The Halophile Protein Database

Naveen Sharma, Mohammad Samir Farooqi*, Krishna Kumar Chaturvedi, Shashi Bhushan Lal, Monendra Grover, Anil Rai and Pankaj Pandey

Center for Agricultural Bioinformatics, Indian Agricultural Statistics Research Institute, Pusa Campus, New Delhi 110012, India

*COrresponding author: Tel: +911125841721; Fax: +911125841564; Email: samir@iasri.res.in

Citation details: Sharma,N., Farooqi,M.S., Chaturvedi,K.K. et al. The Halophile Protein Database. Database (2014) Vol. 2014: article ID bau114; doi:10.1093/database/bau114

Received 12 March 2014; Revised 20 September 2014; Accepted 4 November 2014

Abstract

Halophilic archaea/bacteria adapt to different salt concentration, namely extreme, moderate and low. These type of adaptations may occur as a result of modification of protein structure and other changes in different cell organelles. Thus proteins may play an important role in the adaptation of halophilic archaea/bacteria to saline conditions. The Halophile protein database (HProtDB) is a systematic attempt to document the biochemical and biophysical properties of proteins from halophilic archaea/bacteria which may be involved in adaptation of these organisms to saline conditions. In this database, various physicochemical properties such as molecular weight, theoretical pI, amino acid composition, atomic composition, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (Gravy) have been listed. These physicochemical properties play an important role in identifying the protein structure, bonding pattern and function of the specific proteins. This database is comprehensive, manually curated, non-redundant catalogue of proteins. The database currently contains 59 897 proteins properties extracted from 21 different strains of halophilic archaea/bacteria. The database can be accessed through link.

Database URL: http://webapp.cabgrid.res.in/protein/

Introduction

The halophilic archaea/bacteria live in a variety of saline habitats. Halophilic microorganisms are traditionally defined as organisms that optimally grow in NaCl concentrations of above 0.2 M. Some of these halophilic microorganisms grow in NaCl concentrations of above 5 M. Halophilic organisms mostly fall in three classes with reference to salinity level optimal for their growth: halotolerant (1–6%), moderate (6–15%) and extreme (15–30%). Aerobic halophilic archaea have been extensively studied with reference to their physiology, ecology, biochemistry and bioinformatics.

Proteins can exist in globular or fibrous form depending on their function. A polypeptide is a single linear polymer
It is clear that, upon lowering the salt concentration considerable changes take place in the structure of the protein. Several membrane-bound enzymes are inactivated and several membrane-bound enzymes are inactivated disintegrate to give slowly sedimenting fragments, which makes the surface more hydrophilic and more flexible. This in turn promotes nonspecific electrostatic interactions with salts in solution (1, 2). Acidic amino acids cluster on the surface of dihydrofolate reductase, proliferating cell nuclear antigen (PCNA) from Halofex volcanii (3, 4) and glucose dehydrogenase from H. mediterranei. Interactions between acidic residues on surface and hydrated salt ions not only prevent protein aggregation (5) but also maintain the functionality of the protein. Another strategy which increases hydration on the surface of these proteins is making these surfaces deficient in protein hydration. This makes the surface more hydrophilic and more flexible. This in turn promotes nonspecific electrostatic stabilization is the key factor of halophilic adaptation of proteins. Ion pair or salt bridge is an important determinant of stability of proteins (9, 10). This is more so in the case of proteins adapted to extreme environmental conditions such as high salt or temperature. Interaction energy of salt-bridge could impart stability (11–13) or be destabilising for the protein (14–16) as shown by both theoretical and experimental studies. Mostly, halophilic enzymes function at 1–4 M salt concentration. This range is required for the stability and activity of halophilic enzymes (17).

Every protein has specific physicochemical properties. The deleterious effects of monovalent salts at multi molar concentrations on biological macromolecules from various organisms have long been noted (11, 18) and seems to be caused to a large extent by dissociation of groups, subunits, etc., which are involved in ionic linkages. If such ionic bonds are lacking in halophilic cell constituents, the physical chemistry of these structures must be unusual. In general halophilic structures were indeed found to be stable only in the presence of at least 1 M salt. In addition, most systems required or were stimulated by salt at concentrations near or even above this value. Thus, rather than being destroyed at high salt concentration, the macromolecular structures responsible for biological activity in halophiles appear, in fact, to be dependent on the presence of salts. A dramatic example of this unique salt dependence is the behaviour of the cell envelope of the halobacteria, when the salt concentration is lowered. Under these conditions, cells (12, 13) and isolated cell envelopes (14–16, 18–21) disintegrate to give slowly sedimenting fragments, and several membrane-bound enzymes are inactivated (22, 23). It is clear that, upon lowering the salt concentration considerable changes take place in the structure of the cell envelope and its constituents (24). The knowledge available in our database can be compared with non-halophilic archaea/bacteria and conclusions about fundamental mechanisms of survival in halophilic archaea/bacteria can be drawn in light of the above studies.

The current database contains 21 halophilic archaeal/bacterial strains. This database consist information about the physical and chemical properties of halophilic archaeal/bacterial proteins, such as theoretical PI, molecular weight, negative and positive charge, half-life of protein and amino acid index have been populated.

### Materials and Methods

#### Source of data

The protein sequences of different halophilic strains were downloaded from NCBI website http://www.ncbi.nlm.nih.gov/protein/?term=Halophiles+archaea. The Bioperl script was run for all halophilic strains. All biochemical composition such as Number of Amino acids, Instability index, Half-life, Number of atoms, Gravy, Aliphatic index were extracted through Bioperl script as shown in Figure 1.

pl or isoelectric point is the pH at which the net charge on the protein is zero. pl can be directly affected by the

| S. NO | Strain name                  | Total number of protein information |
|-------|-----------------------------|------------------------------------|
| 1     | Azotobacter vinelandii      | 10 414                             |
| 2     | Bacillus cereus ATCC 10987  | 10 691                             |
| 3     | Halobacterium salinarum     | 16                                 |
| 4     | Halofex mediterranei ATCC 33500 | 02                             |
| 5     | Natronomonas pharaonis DSM 2160 | 447                           |
| 6     | Cellulosimicrobium cellulans | 08                                 |
| 7     | Halofex volcanii            | 462                                 |
| 8     | Haloarcula vallismorts ATCC 29715 | 510                         |
| 9     | Chromohalobacter saloxigen DSM 3043 | 6359                        |
| 10    | Halofex denitrificans ATCC 35960 | 490                             |
| 11    | Halorubrum saccharovorum DSM 1137 | 6009                         |
| 12    | Halorubrum distributum JCM 10118 | 467                             |
| 13    | Bacillus cereus G9241       | 2480                                |
| 14    | Salinibacter ruber DSM 13855 | 5287                                |
| 15    | Bacillus cereus E33L        | 9914                                |
| 16    | Chromohalobacter sp. HS2    | 16                                 |
| 17    | Halorubrum lacusprofundi ATCC 49239 | 459                        |
| 18    | Halorubrum trapanicum       | 01                                 |
| 19    | Salinibacter ruber M8       | 5735                                |
| 20    | Halomonas elongata DSM 2581 | 127                                 |
| 21    | Chromohalobacter beijerinckii | 03                             |
| Total |                             | 59 897                              |
reduction of disulphide bonds in the proteins. The molecular weight is the elementary biophysical parameter and has direct correlation with the volume of the molecule. It influences the protein structure, which is functionally very important. The difference between the total number of positively (Arg + Lys) and negatively (Asp + Glu) charged amino acids in the protein gives the net charge of a protein. The pattern of hydrophobicity and net charge on the protein represents a unique structural feature of the proteins (20).

The half-life of a protein is defined as the time required for half of the total amount of protein in a cell to disappear after its synthesis. The in vivo stability of the protein is largely determined by the amino acids present at N-terminal of the protein and is given by the N-end rule (25–27). The instability index is an indicator of stability of a protein in vitro. The proteins with instability index smaller than 40 are predicted as stable, whereas, a value above 40 indicates instability of the protein (28) (http://web.expasy.org/protparam/protparam-doc.html). The formula of instability index (II) is as follows:

\[
i = L - 1 \\
II = \left( \frac{10}{L} \right) \times \sum_{i=1}^{L-1} DIWV(x[i] \times [i+1])
\]
Figure 3. Data flow diagram.

Figure 4. Screenshot of the Halophile Protein Database (HProtDB) home page.
where: \( L \) is the length of sequence

\[
\text{DIWV}(x[i]x[i+1]) \text{ is the instability weight value for the dipeptide starting in position } i.
\]

The relative volume occupied by aliphatic side chains (alanine, valine, isoleucine and leucine) is defined as the aliphatic index of a protein. The aliphatic index may influence thermostability of globular proteins. The sum of hydrophathy values of all the amino acids, divided by the number of residues in the protein sequence gives the GRAVY value.

**Database architecture**

In order to store the information about protein properties of different strains of halophilic archaea/bacteria, open source database software MySQL (version 5.1.3.6) was utilized. The data is stored in the form of associated tables, which also follows Relational Database Management System (RDBMS) concepts. MySQL is feature-rich database software that provides speedy data access, ease of use, portability and also supports most of ANSI SQL commands. The data consistency and non-redundancy were maintained by employing normalization techniques on the developed database. HTML and PHP were used to render a dynamic web interface and the appropriate database connectivity techniques were utilized for quick and easy information retrieval. The viewing of the data is freely available along with a facility to download data. This web application has been hosted using an open source WAMP Server (version 2.0i, windows web development environment) which also provides multiuser access facility. WAMP server allows hosting web applications developed using PHP.
and MySQL over Apache2 web server. Figure 2 depicts the architecture of HProtDB.

The spectrum of the database comprises of database tables for user management, protein, biochemical and biophysical properties of proteins. Besides, fields of the tables cover details of all attributes of the concerned parameter. A primary key in each table is identified for uniquely defining a record. Similarly, the foreign keys were identified from other tables for setting relationship among different entities. Some of the tables were master tables, which were meant for providing the real world values to fields in different tables, while building the queries and presenting the reports.

Figure 3 shows the Data Flow Diagram (DFD) of the HProtDB. The whole system has been depicted in such a way so that the continuity of information flow should not be lost at the next level. This DFD shows all the processes together with the data stores.

The home page of the database is depicted in Figure 4. The different tables on the home page provide links to general information, such as protein, amino acids, microbes and other modules related to data entry and retrieval. The search facility (Figure 5) enables the user to search the biochemical and physical properties of the desired protein either through accession number or protein names given in the dropdown list. The user has to select the desired protein, and subsequently all information related to the protein gets extracted from the database and displayed on the screen. The data retrieval option on the home page also provides the user to search for any specific halophilic archaea/bacteria records. This option provides the list of strains and clicking on a particular strain gives the protein

Figure 6. Snapshot of different strains list.
and protein properties. In this way, user can access any or all 21 different strains of halophilic archaea/bacteria (Figures 6–8).

**Results and discussion**

We have constructed a database which provides biochemical/biophysical properties of the proteins from halophilic archaea/bacteria. The study of these properties may lead to elucidation of mechanisms for salt tolerance. Identifying salt-tolerant proteins in halophilic bacteria and transfer of such proteins to other agriculturally important bacteria such as *Rhizobium, Azotobacter, Cyanobacteria* etc. will be useful from applied point of view as the engineered microbes may be able to adapt in saline conditions. The information in our database may also be useful for designing synthetic proteins with optimal physicochemical proteins which may be of use in saline conditions.

**Conclusion**

The HProtDB lists various physicochemical properties of the proteins of halophilic archaea/bacteria. Halophilic archaea/bacteria are excellent models for study of
osmoregulatory mechanisms that permit these organisms to grow in saline environments. The information in the database might prove useful in elucidating the fundamental mechanisms for salt tolerance and for identifying the characteristics of the genes involved in salt tolerance. These may prove useful in identifying and annotating novel salt tolerant genes (29).

Funding
Funding for open access charge: Centre for Agricultural Bioinformatics, ICAR - Indian Agricultural Statistics Research Institute.

Conflict of interest. None declared.

References
1. Kennedy, S.P., Ng, W.V., Salzberg, S.L. et al. (2001) Understanding the adaptation of Halobacterium species NRC-1 to its extreme environment through computational analysis of its genome sequence. Genome Res., 11, 1641–1650.
2. Paul, S., Bag, S.K., Das, S. et al. (2008) Molecular signature of hypersaline adaptation: insights from genome and proteome composition of halophilic prokaryotes. Genome Biol., 9, R70.
3. Pace, C.N. (1990) Conformational stability of globular proteins. Trends Biochem. Sci., 15, 14–17.
4. Horovitz, A. and Fersht, A.R. (1992) Co-operative interactions during protein folding. J. Mol. Biol., 224, 733–740.
5. Dill, K.A. (1990) Dominant forces in protein folding. Biochemistry, 29, 7133–7155.
6. Winter, J.A., Christofi, P., Morroll, S. et al. (2009) The crystal structure of Haloferax volcanii proliferating cell nuclear antigen reveals unique surface charge characteristics due to halophilic adaptation. BMC Struct. Biol., 9, 55.

7. Mevarech, M., Frolov, F. and Gloss, L.M. (2000) Halophilic enzymes: proteins with a grain of salt. Biophys. Chem., 86, 155–164.

8. Marqusee, S. and Sauer, R.T. (1994) Contribution of a hydrogen bond/salt-bridge network to the stability of secondary and tertiary structures in lambda repressor. Protein Sci., 3, 2217–2225.

9. Pfeil, W. (1986) Unfolding of proteins In: Hinz H.J. (ed.). Thermodynamic Data for Biochemistry and Biotechnology. Springer-Verlag, Berlin, pp. 349–376.

10. Stickler, D.F., Presta, L.G., Dill, K.A. et al. (1992) Hydrogen bonding in globular proteins. J. Mol. Biol., 226, 1143–1159.

11. Jencks, W.P. (1969) Catalysis in chemistry and enzymology. McGraw-Hill Book Co., New York.

12. Von Hippel, P.H. and Schleich, T. (1969) The effects of neutral salts on the structure and conformational stability of macromolecules in solution, In: Timasheff S.N. and Dasman D. (eds). Structure and stability of biological macro molecules. Marcel-Dekker Inc., New York, pp. 416–574.

13. Abram, D. and Gibbons, N.E. (1961) The effect of chlorides of monovalent cations, urea, detergents and heat on morphology and the turbidity of suspensions of red halophilic bacteria. Can. J. Microbiol., 7, 741–750.

14. Brown, A.D. (1963) The peripheral structures of gram-negative bacteria cation-sensitive dissolution of the cell membrane of the halophilic bacterium, Halobacterium halobium. Biochim. Biophys. Acta., 75, 425–435.

15. Brown, A.D. (1964) Aspects of bacterial response to the ionic environment. Bacterial. Res., 28, 296–329.

16. Brown, A.D. (1964) The development of halophilic properties in bacterial membranes by acylation. Biochim. Biophys. Acta., 93, 136–142.

17. Mevarecha, M., Frolova, F., Glossh, L.M. (2000) Halophilic enzymes: proteins with a grain of salt. Biophys. Chem., 86, 155–164.

18. Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. J. Mol. Biol., 157, 105–132.

19. Larsen, H. (1967) Biochemical aspects of extreme halophilism. Adv. Microb. Physiol., 1, 97–132.

20. Dao-pin, S., Anderson, D.E., Baase, W.A. et al. (1991) Structural and thermodynamic consequences of burying a charged residue within the hydrophobic core of T4 lysozyme. Biochemistry, 30, 11521–11529.

21. Kushner, D.J. and Onishi, H. (1966) Contributions of protein and lipid components to the salt response of envelopes of an extremely halophilic bacterium. J. Bacteriol., 91, 653–660.

22. Onishi, H. and Kushner, D.J. (1966) Mechanism of dissolution of the extreme halophile Halobacterium cutirevatum. J. Bacteriol., 91, 646–652.

23. Hochstein, L.I. and Dalton, B.P. (1968) Salt specificity of a reduced nicotinamide adenine dinucleotide oxidase prepared from a halophilic bacterium. J. Bacteriol., 95, 37–42.

24. Lanyi, J.K. (1969) Studies of the electron transport chain of extremely halophilic bacteria, Salt dependence of reduced diphosphopyridine nucleotide oxidase. J. Biol. Chem., 244, 2864–2869.

25. Bachmair, A., Finley, D., Varshavsky, A. (1986) In vivo half-life of a protein is a function of its amino-terminal residue. Science, 234, 179–186.

26. Gonda, D.K., Bachmair, A., Wunning, I. et al. (1989) Universality and structure of the N-end rule. J. Biol. Chem., 264, 16700–16712.

27. Tobias, J.W., Shrader, T.E., Rocap, G. et al. (1991) The N-end rules in bacteria. Science, 254, 1374–1377.

28. Guruprasad, K., Reddy, B.V.B. and Pandit, M.W. (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng., 4, 155–161.

29. Lieberman, M.M. and Lanyi, J.K. (1971) Studies of the electron transport chain of extremely halophilic bacteria, mode of action of salts on cytochrome oxidase. Biochim. Biophys. Acta., 245, 21–33.