

VILNIUS UNIVERSITY

**Gediminas Alzbutas**

**SALT RESISTANCE MECHANISM OF  
HALOTOLERANT / HALOPHILIC  
PROKARYOTIC DNASES AND  
HALOTOLERANCE INDUCTION FOR BOVINE  
DNASEI**

Summary of doctoral dissertation  
Physical sciences, biochemistry (04 P)

Vilnius, 2016

This study was carried out during PhD studies at the Institute of Biotechnology, Vilnius University, in 2011–2015 and at UAB „Thermo Fisher Scientific Baltics“. Dissertation is defended extramurally.

Scientific supervisor:

dr. Arūnas Lagunavičius (UAB „Thermo Fisher Scientific Baltics“, physical sciences, biochemistry – 04 P).

### **Evaluation board of dissertation of Biochemistry trend:**

#### **Chairman:**

dr. **Vaida Šeputienė** (UAB „Thermo Fisher Scientific Baltics“, Physical sciences, biochemistry – 04 P).

#### **Members:**

dr. **Artūras Jakubauskas** (Vilnius University Hospital Santariskiu Klinikos, Physical sciences, biochemistry – 04 P),

dr. **Gražvydas Lukinavičius** (Max Planck Institute for Biophysical Chemistry, Physical sciences, biochemistry – 04 P),

prof. dr. **Elena Servienė** (Nature Research Centre, Biomedical Sciences, biology – 01 B),

dr. **Giedrius Vilkaitis** (Vilnius University, Physical sciences, biochemistry – 04 P).

The thesis defense will take place at Joint Life Science Center, R401, Vilnius University (Saulėtekio Avenue 7, LT-10222, Vilnius, Lithuania) on 7<sup>th</sup> of December, 2016, at 10 a.m.

The summary of doctoral dissertation was sent on 7<sup>th</sup> of November, 2016.

The thesis is available at the Library of Vilnius University and at website of Vilnius University: <http://www.vu.lt/naujienos/ivykiu-kalendorius>

VILNIAUS UNIVERSITETAS

Gediminas Alzbutas

**PROKARIOTINIŲ DNAZĖS I HOMOLOGŲ  
PAKANTUMO DRUSKINGUMUI /  
HALOFILIŠKUMO MECHANIZMAI BEI JAUČIO  
DNAZĖS I ATSPARUMO JONINEI JĖGAI  
DIDINIMAS**

Daktaro disertacijos santrauka  
Fiziniai mokslai, biochemija (04 P)

Vilnius, 2016

Disertacija rengta 2011–2015 m. studijuojant doktorantūroje Vilniaus universiteto Biotechnologijos institute bei UAB „Thermo Fisher Scientific Baltics“ ir ginama eksternu.

Mokslinis konsultantas:

dr. Arūnas Lagunavičius (UAB „Thermo Fisher Scientific Baltics“ (fiziniai mokslai, biochemija – 04 P).

**Disertacija ginama Vilniaus universiteto Biochemijos mokslo krypties taryboje:**

**Pirmininkas:**

dr. **Vaida Šeputienė** (UAB „Thermo Fisher Scientific Baltics“, fiziniai mokslai, biochemija – 04 P).

**Nariai:**

dr. **Artūras Jakubauskas** (Vilnius universiteto ligoninės Santariškių klinikos, fiziniai mokslai, biochemija – 04 P),

dr. **Gražvydas Lukinavičius** (Makso Planko biofizikinės chemijos institutas, fiziniai mokslai, biochemija – 04 P),

prof. dr. **Elena Servienė** (Gamtos tyrimų centras, biomedicinos mokslai, biologija – 01 B),

dr. **Giedrius Vilkaitis** (Vilniaus universitetas, fiziniai mokslai, biochemija – 04 P).

Disertacija bus ginama viešame Biochemijos mokslo krypties posėdyje 2016 gruodžio 7 d. 10 val. Vilniaus universiteto Jungtinio gyvybės mokslų centro R401 salėje.

Adresas: Saulėtekio al. 7, LT-10222, Vilnius, Lietuva.

Disertacijos santrauka išsiuntinėta 2016 lapkričio 7 d.

Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekoje ir VU svetainėje adresu: <http://www.vu.lt/naujienos/ivykiu-kalendorius>

# Contents

Introduction	3
Materials and Methods	6
Results and Discussion	12
Conclusions	32
List of publications	33
Curriculum Vitae	35
Acknowledgements	36
Reziümé	37
Bibliography	38

# Introduction

Eukaryotic DNaseI is commonly used to clear DNA contamination from RNA samples; however, it is very salt-sensitive. Over the course of evolution the DNases of halotolerant / halophilic organisms have adapted to work at high salt concentration. Several nucleases resistant to ionic strength were discovered in the past decades [1–4]. However, these studies didn't reveal any mechanism of halotolerance.

In 1998 Pan and Lazarus published their attempts to design an eukaryotic DNaseI, which would retain its activity at elevated ionic strength [5]. Since high ionic strength hinders the interaction between the enzyme and the DNA, the authors tackled this problem by introducing additional positive residues onto the DNA-binding surface of the nuclease catalytic domain [5].

Meanwhile, the analysis of the sequence data from halotolerant / halophilic prokaryotes indicated that evolution used a completely different approach: the interaction between the enzyme and the substrate was stabilized by an additional C-terminal DNA-binding domain within the enzyme. In this dissertation it is revealed that many DNases from halotolerant / halophilic species are multi domain proteins. This fact led to the hypothesis that in some cases a fusion of an additional domain to the DNase domain was the key factor in evolution, which enabled the activity of bacterial DNases at high ionic strength. In this study the hypothesis was experimentally proved by analysing halotolerance of one DNase from *Thioalkalivibrio sp. K90mix* (DNaseTA) and its mutants.

DNaseTA is comprised of two domains: one domain is DNaseI-like and the other is a DNA-binding domain comprising two HhH (helix-hairpin-helix) motifs. It was decided to mimic the evolutionary step that created the natural fusion. The research revealed that this domain originated from ComEA (competence DNA receptor) proteins and through the course of evolution was fused with the DNaseI-like domain. In this study the domain organization of DNaseTA was mimicked by creating two fusion proteins comprising bovine DNaseI and a DNA-binding domain. The following DNA-binding (HhH)<sub>2</sub> (comprising two tandem HhH motifs) domains were fused to the C-terminus of bovine DNaseI:

1. The C-terminal domain of DNaseTA, hereafter referred to as DT, from an extremely halotolerant bacterium *Thioalkalivibrio sp. K90mix*. This fusion hereinafter is abbreviated as DNaseDT.
2. The C-terminal domain of a relatively well characterized competence protein ComEA, hereafter referred to as BS, from *Bacillus subtilis* [6]. This fusion hereinafter is abbreviated as DNaseBS.

Both fusions with additional DNA binding domains were demonstrated to be more salt tolerant than bovine DNaseI. Literature analysis revealed that similar

approach to enhance DNA-binding had been used for several DNA polymerases: Phage phi 29 DNA polymerase [7], Taq and Pfu DNA polymerases [8,9]. However in this dissertation it was demonstrated for the first time that an additional DNA binding domain could be used to enhance properties of eukaryotic DNaseI.

## **Goal of the dissertation**

To analyze mechanisms of adaptation to high ionic strength of prokaryotic DNaseI-like nucleases and to prove that the revealed mechanisms can be successfully applied to induce salt tolerance for a non halotolerant / halophilic enzyme.

## **Specific tasks of the dissertation**

1. Create phylogenetic tree of prokaryotic DNaseI-like proteins.
2. Analyze potential relation between domain structure of prokaryotic DNaseI-like nucleases and adaptation to high ionic strength.
3. Experimentally test relation between domain structure and resistance to ionic strength.
4. Computationally compare electrostatic surface properties of bovine DNase I and several its homologs from halotolerant / halophilic prokaryotes and to evaluate conservation level of surface residues for one of the homologs.
5. Track evolutionary origin of the DNA-binding domain comprising two HhH motifs that are found in DNaseI-like bacterial nucleases.
6. Choose domains to be fused with bovine DNaseI and construct corresponding chimeric nucleases.
7. Experimentally analyze salt tolerance of the constructed chimeric nucleases.
8. Computationally analyze properties and interaction with DNA of the domains, which were fused to bovine DNaseI.

## **Scientific novelty and practical value**

- It was revealed that domain structure of prokaryotic DNaseI-like nucleases played important role in adaptation to high salt environments.
- Analysis of DNaseI-like prokaryotic proteins revealed that the C-terminal domain comprising two duplicate HhH motifs was found only in DNaseI-like nucleases from halotolerant / halophilic bacteria.
- For the first time it was demonstrated that it was possible to enhance DNaseI properties by creating fusions with DNA binding domains (patent pending).

- The natural evolutionary step that fused DNA-binding domain and DNaseI-like domain resulting in DNaseTA-like nucleases from halotolerant / halophilic bacteria was successfully mimicked *in vitro*.

## **The approbation of the results of the thesis**

The original results of the dissertation were represented in 2 manuscripts published in journals listed by Thompson Reuters ISI with citation index. One patent application was submitted. The results were personally presented in two poster presentations at international conferences.

# Materials and Methods

## Analysis of microbial DNaseI family proteins and the resistance of corresponding micro-organisms to salt

Initially, IPR016202 protein family sequences were collected from the InterPro database (accessed in June, 2014) [10]. The sequences were matched to UniRef90 clusters [11] and subsequent analysis was performed on representative sequences from these clusters. All non-prokaryotic sequences were discarded. Remaining sequences were subjected to phylogenetic analysis and domain detection. A maximal molar NaCl concentration allowing growth of a corresponding organism was inferred for each analysed sequence. In some cases, the maximal NaCl concentration value was found in the published data, in other cases, the salt tolerance was inferred based on a living environment or a cultivation medium. Six arbitrary selected salt tolerance categories were used. The first category, where the maximum salinity is indicated as "close to 0" encompasses the organisms that were not considered in the literature as being salt tolerant or their living environment/growth medium does not imply salt tolerance. The second one (" $< 0.8$ ") encompasses slightly halotolerant / halophilic species. An organism was assigned to this category if the corresponding concentration of NaCl was explicitly indicated in the literature or the microorganism was collected from marine habitats. The assignments to the other four categories ("0.8-0.9", "1.0-1.4", "1.5-2.0", "3.4-5.1") corresponding to medium-extreme halotolerant / halophilic species were based on explicit statements in the literature.

Domains in the sequences of bacterial DNases were detected using InterProScan 5.4-47.0 [12]. Phylogeny analysis of the corresponding sequences was performed using Phyrn-1.7.2 package [13]. 5000 replicates of the distance matrices were generated for bootstrapping. The corresponding neighbour-joining trees were calculated and a consensus tree was produced using the NEIGHBOR and CONSENS programs from the PHYLIP package [14]. ETE 2.2 package was used for the visualization of the tree and supplementary information [15].

A secretion signal search was performed in the sequences of DNases from organisms that can grow in 1.5 M or higher NaCl concentrations. A secretion signal was detected using three programs: 1) SignalIP 4.1 [16] was used for gram-negative bacteria sequences; 2) PRED-SIGNAL [17] was used for an archaeal sequence (*Methanohalobium evestigatum*); 3) Philius [18] was used for the prediction of the presence of a secretion signal and protein type.

## Cloning, expression and purification of proteins

### Cloning of DNase from *Thioalkalivibrio sp. K90mix* and its mutants

Gene of the recombinant DNase from *Thioalkalivibrio sp. K90mix* (DNaseTA) was de novo synthesized by DNA 2.0 (California, USA) and codons were optimized for expression in *E.coli*. Protein sequence was taken from an Uniprot entry (accession code D3SGB1), the secretion signal sequence was excluded. The gene was cloned into a pLATE31 vector (Thermo Fisher Scientific, #K1261). The coding sequence of the C-terminal His<sub>6</sub>-tag originated from the vector.

Two mutants of the his-tagged DNaseTA were constructed: an active site mutant with the inactivating mutation (H134A [19]), denoted as DNaseTA H134A and a mutant with the removed C-terminal domain denoted as DNaseTA  $\Delta$ C. The coding sequence of the C-terminal His<sub>6</sub>-tag originated from the vector and was added during cloning into a pLATE31 vector (Thermo Fisher Scientific, #K1261). DNaseTA H134A was generated by two-steps megaprimer PCR. Both PCR reactions were performed with 2x Phusion High Fidelity PCR Master Mix (Thermo Fisher Scientific, #F-548S). DNaseTA  $\Delta$ C mutant was constructed via single PCR step. The PCR products were cloned into a pLATE31 vector (Thermo Fisher Scientific, #K1261).

### Cloning of bovine DNaseI and its mutants

Coding DNA of mature, without signal peptide, wild type bovine DNaseI (NCBI reference sequence NM\_174534.2) was cloned into a pLATE51 vector using aLICator™ LIC Cloning and Expression system (Thermo Fisher Scientific, #K1271). The sequence of the N-terminal His<sub>6</sub>-tag and associated linker was added during cloning and originated from the vector.

Coding sequence fragments of DNaseDT and DNaseBS for cloning were generated by two-step megaprimer PCR. All PCR reactions were performed with 2x Phusion™ Flash High-Fidelity PCR Master Mix (Thermo Fisher Scientific, #F-548S). The PCR products were cloned into a pLATE51 vector using aLICator™ LIC Cloning and Expression system (Thermo Fisher Scientific, #K1271). DNase from *Thioalkalivibrio sp. K90mix* cloned to a vector pLATE31 [20] was used as a template for the first PCR reaction when DNaseDT was constructed, while *Bacillus subtilis subsp. subtilis str. 168* (ATCC 23857) genomic DNA was used in the corresponding reaction for DNaseBS. The resulting fragments were gel-purified and together with the third primer were used in the second PCR reactions where bovine DNaseI cloned to a vector pLATE51 was used as a template. The resulting fragments were gel-purified and cloned into a pLATE51 vector. The sequences of the N-terminal His<sub>6</sub>-tag and associated linker were added during cloning and originated from the vector.

## Purification and expression of proteins

The constructs were cloned using *E. coli* ER2267 strain (New England Biolabs). The bacteria were grown on LB agar supplemented with 2 % (w/v) glucose and carbenicillin. For the protein expression the constructed recombinant plasmids were transformed into *E. coli* ER2566 strain (New England Biolabs). Bacteria were grown in LB broth supplemented with glucose 1 % (w/v) and carbenicillin (50 µg/ml). A preculture was grown till OD<sub>600</sub> 0.3 at 37° C and then used for inoculation. A main culture was inoculated with 1/40 of the preculture and grown at 37° C until the induction. The expression was induced by addition of IPTG (1 mM) when OD<sub>600</sub> reached 0.8-0.9 and subsequently grown at 23° C for 16 h. Before the induction the culture was cooled on ice. Bacteria were lysed chemically and expressed proteins were purified in one step using nickel affinity spin columns (Thermo Fisher Scientific, #88225). The washing buffer had following composition: 2 % Triton X-100; 20 mM Tris-HCl, pH 7.5; 0.5 M NaCl; 5 mM CaCl<sub>2</sub>; 20 mM imidazole, pH 7.8. The elution buffer had following composition: 2 % Triton X-100; 20 mM Tris - HCl, pH 7.5; 0.5 M NaCl; 5 mM CaCl<sub>2</sub>; 250 mM imidazole, pH 7.8. The eluates were dialysed against buffer having following composition: 50 mM Tris-acetate, pH 7.5; 10 mM CaCl<sub>2</sub>; 50 % glycerol. The concentration of the purified enzymes was assessed by SDS-PAGE and subsequent densitometry.

## Phylogeny of C-terminal (HhH)<sub>2</sub> domain in DNaseI-like nucleases

### Selection of sequences

The goal of the analysis was to track the origin of the (HhH)<sub>2</sub> domains that are detected in bacterial DNaseI-like proteins [20]. Initially, a set of sequences of bacterial proteins that harbour a (HhH)<sub>2</sub> domain (H3TH\_StructSpec-5'-nucleases superfamily, CDD v3.12 [21]) and an exonuclease/endonuclease/phosphatase domain (EEP superfamily, CDD v3.12 [21]) were collected. The sequence fragments that correspond to a (HhH)<sub>2</sub> domain were used to construct a hidden Markov model (HMM) with HMMER v3.1b1 package [22] and a search was run against UniProtKB [11] via HMM webserver [23]. The bacterial sequences in UniProtKB [11] that matched the HMM with e-value lower than 10<sup>-13</sup> were clustered at 70 % identity using CD-HIT [24, 25]. The clustering cut off and selection stringency were manually adjusted to keep up to ~100 of final clusters. A protein from Chinese hamster *Cricetulus griseus* (UniProt AC G3HTJ3) was added to the analysis as an out-group. This eukaryotic protein harbours DNaseI-like domain and two tandem (HhH)<sub>2</sub> domains. The sequence of the domain that best matched the HMM was used for the construction of a phylogenetic tree.

## Tree building

Only the fragments in the collected sequences that matched the HMM were used to construct the phylogenetic tree. The sequence fragments were aligned using PROMALS3D server [26,27]. The tree was build using RAxML 8.1.22 program [28]. Automatic protein model selection using maximum likelihood (ML) score based criterion was used. At first, a Bootstrap search using extended majority-rule consensus tree criterion was conducted. After the Bootstrap analysis a search for the best-scoring ML tree was done. This tree was used for further analysis.

## Tree annotation

CDD v3.12 [21] (server accessed in December, 2014) was used to detect superfamily and multidomain hits in the full length sequences, which regions matching (HhH)<sub>2</sub> domain were used to construct the phylogenetic tree. ETE 2.2 package was used for the visualization of the tree, the detected domain search hits and supplementary information [15]. The root was placed according to the out-group mammalian (*Cricetulus griseus*) protein.

## Molecular modelling of DNA binding domains

### Homology modelling

Spatial structures of BS (from *Bacillus subtilis*) and DT (from *Thioalkalivibrio sp. K90mix*) were modelled using I-Tasser 3.0 [29] and the best models (in terms of C-score) where used for further analysis. The low quality N-terminal unstructured fragments (inter domain regions) were discarded from the models based on expected accuracy along the sequence (I-Tasser 3.0 output). The N-terminal and C-terminal regions were further trimmed to remove unstructured terminal coils that do not match in structural superimposition of the two domains. Structural superimposition, manual inspection and trimming of the models was done using Pymol [30].

### Protein-DNA complex modelling

The resulting domain structures were superimposed (using MultiProt 1.6 [31]) with homologous domain from a crystallographic structure PDB ID: 3E0D [32] (one of templates used in homology modelling) and a 19 bp length DNA fragment close to the superimposed domains was „excised“ from the PDB structure for further modelling. The geometry of this fragment was optimised using 3D-DART server (accessed in December, 2014) [33]. Then it was docked to the two DNA-binding domains using HADDOCK server (accessed in December, 2014) with „Refinement“ interface [34]. The resulting complexes were used for further electrostatic calculations.

## Electrostatic calculations

The two modelled DNA-protein complexes were superimposed using Pymol [30] and used for electrostatic calculations using APBS tools 1.4.0 [35]. Results were visualized in Pymol [30]. The structures were prepared for electrostatics calculations by adding hydrogens with PDB2PQR 2.0.0 [36] using Amber force field [37]. APBS tools were used to predict changes in polar solvation term of free energy that occur when protein-DNA complex is formed. The grid parameters for the calculations were same for both complexes. Nonlinear Poisson-Boltzmann equation was used, all other parameters were default. The calculations were done in accordance with the published analogous calculations on RNA-protein interactions [38]. APBS output files (density maps) were used to infer changes in local ion and charge concentrations upon formation of complexes between DNA and the two DNA-binding domains. This was done using auxiliary tools from the APBS package.

## Molecular modelling of DNase domains

The protein sequence of DNaseTA was downloaded from the corresponding entry in UniproKB database [39] (<http://www.uniprot.org/uniprot/D3SGB1>). HH-blits [40] from HH-suite 2.0 was used to identify the protein domains in DNaseTA. The modelling of the DNase domain was performed using I-Tasser server [29]. The models were refined with Kobamin [41], a knowledge-based potential refinement program. The best quality model was selected using Prosa-web [42]. Prosa-web was used for backbone and Qmean [43] for side chain quality assessment. The electrostatic surface potential of DNase domain of DNaseTA was calculated with APBS tools [35] and visualized in Pymol [30]. The structure was prepared for electrostatics calculations by adding hydrogens with PDB2PQR [36] using Amber force field [37]. The range from -5 kT/e in red to +5 kT/e in blue was chosen for surface colouring. The model for a DNase from *Methanohalobium evestigatum* was created and analysed in a similar manner. For evaluation of the surface residues conservation level proteins homologous to DNaseTA were identified using PSI-Blast [44] and Jackhammer [45], profile-profile alignment search tools. Homologous proteins were aligned using multiple alignment program MAFFT [46] with L-INS-i [47] option. Lastly, multiple-aligned sequences were imported into a ConSurf [48] program for mapping of conserved amino acids. In order to indicate potential ion and DNA binding residues in DNaseTA, the DNase domain of DNaseTA, DNaseI structures PDB ID: 4AWN, 2A3Z, 3DNI (known positions of ions) and PDB ID: 1DNK (known positions of DNA) were superimposed. The superimposition was performed using Dali Server [49] and mapped in Jalview Waterhouse et al. [49].

## Activity assays

### Digestion of long DNA substrate

2 µg pUC19 DNA cleaved with SmaI was digested with 2.5 nM of enzyme. Two ranges of NaCl concentration were explored: 0 - 1 M in 0.1 M increments and 0 - 4 M in 0.4 M increments. The reactions were performed for 10 min at 37° C in 100 µl of the reaction mixture with 10 mM Tris-HCl, pH 7.5 and varying amount of CaCl<sub>2</sub> and MgCl<sub>2</sub>. ZipRuler Express DNA Ladder 1 (#SM1373) was used as a molecular weight standard to evaluate the degradation of the DNA substrate.

### Digestion of short DNA substrate

10 nM 16 bp DNA (2 nM were labelled with <sup>33</sup>P at 5'-end) was digested with 0.66 nmol of enzyme at 37° C in 100 µl of a reaction buffer: 10 mM Tris-HCl, pH 7.5; 10 mM CaCl<sub>2</sub>; 10 mM MgCl<sub>2</sub>. 9 µl of the reaction mixtures were removed at 1, 2, 4, 8, 16, 32, 64, 128, 192 minutes after start. The samples were mixed with 9 µl of 2x RNA loading dye (Thermo Fisher Scientific, #R0641), heated for 5 min at 95° C and analysed by denaturing PAGE. The half-life of the substrate digestion was estimated during subsequent densitometry.

### Digestion of fluorescently labeled short DNA substrate

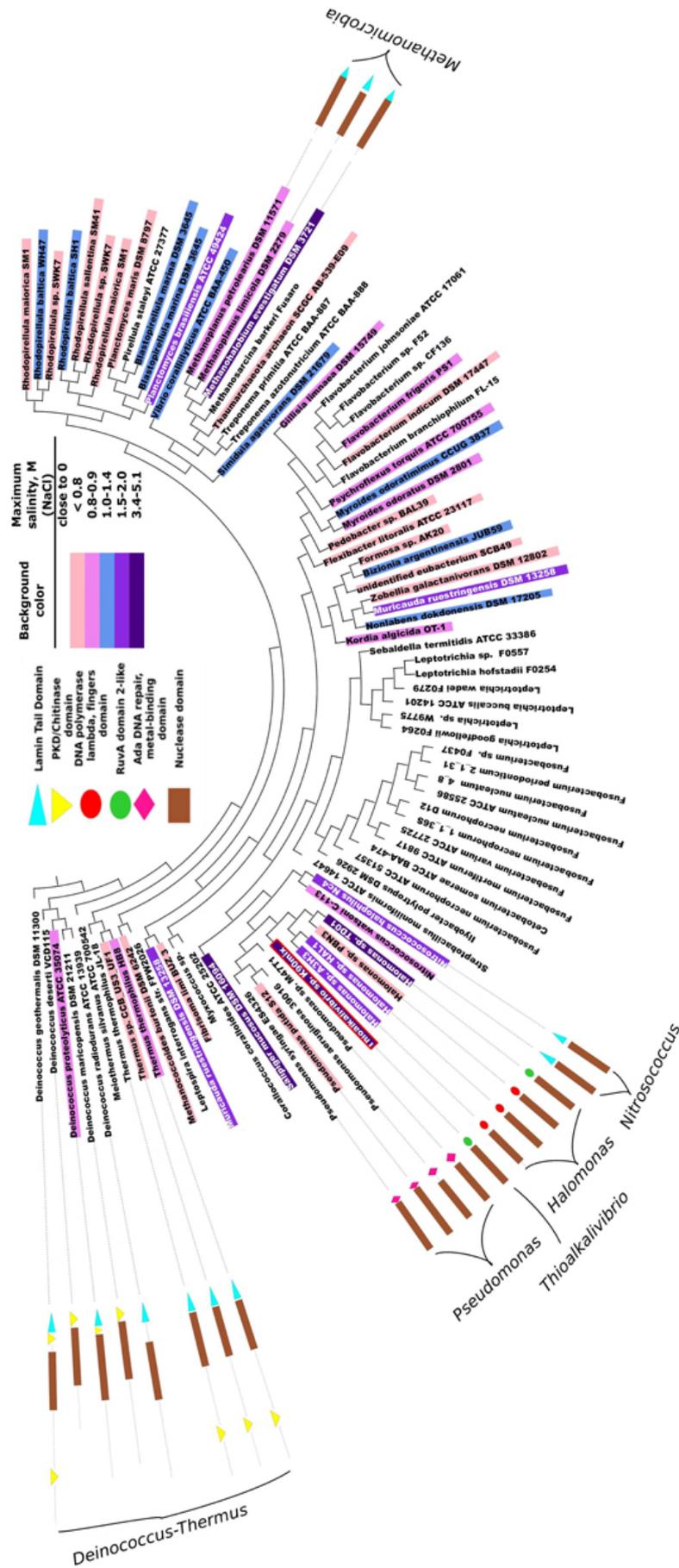
Activity of DNaseI and its fusion variants at relatively low salt concentrations was evaluated by analysing digestion of fluorescently labelled DNA duplex (30 bp). Reactions were prepared in the following buffer: 10 mM Tris – HCl, 3 mM EDTA, 1 % Triton X-100, 1 mg/ml BSA. 0.2 µM of DNA duplex and 0.044 nM concentrations of relevant DNase enzyme (DNaseI, DNaseBS or DNaseDT) were used. Reactions were started by addition of 10X start solution containing 40 mM CaCl<sub>2</sub> and 100 mM Mg acetate. Before start of fluorescence monitoring 25 µl start solution was added to 225 µl of reaction mix with enzyme. The fluorescence was monitored in 12 seconds intervals for 10 minutes and for each curve a maximum fluorescence change rate was calculated (Gen5™ Reader Control and Data Analysis Software), which was proportional to enzymatic activity. Fluorescence was monitored and start solution was distributed across reaction mixes by Synergy 2 Multi-Mode Reader (BioTek). Reaction mixes were prepared and fluorescence measures were taken at ambient temperature (~23° C). Such experiments were performed by varying amounts of NaCl in the final reaction buffer (0, 50, 100 mM) in order to estimate activity changes due to increase in ionic strength.

# Results and Discussion

## Results

### **DNaseI family sequences properties and halotolerance of corresponding micro-organisms**

The purpose of phylogenetic and domain structure analysis of bacterial DNases was to identify the domains, which would be potentially related to halotolerance or halophilicity. The analysis of the InterPro 4.7 database [10] revealed that there are about 300 prokaryotic proteins belonging to DNaseI family. Clustering these sequences at 90% sequence identity level resulted in 86 clusters (Uniref90). Manual inspection of available data in literature revealed that more than a half of the representative sequences originated from halotolerant / halophilic organisms. The summary of this data is given in Fig. 1. In this figure species corresponding to 86 representative sequences are indicated along with the inferred maximum salinities (concentration of NaCl) at which microbial growth occurs. In total data on 82 prokaryotic organisms have been collected. 39 of them were considered to be of low salt tolerance, 16 were classified as being slightly halotolerant / halophilic ( $\geq 0.3$  and  $< 0.8$  M NaCl), 23 were classified as being medium halotolerant / halophilic ( $\geq 0.8$  and  $< 3.4$  M NaCl) and 4 species were classified as extremely halotolerant / halophilic ( $\geq 3.4$  M NaCl).



**Figure 1. Phylogeny analysis of bacterial proteins of DNaseI family and their domain organizations.** Each leaf corresponds to one UniRef90 cluster (group of sequences with at least 90% pairwise identity). The domain structure is presented only if the representative protein of the cluster has other domains in addition to the nuclease domain. Background colours indicate an inferred maximum concentration of NaCl at which the corresponding microorganism still grows. The more blue background – the more salt tolerant the microorganism is. DNA polymerase lambda fingers domain and RuvA domain 2-like domain are similar (red and green circles) as both correspond to duplicate HhH (helix-hairpin-helix) motifs. Red border indicates a DNase from *Thioalkalivibrio* sp. *K90mix*, which was selected for experimental analysis.

**Table 1. The secretion signal in DNases from the most salt tolerant organisms**

Representative sequence	Representative species/label used in phylogeny tree	Secretion signal		Protein type prediction by Philius			Additional domain	Maximum NaCl, M
		SignalIP 4.1 <sup>1</sup>	PRED-SIGNAL <sup>*1</sup>	Type	Confidence			
D7E828	<i>Methanohalobium evestigatum</i> DSM 3721	-	Y	Globular with Signal Peptide	0.99	Lamin tail	5.1 <sup>1</sup>	
D3SGB1	<i>Thioalkalivibrio sp. K90mix</i>	Y	-	Globular with Signal Peptide	0.99	duplicate HhH <sup>*2</sup>	4 <sup>2</sup>	
F7SPZ3	<i>Halomonas sp. TD01</i>	Y	-	Globular with Signal Peptide	0.98	duplicate HhH <sup>*2</sup>	3.42 <sup>3</sup>	
S9QWK2	<i>Salipiger mucosus</i> DSM 16094	N	-	Globular with Signal Peptide	0.71	-	3.42 <sup>4</sup>	
G4F511	<i>Halomonas sp. HAL1</i>	Y	-	Globular with Signal Peptide	0.99	duplicate HhH <sup>*2</sup>	2 <sup>5</sup>	
F0SFA3	<i>Planctomyces brasiliensis</i> ATCC 49424	N	-	Transmembrane	0.68	-	1.72 <sup>6</sup>	
DSKCF8	<i>Nitrosococcus halophilus</i> Nc4	Y	-	Globular with Signal Peptide	0.99	Lamin tail	1.60 <sup>7</sup>	
G2PLU4	<i>Muricauda ruestringensis</i> DSM 13258	N	-	Globular	0.99	-	1.54 <sup>8</sup>	
G2PQ99	<i>Muricauda ruestringensis</i> DSM 13258	N	-	Globular with Signal Peptide	0.99	-	1.54 <sup>8</sup>	
T2LFY5	<i>Halomonas sp. A3H3</i>	N	-	Globular with Signal Peptide	0.91	duplicate HhH <sup>*2</sup>	1.5 <sup>9</sup>	

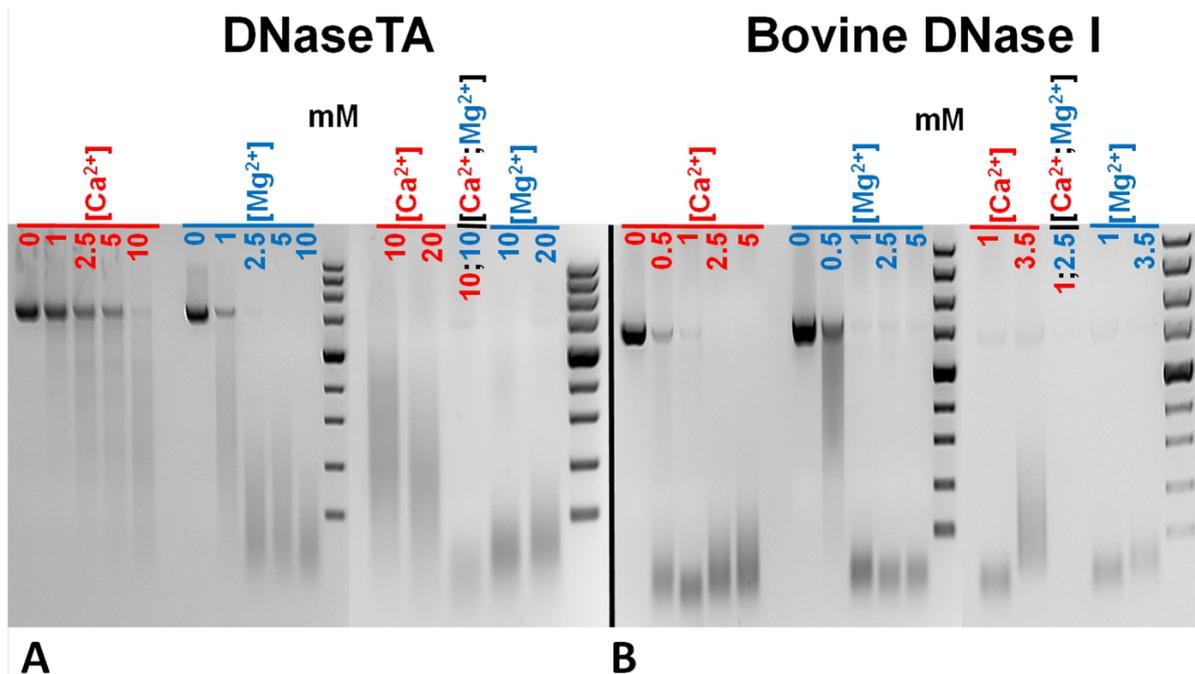
<sup>1</sup> [52] <sup>2</sup> [53] <sup>3</sup> [54] <sup>4</sup> [55] <sup>5</sup> [56] <sup>6</sup> [57] <sup>7</sup> [58] <sup>8</sup> [59] <sup>9</sup> [60] <sup>\*1</sup> resulting decision on the existence (Y) or not existence (N) of the secretion signal sequence. <sup>\*2</sup> corresponds to RuvA domain 2-like domain or DNA polymerase lambda fingers domain

The phylogeny of the sequence fragments, which correspond to the nuclease domain of the proteins, was analysed and the topography of the resulting phylogeny tree is given in Fig. 1. All the domain names used in the figure are the same as used in InterPro 4.7 database, except for the nuclease domain. Exo/endo/phospho and DNase domain was named as „nuclease“ domain (note that all analysed proteins belong to DNaseI family). This analysis indicates that bacterial DNases apart from their nuclease domain might have 5 types of other domains: Lamin tail, PKD/Chitinase, Ada DNA repair/metal binding, RuvA domain2-like and DNA polymerase lambda fingers. The last two domains: the RuvA domain 2-like [50] and the DNA polymerase lambda fingers [51] are similar (shown as red and green circles in Fig. 1) as both of them correspond to the duplicate HhH (helix-hairpin-helix) motifs [10]. DNases of *Deinococcus-Thermus* phylum stand out as a distinctive group: the proteins are longer compared to other bacterial DNases and contain significant sequence fragments that cannot be assigned to any functional domain. These proteins have either PKD/Chitinase or Lamin tail domain, or both.

If a protein is secreted, then it should be adapted to the surrounding environment. Thus additional search for secretion signal was performed on the sequences originating from the organisms which can grow at concentration of NaCl  $\geq 1.5$  M. The results are given in Table 1.

## DNaseTA requirement of divalent ions and resistance to ionic strength

Approximate optimal concentrations of divalent cations for the bovine DNaseI (a reference enzyme) and the DNase from *Thioalkalivibrio sp. K90mix* (DNaseTA) have been estimated. It was observed (Fig. 2) that both enzymes required Ca<sup>2+</sup> or Mg<sup>2+</sup> for catalytic activity and the maximum activity was achieved when both of these ion species were present. Data shows that DNaseTA requires significantly higher concentrations of divalent ions than bovine DNaseI:  $\sim 10$  mM Ca<sup>2+</sup> and  $\sim 10$  mM Mg<sup>2+</sup> is the optimal combination for the DNaseTA (Fig. 2 A), while for the bovine DNaseI the respective concentrations are  $\sim 1$  mM and  $\sim 2.5$  mM (Fig. 2 B).

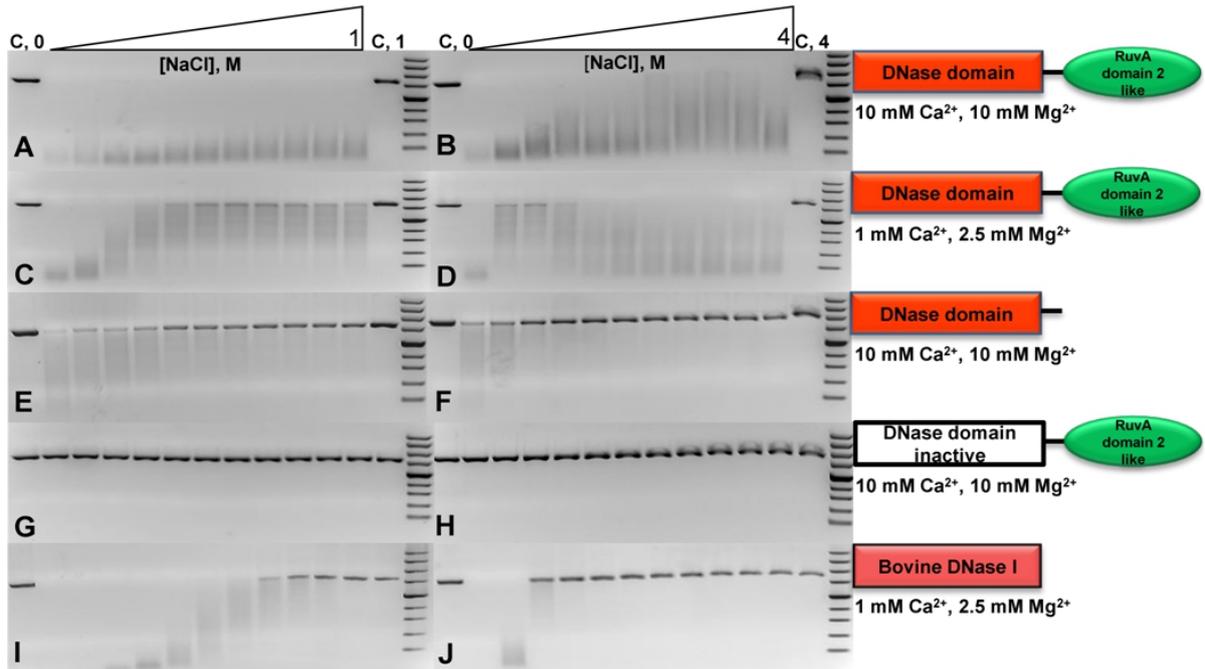


**Figure 2.** Different requirements for divalent ions of DNase from *Thioalkalivibrio sp. K90mix* (DNaseTA) and bovine DNaseI. Several different concentrations of  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  were screened evaluating approximate optimum for DNaseTA (A) and bovine DNaseI (B). It was found that both enzymes require  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  for activity, however, DNaseTA requires significantly higher concentrations. The optimal concentrations are  $\sim 10$  mM  $\text{Ca}^{2+}$  and  $\sim 10$  mM  $\text{Mg}^{2+}$  for DNaseTA and, respectively,  $\sim 1$  mM and  $\sim 2.5$  mM for bovine DNaseI.

Influence of increasing salt (NaCl) concentration on DNA hydrolysis by DNaseTA and its two mutants have been analysed. Digestion of two DNA types was assayed: i) long plasmid, ii) short duplex. This was done in order to double-check the findings as the data on the digestion of the long plasmid were qualitative agarose gel images and it was accompanied by quantitative data on the short substrate digestion.

The digestion of the long substrate by the DNaseTA was assayed under two conditions in terms of divalent cations: 1) the near optimum combination of  $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$  concentrations for DNaseTA (Fig. 3 A, B). 2) the lower concentrations of these divalent ions, which resembled near optimum conditions for the bovine DNaseI (Fig. 3 C, D). The results indicate that DNaseTA digests DNA in the presence of high salt concentration (up to 4 M NaCl) at both concentrations of divalent ions, however, apparent differences between those two combinations are noticeable at lower ionic strength (up to  $\sim 1$  M NaCl). DNaseTA digests DNA at the lower ionic strength and higher concentration of divalent ions, although the length of final product gradually increase (presumably due to decreasing DNase activity). Contrastingly, under lower concentration of divalent ions the ability of DNaseTA to digest DNA decreases significantly when salt concentration reaches  $\sim 0.4$  M NaCl. At this point some substrate remains even undigested. Further increase in ionic strength, however, reinforces the degradation of DNA

by DNaseTA as the length of the DNA substrate is shortened significantly even at 4M NaCl.



**Figure 3. Activity of DNase from *Thioalkalivibrio sp. K90mix* (DNaseTA), its mutants and bovine DNaseI at different ionic strengths.** Activity of analysed proteins was evaluated by digestion of linearised pUC19 plasmid in the presence of various concentrations of NaCl. Two series of NaCl gradients were used: the first one corresponds to a range from 0 to 1 M with increments of 0.1 M (left electrophoregrams), the second one corresponds to a range from 0 to 4 M with increments of 0.4 M (right electrophoregrams). Activity assays in case of DNaseTA were performed comparing two compositions of divalent ions: a suboptimal one, corresponding to lower (C, D) concentrations, and near optimum one, which corresponded to higher concentrations (A, B). The data on the mutant with the removed C-terminal domain (DNaseTA  $\Delta$ C) is given at the E and F panels. The data on DNaseTA with mutation at DNase active site (DNaseTA H134A) is given at the G, H panels. Analogous data for Bovine DNaseI is given at the I, J panels (the near optimum concentration of divalent cations for DNaseI were used).

The digestion of the long substrate by the mutants of DNaseTA was analysed at the near optimum combination of divalent ions for DNaseTA. The truncated form of DNaseTA with the removed C-terminal domain (DNaseTA  $\Delta$ C) retains its DNase activity as presented in Fig. 3 E and F. This mutant of DNaseTA noticeably digests the DNA substrate only up to  $\sim$ 0.6 M NaCl and the DNA digestion is completely inhibited at NaCl concentrations higher than 1.2 M. The mutant of DNaseTA, which has the inactivating mutation H134A, (DNaseTA H134A) is unable to digest DNA at any of the tested concentrations of NaCl (Fig. 3 G and H).

The data corresponding to the bovine DNaseI is given in Fig. 3 I and J panels, which indicates that the concentration of NaCl above  $\sim$ 0.9 M totally

inhibits DNA digestion. In this case near optimum combination of divalent ions for the bovine DNaseI was used.

The data on the digestion of the short DNA substrate is presented in the Table 2. In this experiment for analysis of DNaseTA and its mutants the near optimum combination of divalent ions for DNaseTA was used. For the analysis of DNaseI - the corresponding near optimum concentration of divalent cations was used. It complements the data on the long substrate digestion and indicates that DNaseTA is able to digest DNA in the presence of 4 M NaCl, the truncated version of the DNaseTA (without C-terminal domain) is active only at low ionic strength, and the DNaseTA active site mutant is inactive at any ionic strength.

**Table 2. The half-life of the radioactive substrate digestion of the recombinant DNase from *Thioalkalivibrio sp. K90mix* (DNaseTA) and its mutants.**

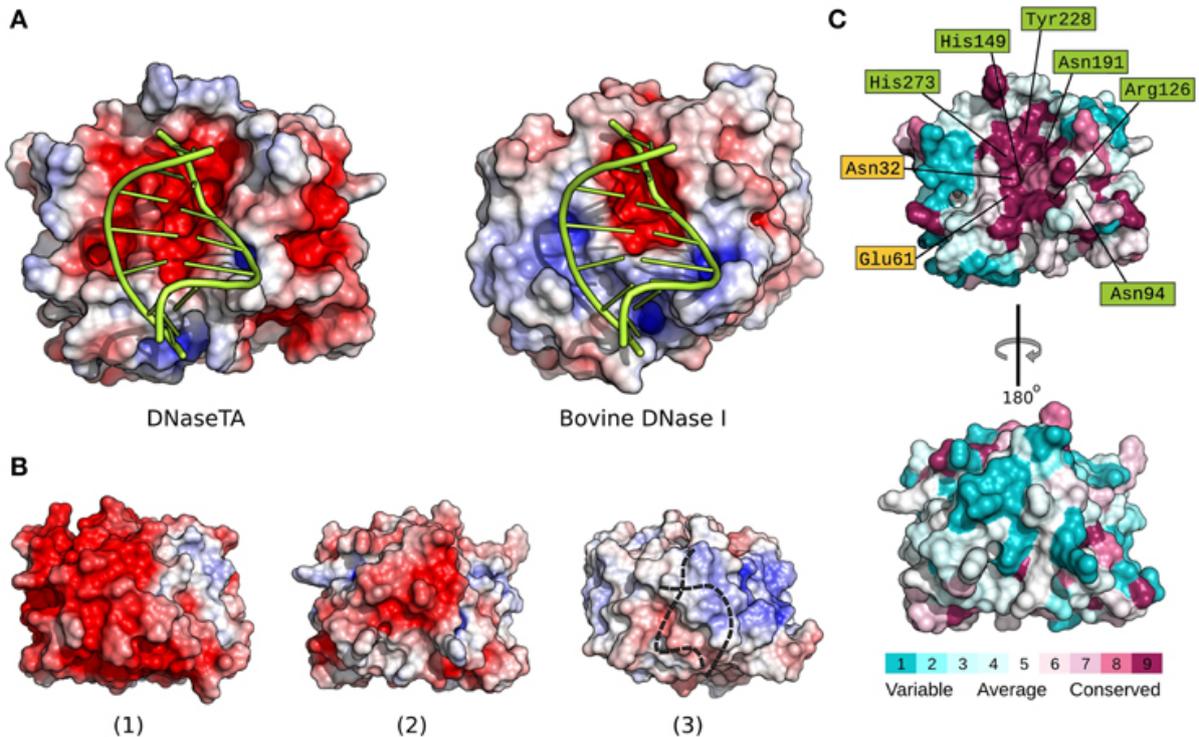
NaCl, M	DNaseTA	DNaseTA $\Delta C$	DNaseTA H134A
	Half-life of substrate digestion $T_{1/2}$ , min		
0	9.43	31.29	ND <sup>1</sup>
0.5	43.39	ND <sup>1</sup>	ND <sup>1</sup>
1.2	60.18	ND <sup>1</sup>	ND <sup>1</sup>
4.0	47.66	ND <sup>1</sup>	ND <sup>1</sup>

<sup>1</sup> ND – not detectable

## Surface properties of DNase domain of DNaseTA

In order to elucidate the apparent differences between DNaseTA and bovine DNaseI regarding the ability to digest DNA at high ionic strength, the electrostatic potential properties and conservation of surface residues of the DNase domain were analysed. The active site residues, which interact with the DNA and are involved in catalysis, are marked in green colour and the residues, which bind magnesium ions, are marked in yellow colour in Fig. 4 C. As we see in Fig. 4 an electrostatic potential (the redder - the more electronegative) of the surfaces indicates that the active site regions of DNaseTA and bovine DNaseI are both negatively charged. However, the DNaseTA has significantly larger electronegative patch in the DNA binding surface than the bovine DNaseI (Fig. 4 A). The fact that the predicted surface of DNaseTA is more electronegative compared to DNaseI is also evident when those parts of protein surfaces which are not facing DNA were compared (Fig. 4 B(2) and Fig. 4 B(3)). Therefore, the surface of DNaseTA is overall more electronegative than the surface of bovine DNaseI. For comparison, along with DNaseTA and bovine DNase in Fig. 4 B(1) the surface electrostatic potential of the DNase from *Methanohalobium evestigatum* (DNaseME) was visualized. *Methanohalobium evestigatum* is the most salt tolerant organism, which was included in the research. The maps demonstrate that the DNaseME has the most electronegative surface out of all three DNases. In Fig. 4 B a protein from the most halotolerant / halophilic microorganism is on

the left (1 – DNaseME), a protein from less halotolerant / halophilic organism is in the middle (2 – DNaseTA) and an eukaryotic protein is on the right (3 – bovine DNaseI). For the references on the salt tolerance see the Table 1. Thus the more halophilic the DNase is, the more electronegative surface it has.



**Figure 4. Electrostatic surface potential and evolutionary conserved residue maps.** Difference between DNase domain of DNase from *Thioalkalivibrio sp. K90mix* (DNaseTA) and bovine DNaseI electrostatic potential surfaces, which are in contact with DNA, are depicted in the A panel. Here DNA is positioned based on crystallographic bovine DNaseI structure (PDB ID: 1DNK). Electrostatic surface potential maps of DNase domains' sides which are not facing DNA are depicted in the B panel: DNase from *Methanohalobium evestigatum* (DNaseME) (1), DNaseTA (2), and bovine DNaseI (3). Dashed line (3) represents DNA position behind DNaseI. Evolutionary conserved residue map of DNase domain model of DNaseTA is depicted in the C panel. The front (top structure) that faces DNA, and the back (bottom structure) sides of the proteins are visualized. Scales from variable (cyan) to conserved (purple) are shown under the structures.

Alongside with the electrostatic potential the conservation of the DNaseTA surface residues was also mapped (Fig. 4 C). The results show that the positions of the active site and DNA binding residues Asn94, Arg126, His149, Asn191, Tyr228, His273 along with the  $Mg^{2+}$  binding residues Asn32, Glu61 are conserved, while many other surface residues are variable. It is evident that the surface side, which is not facing DNA, is variable and does not show any strong evolutionary conservation (see Fig. 4 C bottom surface). Therefore, the surface residues in the homologs of DNaseTA are mostly variable and thus give room for evolutionary adaptations.

## Origin of DNaseTA C-terminal domain

The origin of the DNaseTA C-terminal domain, which contains two tandem HhH motifs, was analysed. The topography of the resulting phylogenetic tree is given in Fig. 5. All the proteins included in the analysis harbour a two tandem HhH motifs comprising domain, which belongs to a superfamily of H3TH domains of structure-specific 5' nucleases [21] and hereinafter is referenced as a H3TH domain. At the base of the tree there is a group of proteins that harbour multiple H3TH domains from *Thermotogae* phylum, which match multi-domain hits ComEA [21]. This group is denoted as „A“ in Fig. 5. All other proteins emerge from the sister group. This group further splits into two distinct sister groups („B“ and „C“ in Fig. 5). The „B“ group (Fig. 5) contains two experimentally characterized competence proteins: i) ComE from *Neisseria gonorrhoeae* [61], which comprises one H3TH domain; and ii) ComEA from *Bacillus subtilis* [6], which comprises one SLBB (soluble ligand-binding beta-grasp domain superfamily) domain and one H3TH domain. In addition to proteins that have such domain organization, the „B“ group contains plethora of experimentally not characterized proteins, which harbour other domains in addition to a H3TH domain. ComE-like proteins from *Thioalkalivibrio* and *Halomonas* genera that comprise single H3TH domain are localized at the base of the „C“ group (Fig. 5). Analysing proteins that are close to the lineage that leads to DNaseTA, we find proteins from *Bacillales* order that have a H3TH domain fused to a metallo- $\beta$ -lactamase superfamily domain ( $C^1$  in Fig. 5). Hereinafter such proteins are referred as lactamase- $\beta$ -H3TH proteins. Further in the lineage leading to DNaseTA we find two sister groups: i) one ( $C^3$  in Fig. 5) contains lactamase- $\beta$ -H3TH proteins from *Bacillaceae* family; ii) the other one ( $C^2$  in Fig. 5) contains DNaseTA and other proteins that harbour EEP (exonuclease/endonuclease/phosphatase superfamily) domains fused to H3TH domains. In this group ( $C^2$  in Fig. 5) apart from nucleases harbouring a H3TH domain (including DNaseTA) we find one lactamase- $\beta$ -H3TH protein. Thus, the H3TH domain of DNaseTA originated from proteins that were related to the experimentally characterized bacterial competence proteins and shared with them common precursors.

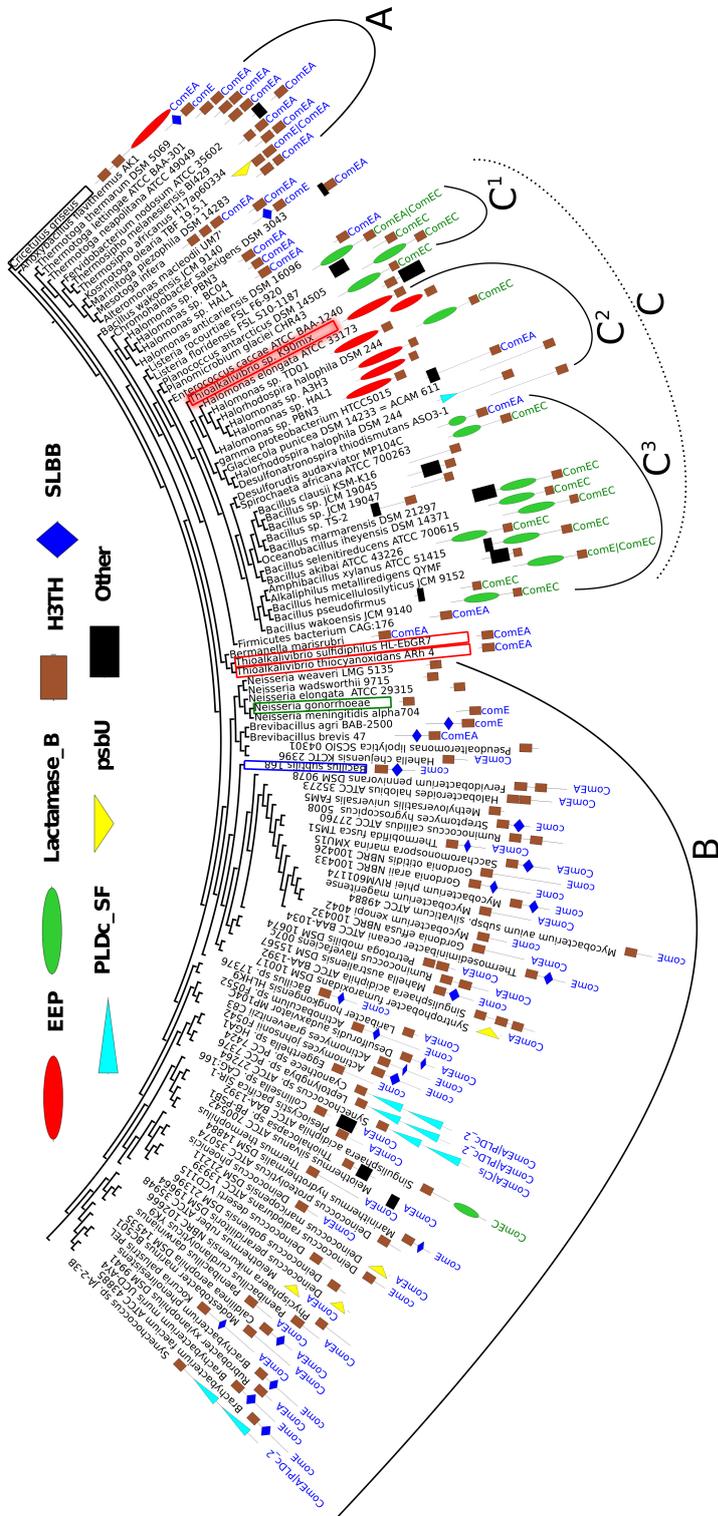


Figure 5. Phylogeny analysis of bacterial proteins that harbour domains similar to (HhH)<sub>2</sub> domains (H3TH domains) found in bacterial proteins that comprise one H3TH domain (H3TH – StructSpec-5'-nucleases superfamily, CDD v3.12 [21]) and one exonuclease/endonuclease/phosphatase domain (EEP superfamily, CDD v3.12 [21]). An experimentally characterized ComEA protein from *Bacillus subtilis* [6] is marked by blue frame. Other experimentally characterized ComE protein from *Neisseria gonorrhoeae* [61] is marked by green frame. Leaf names indicate species of the corresponding organisms. Leafs of proteins from *Thioalkalivibrio* genus are marked by red frames. The experimentally characterized DNase from *Thioalkalivibrio* sp. *K90mix* (DNaseTA) is marked by reddish background [20]. The domain architecture of the corresponding sequence is indicated by figures next to the name of the corresponding species. Next to the figures indicating domain architectures corresponding names of multi-domain CDD v3.12 hits [21] are indicated. Domain architectures are indicated by using superfamily names of CDD v3.12 hits [21], only H3TH\_StructSpec-5'-nucleases superfamily is abbreviated as H3TH. EEP denotes exonuclease/endonuclease/phosphatase superfamily, Lactamase\_B – metallo-beta-lactamase superfamily, SLBB – soluble ligand-binding beta-grasp domain superfamily, PLDc\_SF – catalytic domain of phospholipase D superfamily, psbU – photosystem II 12 kDa extrinsic protein. Hits, which occurred only once, were indicated as Other. Notations A, B, C, C<sup>1</sup>, C<sup>2</sup>, C<sup>3</sup> are used to describe the tree in the main text.

## Salt resistance of DNaseI fusions

The fusion proteins, DNaseBS and DNaseDT, comprise the sequence of the bovine DNaseI fused to the C-terminal DNA-binding domain of *Bacillus subtilis* and *Thioalkalivibrio sp. K90mix*. Only these two fusions were constructed and were found, as given below data indicated, to be functional enzymes.

The activity and salt resistance of the fusions were assessed by digestion of long (Fig. 6) and short DNA substrates (Table 3, Fig. 7). The quantitative data of radioactive substrate digestion half-life (Table 3) complement the data obtained by electrophoregrams (Fig. 6).

**Table 3. Radioactive substrate digestion half-life of wild type bovine DNaseI and its fusion variants.** DNaseDT denotes the fusion with the (HhH)<sub>2</sub> domain of DNase from *Thioalkalivibriosp. K90mix*. DNaseBS –the fusion with homologous domain of ComEA protein from *Bacillus subtilis*

NaCl, M	DNaseI	DNaseDT	DNaseBS
	Half-life of 16 bp DNA substrate digestion T <sub>1/2</sub>		
0	0.61	0.41	0.70
0.5	732.7 <sup>1</sup>	64.41	488.6 <sup>1</sup>
1.2	ND <sup>2</sup>	ND <sup>2</sup>	ND <sup>2</sup>
4.0	ND <sup>2</sup>	ND <sup>2</sup>	ND <sup>2</sup>

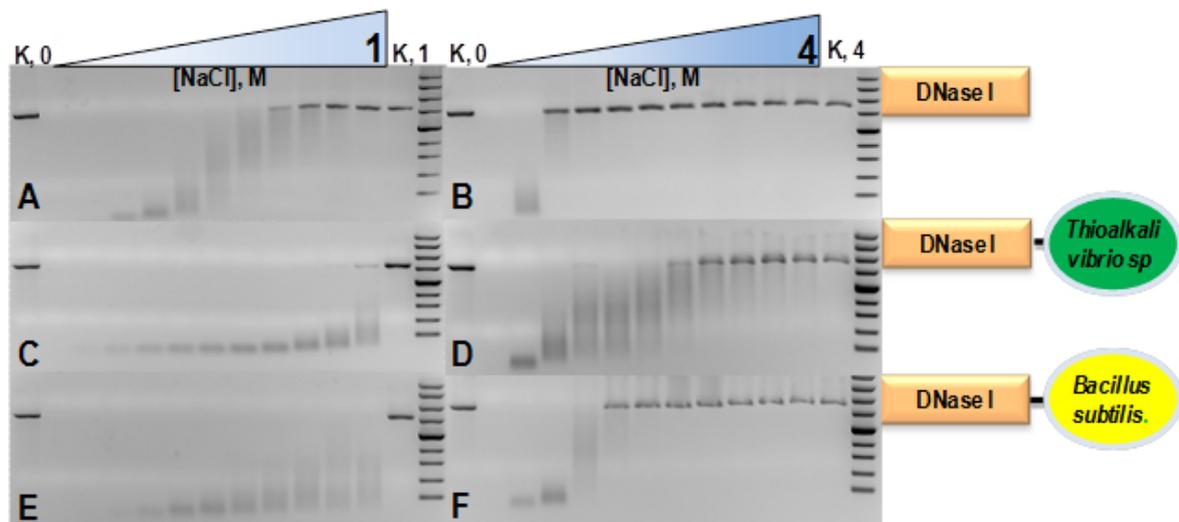
<sup>1</sup> – Half-life lasted longer than the experiment itself and was computationally inferred from collected data.

<sup>2</sup> – DNA digestion not detectable.

As shown in Fig. 6, at 1 M NaCl salt concentration the wild type DNaseI is inactive and does not degrade DNA while both fusion proteins retain activity (the left panels, Fig. 6). DNaseDT retains a detectable DNase activity even at 4 M NaCl. However, DNaseBS is less tolerant to high salt concentrations than DNaseDT as it retains activity up to ~1.6 M NaCl (the right panels, Fig. 6). The data show that at 0.5 M NaCl DNaseDT is notably more active than DNaseBS and bovine DNaseI. The data in Table 3 also reveal that DNaseDT is more active than DNaseBS and DNaseI at 0 and at 0.5 M NaCl while there is no detectable digestion of the radioactive DNA duplex at 1.2 and 4.0 M NaCl.

The data shown in Fig. 7 elucidate differences in activities of the three analysed proteins at relatively low salt concentrations (50 and 100 mM NaCl). In this figure fluorescence curves and corresponding changes in the activities in response to the increased ionic strength are presented. The data in Fig. 7 indicate that up to 100 mM NaCl the digestion of the DNA duplex by DNaseBS is less suppressed by salinity compared to DNaseDT and DNaseI.

Thus, the experimental data indicate that at relatively low salt concentrations DNaseBS is more resistant to the increased ionic strength compared to



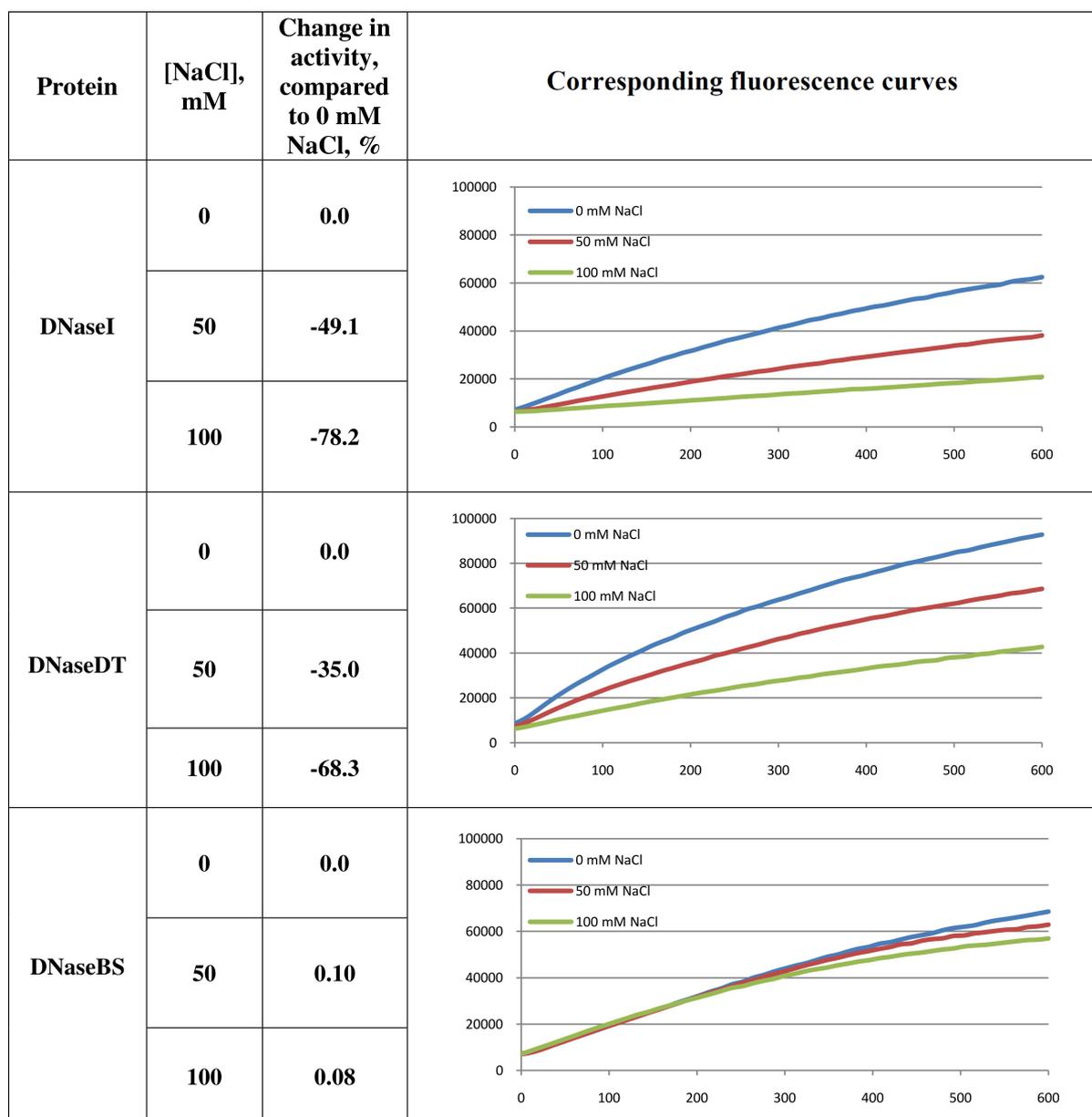
**Figure 6. Activity of bovine DNaseI and its two fusion variants at different ionic strengths.** Two series of NaCl gradients were used: the first one corresponds to a range from 0 to 1 M with increments of 0.1 M (left electrophoregrams), the second one corresponds to a range from 0 to 4 M with increments of 0.4 M (right electrophoregrams). Activity of analysed proteins was evaluated by digestion of linearised pUC19 plasmid in the presence of various concentrations of NaCl. The data on the fusion with the (HhH)<sub>2</sub> domain of DNaseI from *Thioalkalivibrio sp. K90mix* is given at the C and D panels. The data on the fusion with homologous domain of ComEA protein from *Bacillus subtilis* is given at the E, F panels. The data on bovine DNaseI are given at the A, B panels.

DNaseDT, while at higher salinity the situation is *vice versa*. Molecular modelling approach were employed to further elucidate these differences.

### Molecular modelling of the DNA-binding domains

Structural models of the two DNA-binding domains that were fused to bovine DNaseI have been created. The superimposition of the structures along with the corresponding sequences alignments are presented in Fig. 8. The sequences are quite different (28 % identity), however the modelled structures matches relatively well (RMSD of C<sub>α</sub> is 0.87 Å). However, some apparent differences in the structures are evident. The second HhH motif in BS domain starts with a longer helix compared to the DT domain. The additional residues are marked by yellow-colour in Fig. 8. In this figure a potential DNA position is indicated. It is evident that both domains have two positive DNA approaching residues: BS has two lysines, DT – two arginines. The matching electrostatic potential of the DNA contacting surfaces are visualized in Fig. 9 A (BS) and B (DT) panels. This data indicate that DT from *Thioalkalivibrio sp. K90mix* is more compact and has smaller continuous electropositive patch on the surface than BS from *Bacillus subtilis*.

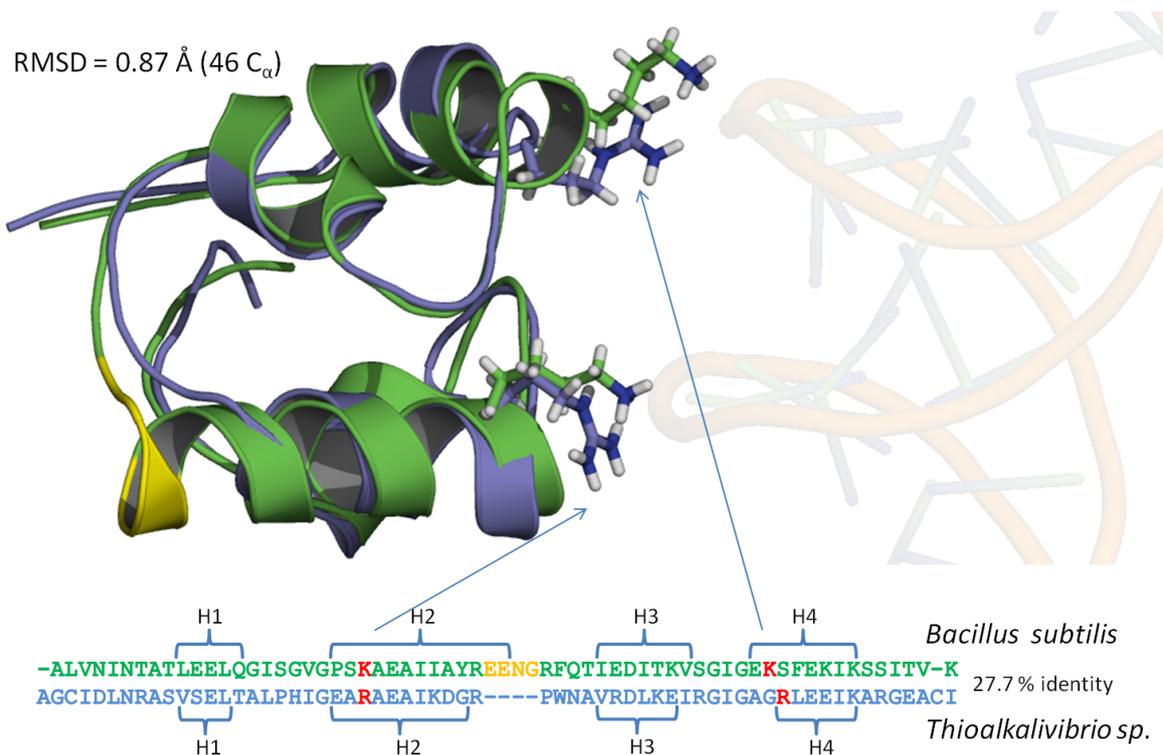
Further complexes that can possibly be formed by the two DNA-binding domains and DNA were modelled. This modelling approach revealed that DT



**Figure 7. Activity of DNaseI and its fusion with ComeEA domain variants at low salt concentration.** The digestion of labelled DNA duplex (30 bp) was observed, corresponding activities at 50 and 100 NaCl mM were compared to the activities in the absence of NaCl. For each curve a maximum fluorescence change rate was calculated, which was considered to be proportional to enzymatic activity. DNaseDT denotes the fusion with the (HhH)<sub>2</sub> domain of DNase from *Thioalkalivibrio sp. K90mix*. DNaseBS denotes the fusion with homologous domain of ComeEA protein from *Bacillus subtilis*.

can form more hydrogen bonds with DNA than BS (Table 4).

The modelled complexes were further analysed using APBS tools 1.4.0 [35]. In order elucidate feasibility of DNA-binding free energy changes that could occur upon the domain binding to DNA (polar solvation term) under several different ionic strength conditions were modelled (Table 5). This data indicate that DNA-binding by both domains is not energetically favorable in terms of polar solvation at 0 mM NaCl. At 100 mM NaCl the modelled polar solvation



**Figure 8. Superimposition of structural models of two (HhH)<sub>2</sub> domains from *Thioalkalivibrio sp. K90mix* (blue colour) and *Bacillus subtilis* (green colour) and corresponding sequence alignment.** DNA phosphate contacting positive residues are indicated by red colour in the sequence alignment and by stick representations in the structural alignment. Yellow colour indicates part of the domain from *Bacillus subtilis*, which has no corresponding residues in the domain from *Thioalkalivibrio sp. K90mix*. Approximate position of DNA is indicated by transparent sticks based on structure PDBID: 3E0D after superimposition with the domains. „H“ indicates helical regions.

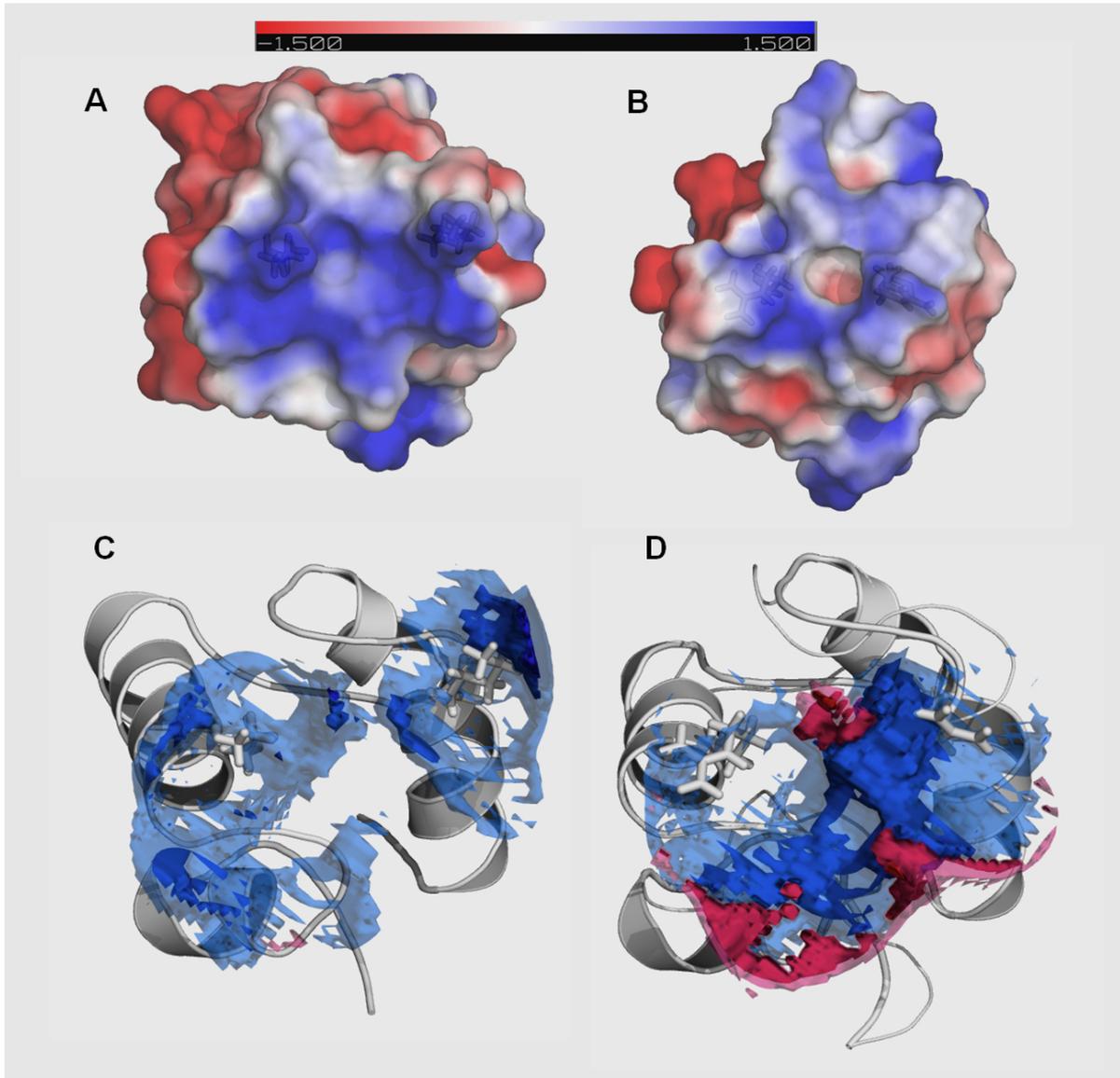
term of the BS (Table 5) is less energetically favourable than DT, but the values are quite close (-6.79 vs -20.45). Further increase in NaCl concentration resulted in much more negative values in case of DT compared to BS. Therefore, at elevated ionic strength DT binding to DNA is much more energetically favorable than BS domain binding (predicted polar solvation term). To further elucidate this phenomenon changes in local ion concentrations upon formation of complexes between DNA and the two DNA-binding domains were analysed. The results (Fig. 9 C, D panels) revealed that upon DT binding to DNA the ion distribution on the protein - DNA-binding surface changed much more than the analogous BS binding. Further analysis of corresponding changes in a „cloud“ of positive sodium ions that accompany negatively charged DNA indicated that upon binding of DT to DNA more sodium ions should be transferred from DNA to solvent compared to the corresponding binding of BS. The data shown in Fig. 9 represents modelled complexes at 400 mM NaCl. Such concentration is quite close to commonly accepted upper limit for reliable electrostatic calculations (500 mM); however, analogous calculations at 150 mM also showed evident differences between the domains (data not shown).

**Table 4. Hydrogen bonding network between (HhH)<sub>2</sub> domains and DNA.** Two domains were analysed: one from an extremely salt tolerant bacterium *Thioalkalivibrio sp. K90mix* (DT), the other one from *Bacillus subtilis* (BS)

Domain	Protein		DNA			Distance, Å
	Residue number	Residue type	Atom	Residue number	Residue type	
BS	44	LYS	NZ	31	ADE O1P	2.89
	73	LYS	NZ	12	CYT O2P	2.91
	73	LYS	NZ	12	CYT O5'	2.94
DT	40	ARG	NE	31	ADE O1P	3.03
	40	ARG	NH2	31	ADE O1P	3.02
	40	ARG	NH2	31	ADE O2P	2.88
	66	ARG	NH1	12	CYT O1P	3.26
	66	ARG	NH1	12	CYT O2P	2.91
	66	ARG	NH2	12	CYT O2P	2.91

**Table 5. Changes in polar electrostatic solvation energy upon complex formation at different NaCl concentrations.** Two domains were modelled: one from an extremely salt tolerant bacterium *Thioalkalivibrio sp. K90mix* (DT), the other one from *Bacillus subtilis* (BS).

Concentration of NaCl, M	Polar solvation energy changes upon complex formation, kJ/mol		Ratio
	<i>BS</i>	<i>DT</i>	
0	29.10	22.15	0.76
0.01	8.49	13.73	1.62
0.1	-6.79	-20.45	3.01
0.4	-94.71	-20851.04	220.15
0.5	-176.66	-898908.16	5088.42



**Figure 9. Electrostatic surface potential of DNA-binding surface and changes in local ion concentration upon binding to DNA.** Electrostatic potential of DNA-binding surface of the domain from *Bacillus subtilis* is shown on the upper-left (A) and corresponding surface of the domain from *Thioalkalivibrio sp. K90mix* is given on the upper-right (B). The range from  $-1.5 \text{ kT/e}$  in red to  $+1.5 \text{ kT/e}$  in blue was chosen for surface colouring. The surface is semi transparent and stick representations of the DNA phosphates contacting residues are visible: lysines - in case of *Bacillus subtilis* domain and arginines in case of *Thioalkalivibrio sp. K90 mix* domain. The changes in local ion concentrations upon formation of a complex with DNA by the domain from *Bacillus subtilis* are depicted in the lower-left (C), the corresponding changes in the case of the domain from *Thioalkalivibrio sp. K90mix* are depicted in the lower-right (D). The DNA phosphates interacting residues are depicted by white sticks. The isocountour surfaces indicate the changes, which occur in the presence of  $0.4 \text{ M NaCl}$ . Four isocountour surfaces are visualized simultaneously. The deep blue represents changes in local ion concentration equivalent to  $-3 \text{ M}$ , the lighter blue represents changes equivalent to  $-2 \text{ M}$ . Similarly deep red indicates changes equivalent to  $+3 \text{ M}$  and the lighter red indicates changes equivalent to  $+2 \text{ M}$ . Isosurfaces equivalent to  $+2/-2 \text{ M}$  overlaps corresponding isosurfaces which represent changes equivalent to  $+3/-3 \text{ M}$ .

# Discussion

In this dissertation the data on the phylogeny of microbial DNases, their domain structure and salt tolerance of the corresponding microorganism (Fig. 1) was studied. The results revealed that two types of domain organizations are found exclusively in DNases from halotolerant / halophilic species: i) the C-terminus of DNase fused to a Lamin tail domain, ii) the C-terminus of DNase fused to a domain containing HhH duplicate motifs (RuvA domain2-like or DNA polymerase lambda fingers). These two types of motifs are not homologous, however, they could be analogous in function. The HhH duplicate domain binds to DNA and is found in many DNA interacting proteins [62]. The Lamin tail domain was discovered in nucleus envelope of the eukaryotic cells and is also associated with the binding to DNA [63,64]. As it is well established, ionic strength has a diminishing effect on electrostatic interactions. Therefore, an additional DNA binding domain should enhance the activity of DNases in the presence of the elevated ionic strength.

Amongst organisms that can tolerate salinity of  $\sim 1.5$  M and more (Fig. 1) nine microorganisms having ten DNases were discovered. Out of the nine organisms eight were bacteria. For adaptation to high salinity bacteria usually employ "salt-out" strategy using compatible solutes [65] and their intracellular proteins might not be adapted to high salt concentrations. However, if proteins from halotolerant bacteria have secretion signals then, most likely, they enter a secretory pathway and, if are secreted, they should be evolutionary adapted to elevated ionic strength. Therefore, it was searched for the presence of a secretion signal (Table 1) and it was found that five of the DNases had an evident secretion signal. Three of these DNases had a domain with duplicate HhH motifs that was fused to a nuclease domain. For further investigations one such DNase, which has a C-terminal domain with duplicate HhH motifs, have been selected. It was examined if this domain was responsible for the resistance to high ionic strength. The chosen DNase was from an extremely halotolerant species – *Thioalkalivibrio sp. K90mix*.

The first task in performing experimental assays with this enzyme was to evaluate the need for divalent cations. The data (Fig. 2) indicates that higher concentrations of divalent ions are required for the optimal activity of DNaseTA compared to its bovine counterpart. The analysis of the bovine DNase [66] showed that the divalent ions were essential for DNA binding, as without bound ions the DNA binding surface in eukaryotic DNase was electronegative and thus unable to bind DNA. The electrostatic surface potential of the bovine DNaseI and the DNase from *Thioalkalivibrio sp. K90mix* are visualized in Fig. 4. These data might explain why the prokaryotic DNase requires higher concentration of divalent ions. The DNA binding surface of the DNase from *Thioalkalivibrio sp.*

*K90mix* has a much larger patch of electronegative area (Fig 4). This implies that more divalent ions have to be bound to DNaseTA in order to make the surface electro-positive and enable the binding to DNA.

Further the capability of DNase from *Thioalkalivibrio sp. K90mix* to digest DNA substrates in a series of buffers where NaCl concentration was gradually increased (Fig. 3) was explored. In these assays two compositions of divalent ions were compared: one suboptimal, with lower concentrations, and the second, using near optimal concentrations. Interestingly, the concentration of divalent ions remarkably influences the resistance of the DNase to the ionic strength in buffers with NaCl concentration up to  $\sim 1$  M. When the concentration of divalent ions was low, gradual increase in a salt concentration revealed two peaks of the activity: one in the absence of NaCl and the second, when the concentration of NaCl was above 1.2M (Fig. 3 C, D). In contrast, when the concentration of divalent ions was higher, there was no activity suppression by moderate concentrations of NaCl and a recovery by higher concentrations. This implies that at the concentrations of NaCl  $\sim 1.2$  M and higher, divalent ions aid the enzyme to maintain its activity and resist to the increasing ionic strength. As increasing ionic strength weakens DNA – protein interaction, the data suggests that at the higher concentration of divalent ions the prokaryotic DNase has higher affinity to DNA than at the lower concentrations. Thus the dual peak of activity implies that the increasing salt concentration gradually suppresses the activity and the DNA binding domain is engaged only at higher salt concentrations thus rescuing the enzyme’s ability to bind DNA.

The experimental data (Table 2, Fig. 3) implies that the C-terminal domain is indeed important for the adaptation of prokaryotic DNases to high ionic strength. This notion was investigated by creating two mutants of the prokaryotic DNase from *Thioalkalivibrio sp. K90mix*: the first mutant was generated by removing the C-terminal domain of the enzyme, the second mutant harboured an inactivating mutation in the active centre of the DNase domain. The mutant with inactivating mutation in the active site of the DNase domain was created in order to test the hypothesis that the C-terminal domain acted only as a facilitator for the DNase domain, but could not catalyse DNA hydrolysis by its own. The mutant with the removed C-terminal domain was created to test if this domain was required for the DNase’s activity at high ionic strength. The catalytic activity impairment of the active site mutant and the diminished salt tolerance of the mutant with the removed C-terminal domain was reliably confirmed by the data on digestion of short and long DNA substrates (Table 2, Fig. 3). Therefore, DNA hydrolysis by the prokaryotic DNase is catalysed by the active site in the DNase domain and the C-terminal domain acts as a facilitator at high salt concentrations (at least up to 4 M NaCl). Contrastingly, the bovine DNaseI is completely inhibited by NaCl concentrations above 1.2 M.

The modelled structures of DNases from two extremely halotolerant organisms *Thioalkalivibrio sp. K90mix* and *Methanohalobium evestigatum* as well as

the subsequent electrostatic calculations indicated that the surfaces of these proteins are more electronegative compared to their eukaryotic counterparts (Fig. 4). The adaptation of the enzyme to saline environments results in accumulation of negative amino acids on the surface [67, 68]. The data presented in Fig.4 B clearly illustrates this trend: the extremely halophilic DNase from *Methanohalobium evestigatum* has the largest electronegative surface, the moderate halophilic DNase from *Thioalkalivibrio sp. K90mix* has a smaller area of the electronegative surface and the non-halotolerant bovine DNaseI has almost no electronegative areas on the surface of the protein side that is not facing DNA. Therefore, the prokaryotic DNaseI homologs tend to accumulate negatively charged surface residues during the adaptation to high salt concentrations and follow the tendency observed in other proteins [67, 68].

A remarkable fact is that we observe accumulation of negatively charged residues in the DNA binding pocket of the protein (Fig.4 A) when we compare DNaseTA (from extremely halotolerant organism) and bovine DNaseI. This is quite remarkable as some DNA binding proteins (TATA-binding protein, ribosome elongation factor) do not acidify during adaptation to high salinity [69] as such adaptation might interfere with the catalytic efficiency by disrupting contacts with DNA. Therefore, in the case of DNaseTA evolution was forced to find a compromise between a necessity to enhance DNA binding under high ionic strength, which hindered electrostatic interactions, and a necessity to adapt protein surface to high salt concentrations via additional negative charges. The compromise was to enhance DNA binding via an additional DNA binding domain, which was experimentally proved to be responsible for the halotolerant properties of the DNase from *Thioalkalivibrio sp. K90mix*.

Therefore in this dissertation it has been demonstrated that a DNase from an extremely halotolerant bacterium *Thioalkalivibrio sp. K90mix* retains activity at high salt concentrations and a (HhH)<sub>2</sub> DNA-binding domain plays the key role here [20]. As discussed above this nuclease is a natural fusion of EPE (exonuclease/endonuclease/phosphatasesuperfamily) and H3TH (two tandem HhH motifs) domains. Subsequently, it was decided to construct analogous fusions of bovine DNaseI that would potentially have enhanced halotolerant properties compared to the wild type bovine DNaseI. DNaseTA has been chosen as an obvious „donor“ of a H3TH domain that could be fused to bovine DNaseI. Additionally, the origin of the H3TH domain in DNaseTA was tracked in order to choose other fusion partner.

The phylogenetic analysis (Fig. 5) shows that the H3TH domain originates from proteins that were related to bacterial competence proteins ComE/ComEA, which harboured single H3TH domain or combination of H3TH and SLBB domains. Such domain organization having protein could be a predecessor of the H3TH domain in bacterial DNases, including DNaseTA. The published data reveals two relatively well characterized ComE/ComEA proteins: i) ComEA protein from *Bacillus subtilis*, which is a cell surface DNA receptor (H3TH domain)

with the N-terminal fragment (SLBB superfamily domain) anchoring to membrane [6, 21], and ii) ComE protein from *Neisseria gonorrhoeae* [61], which comprises single H3TH domain and serves as a part of the machinery that transfers external DNA through periplasm. These proteins were considered as potential „donors“ of the H3TH domain for bovine DNaseI and the ComEA protein from *B. subtilis* was chosen due to its domain organization. This protein comprises two domains; therefore, it was possible to use the naturally occurring inter-domain linker for the fusion.

Interestingly, the phylogenetic tree shown in Fig. 5 indicates that the H3TH domain in the bacterial DNases is closely related to homologous domains in a distinct group of proteins that have a H3TH domain fused to a metallo- $\beta$ -lactamase superfamily domain (lactamase- $\beta$ -H3TH proteins). A H3TH domain of one lactamase- $\beta$ -H3TH protein is more closely related to the homologous domains of the bacterial DNases than to the homologous domains of other lactamase- $\beta$ -H3TH proteins. Thus, it is likely that these proteins inherited the H3TH superfamily domain from the bacterial competence related proteins and then lactamase- $\beta$ -H3TH proteins served as „donors“ of the H3TH domain for the nucleases. Ability to catalyse DNA hydrolysis by several metallo- $\beta$ -lactamase domain having proteins was experimentally demonstrated [70–72]. Therefore, it is quite possible that the metallo- $\beta$ -lactamase domain catalyses hydrolysis of nucleic acids, while the accompanying H3TH domain enhances nucleic acid binding properties as it is in the case of DNaseTA. However, this statement requires experimental verification and is out of the scope of this study. The protein from Chinese hamster, which was used in the phylogeny analysis as an outgroup, is another example of potential evolutionary convergence. In this protein two H3TH domains are fused to the EEP domain (DNaseTA has one EEP and one H3TH domain). Thus, evolution has several times combined DNA binding domains with the domains that are similar to DNaseI.

Two H3TH domains have been selected to be fused with eukaryotic DNaseI: i) the DNaseTA domain, which has been evolutionary fused with DNaseI-like domain resulting in extremely salt tolerant bacterial DNase (the resultant fusion is abbreviated as DNaseDT); ii) the domain of ComEA protein from *B. subtilis*, which is also a multi-domain protein (the resultant fusion is abbreviated as DNaseBS). In order to create a fusion protein comprising several domains it is critical to select the proper inter-domain linker [73, 74]. In this study the natural inter-domain linkers of the two multi-domain proteins have been employed. This approach proved to be correct since only one set of the fusions have been tested and it was succeeded in creating functional enzymes.

In this study both of the fusions were shown to be more salt tolerant than bovine DNaseI, albeit to different extent: the H3TH domain of DNaseTA enabled detectable catalytic activity even at 4M of NaCl, while the fusion harbouring the H3TH domain of ComEA protein from *B. subtilis* was evidently less salt tolerant at this concentration (Fig. 6). The differences could be partially explained

by the origin of the two H3TH domains. The H3TH domain harboured by DNaseBS originates from the ComEA protein of *Bacillus subtilis*. This protein is located on the outer cell surface; thus, it is exposed to the outside environment acting as a DNA receptor [6,75]. *Bacillus subtilis* can tolerate some fluctuations in salinity, but grows best in low salinity environment and is not a high salt tolerant bacterium [76]. The natural environment of this bacterium is soil. In soil the salinity increases episodically in conjunction with desiccation. In contrast, the H3TH domain harboured by DNaseDT originates from an extremely salt tolerant bacterium *Thioalkalivibrio sp. K90mix*. The natural environment of this bacterium is soda lakes - an extremely saline habitat [53]. Thus, it is likely that the salt tolerances of the two fusions DNaseDT and DNaseBS differ because the corresponding H3TH domains were evolutionary adapted to bind DNA at different salinity.

The results of molecular modelling provide some insights on molecular basis of the apparent differences in the salt tolerance of the two fusions. Both H3TH domains bind DNA phosphates with two positive residues, albeit the H3TH domain from *Thioalkalivibrio sp. K90mix* potentially forms more hydrogen bonds (Table 4). The differences became more evident when changes in polar electrostatic solvation energy upon complex formation were modelled at increasing NaCl concentrations (Table 5). The modelling suggested that upon binding of the H3TH domain from *Thioalkalivibrio sp. K90mix* to DNA more sodium ions should be released to the solvent, compared to the corresponding binding of the H3TH domain from *Bacillus subtilis*. This nicely agrees with the experimental data on DNA-binding by two DNA polymerases: more sodium ions were released in case of a more salt tolerant polymerase [77].

However, the collected data still do not allow to explain why DNaseBS is more resistant to ionic strength than DNaseDT at relatively low salt concentrations (up to 100 mM NaCl) (Fig. 7). This difference could be explained by presuming that the domain from *Bacillus subtilis* efficiently interacts with DNA at low salt concentrations while the homologous domain from *Thioalkalivibrio sp. K90mix* acts in its full potential only at high ionic strength [20].

Proteins tend to accumulate negatively charged surface residues during the adaptation to high salt concentrations [67,68]. However, in this case the surface electrostatic potential of the H3TH domain from the extremely salt tolerant bacterium (DT) is only slightly less electro-positive than the homologous domain (BS) from the less salt tolerant bacterium (Fig. 9) - no extensive acidification is evident. This is in agreement with the analysis done by Becker et al. [69]. In this study acidification of nucleic acid binding pocket was not observed in TATA-binding protein and ribosome elongation factor, when proteins from halophilic and mesophilic organisms were compared.

# Conclusions

1. C-terminal DNA binding domains that are found in many DNaseI-like nucleases from halotolerant / halophilic prokaryotes might be an evolutionary adaptation enabling DNA hydrolysis at high ionic strength. Prokaryotic DNaseI-like nucleases that have C-terminal (HhH)<sub>2</sub> domain are found only in halotolerant / halophilic bacteria.
2. The C-terminal (HhH)<sub>2</sub> domain in DNaseI-like nuclease from an extremely halotolerant bacterium *Thioalkalivibrio sp. K90mix* is the key factor enabling DNA digestion at high salt concentrations.
3. DNA binding interface and the remaining surface of DNaseI-like nuclease from *Thioalkalivibrio sp. K90mix* contain more negative residues compared to its non halotolerant / halophilic homolog. It is likely that the DNA binding domain alleviated accumulation of the negative residues at DNA binding surface of the nuclease domain during adaptation to high salinity.
4. Origin of the C-terminal (HhH)<sub>2</sub> domain in DNaseI-like nucleases from halotolerant bacteria is related with bacterial competence proteins ComE/ComEA. It is likely that this domain originated from ComE/ComEA family proteins: lactamase- $\beta$ -H3TH proteins could inherit the domain from the bacterial competence proteins and then „donate“ the domain to the nucleases.
5. Resistance to ionic strength of a non halotolerant / halophilic DNaseI was successfully increased by fusing a (HhH)<sub>2</sub> domain. In this way the evolutionary step that occurred during adaptation to high salinity and formed the chimeric bacterial DNaseI-like nucleases was successfully mimicked *in vitro*.
6. Molecular modelling data suggests that differences in tolerance for high ionic strength between the two created DNaseI fusions could be due to differences in hydrogen bonding between the fused domains and DNA and different ability to transfer cations to solvent during DNA-protein complex formation.

# List of publications

## The thesis is based on the following original publications:

- 1 Alzbutas, G., Kaniusaite, M., & Lagunavicius, A. (2016). Enhancement of DNaseI Salt Tolerance by Mimicking the Domain Structure of DNase from an Extremely Halotolerant Bacterium *Thioalkalivibrio* sp. K90mix. *PloS one*, 11(3), e0150404.
- 2 Alzbutas, G., Kaniusaite, M., Grybauskas, A., & Lagunavicius, A. (2015). Domain organization of DNase from *Thioalkalivibrio* sp. provides insights into retention of activity in high salt environments. *Frontiers in microbiology*, 6.

## Patent:

- 1 Alzbutas, G., Lagunavičius, A., Kaniušaitė, M. Improved Deoxyribonuclease Enzymes. International Application No.: PCT/EP2015/062222., June 2nd, 2014.

## Other publications:

- 1 Povilaitis, T., Alzbutas, G., Sukackaite, R., Siurkus, J., & Skirgaila, R. (2016). In vitro evolution of phi29 DNA polymerase using isothermal compartmentalized self replication technique. *Protein Engineering Design and Selection*.
- 2 Strepetkaitė, D., Alzbutas, G., Astromskas, E., Lagunavičius, A., Sabaliauskaitė, R., Arbačiauskas, K., & Lazutka, J. (2015). Analysis of DNA Methylation and Hydroxymethylation in the Genome of Crustacean *Daphnia pulex*. *Genes*, 7(1), 1.
- 3 Glemzaite, M., Balciunaite, E., Karvelis, T., Gasiunas, G., Grusyte, M. M., Alzbutas, G., ... & Lubys, A. (2015). Targeted gene editing by transfection of in vitro reconstituted *Streptococcus thermophilus* Cas9 nuclease complex. *RNA biology*, 12(1), 1-4.
- 4 Šulčius, S., Alzbutas, G., Kvederavičiūtė, K., Koreivienė, J., Zakrys, L., Lubys, A., & Paškauskas, R. (2015). Draft genome sequence of the cyanobacterium *Aphanizomenon flos-aquae* strain 2012/KM1/D3, isolated from the Curonian Lagoon (Baltic Sea). *Genome announcements*, 3(1), e01392-14.
- 5 Baranauskas, A., Paliksa, S., Alzbutas, G., Vaitkevicius, M., Lubiene, J., Letukiene, V., ... & Skirgaila, R. (2012). Generation and characterization of new highly thermostable and processive M-MuLV reverse transcriptase variants. *Protein Engineering Design and Selection*, 25(10), 657-668.

## Conference presentations:

- 1 Gediminas Alzbutas, Milda Kaniušaitė, Arūnas Lagunavičius, Algirdas Grybauskas. Design of Halotolerant Bovine DNase Guided By Phylogenetic and Domain Struc-

- ture Studies of Bacterial DNases. 5th ICBE - International Conference on Biomolecular Engineering. USA. January 11-14 d., 2015.
- 2 Gediminas Alzbutas, Milda Kaniušaitė, Arūnas Lagunavičius, Algirdas Grybauskas. Domain organization of DNase from halophile *Thioalkalivibrio* sp. XIIIth International Conference of Lithuanian Biochemical Society. Lithuania. June 18-20 d., 2014.

# Curriculum Vitae

**Name** Gediminas Alzbutas  
**Date of birth** 1982-12-13  
**Address** Department of Scientific Research and Development,  
UAB „Thermo Fisher Scientific Baltics“  
V.A. Graičiūno 8, Vilnius LT-02241, Lithuania  
**Phone** +370 615 715 42  
**e-mail** gediminas.alzbutas@thermofisher.com  
**Education**  
2011-2015 PhD studies of Biochemistry at the Institute of Biotechnology, Vilnius University  
2009-2010 M.Sc of Bioinformatics at The University of Manchester  
2005-2007 M.Sc of Molecular Biology and Biotechnology at Vytautas Magnus University  
2001-2005 B.Sc of Biology at Vytautas Magnus University  
**Working experience**  
since 2011 Research Scientist, Department of Research and Development at UAB „Thermo Fisher Scientific Baltics“  
since 2010 Assistant lecturer, Faculty of Mathematics and Informatics, Vilnius university  
2007-2011 Junior Research Scientist, Department of Research and Development at UAB „Thermo Fisher Scientific Baltics“

# Acknowledgements

I would like to express my gratitude to people who helped me during research and writing of my dissertation.

I thank my supervisor dr. A. Lagunavicius for the possibility to work in his group, for scientific ideas, discussions, and constructive criticism.

I thank prof. habil. dr. Arvydas Janulaitis for encouragement to investigate DNaseI homologs.

I thank UAB „Thermo Fisher Scientific Baltics“ for possibility to combine study and work. Thank you to all current and former members of the Department of Scientific Research and Development at UAB "Thermo Fisher Scientific Baltics" for everyday discussions and friendly atmosphere.

Thank you to all of my family, especially my wife Jurga.

# Reziუმэ

Šiame darbe buvo analizuojamos į DNazę I panašios prokariotinės nukleazės. Buvo pastebėta, kad daugeliui DNazės I šeimos baltymų iš halofilinių/ druskingumui pakančių prokariotų yra būdinga panaši domeninė struktūra: nukleazinis domenas sulietas su C-galiniu galimai DNR rišančiu domenu.

Darbo metu buvo parodyta, kad tokia į DNazę I panaši nukleazę iš itin druskingumui pakančios bakterijos *Thioalkalivibrio sp. K90mix* yra aktyvi iki 4 M NaCl. Šios nukleazės mutantų savybės, nustatytos eksperimentiškai, parodė, kad esminis faktorius, leidžiantis šiam fermentui hidrolizuoti DNR didelėje joninėje jėgoje, yra C-galinis DNR rišantis domenas, kurį sudaro du HhH motyvai. Molekulinis modeliavimas leido daryti prielaidą, kad papildomas DNR rišantis domenas įgalino kaupti neigiamo krūvio aminorūgštis DNazinio domeno DNR surišimo paviršiuje, bakterinėms DNazėms prisitaikant prie didelio druskingumo.

Filogenetinė analizė parodė, kad dvigubo HhH motyvo domenas, randamas į DNazę I panašiose nukleazėse iš halofilinių/ druskingumui pakančių bakterijų, kilo iš kompetencijos sistemos ComEA/ComE baltymų, kurie lemia gebėjimą įsisavinti ekstraląstelinę DNR. Šiame darbe pavyko *in vitro* imituoti evoliucini žingsnį, kuomet druskingoje aplinkoje gyvenančiuose prokariotuose susiformavo chimeriniai baltymai, turintys DNazinį ir dvigubo HhH motyvo domenus. Buvo sukurti analogiškos domeninės organizacijos jaučio DNazės I liejiniai, pasižymintys didesniu atsparumu joninei jėgai nei DNazė I. Abiejų liejinių atsparumas druskingumui buvo skirtingas. Šį skirtumą galima aiškinti remiantis molekulinio modeliavimu, prognozuojančiu skirtingą prielietų domenų sąveiką su DNR ir gebėjimą nustumti katijonus nuo DNR, vykstant baltymo – DNR sąveikai.

# Bibliography

- [1] Kamekura, M. and Onishi, H. Properties of the halophilic nuclease of a moderate halophile, *Micrococcus varians* subsp. *halophilus*. *J Bacteriol*, vol. 133, (1978), pp. 59–65.
- [2] Kamekura, M. and Onishi, H. Halophilic nuclease from a moderately halophilic *Micrococcus varians*. *J Bacteriol*, vol. 119, (1974), pp. 339–344.
- [3] Onishi, H.; Mori, T.; Takeuchi, S.; Tani, K.; Kobayashi, T.; and Kamekura, M. Halophilic Nuclease of a Moderately Halophilic *Bacillus* sp.: Production, Purification, and Characterization. *Appl Environ Microbiol*, vol. 45, (1983), pp. 24–30.
- [4] Kanlayakrit, W.; Ikeda, T.; Tojai, S.; Rodprapakorn, M.; and Sirisansa-neeyakul, S. Isolation and characterization of extracellular halophilic ribonuclease from halotolerant *Pseudomonas* species. *KASETSART JOURNAL*, p. 179.
- [5] Pan, C.Q. and Lazarus, R.A. Hyperactivity of human DNase I variants. Dependence on the number of positively charged residues and concentration, length, and environment of DNA. *J Biol Chem*, vol. 273, (1998), pp. 11701–11708.
- [6] Inamine, G.S. and Dubnau, D. ComEA, a *Bacillus subtilis* integral membrane protein required for genetic transformation, is needed for both DNA binding and transport. *J Bacteriol*, vol. 177, (1995), pp. 3045–3051.
- [7] de Vega, M.; Lázaro, J.M.; Mencía, M.; Blanco, L.; and Salas, M. Improvement of  $\phi$ 29 DNA polymerase amplification performance by fusion of DNA binding motifs. *Proc Natl Acad Sci U S A*, vol. 107, (2010), pp. 16506–16511.
- [8] Pavlov, A.R.; Belova, G.I.; Kozyavkin, S.A.; and Slesarev, A.I. Helix-hairpin-helix motifs confer salt resistance and processivity on chimeric DNA polymerases. *Proc Natl Acad Sci U S A*, vol. 99, (2002), pp. 13510–13515.
- [9] Pavlov, A.R.; Pavlova, N.V.; Kozyavkin, S.A.; and Slesarev, A.I. Cooperation between catalytic and DNA binding domains enhances thermostability and supports DNA synthesis at higher temperatures by thermostable DNA polymerases. *Biochemistry*, vol. 51, (2012), pp. 2032–2043.
- [10] Hunter, S.; Jones, P.; Mitchell, A.; Apweiler, R.; Attwood, T.K.; Bateman, A.; Bernard, T.; Binns, D.; Bork, P.; Burge, S.; de Castro, E.; Coggill, P.; Corbett, M.; Das, U.; Daugherty, L.; Duquenne, L.; Finn, R.D.; Fraser, M.; Gough, J.; Haft, D.; Hulo, N.; Kahn, D.; Kelly, E.; Letunic, I.; Lonsdale, D.; Lopez, R.; Madera, M.; Maslen, J.; McAnulla, C.; McDowall, J.; McMenamin, C.; Mi, H.; Mutowo-Muellenet, P.; Mulder, N.; Natale, D.; Orengo, C.; Pesseat, S.; Punta, M.; Quinn, A.F.; Rivoire, C.; Sangrador-Vegas,

- A.; Selengut, J.D.; Sigrist, C.J.A.; Scheremetjew, M.; Tate, J.; Thimmajarnathanan, M.; Thomas, P.D.; Wu, C.H.; Yeats, C.; and Yong, S.Y. InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res*, vol. 40, (2012), pp. D306–D312.
- [11] , U.C. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res*, vol. 42, (2014), pp. D191–D198.
- [12] Jones, P.; Binns, D.; Chang, H.Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; Pesseat, S.; Quinn, A.F.; Sangrador-Vegas, A.; Scheremetjew, M.; Yong, S.Y.; Lopez, R.; and Hunter, S. InterProScan 5: genome-scale protein function classification. *Bioinformatics*, vol. 30, (2014), pp. 1236–1240.
- [13] Bhardwaj, G.; Ko, K.D.; Hong, Y.; Zhang, Z.; Ho, N.L.; Chintapalli, S.V.; Kline, L.A.; Gotlin, M.; Hartranft, D.N.; Patterson, M.E.; Dave, F.; Smith, E.J.; Holmes, E.C.; Patterson, R.L.; and van Rossum, D.B. PHYRN: a robust method for phylogenetic analysis of highly divergent sequences. *PLoS One*, vol. 7, (2012), p. e34261.
- [14] Felsenstein, J. PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics*, vol. 5, (1989), pp. 164–166.
- [15] Huerta-Cepas, J.; Dopazo, J.; and Gabaldón, T. ETE: a python Environment for Tree Exploration. *BMC Bioinformatics*, vol. 11, (2010), p. 24.
- [16] Petersen, T.N.; Brunak, S.; von Heijne, G.; and Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*, vol. 8, (2011), pp. 785–786.
- [17] Bagos, P.G.; Tsirigos, K.D.; Plessas, S.K.; Liakopoulos, T.D.; and Hamodrakas, S.J. Prediction of signal peptides in archaea. *Protein Eng Des Sel*, vol. 22, (2009), pp. 27–35.
- [18] Reynolds, S.M.; Käll, L.; Riffle, M.E.; Bilmes, J.A.; and Noble, W.S. Transmembrane topology and signal peptide prediction using dynamic bayesian networks. *PLoS Comput Biol*, vol. 4, (2008), p. e1000213.
- [19] Chen, W.J.; Lai, P.J.; Lai, Y.S.; Huang, P.T.; Lin, C.C.; and Liao, T.H. Probing the catalytic mechanism of bovine pancreatic deoxyribonuclease I by chemical rescue. *Biochem Biophys Res Commun*, vol. 352, (2007), pp. 689–696.
- [20] Alzbutas, G.; Kaniusaite, M.; Grybauskas, A.; and Lagunavicius, A. Domain organization of DNase from *Thioalkalivibrio* sp. provides insights into retention of activity in high salt environments. *Frontiers in Microbiology*, vol. 6. ISSN 1664-302X.
- [21] Marchler-Bauer, A.; Derbyshire, M.K.; Gonzales, N.R.; Lu, S.; Chitsaz, F.; Geer, L.Y.; Geer, R.C.; He, J.; Gwadz, M.; Hurwitz, D.I.; Lanczycki, C.J.; Lu, F.; Marchler, G.H.; Song, J.S.; Thanki, N.; Wang, Z.; Yamashita, R.A.; Zhang, D.; Zheng, C.; and Bryant, S.H. CDD: NCBI’s conserved domain database. *Nucleic Acids Res*.
- [22] hmmscan :: search sequence(s) against a profile database. HMMER 3.1b1

- (27 May 2013) ; <http://hmmer.org/>, 2013.
- [23] Finn, R.D.; Clements, J.; and Eddy, S.R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*, vol. 39, (2011), pp. W29–W37.
  - [24] Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; and Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, vol. 28, (2012), pp. 3150–3152.
  - [25] Li, W. and Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, vol. 22, (2006), pp. 1658–1659.
  - [26] Pei, J. and Grishin, N.V. PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. *Methods Mol Biol*, vol. 1079, (2014), pp. 263–271.
  - [27] Pei, J.; Kim, B.H.; and Grishin, N.V. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res*, vol. 36, (2008), pp. 2295–2300.
  - [28] Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, vol. 30, (2014), pp. 1312–1313.
  - [29] Roy, A.; Kucukural, A.; and Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols*, vol. 5, (2010), pp. 725–738.
  - [30] DeLano, W.L. The PyMOL molecular graphics system.
  - [31] Shatsky, M.; Nussinov, R.; and Wolfson, H.J. A method for simultaneous alignment of multiple protein structures. *Proteins*, vol. 56, (2004), pp. 143–156.
  - [32] Wing, R.A.; Bailey, S.; and Steitz, T.A. Insights into the replisome from the structure of a ternary complex of the DNA polymerase III alpha-subunit. *J Mol Biol*, vol. 382, (2008), pp. 859–869.
  - [33] van Dijk, M. and Bonvin, A.M.J.J. 3D-DART: a DNA structure modelling server. *Nucleic Acids Res*, vol. 37, (2009), pp. W235–W239.
  - [34] de Vries, S.J.; van Dijk, M.; and Bonvin, A.M.J.J. The HADDOCK web server for data-driven biomolecular docking. *Nat Protoc*, vol. 5, (2010), pp. 883–897.
  - [35] Baker, N.A.; Sept, D.; Joseph, S.; Holst, M.J.; and McCammon, J.A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences*, vol. 98, (2001), pp. 10037–10041.
  - [36] Dolinsky, T.J.; Nielsen, J.E.; McCammon, J.A.; and Baker, N.A. PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic acids research*, vol. 32, (2004), pp. W665–W667.
  - [37] Case, D.; Babin, V.; Berryman, J.; Betz, R.; Cai, Q.; Cerutti, D.; Cheatham III, T.; Darden, T.; Duke, R.; Gohlke, H.; et al. Amber 14.

- [38] García-García, C. and Draper, D.E. Electrostatic interactions in a peptide–RNA complex. *J Mol Biol*, vol. 331, (2003), pp. 75–88.
- [39] Consortium, U. et al. Activities at the Universal Protein Resource (UniProt). *Nucleic acids research*, vol. 42, (2014), pp. D191–D198.
- [40] Remmert, M.; Biegert, A.; Hauser, A.; and Söding, J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nature methods*, vol. 9, (2012), pp. 173–175.
- [41] Rodrigues, J.P.; Levitt, M.; and Chopra, G. KoBaMIN: a knowledge-based minimization web server for protein structure refinement. *Nucleic acids research*, p. gks376.
- [42] Wiederstein, M. and Sippl, M.J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*, vol. 35, (2007), pp. W407–W410.
- [43] Benkert, P.; Tosatto, S.C.; and Schomburg, D. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function, and Bioinformatics*, vol. 71, (2008), pp. 261–277.
- [44] Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; and Lipman, D.J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, vol. 25, (1997), pp. 3389–3402.
- [45] Eddy, S.R. Accelerated profile HMM searches. *PLoS computational biology*, vol. 7, (2011), p. e1002195.
- [46] Katoh, K. and Standley, D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, vol. 30, (2013), pp. 772–780.
- [47] Katoh, K. and Toh, H. Recent developments in the MAFFT multiple sequence alignment program. *Briefings in bioinformatics*, vol. 9, (2008), pp. 286–298.
- [48] Celniker, G.; Nimrod, G.; Ashkenazy, H.; Glaser, F.; Martz, E.; Mayrose, I.; Pupko, T.; and Ben-Tal, N. ConSurf: using evolutionary data to raise testable hypotheses about protein function. *Israel Journal of Chemistry*, vol. 53, (2013), pp. 199–206.
- [49] Holm, L. and Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic acids research*, vol. 38, (2010), pp. W545–W549.
- [50] Singh, S.; Folkers, G.E.; Bonvin, A.M.J.J.; Boelens, R.; Wechselberger, R.; Niztayev, A.; and Kaptein, R. Solution structure and DNA-binding properties of the C-terminal domain of UvrC from E.coli. *EMBO J*, vol. 21, (2002), pp. 6257–6266.
- [51] Garcia-Diaz, M.; Bebenek, K.; Krahn, J.M.; Blanco, L.; Kunkel, T.A.; and Pedersen, L.C. A structural solution for the DNA polymerase lambda-dependent repair of DNA gaps with minimal homology. *Mol Cell*, vol. 13, (2004), pp. 561–572.
- [52] Wilharm, T.; Zhilina, T.; and Hummel, P. DNA-DNA Hybridiza-

- tion of Methylophilic Halophilic Methanogenic Bacteria and Transfer of Methanococcus halophilusvp to the Genus Methanohalophilus as Methanohalophilus halophilus comb. nov. *International journal of systematic bacteriology*, vol. 41, (1991), pp. 558–562.
- [53] Muyzer, G.; Sorokin, D.Y.; Mavromatis, K.; Lapidus, A.; Foster, B.; Sun, H.; Ivanova, N.; Pati, A.; d’Haeseleer, P.; Woyke, T.; et al. Complete genome sequence of Thioalkalivibrio sp. K90mix. *Standards in genomic sciences*, vol. 5, (2011), p. 341.
- [54] Tan, D.; Xue, Y.S.; Aibaidula, G.; and Chen, G.Q. Unsterile and continuous production of polyhydroxybutyrate by Halomonas TD01. *Bioresour Technol*, vol. 102, (2011), pp. 8130–8136.
- [55] Martínez-Cánovas, M.J.; Quesada, E.; Martínez-Checa, F.; del Moral, A.; and Béjar, V. Salipiger mucescens gen. nov., sp. nov., a moderately halophilic, exopolysaccharide-producing bacterium isolated from hypersaline soil, belonging to the  $\alpha$ -Proteobacteria. *International journal of systematic and evolutionary microbiology*, vol. 54, (2004), pp. 1735–1740.
- [56] Lin, Y.; Fan, H.; Hao, X.; Johnstone, L.; Hu, Y.; Wei, G.; Alwathnani, H.A.; Wang, G.; and Rensing, C. Draft genome sequence of Halomonas sp. strain HAL1, a moderately halophilic arsenite-oxidizing bacterium isolated from gold-mine soil. *Journal of bacteriology*, vol. 194, (2012), pp. 199–200.
- [57] Schlesner, H. Planctomyces brasiliensis sp. nov., a Halotolerant Bacterium from a Salt Pit. *Systematic and applied microbiology*, vol. 12, (1989), pp. 159–161.
- [58] Koops, H.P.; Böttcher, B.; Möller, U.; Pommerening-Röser, A.; and Stehr, G. Description of a new species of Nitrosococcus. *Archives of Microbiology*, vol. 154, (1990), pp. 244–248. ISSN 0302-8933.
- [59] Bruns, A.; Rohde, M.; and Berthe-Corti, L. Muricauda ruestringensis gen. nov., sp. nov., a facultatively anaerobic, appendaged bacterium from German North Sea intertidal sediment. *Int J Syst Evol Microbiol*, vol. 51, (2001), pp. 1997–2006.
- [60] Koechler, S.; Plewniak, F.; Barbe, V.; Battaglia-Brunet, F.; Jost, B.; Joulian, C.; Philipps, M.; Vicaire, S.; Vincent, S.; Ye, T.; et al. Genome sequence of Halomonas sp. strain A3H3, isolated from arsenic-rich marine sediments. *Genome announcements*, vol. 1, (2013), pp. e00819–13.
- [61] Chen, I. and Gotschlich, E.C. ComE, a competence protein from Neisseria gonorrhoeae with DNA-binding activity. *J Bacteriol*, vol. 183, (2001), pp. 3160–3168.
- [62] Doherty, A.J.; Serpell, L.C.; and Ponting, C.P. The helix-hairpin-helix DNA-binding motif: a structural basis for non-sequence-specific recognition of DNA. *Nucleic Acids Res*, vol. 24, (1996), pp. 2488–2497.
- [63] Bruston, F.; Delbarre, E.; Ostlund, C.; Worman, H.J.; Buendia, B.; and Duband-Goulet, I. Loss of a DNA binding site within the tail of prelamins A contributes to altered heterochromatin anchorage by progerin. *FEBS Lett*,

- vol. 584, (2010), pp. 2999–3004.
- [64] Mans, B.J.; Anantharaman, V.; Aravind, L.; and Koonin, E.V. Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. *Cell Cycle*, vol. 3, (2004), pp. 1612–1637.
- [65] da Costa, M.S.; Santos, H.; and Galinski, E.A. An overview of the role and diversity of compatible solutes in Bacteria and Archaea. *Adv Biochem Eng Biotechnol*, vol. 61, (1998), pp. 117–153.
- [66] Guérout, M.; Picot, D.; Abi-Ghanem, J.; Hartmann, B.; and Baaden, M. How cations can assist DNase I in DNA binding and hydrolysis. *PLoS Comput Biol*, vol. 6, (2010), p. e1001000.
- [67] Sigliocolo, A.; Paiardini, A.; Piscitelli, M.; and Pascarella, S. Structural adaptation of extreme halophilic proteins through decrease of conserved hydrophobic contact surface. *BMC structural biology*, vol. 11, (2011), p. 50.
- [68] Graziano, G. and Merlino, A. Molecular bases of protein halotolerance. *Biochim Biophys Acta*, vol. 1844, (2014), pp. 850–858.
- [69] Becker, E.A.; Seitzer, P.M.; Tritt, A.; Larsen, D.; Krusor, M.; Yao, A.I.; Wu, D.; Madern, D.; Eisen, J.A.; Darling, A.E.; and Facciotti, M.T. Phylogenetically driven sequencing of extremely halophilic archaea reveals strategies for static and dynamic osmo-response. *PLoS Genet*, vol. 10, (2014), p. e1004784.
- [70] Ma, Y.; Pannicke, U.; Schwarz, K.; and Lieber, M.R. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell*, vol. 108, (2002), pp. 781–794.
- [71] Tiefenbach, T. and Junop, M. Pso2 (SNM1) is a DNA structure-specific endonuclease. *Nucleic Acids Res*, vol. 40, (2012), pp. 2131–2139.
- [72] Dominski, Z. Nucleases of the metallo-beta-lactamase family and their role in DNA and RNA metabolism. *Crit Rev Biochem Mol Biol*, vol. 42, (2007), pp. 67–93.
- [73] George, R.A. and Heringa, J. An analysis of protein domain linkers: their classification and role in protein folding. *Protein Eng*, vol. 15, (2002), pp. 871–879.
- [74] Gokhale, R.S. and Khosla, C. Role of linkers in communication between protein modules. *Curr Opin Chem Biol*, vol. 4, (2000), pp. 22–27.
- [75] Provvedi, R. and Dubnau, D. ComEA is a DNA receptor for transformation of competent *Bacillus subtilis*. *Mol Microbiol*, vol. 31, (1999), pp. 271–280.
- [76] Pittelkow, M. and Bremer, E. Cellular Adjustments of *Bacillus subtilis* and Other Bacilli to Fluctuating Salinities. In A. Ventosa; A. Oren; and Y. Ma (editors), *Halophiles and Hypersaline Environments*, pp. 275–302. Springer Berlin Heidelberg. ISBN 978-3-642-20197-4, 2011.
- [77] Datta, K. and LiCata, V.J. Salt dependence of DNA binding by *Thermus aquaticus* and *Escherichia coli* DNA polymerases. *J Biol Chem*, vol. 278, (2003), pp. 5694–5701.