ScienceDirect

View PDF

Download full issue

Current Research in Structural Biology
Volume 7, 2024, 100129

Research Article

Living in trinity of extremes: Genomic and proteomic signatures of halophilic, thermophilic, and pH adaptation

Aidana Amangeldina [a] [b], Zhen Wah Tan [a], Igor N. Berezovsky [a] [b]

Show more

Outline | Share | Cite

https://doi.org/10.1016/j.crstbi.2024.100129

Get rights and content

Under a Creative Commons license ● Open access

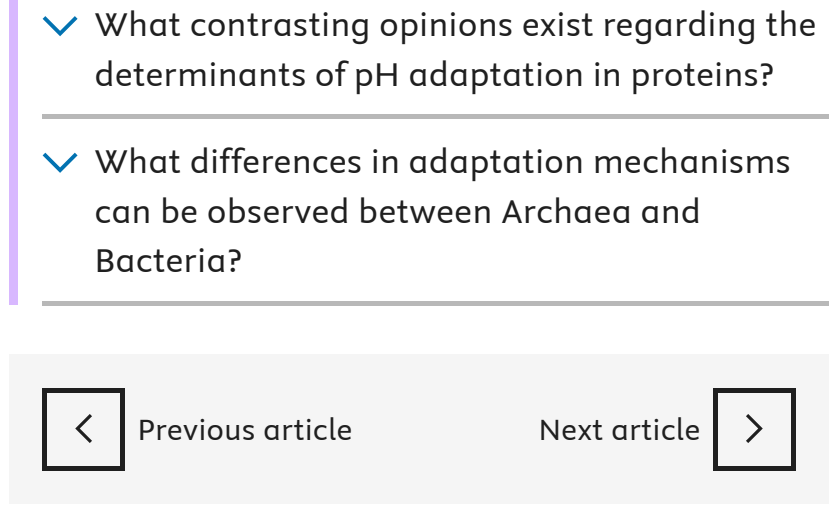
Highlights

- Trinity of extremes: pH, salinity, temperature.
- Genomic/proteomic signatures of molecular adaptation.
- Getting hints on molecular mechanisms of adaptation.

Abstract

Since nucleic acids and proteins of unicellular prokaryotes are directly exposed to extreme environmental conditions, it is possible to explore the genomic-proteomic compositional determinants of molecular mechanisms of adaptation developed by them in response to harsh environmental conditions. Using a wealth of currently available complete genomes/proteomes we were able to explore signatures of adaptation to three environmental factors, pH, salinity, and temperature, observing major trends in compositions of their nucleic acids and proteins. We derived predictors of thermostability, halophilic, and pH adaptations and complemented them by the principal components analysis. We observed a clear difference between thermophilic and salinity/pH adaptations, whereas latter invoke seemingly overlapping mechanisms. The genome-proteome compositional trade-off reveals an intricate balance between the work of base paring and base stacking in stabilization of coding DNA and r/tRNAs, and, at the same time, universal requirements for the stability and foldability of proteins regardless of the nucleotide biases. Nevertheless, we still found hidden fingerprints of ancient evolutionary connections between the nucleotide and amino acid compositions indicating their emergence, mutual evolution, and adjustment. The evolutionary perspective on the adaptation mechanisms is further studied here by means of the comparative analysis of genomic/proteomic traits of archaeal and bacterial species. The overall picture of genomic/proteomic signals of adaptation obtained here provides a foundation for future engineering and design of functional biomolecules resistant to harsh environments.

Graphical abstract



Download: Download high-res image (177KB)
Download: Download full-size image

Beta Powered by GenAI

Questions answered in this article

This is generative AI content and the quality may vary. Learn more.

What is the relationship between amino acid combinations and environmental factors in adaptation?

What complex relationships exist between nucleic acid bias and amino acid compositions in extreme environmental adaptations?

What trends were observed in salinity and pH adaptation predictors?

What contrasting opinions exist regarding the determinants of pH adaptation in proteins?

What differences in adaptation mechanisms can be observed between Archaea and Bacteria?

< Previous article   Next article >

Keywords

Thermophilic adaptation; Halophilic adaptation; PH adaptation; Genomes/proteomes; Protein physics; Molecular adaptation; Protein evolution; Archaea; Bacteria

1. Introduction

The genomic-proteomic compositional determinants of molecular mechanisms of adaptation to extremes of pH, salinity, and temperature in unicellular organisms, archaea, and bacteria, are studied here. Cell viability is obviously dependent on proteins that perform various functions, whereas the passage of genetic information necessary for building the cell machinery also requires the integrity of the genetic material (DNA) and the transcriptional intermediate (RNA). Thus, DNA, RNA, and proteins are the most essential molecules in the cell, and their functioning and physical stability at any given environmental condition determine cell survival and reproduction. As directed by the central dogma of molecular biology, the integrity of the flow of genetic information cannot be guaranteed without stability of any element in the chain: from DNA to RNA to proteins. Some previous studies reported that nucleic acid bias shapes amino acid compositions and thus shapes the evolutionary landscape of proteins (Singer and Hickey, 2000; Bohlin et al., 2013; Tekaia and Yeramian, 2006) while others argue that there is more complex relationship (Goncearenco and Berezovsky, 2014), especially in relation to adaptation to extreme environments (Fukuchi et al., 2003; Goncearenco et al., 2014; Nakashima et al., 2003; Tekaia et al., 2002). Extreme environmental adaptation introduces certain trends for stability on both the nucleic acid (DNA/RNA) and protein levels, such that DNA with biases might encode further for different amino acids, and amino acid biases needed for protein stability may originate trends towards specific codons, affecting, thus, the nucleotide sequences (Goncearenco and Berezovsky, 2014). Therefore, a combination of the redundant genetic code and the availability of several types of amino acids with similar physicochemical properties leads to a complex trade-off working in the mutual environmental adaptation of DNA and proteins (Goncearenco and Berezovsky, 2014; Goncearenco et al., 2014).

Additionally, consideration of adaptation mechanisms should also include the evolutionary perspective (Goncearenco et al., 2014; Berezovsky and Shakhnovich, 2005; Tokuriki et al., 2009) that leaves its marks from the very Origin of Life, determining the genetic code emergence, codon chronology (Trifonov, 2000), and consensus temporal order of amino acids (Trifonov, 2000; Trifonov et al., 2001), The evolution is driven by the polymer nature of nucleic acids and proteins that establishes their basic units (Berezovsky et al., 1999, 2000a, 2017a; Koczyk and Berezovsky, 2008; Svedberg, 1929; Berezovsky, 2003; Berezovsky and Trifonov, 2001), and shapes them on different stages of evolution (Trifonov et al., 2001; Aziz and Caetano-Anolles, 2021), The emergence of functional diversity (Berezovsky et al., 2003, 2017a; Aziz et al., 2016; Tal et al., 2016; Zeldovich et al., 2006; Goncearenco and Berezovsky, 2015), and specific ways of its design (Berezovsky, 2019; Hocker, 2014; Yin et al., 2021) and regulation (Berezovsky et al., 2017b), including allosteric mechanisms (Tee et al., 2020, 2021, 2022; Guarnera and Berezovsky, 2019) are also results of the evolution.

Enzymes from extremophiles with enhanced balance between the stability, activity, and flexibility in order to function (Hou et al., 2023) at extreme pH, salinity, and temperature are instrumental in application in harsh industrial processes. The protein stability is a result of a mutual work of positive and negative components of design (Berezovsky et al., 2007), making as low as possible energy of the protein native state (positive design) and increasing energies of misfolded conformations (negative design). It results in a widening of the gap between the energies of the native state and non-native structures (Berezovsky et al., 2007). The case study of the work of negative and positive components of design in protein thermostability revealed so-called "from both ends of hydrophobicity scale" trend in the amino acid composition (Goncearenco et al., 2014; Berezovsky et al., 2007; Ma et al., 2010). Specifically, increasing usage of strong hydrophobic and charged residues at the expense of polar ones was observed upon increase of the organismal optimal growth temperature (OGT). This compositional bias was shown to contribute to the enthalpy of the protein, providing stronger van der Waals interactions (Berezovsky and Shakhnovich, 2005; Berezovsky, 2003; Berezovsky et al., 1997, 2000b) in the protein core and forming additional stabilizing ion pairs and hydrogen bonds on the protein surface – elements of the *positive design* (Berezovsky et al., 2007). At the same time, increase of the amount of positively charged amino acids contributes to the negative component of design by increasing repulsions between positive charges (Berezovsky et al., 2007) in misfolded conformations and in non-native protein-protein interfaces (Ma et al., 2010; Berezovsky, 2011). Obviously, the fraction of only one sign of charged residues can be used in the negative design to provide massive repulsion in non-native conformation. In protein thermostability of both monomeric proteins (Goncearenco et al., 2014; Berezovsky et al., 2007; Ma et al., 2010; Berezovsky, 2011) and protein-protein interfaces (Ma et al., 2010) the positive charges are apparently a key element of the negative design, increasing the energies of non-native conformations. As a result, the gap between the native state energy and those of misfolds is widening and thermodynamic stability of the protein is increasing (Goncearenco et al., 2014; Berezovsky et al., 2007; Ma et al., 2010; Berezovsky, 2011). Adaptation to extremes of temperature and to other harsh environments is provided by combined contributions of distinct stabilizing interactions (Berezovsky, 2003, 2011; Berezovsky et al., 1997, 1999, 2000b, 2005; Cambillau and Claverie, 2000; Folch et al., 2008; Jaenicke, 1999; Dyson et al., 2006; Pace et al., 2014a, 2014b; Pucci and Rooman, 2014; Pylaeva et al., 2018; Shakhnovich, 2006; Van Dijk et al., 2015; Kajander et al., 2000; Makhatadze et al., 2003; Bresler and Talmud, 1944a, 1944b), presence of which is reflected in a number of compositional and sequence/structure determinants (Goncearenco et al., 2014; Ma et al., 2010; Cambillau and Claverie, 2000; Shakhnovich, 2006; Van Dijk et al., 2015; Chakravarty and Varadarajan, 2000; Mamonova et al., 2013; Gromiha et al., 2013; Gromiha and Suresh, 2008). It was shown, for example, that specific interactions and characteristics of the structure may dominate in adaptation to certain extreme conditions, such as (hyper)thermophilic and psychrophilic adaptation (Pucci and Rooman, 2017), which require better packed and stabilized structures in the former and their less rigid, flexible homologs in the latter (Feller and Gerday, 2003; Goodchild et al., 2004).

A salty environment is characterized by the low water availability at high salt ionic concentrations, providing less contribution to protein folding by hydrophobic effect. It was shown that halophilic adaptation can happen through destabilization of unfolded state by cation exclusion in unfolded states, while electrostatic interactions between the cation and abundant acidic amino acid residues can contribute to the stability of the folded state (Ortega et al., 2015). Putting above results into the terminology of positive-negative design (Berezovsky et al., 2007) discussed above: depletion of positive charges leads to additional repulsion in misfolded conformations constituting the negative component of design, while contribution to the positive one is provided by electrostatic interactions between the environmental cations and negatively charged amino acids. The excess of acidic residues can also support the protein hydration, stabilizing the folded protein form via interactions with hydrated cations at high salt concentrations (Ortega et al., 2015; Deole et al., 2013; Ebrahimie et al., 2011). In the other study it was suggested that hydration of protein with various acidic residue content had the same hydration level, while the role of acidic residues could be in the prevention of protein aggregation (Daronkola and Verde, 2021), which might be another example of the work of negative charges in negative design. At the same time, however, in addition to documented role of acidic amino acid residues in high salt adaptation, it was proposed elsewhere that halophilic adaptation can be based on basic residues (Elevi Bardavid and Oren, 2012).

There is also a diversity of opinions on the determinants of pH adaptation, including opposite conclusions that increased (decreased) positively charged residues and decreased (increased) negatively charged residues are important for high/low pH adaptation (Daronkola and Verde, 2021). The specificity of protein adaptation to different pH environments for each class of proteins was also proposed (Dubnovitsky et al., 2005). Since pH adaptation is related to the change of the charge state of the amino acid, which might lead to disruption of the existing structure, ionizable residues and the ionization state of certain residues seem to play an important role in protein stability and function (Beliën et al., 2009; Xu et al., 2013). For example, replacing residues susceptible to charge-state change with non or less-susceptible ones was found to be a viable strategy for pH adaptation design (Suplatov et al., 2014). It was also shown that replacing a basic amino acid with an acidic one improved the protein stability and catalytic efficiency under an acidic environment (Yang et al., 2013). And vice versa, replacing acidic residues with less acidic or more basic (like glutamic acid with glutamine) led to a change of the pH optima (Suplatov et al., 2014; Fushinobu et al., 1998; Liu et al., 2009). Additional potential strategy for both pH adaptation and salinity adaptation is pKa modulation (Beliën et al., 2009; Xu et al., 2013; Gutteridge and Thornton, 2005; Francois et al., 2006; Andreeva and James, 1991).

The next level of complexity in the study of environmental adaptation is a situation when some extreme environments are coupled (Oren, 2002; Reed et al., 2013), in which certain compositional trends can be related to adaptation to several different extreme environments at the same time. For example, an increased ratio of acidic over basic amino acid residues has been linked not only to psychrophiles (Xia et al., 2018), but also to adaptation to high pH (alkalophilic proteins) (Mamo et al., 2009) and to high salt (halophilic proteins) (Daronkola and Verde, 2021). Another example is Asparagine (N) amino acid residue that is thermolabile and alkali susceptible (Gülich et al., 2002; Walden et al., 2004): increased fraction of charged amino acids and lower content of Asn residues is an adaptation trend common for both thermophilic and alkaliphilic protein stabilization (Manikandan et al., 2006). The elevated number of salt bridges is characteristic for thermophilic and salinity adaptations (Berezovsky et al., 2007; Ma et al., 2010; Dyson et al., 2006; Mamonova et al., 2013; Bandyopadhyay et al., 2007; Dym et al., 1995; Nayek et al., 2014). So far, only limited number of studies account for poly-extremophilic adaptation, where adaptation to one factor might be affecting (enabling or disabling) the other type of adaptation (Alcaide et al., 2015; Popinako et al., 2017; Sriaporn et al., 2021). In particular, it was shown that adaptation to high salinity can be coupled with adaptation to a high pH (Manikandan et al., 2006). An increase in acidic residues and decrease in Lysine, which becomes unstable due to proton loss at high pH and has a too-long hydrophobic chain for the high salt environment, was observed for both alkaliphilic and halophilic proteins (Ortega et al., 2015; Popinako et al., 2017). Another drawback in previous studies of the polyextremophile adaptation is that they are typically performed on individual proteins, not allowing to infer generic trends reflecting mechanisms of adaptation characteristic for the whole proteomes.

We work here with large datasets of genomes/proteomes annotated with data on the optimal growth pH (OGP), salinity (OGS), and temperature (OGT) of corresponding organisms. It allows us to capture the most important signals of adaptation to each extreme environment, to follow trends reflecting the stability tuning of both nucleic acids and proteins, to delineate the causal relation between the compositional trends in amino and nucleic acids, and to find the relationship between adaptation mechanisms evolved in response to different environments. We also considered emergence and development of adaptation mechanisms from the evolutionary perspective, analysing the compositional trends characteristic for mechanisms of adaptation working in groups of archaeal and bacterial organisms.

2. Materials and Methods

2.1. Environmental data

BacDive database was used as the primary source database to build the sets of organisms with available temperature, pH, or salinity and to obtain corresponding lists of BacDive IDs for each environmental factor. The original data consisted of 'general growth' and 'optimal growth' data, represented as 'numeric range' or 'numeric points'. Each environmental entry for one BacDive entry was accessed, and we have selected 'optimum' and 'growth' points. The entry to be used in the analysis for characterizing features of corresponding proteome was selected according to the following priority: the optimal growth (OG) number for each environmental factor is assigned as the average value of optimal growth interval. For those organisms that had entry points given as only one number we use this number as an optimal one (both minimum and maximum values are designated to this number). If only maximum or minimum optimum growth points was available (for example <105 °C), this available number is taken as an OG value. If optimal growth interval is not available, we use the average of the growth value interval, or the available general growth point. The salt concentration measurements in molarity or artificial seawater concentration were converted to NaCl salt percent concentration. Below are examples of the conversion of molarity and artificial seawater concentration into the NaCl percent concentration. First, given molarity = 0.14 M and molar weight of NaCl $\mathbf{MW_{NaCl}} = 58.44\ \frac{g}{mol}$, one obtains $\mathbf{0.14}\ \frac{Mol}{L}\ *\ 58.44\ \frac{g}{mol} = 8.18\ \frac{g}{L}$, or **0.82 %** NaCl concentration. Second, given the reference protocol for preparation of the 35 % artificial seawater (AS (Kester et al., 1967),) that should contain 23.93 g/kg of NaCl (**2.393 %**, slight halophile), the AS = 100 % will show **6.8 %** salt concentration corresponding to moderate halophile.

The organism annotation consists of an organism name, domain, oxygen tolerance, phylum, gram stain. Partially or fully unclassified species and eukaryotic organisms were discarded from the dataset. The species without domain or with several unconventional environmental factor annotations were corrected or cleared out from the dataset. The four datasets were obtained: full pH dataset (pH_set, Suppl. Table S1), full salinity dataset (S_set, Suppl. Table S2), full temperature data (T_set, Suppl. Table S3), and the dataset with all three environmental factor data available for each organism (Env_set, Suppl. Table S4). The classification of organisms according to their degree of adaptation to extreme environments is presented in Table 1. The statistics for all the above groups are provided in Table 2, and the information on overrepresentation of organisms on certain environmental conditions (e.g., OGTs 28, 30, 37 °C etc.) - in Suppl. Table S6.

Table 1. Classification range criteria used for different environmental groups.

Classification Range	Classification group
Temperature	
(, 25] °C	psychrophiles
(25, 50) °C	mesophile
[50–80) °C	thermophile
[80) °C	hyperthermophile
Salinity	
(, 2] NaCl %	nonhalophile
(2, 5) NaCl %	slight halophile
[5, 20) NaCl %	moderate halophile
[20) NaCl %	extreme halophile
pH	
[0, 5.5] pH	acidophile
(5.5, 8) pH	neutrophile
[8, 14] pH	alkaliphile

Table 2. **Summary of datasets for each environmental factor and for environmental set.** Note: BacDive classifies 186 organisms from * as thermophilic, one organism from * as psychrophilic, two organisms from ** as hyperthermophilic.

Dataset type	*Counts*	*Bacteria*	*Archaea*
Full T data (T_set)	***9225***	***8884***	***341***
psychrophiles	648	645	3
mesophile	8052*	7830	222
thermophile	445**	400	45
hyperthermophile	80	9	71
Full S data (S_set)	***2568***	***2489***	***79***
nonhalophile	1239	1217	22
slight halophile	806	801	5
moderate halophile	496	469	27
extreme halophile	27	2	25
Full P data (P_set)	***2958***	***2869***	***89***
alkaliphile	497	488	9
neutrophile	2348	2274	74
acidophile	113	107	6
Env Dataset (Env_set)	**2236**	**2159**	**77**
psychrophiles	238	238	0
mesophile	1870	1807	63
thermophile	118	112	6
hyperthermophile	10	2	8
nonhalophile	1103	1082	21
slight halophile	699	694	5
moderate halophile	408	381	27
extreme halophile	26	2	24
neutrophile	1765	1700	65
alkaliphile	408	400	8
acidophile	63	59	4

2.2. Sequence data

The nucleotide coding sequences and protein sequences were downloaded from the NCBI database (Refseq (Tatusova et al., 2016) or GenBank (Clark et al., 2016)). tRNA and rRNA data has been extracted from GenBank. Non-codon-biased sequences were generated by assigning equal probability to each of the synonymous codons weighted by their encoding amino-acid weight in the proteome. Amino acid composition and dipeptide composition was calculated for each protein for the proteome of each organism.

Proteins were annotated as a membrane or nonmembrane proteins (Olivella et al., 2013), depending on the presence of 'membr' string in the protein name. Further, the missing membrane annotations were extracted using BLAST+, version 2.10 (Camacho et al., 2009). The BLAST search was built on a blastp-fast task with a parameter of e-value equal to 10^{-5} with minimum query coverage of 50 % per high-scoring segment pair. The proteins with a minimum 35 % identity with membrane proteins were annotated as membrane ones and excluded from further consideration because of the distinct amino acid composition between the membrane and nonmembrane proteins.

2.3. Proteomic predictors of adaptation to extremes of pH, salinity, and temperature