexigens* DSM 3043	Periplasmic phosphate binding protein	Periplasmic component	Inorganic ion transport	Transportation
292	*C. salexigens* DSM 3043	Extracellular solute-binding protein	Family 5	Actively transport solutes across the cytoplasmic membrane	A.acid transport and metabolism

Spot	Origin Organism	Homologous Proteins	Family	Rxn(s)	Role in
2816	*M. acetivorans*	Putative cobalt-precorrin 6A synthase	Transferase (EC 2.1.1.152)	Precorrin 6X + Acetate+ S-Adenosyl-L-homocysteine ↔ Precorrin 5 + S-Adenosyl-L-methionine + H_2O	Metabolism of Cofactors and Vitamins Porphyrin metabolism
285	*C. salexigens* DSM 3043	Electron transfer flavoprotein, beta subunit		Accepts electron for various mitochondrial hydrogenases	Mainly respiratory chain
2819	*C. salexigens* DSM 3043	Single-stranded DNA binding protein		Interacting selectively with single-stranded DNA	DNA replication, repair DNA recombination DNA transcription
2821	*E. carotovora subsp. atroseptica* SCRI1043	Hypothetical protein ECA3428			Intracellular trafficking secretion and transport
2822	*S. solfataricus* P2	Hypothetical protein SSO0826		Unknown	
2823	*Halomonas sp.* # 593	Nucleoside diphosphate kinase	Transferase (EC 2.7.4.6)	ATP + ADP ↔ ADP + ATP ATP + UDP↔ ADP + UTP ATP + GDP ↔ ADP + GTP ATP + CDP ↔ ADP + CTP ATP + IDP ↔ ADP + ITP ATP + dADP ↔ ADP + dATP ATP + dGDP ↔ ADP + dGTP ATP + dTDP ↔ ADP + dTTP ATP + dCDP ↔ ADP + dCTP ATP + dUDP ↔ ADP + dUTP ATP + dIDP ↔ ADP + dITP	Nucleotide metabolism
2830	*C. violaceum*	30S Ribosomal protein	RP S20P		Translation

* States that the identification was performed by blast search.