British Biotechnology Journal
14(2): 1-12, 2016, Article no.BBJ.25207
ISSN: 2231–2927, NLM ID: 101616695

SCIENCEDOMAIN *international*
www.sciencedomain.org

Database Analysis of Acidic Proteins from Halophilic Species and Their Corresponding Basic Proteins from Non-halophilic Species

Hiroshi Nakashima[1], Keiko Homma[1], Michiko Yamazaki[1,2], Masao Ishizaki[2] and Kazuhiro Mawatari[1*]

[1]Department of Clinical Laboratory Science, Graduate Course of Medical Science and Technology, School of Health Sciences, Kanazawa University, 5-11-80 Kodatsuno, Kanazawa 920-0942, Japan.
[2]Department of Social and Environmental Medicine, School of Medicine, Kanazawa Medical University, 1-1 Daigaku, Uchinada, Ishikawa, 920-0293, Japan.

Authors' contributions

This work was carried out in collaboration between all authors. Authors HN, KH, MY and KM designed the study. Author HN performed the data analysis and wrote the first draft of the manuscript. Authors KH, MY, MI and KM managed the analyses of the study. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/BBJ/2016/25207
Editor(s):
(1) Nikolaos Labrou, Department of Agr. Biotechnology, Agricultural University of Athens, Greece.
(2) P. Mary Anupama, Department of Chemical Engineering and Biotechnology, Anil Neerukonda Institute of Technology and Sciences, India.
Reviewers:
(1) R. Jasmine, Bishop Heber College, Trichy, India.
(2) Volodymyr Chernyshenko, Palladin Institute of Biochemistry NAS of Ukraine, Ukraine.
Complete Peer review History: http://sciencedomain.org/review-history/14871

Original Research Article

Received 23rd February 2016
Accepted 7th April 2016
Published 1st June 2016

ABSTRACT

Aims: To reveal which amino acid residues determine whether a protein is acidic or basic between orthologous pairs, acidic proteins from halophilic species and corresponding basic proteins from non-halophilic species were compared. Similarly acidic versus acidic protein pairs, and basic versus basic protein pairs were also analyzed.
Place and Duration of Study: Department of Clinical Laboratory Science, Graduate Course of Medical Science and Technology, School of Health Sciences, Kanazawa University, Japan.
Methodology: *Halobacterium* sp. NRC-1 was used as halophilic species and Gram-positive bacterium *Bacillus subtilis*, and radiation resistant bacterium *Deinococcus radiodurans* were used as non-halophilic species. The three species were selected because their proteins were closely

**Corresponding author: E-mail: mawatari@staff.kanazawa-u.ac.jp;*

related each other. The amino acid compositions were compared and the amino acid substitutions were counted for the orthologous protein pairs between *Halobacterium* and *B. subtilis.* Similar comparison was done for the proteins between *Halobacterium* and *D. radiodurans.*

Results: The Asp and Glu residues are determinant whether a protein of *Halobacterium* sp. NRC-1 is acidic or basic. Amino acid substitutions to increase the Asp residues in the acidic proteins of *Halobacterium* from the corresponding proteins of non-halophilic species were almost identical whether the corresponding proteins were acidic or basic. This result suggested that the change of protein charges from basic proteins to acidic ones was same as from acidic proteins to acidic ones. The proteins of *Halobacterium* showed a tendency to have residues with smaller side chain than the proteins of *B. subtilis* / *D. radiodurans.*

Keywords: Orthologous acidic and basic protein pairs; isoelectric point; halophilic and non-halophilic species; acidic and basic amino acid residues; side chain volume.

1. INTRODUCTION

There are many ways to classify proteins according to their structures or functions. Proteins can be classified into groups/families based on amino acid sequence similarity [1] or three-dimensional structural similarity [2,3]; globular, membrane or fibrous proteins based on their shape and solubility; all alpha, all beta, alpha/beta or alpha + beta proteins based on the type of secondary structures present [4]; intracellular or extracellular proteins based on their localization. Classification of proteins into acidic, neutral and basic proteins based on their isoelectric point (pI) is possible.

It is reported that some classifications are related to the amino acid compositions, e.g. membrane or globular proteins [5-7], folding types [8-14], intracellular or extracellular proteins [15,16]. Previously, it was noticed that orthologous proteins among prokaryotes have similar pI values. The pI of a protein is determined by the balance of acidic and basic amino acid residues. Therefore, it seems that the balance of acidic and basic residues is important and conserved for the function of a protein.

The proteins from halophilic species are rich in acidic residues such as Asp and Glu [17,18], therefore, the proteins are biased to acidic proteins. The abundance of acidic residues on the protein surface from halophilic species is a key determinant of adaptation to high salt conditions [19,20]. It is considered that the orthologous acidic proteins from halophilic species must be corresponding to the basic proteins from non-halophilic species, and the comparison would reveal the factors between acidic and basic proteins. The genomic sequence of halophilic *Halobacterium* sp. NRC-1 (*Halobacterium)* [21] has indicated that the proteins are closely related to the proteins of Gram-positive bacterium *Bacillus subtilis* [22], or to the proteins of radiation resistant bacterium *Deinococcus radiodurans* [23].

In this study, comparison was carried out for the orthologous protein pairs between *Halobacterium* and *B. subtilis* for their acidic versus basic protein pairs, together with acidic vs. acidic pairs and basic vs. basic pairs. Similar comparison was done for the proteins between *Halobacterium* and *D. radiodurans.* The amino acid compositions were compared and the amino acid substitutions were counted in using the pairwise alignments.

2. MATERIALS AND METHODS

2.1 Sequence Retrieval

Protein sequences of *Halobacterium* sp. NRC-1 belonging to Archaea, Euryarchaeota, *Bacillus subtilis* belonging to Eubacteria, Firmicutes, and *Deinococcus radiodurans* belonging to Eubacteria, Deinococcus-Thermus were retrieved from the National Center for Biotechnology Information (NCBI) database web site (http://www.ncbi.nlm.nih.gov/) for the comparison of orthologous protein pairs.

Protein sequences of *Halorhabdus utahensis, Halorubrum lacusprofundi, Natronomonas pharaonis, Haloquadratum walsbyi, Bacillus halodurans, Acidobacterium capsulatum*, *Thermus thermophilus, Vibrio parahaemolyticus* and *Picrophilus torridus* were retrieved from the NCBI web site for the pI distribution analysis.

2.2 Orthologous Proteins

The pI value of a protein was estimated by a program developed in-house. The validity of this

program was checked by the comparison of pI values between calculated and experimentally determined proteins. The acidic proteins were selected as pI < 6 and the basic proteins as pI >8. The orthologous protein pairs were identified as the mutual best hit pair in homology search between two organisms using BLASTP program [24]. The orthologous protein pairs between *Halobacterium* acidic proteins and *B. subtilis* basic proteins were selected by following procedures. Firstly, the sequence alignments greater than 30% sequence identity longer than 100 residues between a protein pI < 6 of *Halobacterium* and a protein pI > 8 of *B. subtilis* were selected. Then, the sequence similarity among selected sequences of *Halobacterium* was examined. If there were some sequences which have similarity greater than 30%, only one sequence alignment was left and the other alignments were excluded to avoid the bias of sequences. Similarly, the sequence similarity among selected sequences of *B. subtilis* was examined. The orthologous protein pairs between *Halobacterium* and *B. subtilis* for their acidic vs. acidic and basic vs. basic protein pairs were selected by similar procedure. The orthologous protein pairs between *Halobacterium* and *D. radiodurans* were selected similarly.

3. RESULTS

3.1 Isoelectric Point Distribution of Proteins

Distribution of pI values of 2075 proteins from *Halobacterium* and 4105 proteins from *B. subtilis* is plotted in Fig. 1. The pI profile of *Halobacterium* was consistent with the previous reports [25,26]. In *Halobacterium,* more than 84% of proteins were considered as acidic proteins. Neutral and basic proteins were about 4% and 6%, respectively. Other halophilic species such as *Halorhabdus utahensis, Halorubrum lacusprofundi* and *Natronomonas pharaonis* also indicated a similar pI distribution. Generally, halophilic species have high G+C content more than 60% in their genome, however, G+C content of *Haloquadratum walsbyi* is 47.9%. The proteins of *H. walsbyi* also indicated similar pI distribution with *Halobacterium*.

The proteins of *B. subtilis* were classified as 60% acidic, 8% neutral and 30% basic proteins based on the pI distribution. The pI profile was consistent with the previous report [25]. The pI distribution was examined of following bacteria; radiation resistant *Deinococcus radiodurans*, alkaliphilic *Bacillus halodurans,* acidophilic *Acidobacterium capsulatum*, thermophilic *Thermus thermophilus,* moderately halophilic *Vibrio parahaemolyticus*. All the species mentioned above indicated similar pI distribution with *B. subtilis*. To show the pI distribution clearly, only the pI profiles of *Halobacterium* and *B. subtilis* are shown in Fig. 1.

It is known that the solubility of a protein is the lowest at the pH of its pI. Generally, the pH in a cell of a bacterium is around pH 7. Therefore, small amount of neutral proteins seemed to be appropriate to avoid precipitation. The acidophilic *Sulfolobus acidocaldarius* grows optimally in acidic environment at pH 2-3 but maintains the pH in a cell at about 6.5 [27]. The pI distribution of proteins of thermoacidophilic *Picrophilus torridus* was also examined. The *P. torridus*

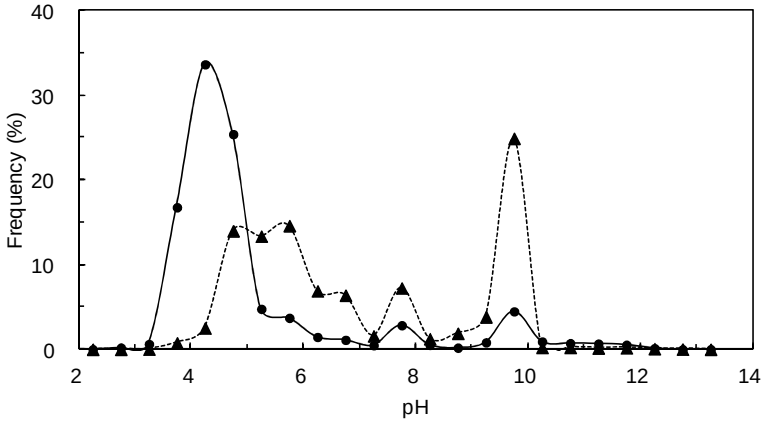

Fig. 1. Distribution of estimated pI values of 2075 proteins from *Halobacterium* (filled circles) and those of 4105 proteins from *Bacillus subtilis* (broken line, filled triangles)

optimally grows at pH 0.7 and its intracellular pH is 4.6 [28]. The ratio of basic proteins of *P. torridus* was 37%, which was a little higher than other species. *Halobacterium* has a neutral intracelluar pH 7.2 [29]. According to the pI profile of *Halobacterium* (Fig. 1), the number of orthologous protein pairs between *Halobacterium* and *B. subtilis* was considered to be small for their acidic vs. basic pairs as well as basic vs. basic pairs.

3.2 Amino Acid Composition

The selected orthologous protein pairs between *Halobacterium* and *B. subtilis* were 100 for acidic vs. acidic pairs, 53 for acidic vs. basic pairs, and 21 for basic vs. basic pairs. As mentioned above, the number of orthologous acidic vs. basic pairs and basic vs. basic pairs was not so large. The five representative orthologous protein pairs are listed in Table 1. Ribosomal proteins were rich in the basic proteins of *Halobacterium.* Some basic proteins of *B. subtilis* were assigned as acidic proteins in *Halobacterium* (see Table 1).

The amino acid composition of orthologous proteins between *Halobacterium* and *B. subtilis* is indicated in Table 2. The content of the Asp residue was 9.98% in the acidic proteins of *Halobacterium.* This high content of Asp is consistent with the reports [17,18]. However, the content of Asp was 4.14% in the basic proteins, therefore, the deviation between acidic and basic proteins was 5.84%. Similar deviations of Glu, Arg, and Lys residues were 4.22%, 1.13% and 0.35%, respectively. This result indicated that whether a protein of *Halobacterium* is acidic or basic is almost determined by the acidic residues Asp and Glu, and the effect of the basic residues Lys and Arg is very small. In the proteins of *B. subtilis,* the Glu residue showed the largest deviation between acidic and basic proteins and Asp residue followed.

To clearly show the differences in amino acid content between *Halobacterium* and *B. subtilis*, the ratios of each amino acid of *Halobacterium* to *B. subtilis* were calculated. The residues with ratios >1.30 were considered favorable and ratios <0.77 were considered unfavorable in *Halobacterium* than in *B. subtilis.* The Asp residue indicated the largest ratio 1.64 (9.98/6.07) and the Lys residue had the lowest ratio 0.28 (1.82/6.49) in the acidic vs. acidic protein pairs. The Asp and Ala residues were more frequently used in *Halobacterium* and the Lys, Ile, Asn and Met were less frequently used in *Halobacterium* commonly in the three protein pairs. The Ala codon is G+C-rich, and the codons of Lys, Ile, Asn and Met are A+T-rich.

Table 1. List of representative orthologous protein pairs between *Halobacterium* and *B. subtilis.*

Halobacterium	*B. subtilis*
Acidic proteins	Acidic proteins
Aspartate aminotransferase	Aspartate aminotransferase
Phosphoglycerate kinase	Phosphoglycerate kinase
Phosphopyruvate hydratase	Phosphopyruvate hydratase
Serine protein kinase	Serine protein kinase
Tryptophan synthase subunit beta	Tryptophan synthase subunit beta
Acidic proteins	Basic proteins
30S ribosomal protein S5P	30S ribosomal protein S5
50S ribosomal protein L11P	50S ribosomal protein L11
DNA-directed RNA polymerase	DNA-directed RNA polymerase
DNA topoisomerase I	DNA topoisomerase I
Lipoyl synthase	Lipoyl synthase
Basic proteins	Basic proteins
30S ribosomal protein S9P	30S ribosomal protein S9
30S ribosomal protein S12P	30S ribosomal protein S12
50S ribosomal protein L2P	50S ribosomal protein L2
50S ribosomal protein L14P	50S ribosomal protein L14
Heat shock protein Hsp4	Heat shock protein HtpX

The genomic G+C content of *Halobacterium* is 65.9% and that of *B. subtilis* is 43.5%. The amino acid bias is consistent with the genomic G+C content. The codon of Asp residue is neutral in G+C content, therefore, the richness of Asp residues in *Halobacterium* cannot be explained by G+C content.

The pI value of a protein is determined by the balance of positively and negatively charged residues. The Lys, Arg and His residues have potential to possess positive charge at their side chains, and Asp, Glu, Cys and Tyr residues have potential to possess negative charge. At pH 7, the Lys, Arg, Asp and Glu residues are in their fully charged form, and the Cys and Tyr residues are in their uncharged form. The His residue is partly positive as it has pK about 6.5. Usually, the content of His is low, so the effect of His residue was neglected. The content of (Lys + Arg) – (Asp + Glu) was simply calculated, which was negative for acidic proteins and positive for basic proteins (see Table 2). This calculated value was roughly correlated with pI value.

The five representative orthologous protein pairs between *Halobacterium* and *D. radiodurans* are listed in Supplementary Table S1. The amino acid composition of orthologous proteins is indicated in Supplementary Table S2. The number of orthologous acidic vs. basic pairs and basic vs. basic pairs was 59 and 15, respectively. In this case, the number of those pairs is not so large too. The genomic G+C content of *D. radiodurans* is 66.6% and that of *B. subtilis* is 43.5%. The content of the Lys and the Arg residues depend on the G+C content, therefore, their content was different between two species. However, it was interesting that the sum of positively charged residues, Lys + Arg was almost identical. For example, the sum of Lys + Arg was 9.14% in *D. radiodurans,* and 9.12% in *B. subtilis* for their basic proteins, respectively.

Table 2. Comparison of amino acid composition of orthologous protein pairs between *Halobacterium* and *B. sublitis*

	Halobacterium	*B. subtilis*	*Halobacterium*	*B. subtilis*	*Halobacterium*	*B. subtilis*
Amino acid	acidic	acidic	acidic	basic	basic	basic
Ala	12.92	8.61	12.15	8.26	13.78	8.92
Cys	0.71	0.89	0.74	0.95	0.48	0.51
Asp	9.98	6.07	7.93	4.89	4.14	2.95
Glu	7.55	8.60	6.16	5.40	3.33	3.29
Phe	2.89	3.38	3.54	4.44	4.26	5.71
Gly	9.09	8.25	9.44	8.47	10.42	8.97
His	2.56	2.23	1.66	1.76	1.65	1.84
Ile	3.87	7.47	4.39	7.43	4.88	8.69
Lys	1.82	6.49	2.41	6.67	2.07	4.18
Leu	7.89	8.78	9.06	10.36	10.60	11.69
Met	1.76	2.50	1.92	2.76	1.75	2.89
Asn	2.21	3.46	2.37	3.36	2.01	2.78
Pro	4.48	3.96	4.42	3.99	4.49	4.38
Gln	2.38	3.42	2.98	2.89	2.31	2.57
Arg	5.63	4.04	5.87	5.19	6.76	4.94
Ser	4.57	5.39	5.34	6.03	4.42	6.62
Thr	6.59	5.40	6.46	5.42	6.60	5.77
Val	9.86	7.46	9.88	8.17	11.85	8.93
Trp	0.68	0.64	1.11	0.98	1.26	1.47
Tyr	2.56	2.96	2.17	2.58	2.94	2.90
Protein pairs	100		53		21	
Average of pIs	4.3	5.3	5.2	8.6	8.7	9.2
(K+R) - (D+E)	-10.08	-4.14	-5.81	1.57	1.36	2.88

3.3 Amino Acid Substitutions

Sequence alignments were used to analyze amino acid substitutions. An example of sequence alignment of 50S ribosomal protein L11 among *Halobacterium, B. subtilis*, and *D. radiodurans* is shown in Fig. 2. The protein from *Halobacterium* was assigned as acidic, and proteins from both *B. subtilis* and *D. radiodurans* were assigned as basic proteins. The amino acid identities among sequences were 41% in *Halobacterium* vs. *B. subtilis*, 37% in *Halobacterium* vs. *D. radiodurans,* 66% in *B. subtilis* vs. *D. radiodurans.* Nine of proline residues and seven of glycine residues were conserved in the three sequences.

The 53 sequence alignments of acidic vs. basic protein pairs were examined. The 7,015 amino acid substitutions were observed in the 11,346 residues examined. Frequently observed ten amino acid substitutions between *Halobacterium* and *B. subtilis* orthologous protein pairs are indicated in Table 3. Five substitutions, Val to Ile, Val to Leu, Leu to Ile, Ala to Ser, and Ala to Val between *Halobacterium* and *B. subtilis* were commonly observed in the top ten amino acid substitutions. This result suggested that frequently observed substitutions between orthologs are independent on the type of proteins, i.e. acidic or basic.

Compared to Val and Ile or Val and Leu, the side chain of Ile or Leu have one methylene group longer than Val. This result suggested that small side chain is favored in *Halobacterium.* To compare the side chain volume clearly, accessible surface area of the residue, R, in the tripeptide Gly-R-Gly [30] was employed. The side chain volume comparison indicated Vla < Ile, Val < Leu, Leu < Ile, Ala < Val, Asp < Glu, and Glu < Lys for the frequently observed substitutions. This result clearly indicated that smaller side chain volume was preferred in the proteins of *Halobacterium.* Same trend of small side chain preference in the proteins of *Halobacterium* was observed in the amino acid substitutions between *Halobacterium* and *D. radiodurans* orthologs (Supplementary Table S3).

Fig. 2. Alignment of 50S ribosomal protein L11 among *Halobacterium, B. subtilis* and *D. radiodurans*. *Halo, B. sub*, and *D. rad* represent *Halobacterium, B. subtilis* and *D. radiodurans*, respectively

Table 3. Top 10 amino acid substitutions between *Halobacterium* and *B. subtilis* orthologous protein pairs.

No.	*Halobacterium* acidic proteins	*B. subtilis* acidic proteins	frequency %	*Halobacterium* acidic proteins	*B. subtilis* basic proteins	frequency %	*Halobacterium* basic proteins	*B. subtilis* basic proteins	frequency %
1	Val	Ile	3.93	Val	Ile	3.33	Val	Ile	3.40
2	Asp	Glu	2.97	Val	Leu	2.35	Val	Leu	3.10
3	Val	Leu	2.25	Leu	Ile	2.01	Leu	Ile	2.67
4	Leu	Ile	2.03	Arg	Lys	1.82	Leu	Val	1.85
5	Ala	Glu	1.83	Ala	Ser	1.67	Ala	Leu	1.85
6	Glu	Lys	1.66	Ala	Val	1.54	Ala	Ser	1.72
7	Ala	Ser	1.66	Ile	Val	1.47	Thr	Ser	1.68
8	Asp	Lys	1.44	Asp	Glu	1.44	Ala	Val	1.68
9	Arg	Lys	1.37	Glu	Lys	1.40	Leu	Phe	1.68
10	Ala	Val	1.36	Leu	Val	1.33	Gly	Ala	1.55
total substitutions			16954			7015			2324
residues			29263			11346			3772

The Asp residues are the main determinant whether a protein of *Halobacterium* is acidic or basic, so a possibility of different substitution patterns of Asp residues was examined according to the type of proteins acidic or basic. The replacements from Asp residues of *Halobacterium* to another residues of *B. subtilis* were counted. The Asp residues in the acidic proteins of *Halobacterium* were replaced in order by Glu > Lys > Asn > Ser > Gly in the basic proteins of *B. subtilis.* The order of substituted residues was same for the acidic proteins of *B. subtilis.* Similar analysis was done using sequence alignments between *Halobacterium* and *D. radiodurans.* The Asp residues in the acidic proteins of *Halobacterium* were replaced by Glu > Gly > Ala > Arg > Gln both in the basic and acidic proteins of *D. radiodurans.* These results indicated that substitution patterns of Asp residues of *Halobacterium* to another residues of *B. subtilis* / *D. radiodurans* were almost identical and there were no differences in the substitution patterns of Asp residues between acidic and basic proteins.

4. DISCUSSION

The difference between acidic proteins and corresponding basic proteins was examined. The pI value of a protein was used to determine whether a protein is acidic or basic. Generally, the pI values are conserved among orthologous proteins of prokaryotes. To examine a large number of acidic vs. basic orthologous protein pairs, acidic proteins from halophilic species and basic proteins from non-halophilic species were employed. Therefore, the difference between halophilic and hon-halophilic proteins might be reflected on the results.

Halophiles can be classified as slightly, moderately or extremely halophilic organisms depending on their optimally growth salt concentration. Halophilic organisms have to adjust osmotic pressure at their salt concentrations they inhabit. There are three ways in adjustment; accumulation of KCl in a cell [26,31], accumulation of organic osmotic solutes [26], and accumulation of acidic proteins with large negative charges. Most of the extremely halophilic organisms accumulate KCl and they have the pI distribution patterns like *Halobacterium.* Moderately halophilic organisms like *Vibrio parahaemolyticus* usually accumulate organic osmotic solutes and indicate the pI distribution patterns similar to that of *Bacillus subtilis.* Albumin is the smallest and most abundant of the human plasma proteins, and plays an important role in osmotic regulation. Albumin has a negative charge of 18 with pI 4.7, and produces a greater osmotic effect than expected for its concentration in plasma [32]. Acidic proteins of *Halobacterium* have similar character like albumin in terms of pI and negative charge, therefore, it is assumed that they have potential to adjust osmotic pressure. The combination of osmotic pressure adjustment is possible.

It is known that the hydrophobic Leu, Ile and Val residues are mostly found in interior regions of globular proteins, and hydrophilic Glu, Asp, Lys and Arg residues are mostly found in surface regions. The proteins of *Halobacterium* showed the preference of both hydrophobic and hydrophilic residues with smaller side chain volume. The small side chain volume of hydrophobic residues in interior of proteins may lead to compact shape. The compact size of proteins might be stable in high salt medium. This assumption need to be validated. The meaning of the small side chain volume of hydrophilic residues in surface regions is not clear. Halophilic malate dehydrogenase tetramer is wider than the similar dogfish lactate dehydrogenase. This is because the large excess of acidic residues on the surface of halophilic enzyme yield negative charge repulsion of interdimer surface [19].

It is considered that the difference of the side chain size in a sequence may be reflected on the molecular weight. The molecular weights of the two sequences in the alignments were compared adjusting the length. The molecular weights of the proteins of *Halobacterium* were a bit lower than those of *B. subtilis* in the acidic vs. basic protein pair alignments, acidic vs. acidic pairs, and basic vs. basic pairs. Similarly, the molecular weights of *Halobacterium* were a bit lower than those of *D. radiodurans.* When compared the whole sequences, the molecular weights of proteins of *Halobacterium* were larger than the corresponding proteins of *B. subtilis.* This is because the lengths of proteins of *Halobacterium* were longer than the corresponding proteins of *B. subtilis.* The proteins of *Halobacterium* are compact than the corresponding proteins of *B. subtilis* or *D. radiodurans* when compared with the same length.

In this study, orthologs between *Halobacterium* and *B. subtilis* together with *Halobacterium* and *D. radiodurans* were compared. The three

organisms belong to different taxonomy and they are remotely located on the phylogenetic tree [33,34]. However, they share considerable sequence similarity [21], even though the genomic G+C content differs considerably; 65.9% in *Halobacterium,* 43.5% in *B. subtilis* and 66.6% in *D. radiodurans.* Amino acid composition is affected on G+C content [7,35]. It is reported that the sequence similarity of orthologous proteins among *Halobacterium, B. subtilis* and *D. radiodurans* is due to laternal gene transfer [21,25]. The proteins of *Halobacterium* have changed to adapt to the high salt conditions. The simple way of adaptation is the change of protein charges from a basic protein to an acidic protein. The Asp residues are main determinant whether a protein is acidic or basic. Therefore, amino acid substitutions to increase the Asp residues might be important process for the adaptation. If gene transferred proteins are acidic, it seems that there is no need to adapt. However, those proteins showed identical substitution patterns of the Asp residues as basic proteins showed. This result suggested that adaptation from basic proteins to acidic proteins is not a special way.

Ribosomal proteins were rich in the basic proteins of *Halobacterium.* Some of the ribosomal proteins of *B. subtilis* were basic proteins, and the corresponding ones were changed to acidic proteins in *Halobacterium*. The atomic structure of the large ribosomal subunit from halophilic *Haloarcula marismortui* was determined [36]. According to the structure, ribosomal protein L2 and L14 have substantial interactions with 23S rRNA, while ribosomal protein L1, L5, L11 and L18 have weak interactions. *Halobacterium* ribosomal proteins L2 and L14 were assigned as basic and L1, L5, L11 and L18 as acidic proteins. The possibility was estimated that the ribosomal proteins which have strong interactions with 23S rRNA remained as basic proteins and proteins with weak interactions changed to acidic proteins. RNA consists of negatively charged phosphates, which may interact with positively charged Lys or Arg residues. If the interactions are essential the Lys or Arg residues would be conserved, if not Lys or Arg residues are allowed to substitute. This scenario of the change from basic proteins to acidic proteins is based on the atomic structural data.

5. CONCLUSION

Most of the proteins of *Halobacterium* showed high content of Asp residues and considered as acidic proteins, however, the content of Asp was not high for the basic proteins. The Asp and Glu residues are determinant whether a protein of *Halobacterium* is acidic or basic. The substitution patterns to increase the Asp residues in the acidic proteins of *Halobacterium* are independent on the character of the corresponding proteins whether they are acidic or basic. The proteins of *Halobacterium* showed a tendency to have residues with smaller side chain than the proteins of *B. subtilis* / *D. radiodurans.*

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Baker WC, George DG, Hunt LT. Protein sequence database. Methods Enzymol. 1990;183:31-49.
2. Orengo CA, Flores TP, Taylor WR, Thornton JM. Identification and classification of protein fold families. Protein Eng. 1993;6(5):485-500.
   DOI: 10.1093/protein/6.5.485
3. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: A structural classification of proteins database for the investigation of sequences and structures. J Mol Biol. 1995;247(4):536-540.
4. Levitt M, Chothia C. Structural patterns in globular proteins. Nature. 1976;261(5561): 552-558.
5. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982;157(1):105-132. DOI: 10.1016/0022-2836(82)90515-0
6. Hirokawa T, Boon-Chieng S, Mitaku S. SOSUI: Classification and secondary structure prediction system for membrane proteins. Bioinformatics. 1998;14(4):378-379.
   DOI: 10.1093/bioinformatics/14.4.378
7. Nakashima H, Yoshihara A, Kitamura K. Favorable and unfavorable amino acid residues in water-soluble and transmembrane proteins. J Biomed Sci Eng. 2013;6(1)36-44.
   DOI: 10.4236/jbise.2013.61006
8. Nakashima H, Nishikawa K, Ooi T. Folding type of a protein is relevant to the amino acid composition. J Biochem. 1986;99(1): 153-162.
   DOI: 10.1006/JMBI.1994.1267

9. Klein P. Prediction of protein structural class by discriminant analysis. Biochim Biophys Acta. 1986;874(2):205-215.
DOI: 10.1016/0167-4838(86)90119-6
10. Dubchak I, Muchnik I, Holbrook SR, Kim S-H. Prediction of protein folding class using global description of amino acid sequence. Proc Natl Acad Sci USA. 1995;92(19): 8700-8704.
11. Bahar I, Atilgan AR, Jernigan RL, Erman B. Understanding the recognition of protein structural classes by amino acid composition. Proteins. 1997;29(2):172-185.
DOI: 10.1002/(SICI)1097-0134(199710)29:2
12. Chou KC. A key driving force in determination of protein structural classes. Biochem Biophys Res Commun. 1999; 264(1):216-224.
DOI: 10.1006/BBRC.1999.1325
13. Taguchi Y, Gromiha MM. Application of amino acid occurrence for discriminating different folding types of globular proteins. BMC Bioinf. 2007;8:404.
DOI: 10.1186/1471-2105-8-404
14. Nakashima H, Saitou Y, Usuki N. Differences in amino acid composition between α and β structural classes of proteins. J Biomed Sci Eng. 2014;7(11): 890-918.
Doi.org/10.4236/jbise.2014.711088
15. Nakashima H, Nishikawa K. Discrimination of intracellular and extracellular proteins using amino acid composition and residue-pair frequencies. J Mol Biol. 1994; 238(1):54-61.
DOI: 10.1006/JMBI.1994.1267
16. Andrade MA, O'Donoghue SI, Rost B. Adaptation of protein surfaces to subcellular location. J Mol Biol. 1998; 276(2):517-525.
DOI: 10.1006/JMBI.1997.1498
17. Fukuchi S, Yoshimune K, Wakayama M, Moriguchi M, Nishikawa K. Unique amino acid composition of proteins in halophilic bacteria. J Mol Biol. 2003;327(2):347−357.
DOI: 10.1016/S0022-2836(03)00150-5
18. Paul S, Bag SK, Das S, Harvill ET, Dutta C. Molecular signature of hypersaline adaptation: Insights from genome and proteome composition of halophilic prokaryotes. Genome Biol. 2008;9(4):R70.
DOI: 10.1186/gb-2008-9-4-r70
19. Dym O, Mevarech M, Sussman JL. Structural features that stabilize halophilic malate dehydrogenase from an archaebacterium. Science. 1995; 267(5202): 1344-1346.
DOI: 10.1126/science.267.5202.1344
20. Elcock AH, McCammon JA. Electrostatic contributions to the stability of halophilic proteins. J Mol Biol. 1998;280(4):731-748.
DOI: 10.1006/JMBI.1998.1904
21. Ng WV, Kennedy SP, Mahairas GG, Berquist B, Pan M, Shukla HD, et al. Genome sequence of *Halobacterium* species NRC-1. Proc Natl Acad Sci USA. 2000;97(22):12176-12181.
DOI: 10.1073/pnas.190337797
22. Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V, et al. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. Nature. 1997;390(6657):249-256.
DOI: 10.1038/36786
23. White O, Eisen JA, Heidelberg JF, Hickey EK, Peterson JD, Dodson RJ, et al. Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. Science. 1999;286(5444):1571-1577.
DOI: 10.1126/science.286.5444.1571
24. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-410.
DOI: 10.1016/S0022-2836(05)80360-2
25. Kennedy SP, Ng WV, Salzberg SL, Hood L, DasSarma S. Understanding the adaptation of *Halobacterium* species NRC-1 to its extreme environment through computational analysis of its genome sequence. Genome Res. 2001;11(10): 1641-1650.
DOI: 10.1101/gr.190201
26. Oren A. Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes. Front Microbiol. 2013;4:1-6.
DOI: 10.3389/fmicb.2013.00315
27. Moll R, Shäfer G. Chemiosmotic H^+ cycling across the plasma membrane of the thermoacdophilic archaebacterium *Sulfolobus acidocaldarius*. FEBS Lett. 1988;232(2):359-363.
DOI: 10.1016/0014-5793(88)80769-5
28. Fütterer O, Angelov A, Liesegang H, Gottschalk G, Schleper B, Dock C, et al. Genome sequence of *Picrophilus torridus* and its implications for life around pH 0. Proc Natl Acad Sci USA. 2004;101(24): 9091-9096.
DOI: 10.1073/pnas.0401356101

29. Tsujimoto K, Semadeni M, Huflejt M, Packer L. Intracellular pH of halobacteria can be determined by the fluorescent dye 2', 7'-bis(carboxyethyl)-5(6)-carboxyfluorescein. Biochem Biophys Res Commun. 1988;155(1):123-129.
DOI: 10.1016/S0006-291X(88)81058-1.
30. Chothia C. The nature of the accessible and buried surfaces in proteins. J Mol Biol. 1976;105(1):1-12.
DOI: 10.1016/0022-2836(76)90191-1
31. Deole R, Challacombe J, Raiford DW, Hoff WD. An extremely halophilic proteobacterium combines a highly acidic proteome with a low cytoplasmic potassium content. J Biol Chem. 2013; 288(1):581-588.
DOI: 10.1074/jbc.M112.420505
32. White A, Handler P, Smith EL. Principles of biochemistry, 5th ed. Tokyo: McGraw–Hill Kogakusha Ltd; 1973.
33. Wolf YI, Rogozin IB, Grishin NV, Koonin EV. Genome trees and the tree of life. Trends Genet. 2002;18(9):472-479.
Doi.Org/10.1016/S0168-9525(02)02744-0
34. Fukami-Kobayashi K, Minezaki Y, Tateno Y, Nishikawa K. A tree of life based on protein domain organizations. Mol Biol Evol. 2007;24(5):1181-1189.
DOI: 10.1093/molbev/msm034
35. Lobry JR. Influence of genomic G+C content on average amino-acid composition of proteins from 59 bacterial species. Gene. 1997;205:309−316.
DOI: 10.1016/S0378-1119(97)00403-4
36. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. Science. 2000;289(5481):905-920.
DOI: 10.1126/science.289.5481.905

APPENDIX

Supplementary table S1. List of representative orthologous proteins between *Halobacterium* and *D. radiodurans.*

Halobacterium	*D. radiodurans*
Acidic proteins	Acidic proteins
Delta-aminolevulinic cid dehydratase	Delta-aminolevulinic acid dehydratase
Aspartate aminotransferase	Aspartate aminotransferase
Acetyl-CoA acetyltransferase	Acetyl-CoA acetyltransferase
Cytidine deaminase	Cytidine deaminase
Phosphoglycerate kinase	Phosphoglycerate kinase
Acidic proteins	Basic proteins
50S ribosomal protein L11P	50S ribosomal protein L11
30S ribosomal protein S5P	30S ribosomal protein S5
3-methyladenine DNA glycosylase	DNA-3-methyladenine glycosidase II
30S ribosomal protein S11P	30S ribosomal protein S11
Glutamine amidotransferase	Glutamine amidotransferase
Basic proteins	Basic proteins
50S ribosomal protein L2P	50S ribosomal protein L2
30S ribosomal protein S9P	30S ribosomal protein S9
NADH dehydrogenase/oxidoreductase	NADH dehydrogenase I subunit H
Prenyltransferase	Prenyltransferase
50S ribosomal protein L2P	50S ribosomal protein L2

Supplementary table S2. Comparison of amino acid composition of orthologs between *Halobacterium* and *D. radiodurans.*

	Halobacterium	*D. radiodurans*	*Halobacterium*	*D. radiodurans*	*Halobacterium*	*D. radiodurans*
Amino acid	acidic	acidic	acidic	basic	basic	basic
Ala	12.79	11.83	12.31	12.27	13.53	12.65
Cys	0.81	0.74	0.68	0.54	0.37	0.33
Asp	9.75	6.18	8.33	4.70	3.65	2.79
Glu	7.71	6.96	6.46	4.88	3.44	2.77
Phe	3.08	2.83	3.62	3.53	4.68	4.79
Gly	8.83	9.42	9.25	10.11	10.84	11.93
His	2.33	2.01	2.21	2.10	1.02	1.26
Ile	4.00	4.47	4.28	4.57	5.50	5.04
Lys	1.80	3.22	2.20	3.54	2.03	2.73
Leu	7.69	10.28	7.95	10.89	11.80	15.72
Met	1.68	2.01	1.73	1.92	2.15	1.93
Asn	2.26	2.28	2.49	2.73	2.33	2.16
Pro	4.47	5.10	4.50	5.06	4.60	4.80
Gln	2.53	3.73	2.97	3.61	2.24	2.00
Arg	5.71	6.58	6.16	7.15	5.89	6.41
Ser	4.53	4.64	5.02	4.75	4.98	4.71
Thr	6.84	5.40	6.98	5.45	6.14	5.51
Val	9.89	8.83	9.44	8.53	10.58	8.15
Trp	0.71	0.94	1.03	1.37	1.70	1.75
Tyr	2.59	2.55	2.39	2.30	2.53	2.57
Proteins	100		59		15	
Average of pIs	4.3	5.3	4.6	9.6	9.1	10.3
(K+R) - (D+E)	-9.95	-0.38	-6.43	1.11	0.83	3.58

Supplementary table S3. Top 10 amino acid substitutions between *Halobacterium* and *D. radiodurans* orthologous protein pairs.

No.	*Halobacterium* acidic proteins	*D. radiodurans* acidic proteins	frequency %	*Halobacterium* acidic proteins	*D. radiodurans* basic proteins	frequency %	*Halobacterium* basic proteins	*D. radiodurans* basic proteins	frequency %
1	Val	Leu	2.59	Val	Leu	2.69	Val	Leu	4.77
2	Val	Ile	2.27	Val	Ile	2.11	Gly	Ala	2.45
3	Asp	Glu	2.24	Ile	Leu	1.71	Ala	Leu	2.41
4	Ile	Val	1.59	Ala	Gly	1.67	Ile	Leu	2.27
5	Ala	Val	1.58	Ile	Val	1.65	Phe	Leu	2.23
6	Ile	Leu	1.49	Leu	Val	1.43	Ala	Gly	2.01
7	Val	Ala	1.46	Val	Ala	1.38	Val	Ala	2.01
8	Leu	Val	1.46	Asp	Glu	1.37	Ala	Val	1.78
9	Ala	Gly	1.43	Asp	Gly	1.37	Val	Ile	1.78
10	Ala	Ser	1.39	Ala	Leu	1.37	Leu	Val	1.69
total substitutions			17304			8257			2242
residues			30137			13578			3669

© 2016 Nakashima et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/14871