UiT The Arctic University of Norway

Faculty of Science and Technology

A functional and structural study of three bacterial nucleic acid-interacting proteins

The story of a Ferric Uptake Regulator, an Oligoribonuclease and an ATP-dependent DNA ligase

Kristel Berg

A dissertation for the degree of Philosophiae Doctor – November 2019

A functional and structural study of three bacterial nucleic acid-interacting proteins

The story of a Ferric Uptake Regulator, an Oligoribonuclease and an ATP-dependent DNA ligase

Kristel Berg

A dissertation for the degree of Philosophiae Doctor

Department of Chemistry

Faculty of Science and Technology

November 2019

To Terje, Malvin and Sigrid

Acknowledgments

This work was carried out at the Norwegian Structural Biology Centre (NorStruct), Department of Chemistry, Faculty of Science and Technology at The University of Tromsø - The Arctic University of Norway. Financial support was granted by the UiT – The Arctic University of Norway. I would also like to acknowledge the financial support and training from BioStruct, the Norwegian Graduate School in Structural Biology and BioCat, the Norwegian Graduate School in Biocatalysis.

My sincerest gratitude goes to my main supervisor Ingar Leiros for giving me the opportunity to work on this project. I highly appreciate the encouragement, humour, and great patience I received from you all these years and I feel grateful to have had such a supportive, understanding and positive supervisor throughout this journey's ups and downs. Your impressive scientific knowledge has been valuable, and I have learnt a lot from you. Thank you also for the immense amount of time and effort you have brought into finalizing the papers and this thesis.

It has been a pleasure to have Hege Lynum Pedersen as my co-supervisor from day one. You have provided me great support and encouragement, both work-related and in life outside of the lab. I appreciate your expert advice and mentoring in the lab. Your friendship, caring personality and infectious laughter has really mattered to me these years.

To my co-supervisors Atle Noralf Larsen and Yvonne Pietrowski, thank you for giving me the opportunity to contribute to your exciting project. Your belief in me and the project was a huge game changer and motivation boost just when I needed it the most. Thank you for your excellent supervision throughout the experiments. I also appreciate your easy-going- and caring personalities and the patience and support you have shown.

I am also sincerely thankful for the opportunity to work on the ligase project under the guidance of my co-supervisor Adele Williamson. Thank you for sharing your knowledge, creativity and your many ideas. You are truly inspiring and a great mentor. Your positive attitude and friendship have also contributed greatly to the nice atmosphere at work. The time and effort you have donated in guiding me through the writing of this thesis, from the other side of the globe, is impressive and highly appreciated.

To all my supervisors – I owe you lots of gratitude for valuable input and expert advice towards finalizing this thesis. You have taught me to be a better scientist and it has been a pleasure to work with you all.

Previous Head of Department Ronny Helland and previous Director of NorStruct Arne Smalås have also provided great support, thank you.

I would like to thank present and former engineers Eva Bjørkeng, Trine Carlsen and Stefan Hauglid for the technical help and practical assistance with the never-ending rounds of protein purifications. Lots of gratitude also goes to Bjarte Aarmo Lund and Marcin Pierechod for valuable advising in my attempts to gain data from Microscale thermophoresis and Biacore SPR analysis.

To all employees at Norstruct, for creating such a pleasant working environment, with parties, laughs, meaningless conversations, camaraderie, and scientific discussions. You all made it easy to go to work every morning, even when I felt science plotted against me.

My office mates have changed throughout the years, but I have been so lucky to share office with Miriam from the beginning. I appreciate how you have been there for me every day, good or bad, and our nights out on laser light-covered dance floors have been awesome. My most recent office mate, Aili, has also been of great support.

To other Norstruct friends, past and present, for supporting me through the joys and hardships both in life and during the course of my studies. I have shared so many laughs with you! I love that you "voluntarily" embraced my crazy ideas. The Norstruct Spice Girls rocks! Additional thanks go to Man Kumari, Susann and Eva for listening, for keeping me sane and for simply being there. I would also like to highlight the smiling face I met by the coffee machine every morning; thank you Kåre-Olav for continuous encouragements, your positive attitude and excellent leading on the Christmas dance floor.

Staying active in this process has been important for my PhD survival. So, thank you Susann and Saana for nerdy running conversations and for happily joining me on treadmill breaks. Thank you, Erik, for fun and mindful climbing breaks, moral support and for always being so positive.

To my new colleagues at the awesome genetics group at NFH, thank you for your patience when I have been a bit absent-minded this first year. Especially my office girls and my group leader, Kim Præbel, have been of great support. To Hege Devold - I appreciate having a close friend upstairs, especially one with a laughter loud enough to remind me about ongoing coffee breaks.

To my great group of friends here in Tromsø, both those close to me since we first met as students about a hundred years ago, and those I got to know more recently – you know who you are and I appreciate each one of you. You bring joy and laughter to my life, with skiing, parties, luxury festival camps, dinners, training, singing (!), random coffee chats and more. Thank you for support, advice and for patiently listening to my frustrations. I look forward to spending more time with you all and contribute more socially. Gratitude also goes to my long-distance friends; Lena, Silvie and Vroni. Thank you for reaching out to me when I was distant in many ways. Our chats have meant more than you know.

Most importantly, I wish to express my deepest appreciation to my family. To my parents, I would not be who I am today without your love, encouragement, and backing. You have always done everything in your power to make sure that I can achieve my dreams. I am also thankful for the unwavering support and joy from my sister, my brother in law, my nephew and niece.

Dear Terje; your love, patience, support and understanding for what a PhD life is about, have meant the world to me and this thesis could not possible be completed without you. You mean the world to me. To my kids, Sigrid and Malvin, to show you anything is possible. You are an endless source of joy and remind me on a daily basis that the most important part of life is not scientific. I love you so much!

Table of Contents

Abstract

Nucleic acid-interacting proteins are essential players in cellular processes of all living organisms. In the present study, we explore three psychrophilic-derived proteins involved in gene regulation, RNA degradation and DNA ligation in bacterial cells, with one common feature; they perform their activity by interacting with nucleic acids.

Iron is essential for all living organisms and functions both as a nutritional and regulatory element, however, its toxic potential demands for a tight control of intracellular concentrations. Exclusively for bacteria, iron homeostasis is mediated by the Ferric uptake regulator (Fur); a global transcription factor that controls expression of a wide variety of genes in an iron-dependent fashion. As a key player in bacterial infections, Fur is an interesting target in the fight against pathogenic bacteria. Although Fur mechanisms are well studied, the Fur regulon (Fur box) is still under debate. This study explores the Fur-DNA interaction in the fish pathogen *Aliivibrio salmonicida* in greater depth, to gain a better understanding of the mechanisms behind the disease cold water vibriosis, caused by this pathogen. Electrophoretic mobility shift assays with mutated variants of the proposed Fur box consensus sequences from Vibrios and *E. coli* highlight important nucleotides involved, and rationalization by structural homology models provides new insights into potential AsFur-DNA interactions. New knowledge about Fur mechanisms in *A. salmonicida* provides potential for future development of antibacterial drugs.

After translation to proteins, a complex machinery of multiple enzymes processes the nucleic acid chain of mRNA to smaller oligoribonucleotides and finally monoribonucleotides. In many bacteria, completion of this mRNA decay depends on the enzymatic activity of Oligoribonuclease (Orn). This work reveals the molecular mechanisms in mRNA degradation in metagenomic Orn (MG Orn), isolated from marine Arctic environments. MG Orn degrades short RNA oligonucleotides with lengths from 2 to 10 nucleotides and the determined three-dimensional structure of the enzyme combined with homology modelling demonstrates how these longer RNA chains fit into the active site of the protein. MG Orn is also strictly dependent on a disulfide bond forming a homodimer for functionality. In light of the bioprospecting aspect of our research, MG Orn shows potential as a target in the search for novel cold adapted enzymes in biotechnological applications.

Ligases are enzymes that join DNA fragments with nicks or overhangs, important for many processes in the cell. Motivated by the potential advantages of DNA ligases operating at low temperatures in biotechnological applications, three minimal Lig E-type ATP-dependent ligases originating from psychrophilic bacteria are characterized; ATP-dependent DNA ligase type Lig Es from *A. salmonicida, Psychromonas sp.* Strain SP041 and *Pseudoalteromonas arctica*. Lig E from *A. salmonicida* shows typical cold adapted behavior in terms of temperature optima and thermal stability, and likely determinants for the adaptation to low temperatures are revealed.

The results presented in this work add knowledge to the nature of the nucleic-acid interacting mechanisms of three proteins originating from psychrophilic bacteria, elucidates features behind cold adaptation and identifies potential use in biotechnological applications and antibacterial drug development.

List of papers

Paper I

Biochemical characterization of Ferric Uptake Regulator (Fur) from *Aliivibrio salmonicida*. Mapping the DNA sequence specificity through binding studies and structural modelling. Berg K, Pedersen HL, Leiros I. Manuscript submitted to BioMetals.

Paper II

Characterization of an intertidal zone metagenome oligoribonuclease and the role of the intermolecular disulfide bond for homodimer formation and nuclease activity. Piotrowski Y, Berg K, Klebl DP, Leiros I, Larsen AN. FEBS Open Bio (2019). Volume 9, Issue 10, pp 1674-1688.

Paper III

Temperature adaptation of DNA ligases from psychrophilic organisms. Berg K, Leiros I, Williamson A. Extremophiles (2019). Volume 23, Issue 3, pp 305–317.

Abbreviations and acronyms

DNA AMP	adenosine monophosphate
ATP	adenosine 5'-triphosphate
bp	base pair
DNA	deoxyribonucleic acid
DSC	differential Scanning Calorimetry
dsDNA	double stranded deoxyribonucleic acid
G, A, C, T	guanine, adenine, cytosine, thymine
ΔG	Gibbs free energy
ΔH	enthalpy
kDa	kilo Dalton
mRNA	messenger RNA
MBP	maltose-binding protein
MD	molecular dynamics
min	minutes
Mw	molecular weight
NAD	nicotinamide adenine dinucleotide
Nt	nucleotide(s)
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	Protein Data Bank
pNP-TMP	*p*-nitrophenyl ester of thymidine 5′-monophosphate
PPi	pyrophosphate
RNA	ribonucleic acid
ΔS	entropy
SDS	sodium dodecyl sulphate
TEV	tobacco etch virus
T_m	thermal unfolding temperature

1 Introduction

1.1 Nucleic acids – DNA and RNA

Nucleic acids are macromolecules built of units called nucleotides that are specialized to store, express and utilize the genetic material found in all living organisms. The term *nucleic acids* cover two varieties of polymers; **deoxyribonucleic acid** (**DNA**) and **ribonucleic acid** (**RNA**). The nucleotides, composed of a nitrogenous base, a pentose sugar, and a phosphate group (fig. 1a), are connected by covalent bonds into a chain of alternating series of sugar and phosphate units [1].

The pentose sugar in RNA is *ribose*, whereas DNA has a *deoxyribose sugar*, differing by the absence of a hydroxyl group at the 2' carbon of the pentose in DNA compared to RNA (fig. 1b). The presence of this hydroxyl group in the pentose ring of RNA allows for an additional hydrogen bond and a greater diversity of secondary structure compared to DNA.

The other chemical difference between RNA and DNA lies in the type of nitrogenous base linked to the sugar. The two purines adenine (A) and guanine (G) are both present in DNA and RNA, but the combination of pyrimidines differs between the two. Both DNA and RNA contain cytosine (C), whereas RNA has a uracil (U) present instead of the thymine (T) found in DNA (fig. 1c). These two bases differ by the presence of a methyl group at the C5 position of uracil [1].

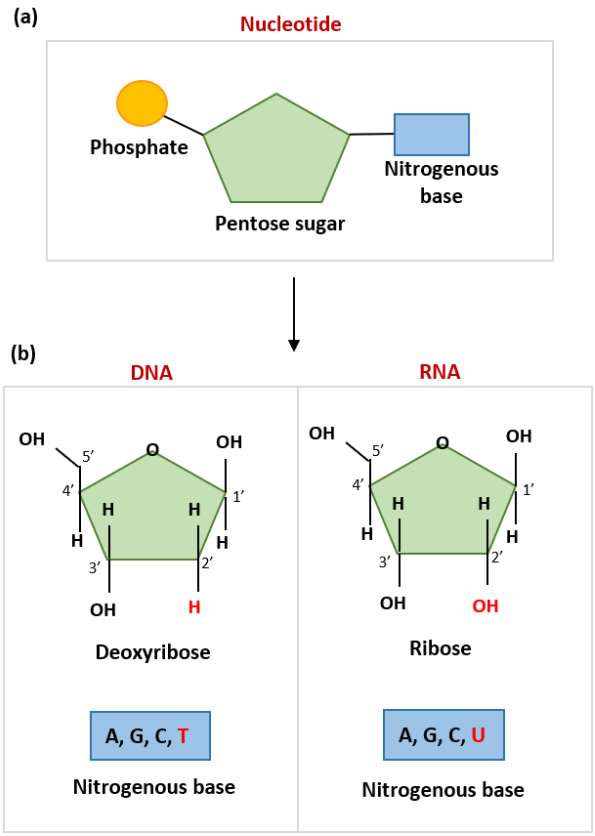


Fig. 1. Comparison of DNA and RNA nucleotides. (a) The building blocks of a nucleotide; a phosphate group (orange), a pentose sugar (green) and a nitrogenous base (blue). **(b)** Nucleotide variations in DNA and RNA are highlighted in red. The top of the panel shows the deoxyribose sugar of DNA to the left and the ribose sugar of RNA to the right, differing by the absence of a hydroxyl group at the 2′ carbon in DNA compared to RNA. The bottom of the panel compares the various nitrogenous bases of DNA and RNA. Thymine (T) in DNA nucleotides is replaced with uracil (U) in RNA nucleotides.

DNA encodes an organism's genetic blueprint and is the storage unit of all the genetic information required to build and maintain an organism. The structure of DNA was first described by Francis Crick and James Watson in 1953 as a twisted structure in the shape of a double helix, with two strands coiled around the same axis [2]. The two strands run antiparallel to each other, stabilized by hydrogen bonds between opposing bases and base stacking (fig 2). The order of the four nucleotides (A, G, C, T) along one DNA strand determines the biological instructions of the genetic code. The other DNA strand

contains the complementary order of the bases due to base-pairing; A pairs with T, while G pairs with C [3].

The double helix of DNA contains two alternating, distinguishable grooves – a wider major groove and a narrower minor groove (fig. 2). The major groove tends to be more involved in protein binding as its size resembles more that of an α-helix in a protein and thus allows for a tighter fit. In addition, DNA-binding proteins can more easily access and recognize the polar and nonpolar groups found within the major groove [4,5].

This Watson-Crick model of the DNA double helix describes the most abundant type of DNA under natural physiological conditions, known as B-DNA. Other alternative forms of DNA, adopted from different ribose sugar confirmations, include the A-DNA and the Z-DNA [6]. Sequence specific protein interactions with the DNA backbone by shape readout often involves the B-DNA conformation, as well as RNA-DNA duplexes and RNA-RNA duplexes [7].

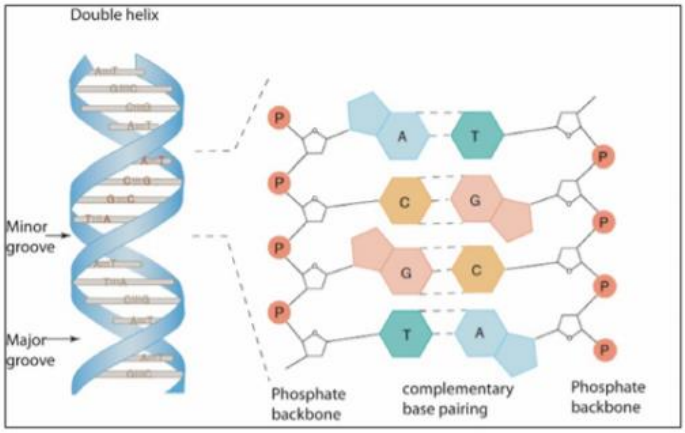


Fig. 2. A simplified model of the common B-DNA double helix, the phosphate backbone and the complementary base pairing. In the structure of DNA to the left, the sticks represent base pairs and the ribbons represent the deoxyribose phosphate backbones of the antiparallel strands in blue. The following schematic representation of the complementary base pairing shows hydrogen bonds between bases represented by dotted lines. The bases are denoted with the letters A for adenine, T for thymine, C for cytosine and G for guanine and the phosphate is denoted with P. The figure is adapted from [8].

In addition to the genetic material found within the cell, extracellular nucleic acids are naturally found within the organism's fluids and in the environment, where they play important biological roles in bacterial ecosystems and in higher organisms. The largest reservoir of extracellular DNA is found in deep-sea sediments, where they serve as a source of energy and nutrition for bacteria in marine and freshwater habitats [9-11]. Uptake and integration of DNA into the cell of naturally competent prokaryotes is further linked to DNA repair, transformation, and generation of genetic diversity, thus recognized as a major force in microbial evolution. In addition to releasing DNA from dying cells, bacteria may also actively produce extracellular DNA to create diversity by horizontal gene transfer [12]. Finally, extracellular DNA is an important component of biofilm formation in various bacteria [12].

1.2 Nucleic acids and proteins – a complex on and off relationship

1.2.1 Nucleic acid interacting proteins in prokaryotes

As the products of the instructions encoded by the DNA, proteins represent the functional components assigned to perform the various activities in the living cell. Proteins exist in many shapes and sizes, reflecting their numerous functions. Many of the proteins within the cell work as enzymes, while others may work as transport molecules across cell membranes or regulate expression of other macromolecules. Nucleic acid interacting proteins recognize and bind to specific or nonspecific sites in DNA or RNA as part of many essential cellular processes in both eukaryotes and prokaryotes; regulation of transcription, translation, DNA replication, repair and recombination and RNA metabolism, all being processes important for our understanding of life on earth. Disruptions of such nucleic acid-protein interactions may lead to serious complications to normal cell function and even survival of the organism.

For DNA recognition, DNA-binding proteins either lack any sequence specificity (non-specific interaction mode) or have specific sequence-recognition requirements (specific interaction mode). For instance, DNA ligases do not exhibit sequence specificity, as they do not discriminate among different nucleotides. The non-specific recognition involves electrostatic interactions between positively charged amino acids and the negatively charged backbone of the DNA, whereas specific DNA recognition involves hydrogen

bonds with bases in the major groove of DNA and non-polar groups recognized by amino acid chains on the protein [13]. The strength of the protein-DNA binding depends both on the type of interaction and specificity, and may thus vary enormously. Nonspecific recognition takes place with essentially the same affinity for all sequences. In comparison, specific DNA binding shows significantly higher affinity to a single sequence. Usually, minor changes in nucleotide sequence are tolerated without loss of protein function. For instance, transcription factors have a preferred sequence with highest affinity, however, individual or even multiple substitutions of the nucleotide sequence does not necessarily affect the ability to recognize and bind, but rather allow for minor variations in affinity. Further, covalent modifications of DNA bases can form more or less favorable interactions and thereby affect the strength of the protein-DNA interaction. These modifications, such as phosphorylation and methylation, are important for the regulation of gene expression [7,14].

All proteins to some degree accommodate nonspecific contact. For instance, in the search of specific binding sites, nonspecific binding with moderate affinity allows proteins to slide along DNA until they encounter their specific high-affinity binding sites and thereby reorient their binding domains relative to DNA in order to establish a stronger and more stable interaction [7].

RNA-binding proteins (RBPs) are key players in regulation of gene expression by being constantly involved in the stabilization or destabilization of mRNAs in response to environmental stimuli. RNA recognition by RBPs depends largely on shape complementarity and interaction with specific bases, hence recognize both sequence and secondary/tertiary structure. The physical forces involved in protein-RNA interactions are similar to interactions with DNA, including electrostatic interactions as salt bridges, dipolar interactions as hydrogen bonds and hydrophobic interactions.

Structures of protein-DNA complexes and amino acid sequence comparisons have identified highly conserved domains or folds defining the nucleic acid interacting function of a protein. Proteins that bind DNA are often composed of two or more domains; a DNA-binding domain (DBD) and an additional domain involved in various functions as ligand binding, protein-protein interactions or enzymatic activity. Described DBDs include helix-turn-helix, winged helix, zinc fingers, leucine zippers and helix-loop-helix. Another common domain that binds both DNA and RNA is the OB-domain (oligonucleotide/oligosaccharide binding folds).

The helix-turn-helix (HTH) domain (fig. 3a) was originally identified as a critical determinant for DNA interaction in bacterial transcription factors [15-17], but have

later been connected to other functions including DNA repair and replication, RNA metabolism and protein-protein interactions. Additionally, HTH is frequently incorporated into the catalytic domains of some enzymes [18]. The helix at the C-terminal end is known as the recognition helix, renders most of the sequence specificity and is embedded into the major groove of the DNA helix upon DNA interaction. The other helix allows for correct orientation of the recognition helix and may interact nonspecifically with the DNA backbone. HTH is the most common DNA binding domain involved in gene regulation in prokaryotes and is identified in the *lac* repressor family, CAP, and many other regulatory proteins [15-17,19,20].

The winged-helix domain represents a subtype of HTH and is involved in establishing protein-DNA-interaction and protein-protein-interactions as well as DNA strand separation or ligation. This domain binds DNA via an additional winged β-sheet with protruding loops [21].

A second important group of nucleic acid-binding domains are the structurally diverse zinc fingers (fig. 3b), characterized by the tetrahedral coordination of one or two zinc ions between a pair of β-strands and an α-helix using conserved cysteine and histidine residues. This type of domain is most commonly found in eukaryotes, although it has been identified in prokaryotes, and shown to have additional roles in RNA packaging, protein folding and assembly, gene regulation and lipid binding [22]. The zinc finger is also identified as a subdomain in NAD^+-dependent ligases [23] and the strictly conserved cysteines coordinating the zinc ion implies the presence of this nucleic acid binding domain in all eukaryotic NAD^+-dependent ligases.

The two α-helices forming the coiled-coil helix dimers of the leucine zipper (fig. 3c) is another common DNA-binding domain, named by the leucines occurring every seven amino acids that mediates dimerization. This domain interacts with the major groove in the DNA via a Y-shaped structure formed by the portion of the two α-helices that are separated from each other in the dimer and typically recognize specific short, inverted repeat sequences via basic residues in its N-terminal end. As for zinc fingers, leucine zippers are present in both eukaryotic and prokaryotic regulatory proteins but are more common in eukaryotes [21].

A fourth important DNA-binding domain is the helix-loop-helix (HLH), which is related to the leucine zipper (fig. 3d). This domain also acts by dimerization and specific DNA binding by basic motifs and consists of a short α-helix connected by a flexible loop to a second α-helix folded against the other, enabling the domain to both bind DNA and to dimerize with a second HLH domain [7].

In the domains described so far, α-helices are the primary secondary structure involved in nucleic acid recognition. The β-sheet is another common domain known to interact with DNA in gene regulation, in which side chains of the amino acids extending from two-stranded antiparallel β-sheet recognize the major groove of the DNA. Thus, the specificity depends on the amino acids that make up the β-sheet [24]. Fig. 3e shows the two-stranded β-sheet that binds to DNA in the prokaryotic *met* repressor protein.

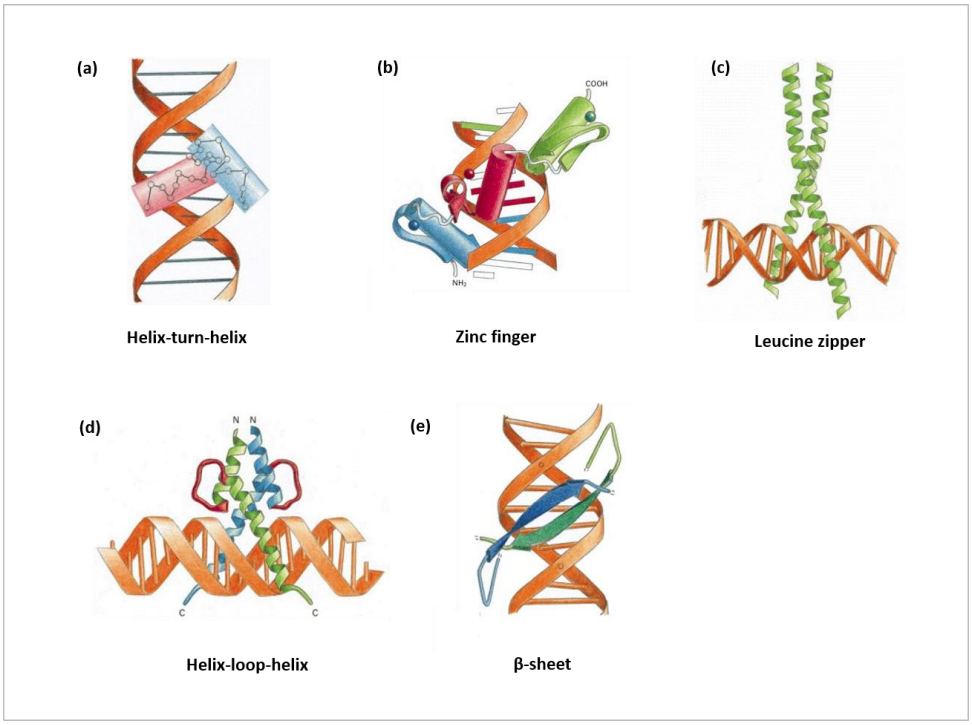


Fig. 3. Common DNA-binding domains in proteins bound to DNA. **(a)** Helix-turn-helix (HTH). **(b)** Zinc finger. **(c)** Leucine zipper. **(d)** Helix-loop-helix (HLH). **(e)** β-sheet. The figures are adapted from [25].

Finally, the OB-domain mediates polynucleotide recognition and binds both DNA and RNA [26]. This domain consists of a five or six-stranded β barrel connected by loops of variable lengths forming the functional unit, which also has the ability to bind ligands and act as an active site. The various sizes of characterized OB-domains (70-150 amino acids) is primarily a result of the variable loop lengths between the well conserved and easily recognized structural elements. The OB-domain is involved in a wide range of

processes that depend on nucleic acid binding; e.g DNA replication, recombination and repair, transcription, translation, cold shock response, and telomere preservation [27]. Studies have revealed that many DNA-damage checkpoints and DNA repair proteins in both eukaryotic and prokaryotic cells possess the OB-fold, including DNA ligases. A general OB-fold topology is shown in fig. 4.

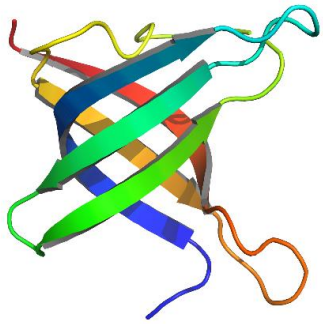


Fig. 4. General nucleic acid binding oligonucleotide/oligosaccharide binding fold (OB-fold) here exemplified by the OB-fold domain of RecO from Deinococcus radiodurans (PDB1W3S).

Oligomers, defined as proteins with more than one subunit at the quaternary structure level, provides an additional level of complexity and play an important role in numerous biological processes [28]. Homo-oligomers describe proteins composed of several copies of identical polypeptide chains, whereas hetero-oligomers have at least one copy of different polypeptide chains. Dimerization or dissociation of dimers (or other oligomers) very commonly affects enzymatic activities as part of allosteric regulation mechanisms [29] and contributes to both conformational and thermal stabilities [30,31]. Furthermore, important genes are regulated by transcription factors that are dependent on oligomerization in response to environmental signals to act on DNA. Forces behind protein oligomerization involve a combination of hydrophobic and polar interactions as hydrogen bonds, salts bridges and occasionally disulfide bonds [32,33].

Metal ions play crucial roles in several metabolic pathways and DNA, RNA and protein synthesis; either functioning as a part of the active site in enzymatic processes or as structural stabilizers in metalloregulators. A range of divalent metals are present in living organisms; Fe^{2+}, Zn^{2+}, Cu^{2+}, Mn^{2+}, Ni^{2+}, Mg^{2+} and Ca^{2+}, the two latter being the

most abundant in living organisms [34]. Inside cells, Mg^{2+} is the most common divalent cation [35]. Mn^{2+} and Ni^{2+} are essential but found at lower concentrations. Divalent metal ions perform their role by using their positive charge for neutralization of phospholipids and nucleic acids, or in specific ligand requirements, relevant for enzymes and transcription factors, respectively. Hydration and specific ligand requirement are also commonly utilized properties of such transition metals.

1.2.2 Gene regulation

As the genome of an organism contains several thousand different genes, all encoding a singular product, it is crucial for the cells to express genes only when their protein product is in demand. Gene regulation in response to the environment in the cell's different stages of life has important implications for the versatility and adaptability, the organizational maintenance, energy conservation and generation of phenotypic variance of the organism.

Cells modulate their gene expression in multiple parts of the process: transcriptional initiation, RNA processing and post-translational modification of a protein [25]. The most efficient way for an organism to regulate genetic expression is at the transcriptional level.

1.2.2.1 Transcription is the hot spot for gene regulation in prokaryotes

Transcription is the first step of translating genes into proteins and involves the step in which a segment of the DNA sequence is copied to an RNA molecule by RNA polymerase through three stages; initiation, elongation and termination, all being potential rate-limiting steps [36]. Several mechanisms regulate RNA polymerase transcription. During the first step of transcription, promotor recognition by RNA polymerase and preinitiation complex formation is one rate-limiting step for gene expression. In prokaryotes, this part of the gene expression process is where most of the regulation takes place due to the lack of the clearly defined nucleus found in eukaryotes, limiting additional gene regulation steps [37]. Regulation of RNA polymerase requires one or more accessory factors for efficient promoter recognition. In prokaryotes, a single accessory factor referred to as sigma locates the RNA

polymerases to its desired promoter as a holoenzyme and thereby affects the specificity and affinity of RNA polymerase promotor binding [37]. In addition, multiple sequence specific DNA-binding proteins regulate binding by RNA polymerase. Repressors and activators, which inhibit or enhance the RNA polymerase-promoter interaction, respectively, are transcription factors that respond to the state of the cell and specifically target the base sequence, often in combination with a ligand [36]. The action of transcription factors, discussed in the following section, allows for unique expression of genes in various cell types in response to a changing environment during development.

1.2.3 Transcription factors and DNA recognition

Transcription factors (TFs) are DNA-binding proteins that regulate gene expression at the transcriptional level. They bind directly to specific regulatory nucleotide sequences upstream of the coding region, or directly to the RNA polymerase molecule, and act either as activators allowing RNA polymerase to bind its promotor and initiate transcription or as repressors that inhibit transcription and subsequent gene expression. These repressors/activators precisely target DNA by specific base recognition, either by themselves or in combination with corepressors or coactivators (ligands), as a response to ligands or small molecule signals to whether the gene expression is needed or not. The well-characterized Lac repressor illustrates this type of gene regulation. When lactose is present, it is converted to allolactose, a small ligand that inhibits DNA binding of the lac repressor, thereby allowing expression of genes involved in lactose metabolism [38].

Various studies of protein-DNA complexes have shed light into the mechanisms underlying the specificity for many TFs. The direct interactions between TFs and specific bases, determined by physical interactions through hydrogen bonds and hydrophobic contacts between amino acid side chains of the TF and the functional groups of the bases, is known as base readout [39]. In addition to recognition of specific bases of their binding site, TFs can also recognize the structural features of the DNA helix, such as sequence-dependent DNA bending and unwinding [53]. This concept of recognizing sequence-dependent DNA structure indirectly is known as shape readout and includes structural readout based on global and local DNA shape features and shape-dependent electrostatic potential. In most cases, the TF-DNA recognition is dependent on both the base- and shape-readout modes [39]. Moreover, TFs commonly form homodimers

which increase their recognition specificity compared to that of a monomer. Dimerization increases the length of recognized sequences and is believed to be important for high-affinity, sequence-specific DNA-binding [40]. However, exactly how target sites are identified *in vivo* and how gene expression is altered remain unresolved.

TF families are diverse and distinctive in their overall protein structure and in which mode they bind DNA, as described for nucleic acid binding proteins in section 1.2.1. They are classified into families based on amino acid sequences and named after their primary function, for instance, the metalloregulator Fur family is named after the ferric uptake regulator (Fur). Other well described prokaryotic TF families include the LacI, AraC, LysR, CRP and OmpR families [41].

1.2.4 Comparing three essential types of nucleic acid binding proteins; a transcription factor, a nuclease and a ligase

Understanding the mechanisms behind nucleic acid binding by proteins and identifying the nucleic acid sequence/structure involved in protein-nucleic acid complexes is vital to interpret the function of these complexes in cellular processes and in disease development.

The Arctic marine-derived nucleic acid binding proteins included in this study are:

1. Ferric uptake regulator (Fur).
2. Oligoribonuclease (Orn).
3. ATP-dependent DNA ligase type Lig E.

In general, DNA or RNA complexes involving these types of proteins are important for regulation of transcription, RNA processing and DNA replication, repair and recombination, thus cover various stages of normal cell development in bacteria. In addition to their various biological roles, the proteins addressed in this work also differ in terms of nucleic acid binding modes, structure and function (Table 1). Although the three individual proteins are presented in more detail in the subsequent sections, a short and general comparison is given here to highlight common aspects as well as differences.

Table 1. General comparison of the different nucleic acid binding proteins investigated in this study.

	Fur	Orn	Lig E
Nucleic acid preference	DNA	RNA	DNA
Protein function	Gene regulation	mRNA degradation	DNA ligation
Enzymatic activity	No	Yes	Yes
Metal ion dependence	Yes*	Yes	Yes
Base recognition	Specific	Non-specific	Non-specific
Oligomeric state	Dimer/Monomer	Dimer	Monomer
Host necessity	Yes	Yes	Unknown?
Pathogenic relevance	Yes	Yes	No

* In general, Fur activity depends on a metal ion, however, it may function without in certain species.

Type of activity

The common feature of all three proteins is their ability to recognize and act on nucleic acids and their involvement in specific DNA/RNA transactions, although the exact biological role of Lig E is not clear. Fig. 5 gives a simplified overview.

Fur is generally described as a metal-dependent DNA-binding regulator that controls the expression of a variety of genes, directly or indirectly, in order to maintain metal homeostasis in bacteria [42-44]. Orn enzymatically degrades small RNA to mononucleotides in the last step of mRNA turnover post translation, a process that influences overall levels of gene expression in bacterial cells [34,45,46]. Thus, both Fur and Orn contribute to regulation of gene expression, but at different stages in the DNA-mRNA-protein pathway; transcriptional initiation and RNA processing respectively. Like Orn, Lig E also has enzymatic activity, ligating breaks in dsDNA using energy derived from the cleavage of ATP [47].

Orn catalyses the breakage of bonds and belongs to the class of enzymes called the hydrolases, while Lig E catalyses the joining of bonds and belongs to the class of enzymes called ligases.

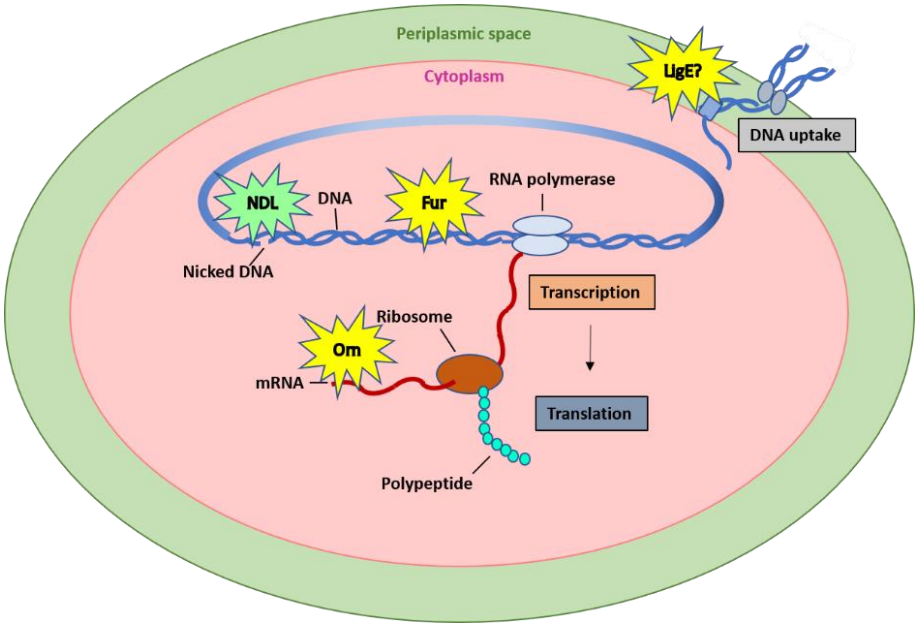


Fig. 5. Overview of various biological processes in the bacterial Gram-negative cell, highlighting the role and location of the proteins in this study (yellow stars); Ferric uptake regulator (Fur), Oligoribonuclease (Orn) and the bacterial ATP-dependent ligase Lig E. The essential NAD^+-dependent ligase (NAD) is added for comparison (green star). Note that the role and location of Lig E is unknown, thus its involvement in DNA uptake remains speculative.

The three proteins are active in various cellular locations as part of different cellular processes (fig.5). Orn is involved in cellular nucleotide recycling in the cytoplasm, Fur controls transcription in the cytoplasm, whereas Lig E has been suggested to be translocated to the periplasm by an N-terminal signal sequence [48-50].

Protein structure and conserved domains for binding

The domain arrangements in proteins are functionally important for their ability to bind nucleic acids. The DNA/RNA binding site of a protein is often found in various discrete conserved domains, as described in section 1.2.1.

The Fur protein includes an N-terminal DNA-binding domain (DBD) with a winged-helix motif and a C-terminal metal-binding domain that is involved in dimerization (DD) (fig 6a). A flexible hinge connects the two domains. Fur is often described as a

homodimeric protein, but may exist in several oligomeric states both in solution and in complex with DNA [51-55]. The dimeric interface of Fur crystal structures is primarily formed by the DD via intermolecular hydrophobic interactions [56]. Conserved key residues in the DBD of the Fur protein recognizes the DNA target through both shape readout and specific base recognition [56]. To enable specific residues to interact with DNA upon binding of Mn^{2+} and DNA, the DBD undergoes both conformational changes and movements of the DBD orientation.

The structure of Orn shows a typical DnaQ-fold with a DEDD domain of five-stranded β sheets flanked by alpha-helices (fig. 6b). As a member of the DEDDh superfamily of exoribonucleases it contains four sequence motifs unique to oligoribonucleases; exo I, exo II exo III and exo IV. Four highly conserved acidic residues and a histidine, essential for binding divalent cations and thus for catalytic activity, are clustered in the active center within the ExoIII domain. Upon RNA substrate interaction, the substrate is bound by a substrate binding surface of one monomer and led into the DEDDh cavity of the other monomer [57].

Lig Es act as monomers and have a minimal modular architecture consisting of a unique arrangement of two discrete domains; catalytic adenylation domain (AD) and oligonucleotide-binding domain (OB) (fig. 6c). In addition, Lig Es have a predicted N-terminal signal directing the mature protein to the periplasmic space of the cell. Lig E type ADLs have open and closed conformational modes and the linker region is involved in coordinating rearrangement of the domains relative to each other to form the C-shaped clamp around the DNA substrate (closed mode), with the nick positioned above the AMP-binding pocket [58]. Further, analysis have shown that Lig E recognizes its DNA substrate using well-ordered and conserved basic residues on the AD and OB domain that contribute to electrostatic interaction between protein and DNA [50].

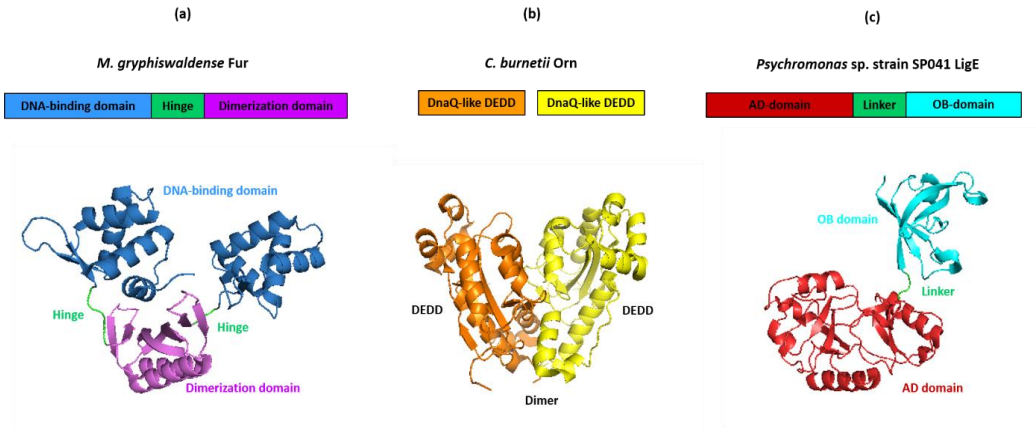


Fig. 6. Conserved structural domains of Fur, Orn and Lig E. (a) Dimeric Fur from *M. gryphiswaldense* (structure from PDB 4RAY). **(b)** Dimeric Orn from *C. burnetii* (structure from PDB 3TR8). **(c)** Lig E from *Psychromonas* sp. strain SP041 (structure from PDB 4D05). All structures are generated in Pymol and colored by domains.

Nucleic acid binding mode

Proteins recognize nucleic acid bases in either a sequence-specific or secondary structure-dependent manner, through major or minor groove interactions, and create complexes with various interaction strengths.

Although Fur binds specific DNA elements within target promotors, it appears to bind rather degenerate substrates with DNA complex formation based on both shape readout recognition and interaction with specific bases. Typical binding sites are the narrow minor groove of palindromic AT-rich sequences. In the typical binding of dimeric Fur to its target, two homodimers are positioned on opposite sides of the double helix and interact with each other [56]. However, other binding modes are demonstrated; dsDNA bound by a single Fur dimer, tetrameric Fur and even cooperative binding of several Fur dimers [56,59,60]. Fur has also been suggested to act upon DNA as a monomer in its apo-form, which further supports its role as a global regulator with a broad substrate affinity [61,62].

In contrast to the regulator Fur that shows affinity for specific nucleic acid sequences and shapes, Orn and Lig E are enzymes that bind in a more unspecific manner, depending on a free 3' hydroxyl group terminus in small RNAs or a free 5'-phosphate in nicks of DNA, respectively.

Orn is generally described as a processive 3' - 5' exonuclease that binds and hydrolyses single stranded small RNAs strictly as a dimer in a non-specific and metal-dependent manner. Also demonstrated is binding affinity for short ssDNA oligos, but with weaker affinity requiring considerably higher enzyme concentrations [63].

Lig E, on the other hand, catalyzes the formation of phosphodiester bonds at single-stranded breaks in dsDNA by esterification of a 5'-phosphoryl to a 3'-hydroxyl group located on opposite sides of a break in the phosphodiester backbone [64]. The 5'-phosphate is absolutely required for discrimination between nicked and linear DNA [65]. Structural studies of two Lig E enzymes showed that the they bind asymmetrically to nicks [58]. The AD domain is involved in positioning the AMP-binding pocket across the nick, whereas OB domain basic residues stretches along the complementary strand in a 5' to 3' direction. Specific side-chain contacts are formed between the OB domain and five of the six complementary nucleotides opposite the nick. In addition to single nicks, some Lig Es are able to recognize and act on 4 base-pair cohesive ends, mismatches at the nick site and to a lesser extent gaps [58].

Cofactors and metal ions

Many cellular processes are dependent on the availability of the appropriate metal cofactor. Fur was originally described as metal-dependent for dimerization and DNA-complex formation, but recent studies have suggested that it may act in both apo- and holo-forms [66]. The classical Fur regulation model involves binding of the divalent cation Fe^{2+} to the monomer, stimulating dimerization and subsequent DNA binding. In vitro, Fur can be activated to bind DNA by a range of divalent metals; Fe^{2+}, Fe^{3+}, Co^{2+}, Mn^{2+} and Zn^{2+} [67]. The residues mediating metal binding are conserved among Fur proteins.

Where Fur utilizes metals as a ligand solely for structural purposes, Orn and Lig E both depend on metal coordination in their active site for catalytic activity. Orn and other DEDD family exonucleases share common active site geometry with the four acidic side chains coordinating the cofactor, either Mn^{2+} or Mg^{2+}, but preferably the latter [68]. Similarly, Lig E is strictly dependent on Mg^{2+} or Mn^{2+} for activity with with Mg^{2+} being the preferred cofactor [49].

Biological importance for the organism

Strict regulation of free iron levels in the cell is crucial for survival in bacteria, and iron homeostasis is regulated primarily by Fur. Regulation by Fur is restricted to bacteria, mostly Gram-negative Proteobacteria (Gram-negative). Exceptions have been identified in certain Gram-positive bacteria in which the diphtheria toxin repressor (DtxR) performs the same role [69]. Although not essential for bacterial survival, Fur-depleted strains have shown impaired cell growth and reduced ability to cause infection [70-72].

Orn hydrolysis of RNA oligonucleotides is essential for completion of the life cycle, as accumulating oligoribonucleotides in the cell reduces cell viability. However, cell viability in the absence of Orn has been documented for some bacteria [73], with the cost of slow cell growth [74]. In addition, some bacteria encode a functional analogue instead of Orn [75]. In contrast to Fur, Orn has a human homologue, Sfn [63], and is thus not restricted to bacteria. As both Fur and Orn have been shown to play a role in bacterial pathogenesis, they are considered as interesting targets for the search of novel antibiotics and drug development.

ADLs, on the other hand, are essential to eukaryotes and their distribution among bacteria is limited, as most bacteria contain only the housekeeping NAD^+-dependent ligase. The biological function of the minimal ADL type Lig E remains unknown, however it has been speculated that they could be involved in natural competence and DNA uptake based on their putative periplasmic location [48,50]. Both Lig E and Orn have potential uses in biotechnological applications, which underlies interest in understanding these enzymes.

1.3 DNA binding by the Ferric uptake regulator (Fur)

1.3.1 Iron homeostasis in bacteria

Iron is an essential nutrient for growth and host colonization in many bacteria. Despite its considerable supply in nature, the low solubility of the dominating oxidized ferric iron restricts the availability of the nutrient to the bacteria [76]. This

is counteracted by expression of high affinity transporters and siderophores to chelate iron by the bacteria [77,78]. However, when found in excess, iron is potentially harmful. If intracellular levels of free iron are not properly maintained, excessive levels may interact with reactive oxygen species through the Fenton reaction and produce free radicals that effectively damage DNA, RNA, proteins and other cellular components [79,80]. To prevent toxicity and simultaneously fulfill nutrient requirements, the intracellular availability of iron must be tightly controlled, and consequently bacteria have evolved various mechanisms to maintain iron homeostasis. High-affinity uptake systems of iron and iron utilization are controlled at the transcriptional level [81]. In bacteria, the ferric uptake regulator (Fur) is the main iron-sensing transcription factor, also described as a global regulator able to control the homeostasis of various metal cofactors.

1.3.2 Ferric uptake regulator - a global regulator involved in pathogenesis

Fur is a global metalloregulator found in all proteobacteria and controls the expression of a wide variety of genes (more than 100) involved in metal uptake, storage and consumption [42-44]. The most described role of Fur is as a repressor in response to high levels of free iron in the cell. In addition to metal homeostasis, Fur plays a role in regulation of the expression of mRNAs encoding for proteins implicated in energy metabolism, acid and oxidative stress defense, nitrogen metabolism, signal transduction, transposition, redox regulation, cell morphology and motility, virulence, protein glycosylation, flagella biogenesis and others [82,83].

New insights have shown that metals inhabiting the microenvironment at an infection site strongly influence bacterial pathogenesis and host immunity [84]. During microbial invasion there are many battles going on between the host and the pathogen, and the battle for iron is one of them. As iron is an important nutrient for successful infections, the host strive to withhold extracellular free iron as an attempt to starve the pathogen, while the bacteria sense the lack of iron and express a range of high-affinity iron uptake mechanisms, controlled by Fur in many species. Experiments have shown that high iron access is correlated with stronger infections. For instance, clinical studies have led to the consensus that iron deficiency is protective against malaria, and iron supplementation increases malaria risk in the absence of access to adequate health care [85].

The ability of Fur to regulate expression of virulence factors contributes to bacterial pathogenicity in many microorganisms; *Pseudomonas aeruginosa* [86-88], *Yersinia pestis* [89,90], *Francisella tularensis* [91] and *Escherichia coli* [92]. Extensive research has been conducted in the recent years to shed more light into the relationship between Fur and virulence [66,81]. Fur knockout mutants of several bacterial species have demonstrated high variability in phenotypes among species, affecting expression of iron uptake systems, virulence factors, resistance to acid, serum, or oxidative stress resistance, motility and biofilm formation [81]. Interestingly, some pathogens with depleted *fur* genes show impaired ability to invade and infect their hosts compared to their wild type counterparts. Fur deletion mutant strains have been shown to exhibit reduced virulence in infection models; *Staphylococcus aureus* [93], *Helicobacter pylori* [94], *Listeria monocytogens* [95] and *Vibrio cholerae* [96]. In addition, a Fur mutant in *Campylobacter jejuni* showed reduced ability to colonize the gastrointestinal tract of chicks [83]. Further, microarray analyses, electrophoretic mobility assay (EMSA) and footprinting studies have revealed Fur regulated genes involved in pathogenesis in *Neisseria meningitidis* and *Neisseria gonorrhoeae* [97]. It is however worth mentioning that the Fur relevance in pathogenesis is species specific, or even specific for the particular infection model. For instance, Fur mutant strains of the pathogenic *E. coli* and *Vibrio vulnificus* showed similar levels of pathogenicity as their wild-type parental strains in animal models [98,99].

Since no homologues are found in eukaryotes, Fur may serve as a potential target in the development of novel antimicrobial agents.

1.3.3 Fur mechanism

In the classical model of Fur regulation, two Fur monomers bind Fe^{2+} under iron-rich conditions, form a dimeric complex that recognizes the Fur regulon and block transcription of iron-regulated genes (fig. 7, left panel). In contrast, under iron starvation conditions, Fur does not dimerize to bind DNA and expression of iron-acquisition genes initiates as RNA polymerase can access its promotor.

Most commonly, the affinity of Fur for target promoter regions increases upon metal binding, leading to repression or activation of the transcription of the downstream genes. However, the Fur mechanistic repertoire is more complex and has recently been expanded. In some strains, like *H. pylori*, *C. jejuni* and *N. meningitidis*, it has been

demonstrated that Fur may both activate and repress gene expression in the presence or absence of regulatory iron, forming a total of four different modes of gene regulation (fig.7) [66,82,100,101].

Ferrous iron is the preferred metal physiologically, but in vitro studies have shown that Fur from *P. aeruginosa* (PaFur), *Y. pestis* (YpFur) and *E. coli* (EcFur) can be activated by other divalent ions such as Zn^{2+}, Co^{2+}, Ni^{2+}, and Mn^{2+} [67,89,90,102,103]. These divalent metal ions are more stable than Fe^{2+} and therefore more commonly used experimentally.

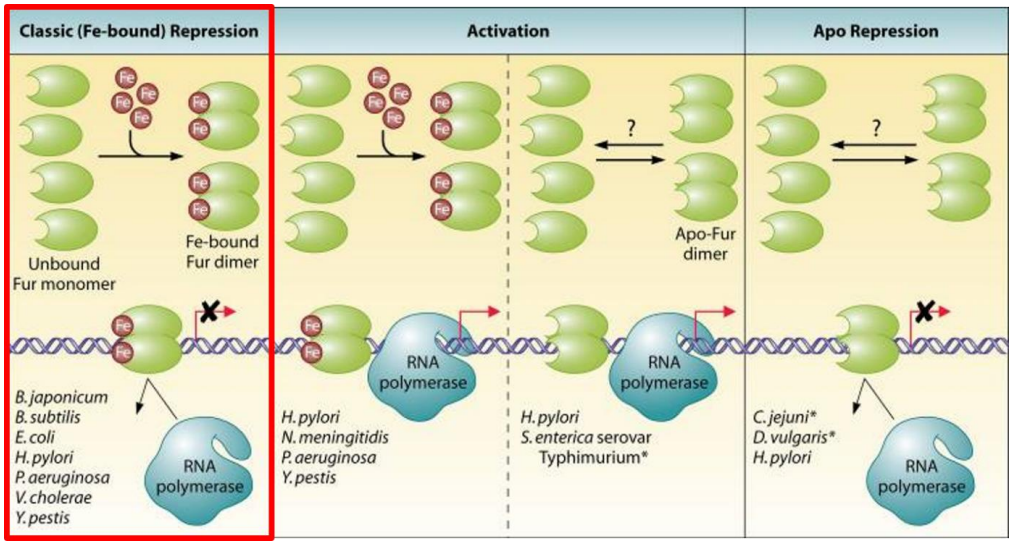


Fig 7. Overview of Fur gene regulation mechanisms; Fe-Fur repression and activation, and apo-Fur repression and activation. The different features of Fur are illustrated both prior to target binding (top) and in complex with DNA. The classical iron-bound Fur repression studied in this work is shown to the left, boxed in red. The bottom of each panel lists some of the organisms that are shown to utilize each type of Fur regulation. An asterisk (*) indicates organisms for which apo-Fur regulation has been indicated but direct interaction remains to be confirmed. The question mark under apo-Fur regulation indicates that it is unknown whether these act upon DNA as monomers or dimers. The figure is adapted from [66].

1.3.4 Fur structure

Crystal structures of Fur from numerous bacterial species have been determined so far: *P. aeruginosa* Fur [104], *H. pylori* Fur [105], *C. jejuni* Fur (CjFur), both including and devoid of any regulatory metals [82,106], *Magnetospirillum gryphiswaldense* Fur (MgFur) [56], *Vibrio cholerae* Fur (VcFur) [107] and *Francisella tularensis* Fur [108]; the latter being the first solved structure of a tetrameric Fur protein with physiological iron bound. Two of the recently solved structures of MgFur represent the first structures of Fur-DNA complexes to be released [56]. In addition, the structure of the DNA-binding domain of EcFur was solved by X-ray crystallography, while its dimerization domain was characterized by nuclear magnetic resonance (NMR) [109].

Fur folds into two distinct domains consisting of an N-terminal helix-winged-helix DNA-binding domain (DBD) linked by a hinge region to a C-terminal dimerization domain (DD) which typically consists of three antiparallel β-strands and two α-helices [104,107] (fig. 6a). The metal-bound Fur homodimer exhibits similar overall structures among different species with a canonical V-shaped conformation, despite relative low sequence homology (24% to 49%). Apo-CjFur shows the characteristic canonical V-shaped dimer, with two zinc ions per monomer, but sets out from other known Fur structures by a 180° reorientation of apo-CjFur's DBD via the hinge region [82]. Metal ions mediate the binding of Fur to its DNA target, and two or three functional metal binding pockets are usually found within different structural sites (designated S1, S2 and S3). These metal-ion binding sites and the residues coordinating them are rather diverse in bacterial species and until recently their roles were poorly understood [82,104,105,107,109].

Recently, the first Fur-DNA complexes and apo-Fur structure without transition metals were solved by Deng *et al* [56], bringing new insight into the field of metal binding, molecular mechanisms and structural basis of Fur-DNA binding (fig. 8). In contrast to all previously solved Fur structures, in which zinc ions were used, MgFur is bound to Mn^{2+} for reversible binding. As Zn^{2+} is bound and coordinated different from Fe^{2+}, incorporation of Mn^{2+} might reflect a truer physiological state of Fe^{2+} binding and furthermore, residues coordinating Mn^{2+} are conserved among Fur proteins. The holo-Fur structure identified two Mn^{2+} binding sites per monomer; S1 and S2, the first linking the DBD and the DD domains and the second is found almost entirely in the DD domain. Analysis of the apo-Fur dimer structure and binding studies of wild-type MgFur and Fur mutated at S1 or S2 in each monomer revealed possible functions of these transition metal sites. Dimerization or secondary structural characteristics of apo-

MgFur is independent of metal ions and both metal sites are thus more involved in regulation and modulation of Fur activity. S1 is essential for DNA binding *in vitro*, whereas S2 only slightly affects DNA binding and its role remains unclear. However, these two sites were suggested to both act in gene regulation as an "on-off switch" (S1) and "fine tuner" (S2), the latter ensuring repression of genes less sensitive to iron [56].

The binding of two Mn^{2+} ions in the holo-Fur homodimer stabilized the hinge area and induced profound conformational changes of the DBD compared to the apo-Fur structure (fig. 8). Furthermore, binding of DNA and Mn^{2+} in the Fur-DNA complex structure triggered both conformational changes (stabilized by Mn^{2+}) and reorientation of DBD, with DNA binding residues in closer proximity to DNA (fig. 8c).

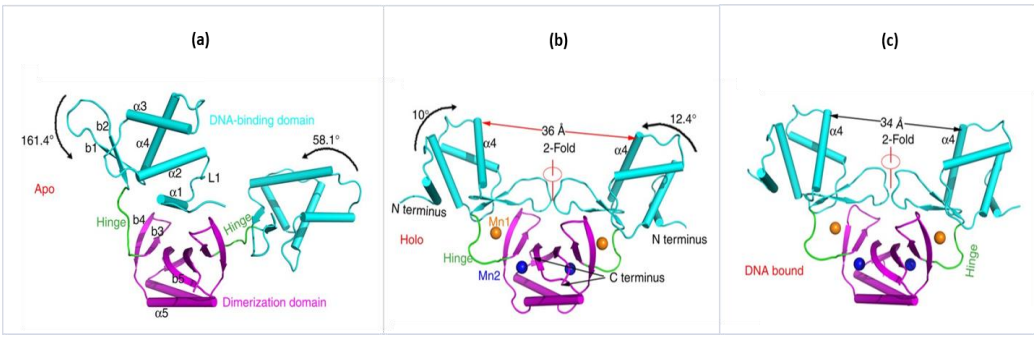


Fig. 8. Tube representation of the different structural conformations of MgFur in apo-form, holo-form and DNA bound-form. **(a)** Apo-Fur showing different conformation of the two monomers. **(b)** Holo-Fur activated by manganese at S1 (orange) and S2 (dark blue). **(c)** DNA-bound Fur-dimer. Holo-Fur and DNA bound Fur show related conformations of their monomers, around a twofold rotation axis. Structural domains are labeled by colours; N-terminal DBD (cyan), hinge (green) and C-terminal DD (magenta). The figure is adapted from [56].

1.3.5 The Fur regulon (Fur-box)

Fur proteins bind with high affinity to specific Fur binding sites known as Fur boxes, located between the -35 and -10 sites at the promotors of Fur-regulated genes [110].

Based on sequence alignment of numerous iron-controlled promoters of various origins and DNase I footprinting experiments of EcFur, the classical Fur box (fig. 9a) was first described as a 19 bp (9-1-9) inverted repeat recognized by a single dimer; 5'-GATAATGATAATCATTATC-3' [51,111]. However, during the last decade Fur boxes from

other genera have also been characterized and described [53,111-114]. In contrast to the A/T centered classical *E. coli* Fur box, many of these are centered on a degenerate nucleotide positioned 3 nt downstream.

As further studied indicated that several Fur dimers may bind two or more overlapping Fur boxes and even polymerize along the DNA outside of the Fur box region [53], it became apparent that the classical 19-bp Fur box sequence did not fit with the tendency of Fur to polymerize. Escolar *et al*. suggested a revised model where two overlapping dimers of Fur interact with the AT-AT core within three repeated arrays of GATAAT in *E. coli* (the third array slightly imperfect) [53,111]. Fig 9b illustrates how the GATAAT hexamer can be interpreted as three direct repeats in a tandem array F-F-F fashion, or rather with the last hexamer inverted to an F-F-R arrangement or with mismatches. Alternatively, as found most frequently in natural Fur binding sites, the F-F-R model can be arranged with a single bp separating the two direct repeats and the inverted repeats, as illustrated by the hexamer model F-F-x-R in fig. 9c [111], although how Fur dimers bind is unclear.

New insight to how Fur recognize its target was suggested by Lavrrar *et al* [115]. Two Fur dimers were predicted to interact with a 13 bp overlapping 6-1-6 motif (fig. 9d) from opposite faces of the helix [115]. Similarly, based on alignment of *Bacillus subtilis* Fur boxes and DNase footprinting, a slightly extended Fur box consensus sequence (21 bp) consisting of two overlapping 7-1-7 motifs was proposed (fig. 9e) [112,116]. A similar binding model was shown in the crystal structure of DtxR bound to its operator site [69,117].

Further studies by Baichoo *et al* indicated that this 7-1-7 heptamer is a minimum, since single 6-mer or 7-mer nucleotides showed no affinity to Fur and Fur boxes with two 6-mers interacting weakly [116]. Similarly, a minimum of three repeats of the hexamer motif GATAAT was required for Fur binding to the *E. coli* Fur box [53]. These results demonstrated a shift of focus from specific sequences/lengths of the Fur box towards the functional pattern within the sequence and introduced the AT-rich consensus hexamer NATA/TAT as the main unit of interaction with Fur, regardless of orientation and number. Thus, Fur appears to have a broad substrate-binding ability, supported by the range of sequence identities between proposed Fur boxes in various bacteria (50% to 80%) [112,118,119].

The high degree of degeneration in the Fur box and the ability to bind DNA at different ratios was demonstrated by gel shift-based assays and crystal structures of MgFur in complex with two different DNA targets; the *feoAB1* operator mutated to a near-

perfect inverted repeat and the typical 7-1-7 hexamer *P. aeruginosa* Fur box (identical in sequence to the *E. coli* Fur box) [56]. Although Fur was shown to bind these two different targets as one dimer and two dimers, respectively, both complexes showed contacts with both DNA strands using its DBD-domain. An AT-rich narrower minor groove with enhanced negative electrostatic potential, bound by the positively charged Lys15, as well as an important G base and a conserved T base, were highlighted as essential for MgFur interactions. Additional *in vivo* experiments indicated that specific Fur-DNA contacts may be directly connected to DNA shape instead of being base specific. Thus, shape readout, rather than the specific Fur box sequence, may define Fur box recognition by Fur.

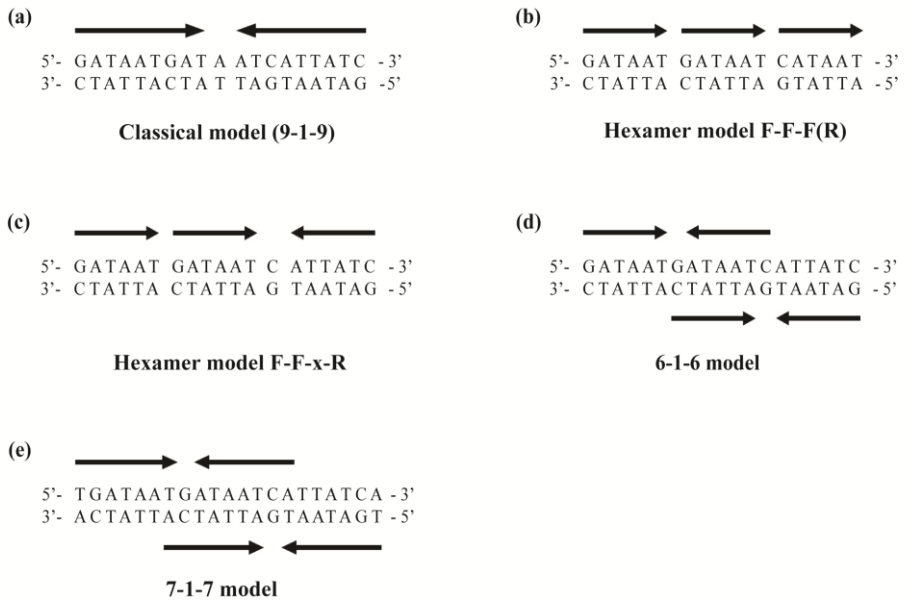


Fig. 9. Alternative arrangements of the Fur box, illustrated by the *E. coli* consensus with GATAAT as the minimal recognition unit (Fur consensus NATA/TAT). Arrows mark inverted repeats/repeated arrays. **(a)** The 19 bp classical model suggests two inverted repeats with an A:T basepair in between, binding a monomer each. **(b)** The 18 bp hexamer model contains a minimum of three direct repeats of the hexamer GATAAT, where the AT-AT pattern within each hexamer was suggested to interact with Fur. The last hexamer may be reversed or imperfect. **(c)** The 19 bp hexamer model is described as repeated arrays of three or more copies of GATAAT motifs, recognized by two hexamers in the forward direction and one hexamer at the reverse orientation, separated by one base pair. **(d)** The hexamer model can be viewed as a 6-1-6 arrangement, where two overlapping hexamer inverted repeats binds Fur dimers at opposite faces of the double helix. **e**) The 21 bp 7-1-7 model defines the Fur box as two overlapping heptamer inverted repeats, also recognized by two Fur dimers at opposite faces.

1.3.6 Characterization of Fur from the fish pathogen *Aliivibrio salmonicida*

Vibrios belong to the *Vibrionaceae* family (γ-proteobacteria). Since many Vibrios are pathogenic to mammals, this group of bacteria has received significant attention in research. Aquaculture has faced challenges due to diseases caused by fish pathogens like *Vibrio Anguillarum*, *Aliivibrio salmonicida*, *Vibrio wodanis* and *Vibrio harveyi*. The Gram-negative *A. salmonicida* is the causative agent of cold water vibriosis in farmed Atlantic salmon, Atlantic cod and rainbow trout [120]. The pathogenesis of this fish pathogen is a rather unexplored field; however, genome analysis has revealed genes encoding for several iron acquisition systems (including a *fur* encoded gene), shown to be important for the virulence of many pathogens, including Vibrios. Thus, *A. salmonicida* is an interesting model organism for studying mechanisms behind iron homeostasis and better understand Vibrio virulence and pathogenesis.

The global regulator Fur from *A. salmonicida* (AsFur) has previously been recombinantly expressed and studied experimentally and computationally by our research group, providing new insight into potential AsFur-DNA interactions [121,122]. AsFur consists of 147 amino acids with a theoretical molecular weight of 16.6 kDa and a theoretical pI of 5.75. Sequence comparisons show 83% identity to VcFur and 79% identity to EcFur. AsFur exists mainly as a homodimer in solutions.

Structure modelling and molecular dynamics simulations proposed a Vibrio Fur box consensus that was specifically recognized and bound by purified AsFur in EMSA studies [123]. The proposed 19 nt inverted repeat Vibrio consensus, 5'-AATGATAATAATTATCATT - 3', is conserved among diverse bacterial species and fits with both the 7-1-7 model and the hexamer F-F-x-R model in fig 9, as well as proposed NATA/TAT hexamer units. In addition, AsFur was shown to bind several other promotor regions predicted to be Fur-regulated, depending on the number of Fur boxes and their predicted strengths. Finally, EMSA experiments indicated that AsFur polymerized on DNA, which is in line with the observed binding of Fur outside the Fur box region in Vibrios [121].

Molecular dynamics and binding free energy calculations further identified unrecognized Fur amino acid residues and Fur box nucleotides that might be directly involved in Fur dimerization and DNA interactions [122]. The amino acid residues R57, Y56 and R70 demonstrated a significant DNA binding potential, and the first two have been shown to contribute to DNA binding in MgFur [56]. In the search for Vibrio-specific Fur box nucleotides, two new nucleotides (A14 and C16) were proposed to interact

with AsFur binding via both strands, whereas T13 was shown to contribute from one strand. Combined with the recently published three-dimensional structure of MgFur-DNA complexes, these results provide potential for further characterization of Fur-DNA complexes by substitutions of specific residues in the Fur protein or nucleotides in the Fur box.

1.4 The role of oligoribonucleases (Orn) in mRNA decay

1.4.1 mRNA metabolism in prokaryotes

Messenger RNA (mRNA) is a single-stranded molecule that carries genetic code from DNA to the ribosome, from where it serves as a template in the translation into functional proteins (fig. 5). Messenger RNA metabolism includes all the processes required for RNA synthesis, maturation and degradation in living cells.

Messenger RNA molecules in prokaryotic and eukaryotic cells differ significantly in many aspects. In prokaryotes, each mRNA can carry genetic code for synthesis of several proteins, transcribed by several genes of an operon. Unlike eukaryotes, where mRNAs transports to the ribosomes in the cytoplasm after transcription, prokaryotes lack a distinct nucleus and start translation on mRNA before transcription is completed. Hence, these two processes are coupled in prokaryotes. Next, prokaryotic mRNAs are constantly broken down by ribonucleases (RNases) and considering the short life-time (average half-life of about two minutes), one end of mRNA is being degraded simultaneously with the other end being translated. This mRNA instability allows transcriptional control of gene expression and is a vital process that enables prokaryotes to quickly respond to changes in the environment by synthesizing different proteins when needed [124].

The rapid enzymatic degradation of mRNAs is called mRNA decay. A series of ribonucleases are utilized in mRNA turnover by multiple pathways. Many studies reviewed in [125] have contributed to understanding the molecular mechanisms behind mRNA decay, but there are still unresolved fundamental aspects related to mRNA decay and gene regulation. Interestingly, many features of this process differ between prokaryotes and eukaryotes as well, for example different functions of

polyadenylation of mRNAs in prokaryotes and composition and biochemical mechanisms of mRNA degrading protein complexes [126].

Enzymes of the RNase superfamily are hydrolytic enzymes that degrade RNA molecules into smaller components, at both transcriptional and translational level, and play an important role in all aspects of RNA metabolism. A range of RNA-modifying and RNA-processing enzymes are involved in processes from RNA synthesis to decay. Here, focus is on RNases involved in mRNA degradation.

RNases act directly on RNA molecules and catalyse the exo- or endoribonucleolytic cleavage of phosphodiester bonds, classifying them as exoribonucleases or endoribonucleases, respectively. Based on extensive sequence analysis and catalytic properties, eight 3'- 5' exoribonucleases are described in *E. coli* [127]; polynucleotide phosphorylase (PNPase), RNase PH, RNase BN, RNase II, RNase R, RNase D, RNase T and oligoribonuclease.

Studies of the two distantly related model organisms *E. coli* and *Bacillus subtilis* have proposed potential pathways and enzymes involved, but also illustrate the wide diversity of mRNA decay mechanisms in prokaryotes [128-130]. Messenger RNA degradation is initiated by the endo- and exoribonucleases. In the model prokaryotes studied, the three endoribonucleases RNase E/G, RNase J and RNase Y are responsible for mRNA cleavage into two smaller fragments with new unprotected 3' ends. All sequenced prokaryotic genomes contain at least one of these nucleases [124], thus reflecting mRNA turnover adaptation through different ecological niches throughout evolution. Next, the cleaved mRNA fragments serve as substrate for further degradation by exoribonucleases [131], represented by RnaseII, PNPase, RNase R in *E. coli*, in a 3' -5' direction. The resulting nanoRNA fragments in the range 2-5 nt are finally converted to mononucleotides by oligoribonucleases, completing the mRNA decay process [124]. For the Gram-positive *B. subtilis*, the nanoRNAs are degraded by NrnA, NrnB or an unidentified enzyme [132]. Fig. 10 provides an overview over the mRNA decay pathway as described for *E. coli*.

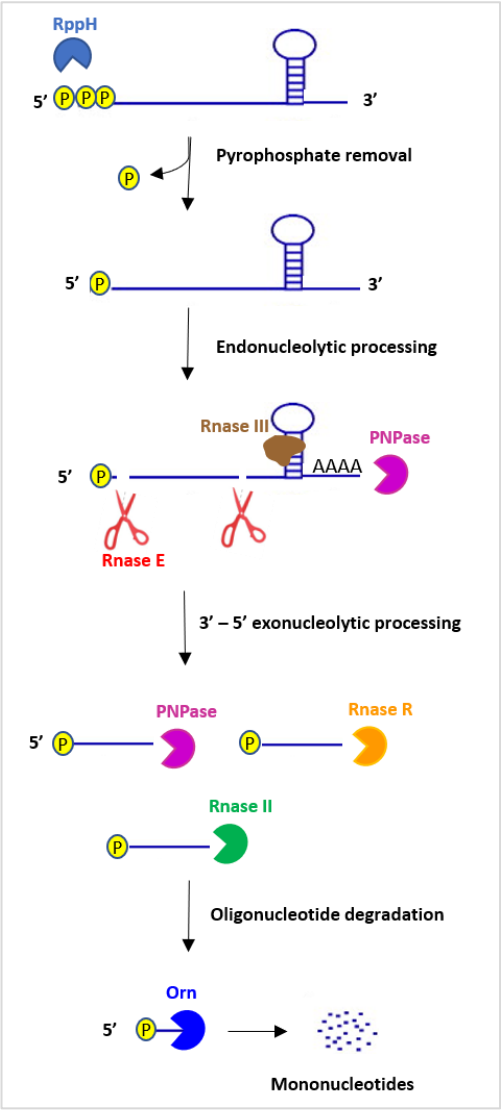


Fig. 10. Simplified mRNA decay pathway in *E. coli*. As the first step, the pyrophosphohydrolase RppH catalyses removal of the pyrophosphate removal from the 5′- termini of mRNA, followed by cleavage of double stranded mRNA by RNAse III or PNPase attracted by an additional poly(A) tail and endoribonucleolytic cleavage performed by 5′ monophosphatedependent endoribonucleases RNase E. Next, decay of cleaved RNAs is mediated by the 3′ - 5′ exoribonucleases PNPase, RNase R, and RNase II. Finally, the 3′-5′ exoribonuclease Orn degrades oligonucleotides into mononucleotides.

1.4.2 Oligoribonucleases degrade small RNAs in prokaryotes

The final step in the mRNA decay is crucial, as accumulation of nanoRNA fragments of 2-5 nucleotides in length can be deleterious and negatively affect cell viability [46]. NanoRNAse refers to the RNAse specifically involved in the final degradation of these nanoRNAs to mononucleotides in the mRNA decay pathway and is therefore essential in many bacteria. The first nanoRNAse was identified in *E. coli* in 1975 as oligoribonuclease (Orn), because of its ability to degrade oligoribonucleotide chains shorter than 5 nucleotides [45]. Orn converts nanoRNA (2-5 nt) to mononucleotides and is important for mRNA decay in cells [46].

Orn is encoded mainly by γ - and β - Proteobacteria and firmicutes, but an Orn homologue is present in almost all eukaryotes, from where it probably originates [127,133]. Described homologues involve the human homolog Small Fragment Nuclease (Sfn) and the *S. cerevisiae* homolog Ynt20 [134].

Where Orn is essential for *E. coli*, being required for cell viability [46], characterization of Orn from *P. aeruginosa* indicated a non-essential role of the protein, as cells remained viable in its absence. In addition, *P. aeruginosa* Orn could rescue the growth defect of an *E. coli* Orn mutant [135].

1.4.3 Roles of Orn in bacterial pathogenesis

Studies of Orn during the recent years have highlighted the importance of Orn in bacteria, with emphasis on its role in bacterial pathogenesis. Orn affects transcription initiation and intracellular turnover of the bacterial second messenger cyclic-di-GMP, with implications for motility, virulence and biofilm formation of bacteria [135,136]. Low level of cyclic-di-GMP promotes motility and expression of virulence factors associated with cytotoxity, whereas high levels suppress these factors [137]. Orn has been demonstrated to be the main enzyme in the hydrolysis of 5'-phosphoguanylyl-(3'-5')-guanosine (pGpG), a cyclic-di-GMP degradation product. Furthermore, the *P. aeruginosa* Orn depletion mutant shows reduced toxicity *in vitro* and *in vivo*, mainly by regulation of the type III secretion system (T3SS) [138], indicating a role of Orn in the bacterial response to environmental stimuli. T3SS, a complex protein secretion- and delivery- machinery for bacterial effector molecules, has been correlated with infection in both animal models and human patients. Infection models in mice show weakened

P. aeruginosa virulence without functioning T3SS. [138]. Other research has also highlighted the importance of the T3SS for bacterial survival and proliferation within the host [139]. Repression of T3SS and reduced mobility while enhancing biofilm formation is mainly a consequence of the elevated intracellular cyclic-di-GMP levels in Orn mutants. Moreover, since nanoRNAs may serve as primers for transcription initiation, the enzymatic action of Orn is indirectly involved in global gene expression. [140]. Next, the crucial host cell-bacteria contact for infection is strongly dependent on Orn, as expression of genes involved in bacterial motility and adherence are repressed in Orn mutants [138].

Finally, research reports a novel role of Orn in antibacterial drug resistance [141] and describes how Orn is essential for the tolerance of *P. aeruginosa* to quinolones. *P. aeruginosa,* which is intrinsically resistant to antibiotics, showed drastically increased bacterial susceptibility to quinolones antibiotics in Orn mutants. Complementation with an *orn* gene (Δ*orn*/*orn*) restored bacterial resistance.

1.4.4 Orn mechanism and structure

As RNases and DNases, Orn belongs to the DnaQ-like (or DEDD) exonuclease superfamily that hydrolyze phosphodiester bonds of nucleic acids [57]. Orn catalyzes the removal of nucleoside monophosphates at the RNA-termini in the 3'- 5' direction. The nuclease activity is dependent on divalent cations, preferably manganese (Mn^{2+}). The four conserved acidic residues characteristic of the DEDD motif are clustered around the active site, serving as ligands for a divalent cation. This group of enzymes can further be divided into the DEDDy and DEDDh subfamilies. Orn is classified as a DEDDh exonucleases because of the presence of the conserved and catalytic histidine [127] involved in attacking the phosphate ester bond via deprotonation of water molecules [34].

Orn hydrolysis is processive in the 3' - 5' direction. Characterization of *E. coli* Orn showed a higher affinity for a 5mer RNA than smaller substrates but the reaction rate was inversely proportional to the length of the chain [142]. As RNA fragments are degraded to shorter molecules, degradation rates increase. The human Orn homolog Sfn exhibits a 3' - 5' nuclease activity on both short RNA and single stranded DNA oligomers. Orn has shown a similar ability to degrade short DNA oligomers, suggesting additional roles in DNA repair, however this activity requires considerable higher

enzyme concentration than RNA degradation. Sfn degrades short RNA ~ 4-fold more efficiently than ssDNA [143].

Orn is a relatively small protein of approximately 20 kDa that requires divalent cations for nuclease activity, a type of two-metal-ion catalysis common among nucleases [144]. Orn usually exists as a homodimer in solution [68,145]. An exception was found for the human Orn *E. coli* homologue which has been shown to exist as both tetramer and homodimer in solution [68,143].

Several crystal structures of Orn have been determined recent years, including Orn from *Colwellia psychrerythraea* (PDB code 6A4A)*, Xanthomonas campestris* (PDB code 2GBZ) [146], *Haemophilus influenzae* (PDB code 1J9A, to be published), *Coxiella burnetii* (PDB code 3TR8) [147], *Acinetobacter baumannii* (PDB 5CY4) and *E. coli* (PDB code 1YTA) [148]. All Orn homologs are structurally similar and topologically arranged into an α + β - fold containing five-six β-strands and nine-ten α-helices.

Studies of the homolog *X. campestris* Orn gave more detailed insight into the mechanism of binding and hydrolysis of a 5-mer RNA substrate [146]. *X. campestris* Orn forms a dimer in the crystal-structure with a 2-fold crystallographic symmetry (Fig. 11c). The overall architecture of this Orn folds into the DnaQ family fold and is similar to other reported 3′ - 5′ RNases (fig. 11a). The active site conserved DEDD motif (Asp15, Glu17, Asp115 and Asp166) binds one divalent cation per monomer (Mg^{2+}) (Fig. 11b). Residues from both monomers were found to interact with the modelled U5 oligonucleotide substrate. *X. campestris* Orn exploits a similar oligonucleotide cleavage mechanism as those employed by enzymes in the DEDDh family, however the functional details behind this cleavage mechanism are not fully understood. To explain the processive manner of Orn cleavage, it is proposed that the highly basic RNA substrates bind Orn by electrostatic forces exhibited by the active site acidic residues and the Mg^{2+} ion [146]. Additionally, *X. campestris* Orn showed a novel structural feature. Different from other described 3′ - 5′ RNases, the helix H is oriented opposingly, possibly to accommodate oligoribonucleotide substrates better, with less steric hindrance [146]. The formation of a stable homodimer is believed to be important, which is achieved by intermolecular hydrophobic interactions, H-bonds, salt-bridges and a disulfide bond. The disulfide bond of Cys113-Cys113 connecting the two monomers is believed to play a major role in stabilizing the homodimer (Fig 11c) [146], although it is unknown whether its contribution is essential for homodimer formation and stability, and thus Orn activity. The cysteines involved do not show conservation among Orn amino acid sequences and most Orn structures form stable homodimers without the presence of disulfide bonds in their dimer interface.

Finally, the structural studies of *X. campestris* Orn highlighted the importance of the well conserved ExoI, ExoII, ExoIII, and ExoIV motifs that are unique to Orn, as the residues involved may play important roles in stabilizing the Orn dimer or in interacting with the U5 substrate [146].

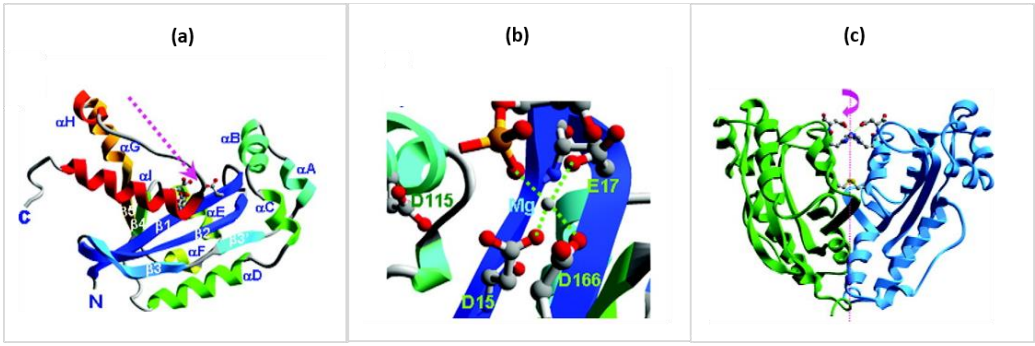


Fig. 11. Ribbon presentation of *X. campestris* Orn structure with metal binding site. (a) Orn monomer with domain architecture labeled with colours from the N-terminus (blue) to the C-terminus (red). The nine α helices (αA–αI) and five β sheets are indicated (β1–β5). The pink dotted line marks the cleft accommodating the four active site acidic residues shown in ball-and-stick. **(b)** Details of the Orn monomer catalytic site with the metal coordinating residues Asp15, Glu17, Asp115, and Asp166 (DEDD motif). **(c)** Orn dimer (monomer 1 in green, monomer 2 in blue) with residues in the dimer interface shown in ball-and-stick (H-bond between Asp24 and His120 and Cys113 in inter-subunit disulfide bond). The pink dotted line shows the non-crystallographic 2-fold symmetry. The figures are adapted from [146].

1.5 ATP-dependent ligases

DNA ligases are DNA joining enzymes essential for survival of all organisms, due to their critical roles in cellular processes like DNA replication/repair and recombination. The ligation reaction covalently joins breaks in the phosphate backbone of DNA, and acts on single nicks or double-strand breaks with blunt or cohesive ends. DNA ligases use ATP or NAD^+ as an energy source, and based on this cofactor requirement can be classified into two groups; ATP-dependent ligases (ADLs) use ATP and are found in most living organisms, including some bacteria, while NAD^+-dependent ligases use NAD^+ as a cofactor and are present in bacteria exclusively [149-151]. In the cases where accessory ADLs are identified in bacteria, it is always in addition to the essential NDLs, rather than instead of [152]. The first bacterial ADL to be characterized was encoded by the

respiratory pathogen *H. influenzae* [153]. Eukaryotes, archaea and many viruses possess at least one ADL that is essential for DNA replication (by joining Okazaki fragments), and some encode multiple forms with dedicated roles in DNA repair [154].

Apart from cofactor differences, the chemical reactions catalyzed by ADLs and NDLs are identical [47,155]. DNA ligases catalyze the formation of the phosphodiester bond between the 5' end of one DNA strand and the hydroxyl group at the 3' end of the other DNA strand, producing an intact sugar-phosphate backbone. The three nucleotidyltransfer steps in the enzymatic reaction mechanism are shown in fig. 12. First, the enzyme is activated through a nucleophilic attack by a lysine residue on the adenosine cofactor ATP or NAD^+, releasing the nicotinamide mononucleotide for NDLs or PPi for ATP-dependent ADLs (fig. 12a). Next, the nucleophilic 5'- phosphate of the DNA attacks the phosphoramide bond to form an adenylated-DNA intermediate (fig 12b). The final step involves attack of the 3'-nucleophilic hydroxyl group to the new pyrophosphate bond, formation of a phosphodiester bond between the 5' and the 3' position of the DNA and release of AMP (fig 12c). All three chemical steps depend on the divalent metal ion Mg^{2+}, although Mn^{2+} can substitute for Mg^{2+} in all catalytic steps performed by DNA ligase with similar kinetic rates, suggesting a flexible active site with regards to metal cofactors. The stoichiometry of the metal cofactor varies between one and two ions per ligase molecule, depending on the type of ligase (NAD vs ATP) and the reaction step [156-158]

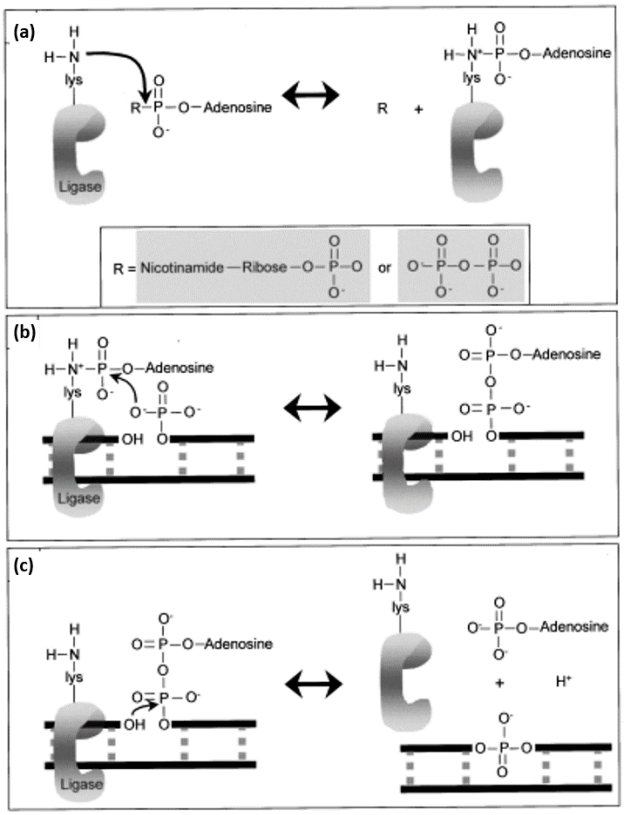


Fig. 12. DNA ligase reaction mechanism. **(a)** The ligase reacts with ATP or NAD^+ via the amino group of lysine and forms a phosphoamide-linked AMP. **(b)** The phosphate at the 5' - OH nicked DNA attacks the activated phosphoryl group of the AMP and forms an adenylated DNA intermediate. **(c)** Attack of the 3' - OH end forming a new phosphodiester bond and releasing AMP. The figure is adapted from [149].

1.5.1 DNA ligase architecture

ADLs show significant variation in amino acid sequence, structural motifs and size, from the large human ADL at over 100 kDa to the much smaller marine gamma proteobacterium *Alteromonas mediterranea* ADL (<30 kDa), with the latter representing the minimal scaffold for DNA binding and ligation activity [58]. Since the first X-ray crystal structure of an ADL was solved in 1996 from bacteriophage T7 [159], numerous structural analyses of bacterial, archaeal and eukaryotic ADLs have followed [160-167], and the wide variety of domains and gene arrangements between

the different classes of ligases has become evident. All of these arrangements are based around a common architecture of the two essential, but distinct catalytic core domains [159,168]: the adenylation domain (AD) which includes the residues directly involved in catalysis, and the smaller oligonucleotide/oligosaccharide binding domain (OB) that is essential for activity [64,168]. These two domains are joined by a flexible linker that allows them to reorient during DNA binding. The core catalytic domains include six conserved motifs (I, III, IIIa, IV, V, and VI) which contact AMP and play essential roles in one or more steps of the ligation pathway [169] and define the covalent NTase enzyme superfamily of polynucleotide ligases and RNA-capping enzymes [170]. Motif I (KxDGxR) covalently attaches AMP via the lysine in the first step of the ligase reaction. Diversity in ligase structure and size arises from the presence of additional domains at the N- or C- termini of this core scaffold, or in some cases, additional loops within the AD and OB domains.

1.5.2 Bacterial ATP-dependent ligases (b-ADLs)

Compared to the relatively homologous ADLs in archaea, those encoded by bacteria are more diverse in terms of size and structural domain arrangements.

The Pfam database identifies a range of structures with conserved domains linked to the core AD domain, with much of this diversity found in bacterial and viral ADLs. This domain architecture of proteins divides bacterial ADLs (b-ADLs) into four main groups based on size and enzymatic domain associations; Lig B, Lig C, Lig D and Lig E, with Lig B being the most widely distributed [50]. Functions of these domains of b-ADLs include enhanced DNA binding and additional enzymatic activities involved in DNA metabolism. The three classes Lig B, Lig C and Lig D appear to have descended from a common ancestor within bacterial species, whereas the recently identified Lig E group, the second most distributed, possess unique genetic and structural features and appears more related to viral and bacteriophage proteins [50,171,172]. Biochemical characteristics, domain structure and inferred extracellular location set this group apart from the other b-ADLs, suggesting that horizontal transfer of Lig E encoding genes into bacteria took place in a separate event from other b-ADLs. Comparison of the different groups of b-ADLs and the clear groupings of their functional domains based on Pfam architectures are illustrated in fig. 13 [50].

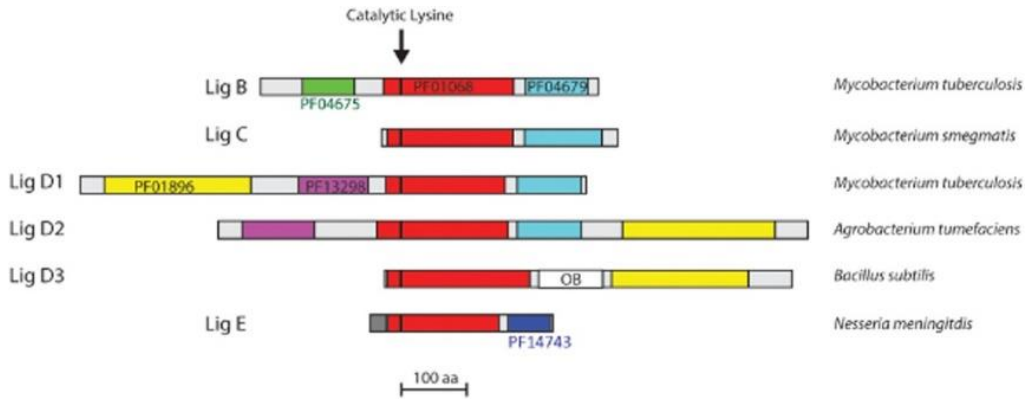


Fig. 13. Bacterial ATP-dependent DNA ligase (b-ADLs) domain structure assigned by Pfam domain identifier. The adenylation domain (AD, red) and the oligogonucleotide domain (OB, dark blue and cyan) form the core catalytic domains found in all types of b-ADLs. Note that the OB domain of Lig D3 (white) was predicted by HHpred instead of Pfam. Other conserved domains include phosphoesterase domain (purple), DNA primase/polymerase (yellow), N-terminal DNA binding domain (green). The periplasmic leader sequence (dark gray) was predicted by the signalP server [173]. The representative sequences analysed for each type of ADLs are *M. tuberculosis* Lig B (gi 148507067), *M. smegmatis* Lig C (gi 118472839), *M. tuberculosis* Lig D1 (gi 15608078), *A. tumefaciens* Lig D2 (gi 159186474), *B. subtilis* Lig D3 (gi 16078405) and *N. meningitidis* Lig E (gi 121051388). Domains are drawn to scale and the position of the catalytic lysine in the active site is marked with an arrow. The figure is adapted from [50].

The function of this multitude of different b-ADLs with different substrate requirements is not fully understood. To date, none have been shown to be essential suggesting they play no role in DNA replication. The large Lig D and smaller Lig C variants require interaction partners for optimal activity and are involved in double-strand break joining and base excision repair respectively [171,172]. Lig B and Lig E are able to efficiently seal single nicks in the DNA backbone, and further research is required to fully understand the roles of these ADLs in their host bacteria.

1.5.3 Lig E type ligases

The minimal and phylogenetically-distinct Lig E-type DNA ligase may serve as a model for the minimal functional unit of the ATP-dependent ligases as they possess only the core unit of AD and OB domains. Lig E show high end-joining activity on singly nicked DNA, some activity on double-stranded breaks, both cohesive and blunt ended, as well as mismatched and gapped substrates [48,49,153]. All Lig E-type ADLs possess a predicted N-terminal signal sequence proposed to direct them to the periplasmic space of the bacterium, followed by its proteolytic removal. Proposed biological functions of such secreted ligases include competence and DNA uptake in the periplasm [48].

The majority of the minimal Lig E-type DNA ligases lack any appending domains or unstructured loops needed for complete encirclement of the DNA duplex during catalysis. The recently solved crystal structures and site directed mutagenesis analysis of Lig E from *Psychromonas sp* strain SP041 (Psy-Lig, fig. 14a) [166] and *Alteromonas mediteranea* (Ame-Lig, fig 14b) [58] bound to nicked adenylated DNA describe the mechanisms behind the partial encirclement of Lig E to nicked DNA. In contrast to some larger AD ligases, no ordering of unstructured regions upon DNA interaction was observed and the main region of plasticity during the binding process was the linker region. For the Ame-Lig structure, the AD and OB domains were reoriented around the linker region involving alteration of the side-chain contacts on the surface of the conserved domains to form a C-shaped clamp around the DNA in a so-called closed state, ensuring tight binding in the absence of appending domains or flexible loops able to completely encircle the substrate (fig. 14b).

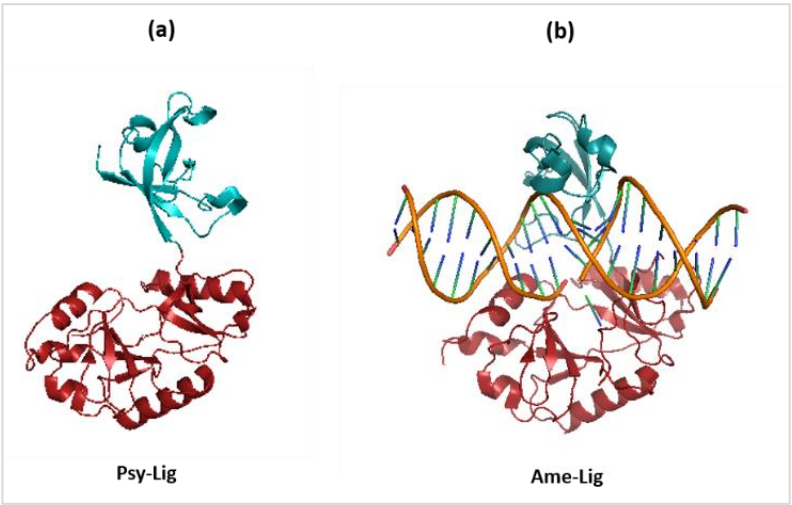


Fig. 14. Cartoon structures of Lig E in bound and unbound state. **(a)** Lig E from *Psychromonas sp* strain SP041 (Psy-Lig) in an open conformation (PDB D405). **(b)** Lig E from *A. mediteranea* (Ame-Lig) in complex with a 21-bp nicked DNA-adenylate (orange) (PDB 6GDR), forming a C-shaped clamp around the DNA in the closed conformation. Individual domains are coloured in PyMol; the adenylation domain (AD) in red and the oligonucleotide binding domain (OB) in cyan.

1.5.4 Applications of cold-adapted ATP-dependent ligases

DNA ligases adapted to low temperatures offer novel potential advantages in biotechnological applications such as genetic engineering and new-generation DNA sequencing methods [169,174,175]. As tools for DNA joining, bacteriophage-derived DNA ligases, such as the T4 and T7 ligases and *E. coli* ligase, are dominating the molecular biology market. However, their optimal activity at elevated temperatures above 15˚C is a disadvantage to the ligation process, as residual nucleases may interfere. Thus, cold adapted DNA ligases able to increase yields of product at lower temperatures, while suppressing contaminating nuclease activity, is highly sought for by this industry. Furthermore, highly active cold-active ligases with better activity rates enables experiments to be carried out with smaller amounts of enzyme.

This is can be exemplified by Golden Gate DNA cloning, one of the easiest and quickest cloning methods, where restriction and ligation of multiple compatible DNA fragments can be performed in one single step. Cutting by type IIS restriction endonucleases

allows for the creation of overhangs, typically ligated by T4 ligases. Since T4 DNA Ligase does not ligate cohesive ends ideally, as it possesses significant activity on some mismatch-containing pairings, especially at lower ligation temperatures, the procedure could benefit from enzymes acting at lower temperatures. Furthermore, the current protocol, which involves temperature cycling between 37°C and 16°C and final enzyme heat inactivating at 60 °C, could potentially benefit from cold-adapted ligases. Increased ligase activity-rates at low temperature increases yields, with shorter incubation time and decreased enzyme concentrations. Finally, short base-pair overhangs will be stabilized due to the low melting temperature, which makes the assembly more efficient.

1.6 Proteins adapted to the cold

1.6.1 Microorganisms and their environment

Extreme environments, outside the limited range where humans and most other eukaryotes can survive, are surprisingly common on our planet [176]. Many organisms live and thrive at extreme physiological conditions, ranging from hot springs, frozen sea water, extreme salt solutions, thousands of meters deep inside the earth's crust, at high pressure deep inside the ocean, and from extreme acid (pH 0) to extreme basic conditions (pH 12.8) [177,178]. Such organisms, mostly unicellular and prokaryotic, classifies as extremophiles.

Various microorganisms have a wide range of tolerances and optima a with respect to temperature, pH, salt concentration, radiation, desiccation and pressure. Temperature is one of the most important physical factors necessary for survival, and in the last part of the 20th century, modern molecular biology methods have revealed that microorganisms inhabit both extreme ends of the temperature scale. Low temperatures represent the most common extreme environments on earth. Over 85% of the biosphere experiences temperatures of less than 5°C during the year , and over 75% is permanently cold (less than 5°C year round) [179], for instance the northern and southern polar areas, alpine regions and the deep sea.

Microorganisms are usually divided into psychrophiles, mesophiles, thermophiles and hyperthermophiles, according to their ability to grow at low, medium, high and

extremely high temperatures, respectively (fig. 15). Psychrophiles and hyperthermophiles, at the lower and higher end of the temperature scale, are classified as extremophiles, and have been detected in earth's most extreme environments. Some organisms frequently found in cooler environments thrive in temperatures that overlap the two thermal groups recognized as psychrophiles and mesophiles; these are known as psychrotolerant or psychrotrophs and can grow from around 4 °C, albeit with lower growth rates, and up to 20-30 °C. Often, these are bacteria also found in mesophilic environments, transferred to colder environments due to horizontal gene transfer [180].

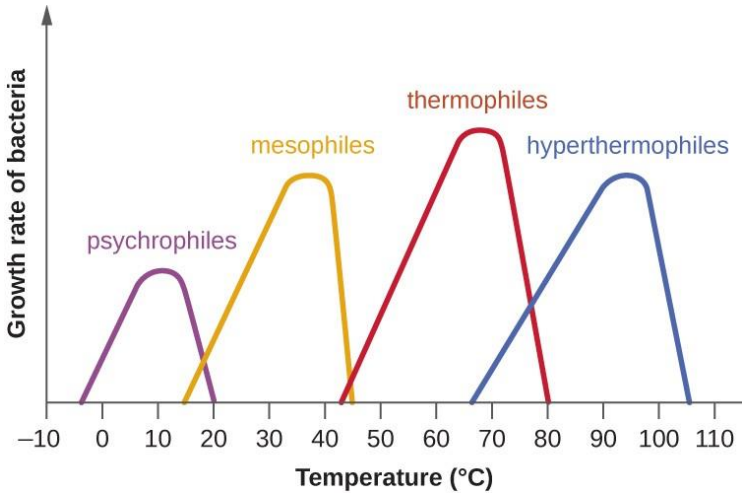


Fig. 15. Adaptation of microorganisms to different thermal environments plotted as temperature against growth rate. Based on their optimal growth temperature, microorganisms are classified as psychrophiles (<20˚C), mesophiles (20-40˚C), thermophiles (40-80˚C) and hyperthermophiles (>80˚C). The top of the curves indicates optimum growth temperatures, whereas the startpoint and endpoint of the curves illustrate minimum and maximum growth temperatures. Adapted from [181].

1.6.2 Cold-adaptation strategies

A considerable fraction of living organisms is found in the sea at temperatures lower than 15°C. As early as 1887, microorganisms isolated from fish were reported to grow and reproduce at 0°C [182]. Psychrophiles are organisms that have evolved towards thriving at low temperatures, close to the freezing point of water [183]. Aerobic and anaerobic bacteria are found living in permafrost soil and in sea ice at down to −20 °C, in restricted areas combined with high salt concentrations. Most cellular activities are affected as the temperature is reduced; enzyme activity rates, decreased membrane fluidity, increased water viscosity, limited salt solubility and changed pH of biological buffers [184,185]. To survive and thrive at these conditions, psychrophilic microorganisms have adapted physiologically. Temperature-dependent changes in membrane lipid fatty acid composition have been documented, as a higher content of unsaturated fatty acid keeps the membrane fluid when it would normally freeze [186]. In addition, many organisms have developed antifreeze and ice-nucleating molecules, cold-shock proteins, regulation of ion-channel permeability, seasonal dormancy and modification of enzyme kinetics. To maintain enzyme catalytic rates, cold-adapted organisms have developed typical strategies, including increased enzyme concentrations [187], seasonal expression of isoenzymes [188-190] and enzymes evolved to depend less on temperature [184]. Research on cold adaptation has shown particular interest in the increased catalytic potential exhibited by enzymes active at low temperature compared to mesophilic homologues. However, the molecular mechanisms behind cellular adaptations are not completely understood and still under investigation.

1.6.3 Cold adapted proteins

Enzymes isolated from prokaryotic and eukaryotic organisms living in cold environments are often called cold-adapted or psychrophilic [191]. Expression of such enzymes, specifically adapted to function optimally at persistently cold habitats, is one of the main physiological cold adaptation strategies used by cold adapted organisms studied so far [192].

Living and thriving at low temperatures require that both enzyme kinetics and protein stability are adapted accordingly. The molecular basis that enables proteins to remain

folded and functional under these extreme temperatures are still only partially elucidated, but it has become evident that the ways to reach this goal are rather diverse. Cold adaption strategies may vary between different families of enzymes, and studies of psychrophilic enzymes have identified subtle and unique structural adjustment compared to mesophilic counterparts. However, some typical traits are identified and will be described in the subsequent sections. Nevertheless, understanding how psychrophilic enzymes sustain or even increase their activity level as they face an exponentially decrease of the chemical reaction rates at lower temperatures is the crux of cold adaptation. In general, it is now widely accepted that the trade-off between thermostability and rate of catalysis, through improved global or local flexibility of cold-active enzymes, enables high specific activity at low temperatures with a lower energy cost [183,193,194].

1.6.3.1 Enzymatic activity of cold adapted enzymes

To understand how enzymes modify their enzyme kinetics as a cold adaption strategy, basic enzyme kinetics and parameters relevant for enzyme catalysis will be briefly discussed.

The following equation (1) describes single-substrate enzymatic reactions with enzyme (E) and substrate (S), generating product (P), following steady-state kinetics.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P \qquad (1)$$

k_1 is the rate constant for formation of ES complex from E + S, and k_{-1} is the rate constant for the dissociation of ES to free E and S. k_2 is the rate constant for conversion of ES complex to P and subsequent release of P from E [195]. E and S are assumed to be in equilibrium between the free and the bound form. In this model, k_2 can be replaced by the more general rate constant k_{cat}, defined as the maximal enzyme reaction rate at a given temperature. It gives a direct measure of the catalytic production under optimum conditions as the number of substrate molecules that are transformed by one molecule of enzyme per unit of time and it also known as the turnover number or specific activity.

The temperature dependence of chemical reactions rates can be inferred from the Arrhenius equation [196].

$$k_{cat} = Ae^{-Ea/RT} \qquad (2)$$

where k_{cat} is the reaction rate, A is the pre-exponential factor, Ea is the activation energy, R is the universal gas constant (8.31 $JK^{-1}mol^{-1}$) and T is the absolute temperature in Kelvin. According to this equation, a decrease in temperature is accompanied by a decrease in reaction rate and molecular motions, explaining the decrease in reaction rates in cold environments. Conversely, a decrease in Ea will increase the reaction rate. For psychrophilic enzymes, the reaction rate is less affected by a decrease in temperature compared to mesophilic enzymes, possibly due to a decrease in the activation energy required for catalysis, E_a, as a cold adaptation strategy [183].

To compensate for the slower reaction rates at low temperatures, psychrophilic enzymes show activity rates up to 10 times higher at low and moderate temperatures than their mesophilic homologues, as illustrated by the temperature optima of a psychrophilic and a mesophilic enzyme in fig. 16.

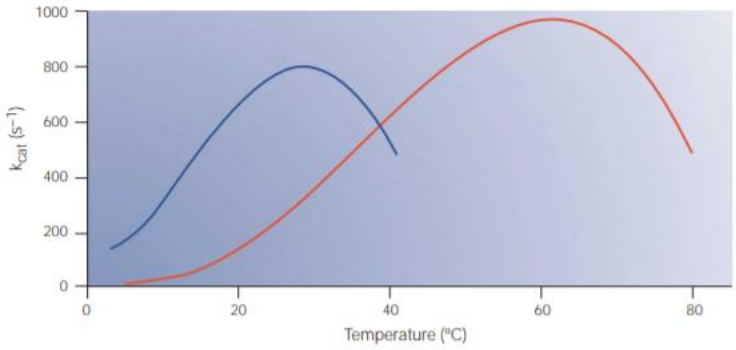


Fig. 16. Temperature dependence of enzyme activity. The activity of psychrophilic enzymes (blue line) and mesophilic enzymes (red line) measured at various temperatures. The optimal temperature for activity is shifted towards lower temperatures for psychrophiles (blue line) and show up to tenfold better activity at lower temperatures (up to 20–30°C) compared to their mesophilic homologues (red line). The activity data are collected from studies on α-amylases from the psychrophilic *Pseudoalteromonas haloplanktis* and the mesophilic *Bacillus amyloliquefaciens*. The picture is adapted from [183].

The free energy of activation, ΔG#, enthalpy of activation, ΔH#, the entropy of activation, ΔS# and the absolute temperature T, are related according to the classical Gibbs-Helmholtz equation.

$$\Delta G\# = \Delta H\# - T\Delta S\# \quad (3)$$

Reactions catalyzed by cold active enzymes often have lower ΔG# values than reactions that are catalyzed by mesophilic enzymes, indicating that the enzymes show increased activity. Furthermore, lower enthalpy values, ΔH#, suggests that the catalysis driven by cold active enzymes is less temperature dependent. A decreased activation enthalpy is now well documented as a relevant cold adaptation strategy. Most psychrophilic enzymes show a lower value of enthalpy for activation compared to mesophilic enzymes, demonstrating reaction rates less affected by decreasing temperatures [197].

Structurally, this kinetic adaptation to lower temperatures is achieved through reduction of the number of enthalpy-related interactions that need to be broken during catalysis, either in the catalytic center or in other areas of the proteins that effects the dynamic properties of catalytic residues from a distance apart [186]. To obtain a better structurally understanding of this cold adapted feature, Isaksen *et al* [198] used computer simulations and high-precision Arrhenius plots to show that the protein surface rigidity outside the active site may control the enthalpy-entropy balance, hence the temperature adaptation of activity.

1.6.3.2 Thermal stability of cold-adapted enzymes

Also characteristic for cold-adapted enzymes is the decreased structural thermostability compared to mesophilic homologues (fig. 17), a common property of nearly all cold-adapted enzymes characterized so far. The overall three-dimensional structure or functional domains tends to unfold at lower temperatures than their mesophilic counterparts (lower melting temperature, T_m), due to weakening of intramolecular forces that contribute to stability.

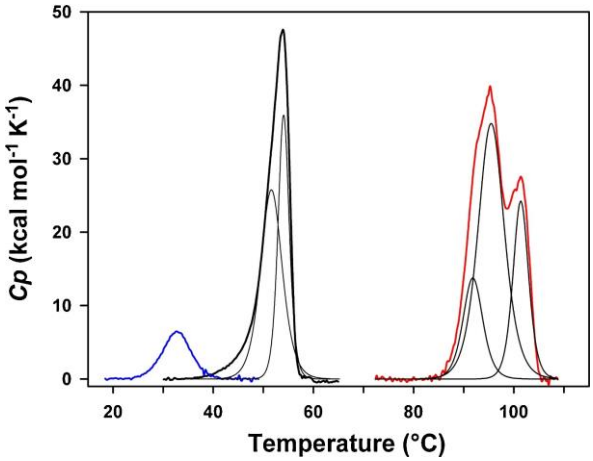


Fig 17. Thermal stability of cold adapted enzymes exemplified by DNA ligase. Thermal unfolding of DNA ligases from the psychrophilic *P. haloplanktis* (blue line), the mesophilic *E. coli* (black line) and the thermophilic *Thermus scotoductus* (red line), as measured by differential scanning calorimetry, show that the psychrophilic enzyme is characterized by a low melting temperature (T_m) compared to its mesophilic and thermophilic homologues. The weak black lines show deconvolution into domains. *P. haloplanktis* DNA ligase denaturates cooperatively devoid of stability domains. The picture is adapted from [199].

Psychrophilic enzymes are often inactivated at temperatures that are well below thermal unfolding temperatures, unlike mesophilic or thermophilic enzymes. For some cold adapted enzymes, the rationale behind this behavior is increased flexibility and consequently more heat-labile active site [200,201]. However, in cases where the active site architecture in cold adapted enzymes and their mesophilic homologues are identical, the low thermal unfolding temperatures are explained by amino acid substitutions elsewhere in the molecule, related to the kinetic adaptations described above [202].

1.6.4 Structural adaptations to the cold

Sequence analysis and comparative studies with mesophilic counterparts have revealed structural features that add flexibility to cold adapted enzymes, although few can be generalized to all enzyme classes. Increased flexibility, either of a selected area or of the overall protein, can be achieved by destabilization of their structure through

a reduced number of intramolecular forces such as salt-bridges, ion-pair networks, hydrogen bonds and aromatic interaction, and increased length of loop regions [203-205]. Further, psychrophiles often show an increased exposure of hydrophobic and non-polar residues to the solvent, combined with a decrease of ion-pairs at the surface, as an entropy-driven destabilizing factor. Also, fewer hydrophobic residues in the core of proteins have been observed [183] and hydrophobic residues in psychrophiles are often smaller than in their mesophilic counterparts, adding cavities and flexibility to the internal packing. This is often expressed by the (Ile)/(Ile+Leu+Val) ratio [206]. Arginine is believed to contribute to stability due to its high potential to form hydrogen bonds within the protein [191]. A decreased arginine content and decreased Arg/(Arg+Lys) ratio has been described in many cold adapted proteins [207], as well as a decrease in the number of prolines in loops.

Comparison of deposited three-dimensional structures with mesophilic and thermophilic homologs show that each cold adapted enzyme uses discrete and diverse structural adjustments to achieve global or local molecular flexibility, increased catalytic efficiency and reduced stability, thus drawing general conclusions from the complex structural adaptations to the cold is now accepted as difficult to achieve [208,209].

This is further complicated by genetic drift by evolutionary pressure; differences in amino acid sequences of enzyme homologues within the same species due to neutral mutations and not necessarily related to cold adaptation. Furthermore, enzymes adapting to new environments are often introduced to additional extremities as high salt, high pressure and high acidity. Hence, relating amino differences to the proper selective pressure can be challenging enough itself [210].

1.6.5 Applications of cold active enzymes

The interest in isolation and cultivation of psychrophiles has risen the last decades after discovery of how certain organisms and their enzymes can be employed in various biotechnological applications and a broad spectrum of cold adapted enzymes is already applied [211]. Properties of special interest is their increased catalytic efficiency and capability to work at low and moderate temperatures at which homologous mesophilic enzymes are not active, combined with the lability observed at elevated temperatures, enabling a rapid and targeted inactivation of the enzyme [212]. In applications that

require large amounts of enzyme, the more efficient psychrophilic enzymes offer economic benefit without compromising quality. Industrial applications include detergent additives for cold washing, additives in food industries, pharmaceuticals and skin care products, cheaper and safer bio-fuel production, fine-chemical synthesis, biotransformation and molecular biology applications [211,213].

In the field of molecular biology, cold active enzymes are widely used within cloning and restriction enzyme digestion, where their implementation has simplified the experimental procedures and lowered the concentrations of both enzyme and substrate required. The typical thermolability observed for psychrophilic enzymes enables heat activation to be performed at low enough temperatures to avoid melting of the double stranded DNA in experimental protocols. This is exemplified by two widely used enzymes in molecular biology applications that origin from marine Arctic environments; The DNA modifying enzyme Shrimp alkaline phosphatase (SAP) used as a psychrophilic alkaline phosphatase [214-216] and Cod Uracil-DNA Glycosylase (Cod UNG) from Atlantic cod, the only UNG fully and irreversibly inactivated by moderate heat [217-219].

Finally, it is worth mentioning that the cold adapted enzymes most frequently reported in the literature between 2010 and 2016 are hydrolases at 91%, which is considerably more than oxidoreductases (4%), transferases (2%), isomerases and ligases (1%) [197]. Likewise, cold-active enzymes identified through metagenomic approaches were all hydrolases, with one exception [220]. This demonstrates that much potential diversity among cold-adapted enzymes remains to be explored.

2 Background and aims of the studies

Nucleic acid binding proteins are important players in all processes in cells; from regulation of transcription, translation, DNA replication, repair and recombination, RNA processing and translocation, hence understanding their mechanisms of action contributes largely to our understanding of general cell biology, cell development and how diseases develop.

The purpose of this study was to explore different nucleic acid interacting proteins originating from psychrophilic bacteria to gain new insight into their biological roles in important cellular processes as gene regulation, RNA degradation and DNA ligation, and ultimately better understand pathogenesis and/or cold adaptation of our model organisms.

By functional and structural studies, three distinct types of nucleic acid binding proteins originating from Arctic marine environments were characterized throughout this study:

1) Ferric uptake regulator (Fur) from *Aliivibrio Salmonicida*
2) Metagenomic Oligoribonuclease
3) ATP-dependent DNA ligases from *Psychromonas spp*. strain SP041, *Aliivibrio salmonicida* and *Pseudoalteromonas arctica.*

<u>Sub-objectives</u>

1) The motivation behind studying Fur was to shed light upon the biological role of Fur from the fish pathogen *A. salmonicida* (AsFur), the causative agent of cold-water vibriosis in farmed salmon. The objective of this project was to extend our knowledge about the AsFur-DNA interaction, based on previous experimentally and computationally studies in our research group, primarily by revealing residues and nucleotides directly involved by functional and structural studies. Fur is a highly interesting target in the design of antibacterial drugs and the overall goal is to provide insight into potential inhibitor binding sites and subsequently engineer small molecule inhibitors of Fur.

   Achieved by:

- Optimize the established purification procedure by identifying conditions where the protein is stable in its active form.

- Design a range of synthetic Fur box-containing DNA oligonucleotides and optimize the binding conditions for interaction with Fur *in vitro* by DNA-binding analyses (EMSA, plasmid protection assay)
- Homology modelling to study Fur-DNA complexes

2) The main motivation behind implementing Orn in this study of nucleic acid binding proteins is the bioprospecting aspect; identification of novel enzymes adapted to low temperatures for potential use in biotechnological applications. Further, this study has relevance to understanding the molecular mechanism of Orn in prokaryotic mRNA degradation as well as studying cold adaptation mechanisms of enzymes.

   Orn is unique for many γ-proteobacteria, in the sense that it is the one and only exoribonuclease that completes the final stage of the mRNA decay and often is essential for cellular physiology and viability.

   Enzymes from Arctic host organisms remain a relatively unexplored field, however characterization of enzymes adapted to cold regions are steadily increasing with the advent of new high-throughput sequencing methods for assessing unculturable microorganisms in the environment. Metagenomics is a culture-independent approach involving DNA extraction of an environmental sample aiming to construct metagenome libraries for isolation of target genes. Our research group use this tool in bioprospecting, as a source for identifying novel cold adapted enzymes. The gene encoding Orn originates from a metagenome library created from the intertidal zone in Svalbard.

   Prior to this work, the gene encoding metagenome Orn (MG Orn) was cloned and tested for solubility.

   Achieved by:

- Establish a protocol for recombinant protein production and purification of stable and active Orn
- Characterize Orn with respect to stability, nuclease activity and specificity; determine enzymatic kinetics and divalent cation requirements by a spectrophotometric assay with a dinucleotide analog substrate, investigate Orn specificity, directionality and salt tolerance by a gel assay with

fluorescence-labeled RNA substrates, and determine thermal stability by the Thermofluor assay.

- Determine the three-dimensional crystal structure to gain a better understanding of Orn catalytic activity and the ability to discriminate between RNA substrates.
- Mutate the cysteines forming the intermolecular disulfide bond involved in dimerization to explore its role in the formation and stability of the homodimer and subsequent nuclease activity.

3) Ligases from marine cold adapted environments have been extensively studied in our research group, with focus on identifying novel minimal bacterial ATP-dependent ligases type Lig E, motivated by their great potential in biotechnological applications.

   In this project, the main goal is to analyze biochemical and biophysical characteristics of Lig E from psychrophilic organisms in an attempt to identify typical cold-adaptation features and thus understand how these enzymes have evolved to function within their extreme environments. To achieve this, we aim to explore and compare the temperature-optima, thermal stabilities and homology models of three minimal Lig E-type ATP-dependent DNA ligases representing obligate psychrophilic species of bacteria; Lig E from *Psychromonas spp*. strain SP041, *Aliivibrio salmonicida*, and *Pseudoalteromonas arctica*, abbreviated Psy-Lig, Vib-Lig and Par-Lig, respectively. All enzymes have been identified by in silico analysis of bacterial genomes. Psy-Lig and Vib-Lig are previously recombinantly produced and characterized and Psy-Lig has been successfully crystallized, whereas Par-Lig, isolated from sandy beach sediment on the Arctic island of Svalbard, will be recombinantly cloned, expressed and purified as part of this project.

Achieved by:

- Establish a protocol for recombinant protein production and purification of stable and active Par-Lig
- Gel-based endpoint assays of Vib-Lig, Psy-Lig and Par-Lig, both with single nicked and overhanging substrates, to compare the temperature optima for ligase activity.

- DCS experiments to compare the melting temperature (T_m) of Psy-Lig, Par-Lig and Vib-Lig.
- Thermofluor assys in various buffer systems for all three enzymes.
- Compare sequence alignments and homology models for Psy-Lig, Par-Lig, Vib-Lig and the mesophilic homologue Lig E from *Vibrio cholera*, to gain further insight into the activity/stability/flexibility relationship and cold adaptation,

3 Summary of papers

3.1 Paper I

This work describes the DNA-binding mode of the ferric uptake regulator (Fur) from the fish-pathogenic *Aliivbrio salmonicida* (AsFur) by biochemical assays and structural modelling, in order to better understand AsFur mechanisms at the molecular level.

By homology modelling, AsFur was aligned with the crystal structures of Fur from *Magnetospirillum Gryphiswaldense Msr-1* (MgFur) in complex with two different oligonucleotide mimics of the Fur box; the *feoAB1* operator and the *E. coli* Fur box.

To evaluate the AsFur-DNA binding mode in EMSA assays, we used oligonucleotides of different lengths and compositions, based on the Vibrio and the *E. coli* consensus sequences as templates. Specific base substitutions identified novel bases involved in the interaction, as well as confirming previously suggested important nucleotides. T13 was clearly important for productive binding by AsFur and structural homology models highlighted the role of the conserved Tyr56 in base-specific major groove interactions with these thymines. Next, A14 and C16, forming part of the first part **G**A**T**AAT hexamer repeat on the complementary strand, were shown to participate in minor- and major-groove nucleotide base interactions with the conserved residues Lys14 and Arg57, respectively, and thus highlighted the importance of both DNA strands in the AsFur-DNA interaction. Furthermore, the importance of an AT-rich core of the Fur box recognized by AsFur was confirmed.

Finally, combined substitutions of important nucleotides identified in this study were not able to reduce the binding capability in a similar fashion as the Vibrio least conserved sequence, indicating that few base-specific contacts are formed by AsFur. Hence, similar to Fur homologs, shape readout is also a driving feature in AsFur-DNA interactions.

In conclusion, EMSA assays of AsFur complexed with oligonucleotides of varying content and lengths combined with structure/function interpretation by AsFur interaction models has identified crucial AsFur-DNA interaction sites, mostly involving conserved residues and nucleotides in the various Fur homologues and their respective proposed consensus sequences. Thus, this study shed new light into the molecular basis of Fur regulation in *A. salmonicida*.

3.2 Paper II

In this paper we have performed biochemical and structural characterization of the Arctic marine oligoribonuclease MG Orn, and by mutagenesis determined the role of an intermolecular disulfide bond in dimer formation and catalytic activity.

The *mg orn* gene from a metagenomic library created from the intertidal zone in Svalbard gene was successfully cloned, expressed and purified to homogeneity as a homodimer. Sequence alignments and phylogenetic analysis revealed that MG Orn show high sequence identity with Orn from *Arenicella xantha* and probably originates from a species within the *Arenicellales* order.

Biochemical characterization showed that MG Orn possess 3' - 5' exonuclease activity on short RNA oligonucleotide substrates from 2mer up to 10mer of length. Further, two different activity assays revealed an absolute metal requirement (Mg^{2+} or Mn^{2+}), a quite narrow pH range (pH 8-9) and a NaCl tolerance up to 250 – 500 mM. By thermal stability assays, MG Orn was also shown to be more heat labile than its counterpart from *Escherichia coli*; a possible cold adapted feature of MG Orn.

The crystal structure of MG Orn was determined to 3.15 Å and showed typical homodimer formation between monomer A and monomer C, connected by an intermolecular disulfide bond, salt bridges, H-bonds and hydrophobic interactions. To investigate the importance of the intermolecular disulfide bond for dimer formation and subsequent catalytic activity, Cys110 was substituted with either Gly or Ala. Both mutants failed to hydrolyze RNA, probably due to hampered dimer formation. Thus, the Cys110 disulfide bond is required for homodimer formation and catalytic function of MG Orn.

Finally, structural modelling of MG Orn with ExoI in complex with ssDNA allowed manual fitting of a 5mer RNA molecule into the active site of MG Orn and showed that a conserved sequence patch (His128-Tyr129-Arg130) in the neighbouring monomer interacts with the 5' end of longer substrates, possibly explaining the structural basis behind hydrolysis of RNA substrates longer than 2mer.

To conclude, we have demonstrated the importance of the disulfide bond connecting the two monomers forming the functional MG Orn homodimer and structurally explained how MG Orn fit longer RNA substrates into its active site.

3.3 Paper III

In this paper we have investigated and compared the temperature optimum, thermal stability, amino acid sequence and structural homology models of the Arctic marine-derived Lig E ADLs from *Psychromonas spp*. strain SP041 (Psy-Lig), *Aliivibrio salmonicida* (Vib-Lig), and *Pseudoalteromonas arctica* (Par-Lig), in an attempt to understand cold adapted properties among DNA ligases.

Gel-based endpoint assays and Differential Scanning Calorimetry (DSC) experiments showed variable activity optima for nicked substrates and thermal denaturation temperatures for the ligases cloned from psychrophilic organisms. While Vib-Lig exhibits classical cold adapted features such as low temperature optimum of activity and decreased thermostability, Psy-Lig and Par-Lig in contrast show activity and stability properties more similar to mesophiles and are therefore not characterized as cold adapted. Relating the enzyme activity with thermal stability also showed that the decrease in activity of Vib-Lig occurs ahead of the beginning of the unfolding transition.

Amino acid sequence comparisons and homology models of the ADLs, including ADL from the presumably mesophilic *Vibrio cholera* (Vch-Lig), show that all three psychrophilic derived enzymes examined exhibit a decreased number of arginines and a general elevated number of hydrophobic residues compared to Vch-Lig. Interestingly, comparisons of Vib-Lig, Psy-Lig and Par-Lig indicate a correlation between a reduced number of Arg residues and decreased thermostability. The arginine residues are mainly substituted by hydrophobic or uncharged amino acids located in non-DNA-binding surface exposed regions, introducing local surface-exposed patches with greater hydrophobicity than Vch-Lig. Catalytic sites and the interdomain linker regions show high conservation between the four enzymes. Finally, Vib-Lig is likley destabilized by fewer hydrogen bonds and its electrostatic surface potential shows that the substrate binding surface is optimized towards binding of the negatively charged DNA.

In summary, the cold adapted characteristics observed for Vib-Lig relative to homologues is reached by subtle structural adjustments distantly from the catalytic center that introduces unique hydrophobic or uncharged surface patches, thus influencing protein flexibility indirectly rather than in specific areas essential for activity.

4 Results and discussion

The results and discussion of three individual nucleic-acid-modifying proteins originating from Arctic marine bacteria are compiled in a common discussion highlighting interesting aspects of this study; protein production and stability challenges, activity on nucleic acids, structural insight, oligomeric states and metal requirements, cold adaptation, biological roles and applications.

4.1 Protein production and stability

4.1.1 Protein production. Three proteins – three strategies.

The three nucleic-acid binding proteins described in this study are extremely diverse in terms of structure and biochemical properties. Since there is no universal protocol for gene expression and purification, choosing an appropriate protocol for each individual recombinant protein is critical.

In particular, downstream applications and behavior/characteristics of the proteins were considered in order to establish individual and applicable expression and purification schemes for our widely different proteins in terms of structure and function, all sharing common goals; homogeneous and stable proteins in satisfying yields with high purity, able to perform *in vitro* biochemical assays and structural studies. This includes evaluating the origin of the proteins and environmental adaptations that may affect their behavior and stability, e.g the low temperature origin in this study. However, despite thorough investigation of the proteins and the literature in the planning process, identifying proper protocols required significant amount of trial and error.

Prior to this study, a large-scale purification procedure of AsFur was established, however, the protein aggregated quickly [121]. Therefore, in this study, significant effort was added to identify conditions where the protein is stable in its active form (discussed in section 4.1.4).

The gene encoding MG Orn originates from an Arctic marine metagenome library collected from the littoral zone and an expression and purification scheme was

developed from scratch; testing various expression plasmids, *E. coli* strains, cultivation media and temperatures.

The ADLs examined in this study are isolated from species inhabiting a consistently low temperature environment; Psy-Lig from *Psychromonas spp* strain SP041, Vib-Lig from *A. salmonicida* and Par-Lig from *P. arctica* (Table 1, paper III) [120,221]. Psy-Lig and Vib-Lig were previously recombinantly produced [49,166], whereas Par-Lig was recombinantly expressed and purified in this study using similar protocols.

Results and challenges are discussed in the following sections.

4.1.2 Protein expression

All our proteins were expressed in the *E. coli* system, which is a rapid, straightforward, readily accessible and relatively inexpensive method for small- and medium scale production. The host growth temperature was lowered to 20 °C prior to induction by IPTG and expression of the three proteins. Lowering the temperature to 15 - 30 °C (slower expression) during the expression period may overcome issues with aggregation into inclusion bodies and enhances the chances for proper folding of the produced proteins [222]. This is particularly important in the production of functional cold-active proteins, as the lower optimal growth temperature of these proteins are incompatible with the mesophilic expression host [223].

For MG Orn, various cultivation strategies were tested. Rosetta 2 (DE3) host cells showed better growth compared to BL 21 Star (DE3) and expression at 20 °C showed sufficient growth of soluble protein compared to 15 °C.

Protein production details for the three proteins are listed in table 2.

Table 2. Protein production specifics for AsFur, MG Orn and Par-Lig

	AsFur	MG Orn	Par-Lig
Fusion tags	No (natural His-tag)	His - MBP	His – MBP - His
Expression vector	pDEST14	pHMGWA	PHMGWA
***E. coli* host**	BL21-CodonPlus® (DE3)-RIL	Rosetta 2 (DE3)	BL21 Star™ (DE3) pLysS
Expression temperature	20 °C	20 °C	20 °C
Expression time	Overnight	4 hours	Overnight

Another critical step to achieve high purity of proteins is the choice of affinity tags as anchoring points to facilitate chromatographic methods for individual recombinant proteins, if needed, as well as to enhance protein solubility and stability, and even increase expression levels [224].

The His-tag (six consecutive histidine residues), added to both Par-Lig and MG Orn constructs, is a very attractive structural affinity tag, as it is short, does not interfere with expression and if desired, can be easily removed by TEV protease following nickel-immobilized affinity chromatography (Ni-IMAC). AsFur has a stretch of His residues within its amino acid sequence, acting as a "natural" affinity tag with ability to immobilize to Ni-IMAC, thus this protein was produced without tags. This highlights the value of studying the amino acid sequence of the target protein in the planning process, as it may reveal unique properties that might assist in purification, e.g size, pI and interesting residues.

The maltose-binding protein (MBP) is one of various fusion partners shown to increase solubility of recombinant proteins and it is suggested that large and bulky fusion tags such as MBP may sterically prevent intracellular binding of DNA which otherwise disturbs the cells' own metabolism and thereby avoiding toxic effects [225]. Work by Niiranen *et al* showed increased yields and improved solubility of MBP-fused proteins from the psychrophilic fish pathogen *A. salmonicida* expressed in *E. coli*, compared to a range of other tested N-terminal fusion partners [226]. Additionally, MBP was also one of the fusion partners with best success rates in terms of protein solubility in work performed by Bjerga *et al*. [223].

Overexpression of Vib-Lig has previously shown moderate toxic effects on *E. coli* cells, both at 37 °C and 22 °C [49], however this issue was overcome by adding the fusion tag

MBP to the construct, with similar cell growth rates as pre-induced cells. For production of the Vib-Lig homolog Par-Lig (paper III) and MG Orn (paper II), small scale expression and solubility testing showed similar positive effects by addition of an MBP fusion partner.

Finally, it has been shown that proteins with extended or natively unstructured regions can be unstable. Previous expression of Vib-Lig showed improved stability without the leader sequence [49], and Par-Lig was successfully expressed in its truncated form as well.

4.1.3 Protein purification

Protein purification was performed at 4 °C, which ensures a low proteolysis rate and promotes structural integrity of proteins, and is especially crucial for our proteins originating from cold habitats.

Choosing the appropriate purification methods and their order can be time-consuming, often by trial and error, however, is crucial for success. It is also advantageous to keep the number of purification steps and buffer exchanges to a minimum.

AsFur, MG Orn and Par-Lig were purified by two or more chromatographic steps to reach the necessary purity for downstream applications, with variable results in terms of purity, stability and yields (fig. 18). A simplified purification scheme is shown in fig. 18a, while fig. 18b shows the purified products of the three proteins as used in biochemical and biophysical studies; AsFur (16.6 kDa), MG Orn (21.5 kDa) and Par-Lig (30.8 kDa).

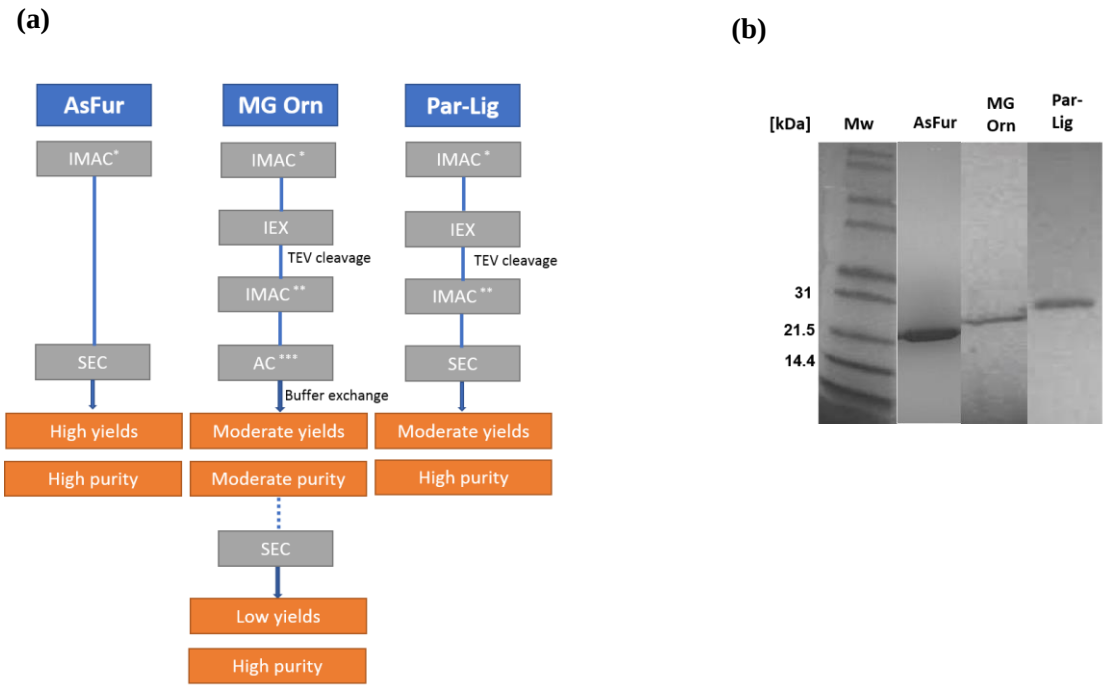


Fig. 18. Purification of AsFur, MG Orn and Par-Lig. (a) Purification steps of the three individual proteins including evaluation of yields and purity. The dotted line indicates an additional purification step utilized for structural studies of MG Orn. An asterisk (*) indicates Immobilized metal ion affinity chromatography (IMAC) by the HisTrap HP column, two asterisks (**) indicate Reverse-IMAC by the HisTrap HP column and three asterisks (***) indicate affinity chromatography (AC) by the HiTrap™ Blue HP column. **(b)** SDS-Page of purified AsFur (16.6 kDa), MG Orn (21.5 kDa) and Par-Lig (30.8 kDa) as eluted from the final purification step, prior to concentration and storage. Ten µl of each protein is added to the gels. The standard marker (Novex Mark 12, Thermo Fisher Scientific) from the purification of MG Orn is shown and each of the other proteins are scaled to the standard in the figure.

The purification of AsFur and Par-Lig mostly followed previously described purification schemes, while the establishment of an MG Orn purification protocol required more testing/trial and error.

The most profound challenge during purification of MG Orn was the vast amount of product lost during TEV cleavage, up-concentration and desalting. Performing dialysis as an alternative desalting procedure to columns increased the yield to some degree. Following the final HiTrap Blue HP affinity step, an additional size exclusion

chromatography step was employed in cases where the highest purity was obligatory, as preparations for crystallization and stoichiometry analysis, although a significant amount of protein was lost in the process.

4.1.4 AsFur stability challenges

The previous difficulties with producing stable AsFur due to rapid aggregation were addressed.

As stated, proper protein function depends on correct structural conformation of the protein, thus solution conditions are also important to consider when designing an optimal expression and purification protocol. Thermal shift assays were performed to improve buffer compositions, salt (fig. 5, paper I) and pH, and various temperatures and additives were tested for storage. We were mostly unsuccessful in finding better suited purification buffers and long-term storage conditions of AsFur (data not shown) but found that by a minor change in the Tris buffer pH from 8.0 to 7.5, aggregation was apparently reduced and AsFur could be stored for approximately 4 weeks at 4 degrees without affecting DNA binding capacities noticeably. Consequently, to achieve reproducible results with active protein, the purification procedure had to be performed regularly throughout this study in order to work with as fresh protein as possible.

More time and effort could potentially improve the expression and purification procedures, e.g psychrophilic expression systems or using removable solubility tags as maltose binding protein (MBP), however, the limited timeframe of this study prevented any further optimization.

4.2 Activity on nucleic acids

Various biochemical assays were utilized to investigate the activity and specificity of our nucleic acid binding protein. This section focuses on new insight into the DNA interaction with the regulator AsFur and into the enzymatic ribonuclease specificity of MG Orn. Ligase activity of Lig E on nicked and overhang substrates are discussed in section 4.5.2.

4.2.1 Opening the Fur box- binding of AsFur to DNA

Although previous characterization of AsFur-DNA interactions have proposed specific amino acids and nucleic acids involved, the structural basis underlying the binding of AsFur to DNA is still controversial [121]. To further evaluate the importance of individual nucleotides or binding sites and thus better understand the exact AsFur-DNA binding mode, we experimentally tested a number of oligonucleotides of different lengths and compositions based on the *E. coli* and Vibrio Fur box consensuses sequences as templates (Table 2, paper I).

EMSA revealed that AsFur is able to effectively interact with both the *E. coli* and the Vibrio Fur box consensus sequences, although with stronger affinity towards the latter. This is not surprising considering the Vibrio origin of AsFur.

The Fur residue T13 has previously been suggested to be involved in specific Fur-DNA recognition in *E. coli* [227] and AsFur [121], by crosslinking experiments and molecular dynamics, respectively. We have experimentally verified the importance of T13 in the AsFur-Vibrio Fur box interaction by substituting the nucleotides T12 and T13 (fig 8E, paper I). Similarly, the effect of A14G/C16A substitutions to the Vibrio consensus sequence (fig 8F, paper I), as well as the corresponding substitutions A17G/C19G to the *E. coli* consensus sequence (fig 8F, paper J), confirms the important role of these nucleotides in AsFur-DNA recognition, in line with previous computational predictions [121].

As described in the introduction, the minor groove of the DNA is often AT-rich. The bases A and T have been considered essential in DNA recognition by Fur, contributing to shape readout mechanism through minor groove interactions. For this reason, we opted to substitute A8 and T9 in the Vibrio Fur box, which clearly highlighted their prominent role in AsFur-DNA binding (Fig. 8B, C and D, paper I). The importance of T9 (corresponding T6 in *E. coli*) was also demonstrated by Escolar *et al* (1998); T_6 of each hexamer unit in the hexamer NATA/TAT (various combinations of three direct repeats of 5-'$G_1A_2T_3A_4A_5$**T_6**-3' units (fig. 9b and c) interacted with EcFur [114]. EcFur was shown to interact through sugar-phosphate bonds flanking T_6 (last base) of the GATAA**T** unit (corresponding to the T9 substituted in the Vibrio consensus sequence). Additionally, the importance of this T_6 base was confirmed by missing T assays, which also indicated involvement of T_3, and T_2 and T_5 on the complementary strand, but to a lesser extent [114].

Compared to the Vibrio least conserved sequence, AsFur show slight increased ability to bind a mutated Vibrio Fur box sequence devoid of the essential nucleotides shown to contribute to AsFur-DNA binding (fig. 8G, paper I). The fact that only the Vibrio least conserved sequence depletes DNA recognition completely may indicate that additional shape-readout mechanisms are also involved in DNA recognition by AsFur, and supports previous studies describing how Fur appears to have a rather broad substrate affinity and is able to recognize numerous diverse genes and degenerate sequences [56] as a global regulator. For instance, computational prediction and identification of Fur-regulated genes and operons in Vibrios identified slightly different binding sites for genes involved in iron homeostasis, virulence and energy metabolism [123].

The main binding elements in the Fur box consensus predicted for various Fur families are rather similar, however as described in the introduction, the functional patterns have been interpreted differently. To study Fur box arrangements, experiments were performed on oligonucleotides arranged as three or four hexamer repeats of the *E. coli* Fur box (GATAAT 19-mer sticky-end and 24-mer quadruple repeat, respectively). These Fur box variants weakened the interaction compared to the corresponding consensus sequence, similar to previous studies on EcFur [59] (fig. 8K and L, paper I). Thus, this study indicates that the 7-1-7 model following the F-F-x-R arrangement seems to be a more valid Fur box model compared to the hexamer repeat.

4.2.2 Biochemical characterization of ORN: a ribonuclease active on short substrates

Given the inconclusive literature data on the substrate preference of Orn, we chose to investigate the optimal RNA substrate length of ORN as part of its general enzyme characterization.

Two different approaches were employed to monitor the nuclease activity of MG Orn; a spectrophotometric assay using the dinucleotide analogue 5'-*para*-nitrophenol ester of 5'-monophosphate (*p*NP-TMP) as substrate and a polyacrylamide gel assay using FAM-labeled RNA substrates of different lengths.

Substrates of various lengths were tested in gel assays both to confirm functionality of MG Orn and to study the substrate specificity. MG Orn shows efficient nuclease activity on both 5mer, 7mer and 10mer RNA substrates (fig. 6 and fig.12, paper II). In addition,

nuclease activity on 20mer RNA using an abundant amount of enzyme was also observed (data not shown). Although substrate affinity was not quantitively determined, these results are consistent with the ability of *E. coli* Orn (Ec Orn) to act on longer substrates (2-5nt) [46]. More extensive data for Ec Orn substrate specificity indicated strongest affinity for short RNA fragments (2-5 nt) and a nuclease activity inversely proportional to the length of the substrate, with a 5mer RNA being most preferable. Kinetic analysis further indicated slower nuclease activity rates as longer RNA fragments are degraded to shorter fragments, subsequently hydrolyzed at higher rates. Thus, Ec Orn showed better affinity for the longer substrates up to 5 nt, but faster reaction rate for shorter substrates [142].

The robust nuclease processing observed by MG Orn and Ec Orn on longer substrates (>2 nt) is in contrast to two recent structural studies of substrate-bound Orn that revealed new insight into RNA-binding structure and function, and suggested a limited RNA substrate length affinity (2-5 nt). Orn from *Colwellia psychrerythraea* strain 34H (Cps ORN) is able to rapidly degrade 5´-fluorescence-labelled 5mer RNA as well as *p*NP-TMP, but not the longer 24mer RNA. It was concluded that Cps Orn has a length specificity for short RNA substrates, however, no intermediate substrate lengths between 5 and 24 nt substrates were functionally analyzed, and it would be interesting to test whether the length preference is limited to short RNAs up to 5 nt, for comparison with MG Orn. In contrast, *V. cholera* Orn (Vc Orn) showed a significant lower affinity for longer substrates than diribonucleotides. The authors therefore propose an in vivo function of Orn as an exoribonuclease targeting diribonucleotide solely and suggest renaming Orn to dinucleotidase [228].

Although RNA is the preferred substrate, MG Orn was also able to act on short ssDNA oligonucleotides *in vitro*, at increased enzyme concentrations (data not shown). Such cross-specificity is common among other RNA specific 3'- 5' exonucleases, and may be expected considering that the DEDD superfamily contains both RNases and DNases [57,63]. The DNase activity observed by the human Orn homolog Sfn raised the possibility that Orn could serve a role in DNA damage repair or recombination *in vivo* [143]. We therefore speculate if the ability of MG Orn to degrade ssDNA indicates a role in DNA repair as well.

Based on the data from biochemical assays we conclude that MG Orn is able to effectively degrade longer substrates. Structural explanation of the MG Orn substrate binding pocket accommodating larger substrates is discussed in section 4.3.3.

4.3 Structural insight

Obtaining a high-resolution crystal structure of a protein provides useful information in understanding how it carries out its function as well as the basis of substrate and other optima. Even more informative is capturing the structure of a protein in complex with its substrate or a substrate-like inhibitor. AsFur and MG Orn were subjected to crystallization trials during this study, however only the three-dimensional structure of MG Orn was successfully determined. In the absence of three-dimensional crystal structures, structural features of AsFur and Vib-Lig were studied by structural modelling.

4.3.1 AsFur crystallization trials

Considerable attempts were made to crystallize apo-AsFur, metal-bound AsFur and AsFur in complex with various Fur box-containing DNA oligomers. The selection of oligomers is crucial for successful crystallization. Based on the binding strengths observed in EMSA, the oligomers L, K and *E. coli* consensus and Vibrio consensus (table 2, paper I) were chosen for complex crystallization.

The experiments included several automated screens with different conditions in 96-wells plates, both in-house made and commercially screens, at 4 degrees and room temperature. Promising conditions were further screened in 24-well plates by the hanging drop method. Although some protein crystals were obtained, most of them diffracted poorly. One crystal diffracted to 2.4 Å in a putative complex with one of the oligomers, however attempts to solve the three-dimensional structure by molecular replacement were unsuccessful. Since at the time no structures of AsFur in complex with DNA were available, great effort was put into reproducing the crystals with similar conditions. Unfortunately, we were not able to reproduce the crystals within the limited time available, but as a DNA bound structure was published soon after, we chose to focus on in-silico studies of homology models.

4.3.2 AsFur homology modelling and DNA specificity

In order to further characterize and validate suggested AsFur-DNA interactions, a homology model of AsFur was built, based on the recent structural models of MgFur in complex with two different oligonucleotides; the *P. aeruginosa* Fur box (identical to the *E. coli* Fur box) and the *feoAB1* operator [56].

Although MgFur and AsFur share only 37 % sequence identity between the 135 residues that can be aligned, the sequence alignment shows conserved sequence patches both in the DD domain and in regions involved in Fur-DNA interactions in the DBD. The AsFur homology model further supports the conservation of residues involved in both shape-readout and base-readout recognition of the Fur box, and similar to MgFur, these residues form matching contacts with both Fur box targets, despite nucleotide sequence variations between these two. Additionally, by comparing these conserved residues with the key nucleotides identified in the EMSA experiments discussed in section 4.2.1, possible AsFur-DNA interactions can be described.

The conserved amino acid Tyr56 has previously been indicated to be involved in base-readout with T13 (corresponding to T16 in the *E. coli* consensus sequence) [121,227]. Our EMSA results (fig. 7 and 8, paper I), combined with the two AsFur models (fig. 10, paper I), further demonstrates the role of Tyr56 in forming hydrophobic and base-specific major groove interactions. Next, as discussed in section 4.2.1, the important role of A14 and C16 in AsFur-DNA binding was evident in the EMSA studies. Our homology model shows that these nucleotides form both minor- and major-groove interactions with Arg57 and Lys14 via the complementary strand (T3' and G1') of the *E. coli* Fur box, highlighting the role of both DNA strands in AsFur-DNA interactions. In line with this, structural analysis of MgFur in complex with DNA combined with mutagenesis and biochemical assays highlights the role of these conserved residues. Lys15 (Lys14 in AsFur) is involved in shape readout through a narrower minor groove in a non-ideal B-DNA form and Tyr56 and Arg57 recognize DNA through base readout [56]. Furthermore, site-specific mutagenesis targeting positively charged residues located in the DNA binding domain of holo-CjFur showed that Tyr68 and Arg69 (corresponding to Tyr56 and Arg57 in AsFur) are involved in interaction with DNA [106].

Central AsFur-DNA interactions confirmed by homology modelling are summarized in fig. 19.

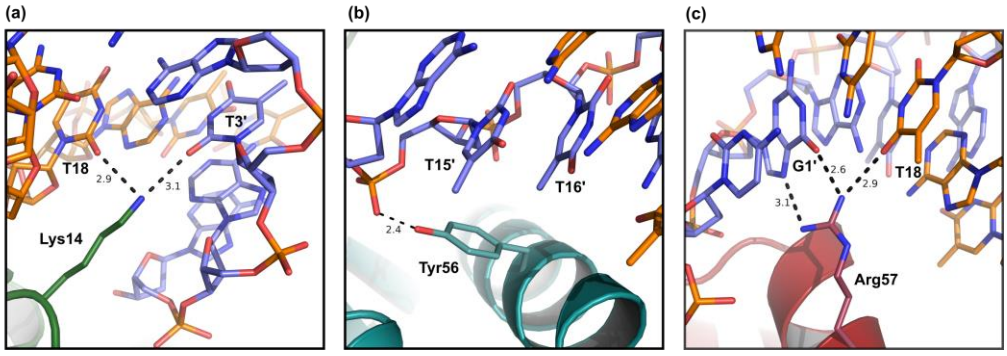


Fig. 19. **AsFur-nucleotide base interactions.** Nucleotide base-interactions observed in the AsFur homology model based on the crystal structure of two dimers of MgFur in complex with the *E. coli* Fur box (PDB4rb1). Nucleotide numbering follows the numbering scheme used for the *E. coli* consensus sequence in Table 1. **(a)** Lys14 in monomer A interacts in the minor groove with T18 on the primary strand and T3' on the complementary strand. **(b)** Tyr56 in monomer B forms hydrophobic interactions in the major groove with T15' and T16' on the complementary strand (identical interactions are formed between Tyr56 in monomer D, generated through a crystallographic symmetry operation, and T15/T16 on the primary strand). **(c)** Arg57 in monomer D interacts in the major groove with T18 on the primary strand and G1' on the complementary strand.

Structural analysis combined with *in vitro* assays has given a deeper understanding of the variety of Fur binding sites and explains how the protein is able to recognize a number of diverse DNA targets and degenerate gene sequences. However, the nucleotide composition of the DNA alone does not explain how Fur precisely identifies its functional binding sites. The many high-resolution structures of general transcription factors that are currently available shows that most DNA-binding proteins use interplay between the base- and shape-readout modes to recognize their DNA binding sites [6,39,56], including MgFur. Deng *et al* identified essential DNA sequence-dependent effects on Fur activity and highlights the important role of shape readout through minor groove electrostatic potential in combination with base readout through direct contacts in the major groove, and even suggested that the latter is of less importance since most MgFur-DNA interactions were sequence independent [56].

Although we have expanded our understanding of the AsFur-DNA interaction at the molecular level, much is still unknown concerning how AsFur recognizes similar functional binding sites in the genome *in vivo*. In this study, important nucleotides in the Fur box are identified by assays of the core binding sites without any flanking

regions. Revealing the specificity of protein-DNA recognition solely through *in vitro* and *in silico* methods excludes other potential determinants affecting DNA recognition, for instance oligomerization along the DNA, and thus does not provide a full understanding of the diverse regulatory networks observed in the cellular environment.

In summary, we have verified the importance of previously proposed sequence-specific nucleotides and amino acids in AsFur-DNA recognition and gained further insight into the structural basis of these interactions.

4.3.3 MG Orn crystal structure and new insight into binding of longer RNA substrates

The crystal structure of MG Orn (fig. 6, paper II) was determined to 3.15 Å resolution by molecular replacement using oligoribonuclease from *X. campestris* (PDB: 2GBZ) as a model (paper II).

As discussed in section 4.2.2, gel assays of MG Orn showed a broad substrate specificity. In order to structurally rationalize the observed *in vitro* cleavage of longer substrates by MG Orn, important conserved residues involved in the stabilization of longer RNA substrates are identified by using *E. coli* Exonuclease I (ExoI in complex with ssDNA, pdb 4JRP) as a template for manual docking of a 5mer RNA into the MG Orn active site. In the docking model, the 3' end of the nucleotide substrate (ssDNA in ExoI; RNA in MG Orn) is coordinated in the active site at overlapping positions in MG Orn compared to ExoI, whereas nucleotides in the 5' end appear to be in tight interaction with a conserved loop sequence patch located in the second monomer of the functional dimer of MG Orn (fig. 8, paper II).

In contrast to the broad substrate specificity for MG Orn, the recently determined co-crystal structures of Cps ORN and Vc Orn both indicate preference for dinucleotides. Co-crystallization of Cps Orn shows that nuclease activity is limited to substrates that are 2-5 nt long, with an active site accommodating two nucleotides [229]. Any additional nucleotides are located outside of the binding pocket in a flexible state, as observed for the 5mer RNA liganded structure. It is argued that Cps Orn possesses a relatively small and short substrate interaction surface compared to other exonucleases, possibly explaining why its activity is limited to short RNAs. Further, where mutation of Tyr129 in Cps Orn indicated that it is not important in the processing

of dinucleotide substrates [229], the model of MG Orn with the 5mer RNA implicates that this residue, completely conserved among 150 Orn homologs, is involved in processing of RNA molecules longer than dinucleotides. The second co-crystal structure; Vc Orn with ligands, reveals a substrate preference for diribonucleotides only. Co-crystal structures with pGpG and other linear diribonucleotides revealed an active site optimized for these, unable to accommodate longer substrates [228].

Thus, modelling of RNA into MG Orn explains how MG Orn can accommodate and process longer RNA substrates *in vitro* by interacting with a conserved sequence patch in the second monomer.

4.3.4 Structural modelling of Lig E

Previous attempts to crystallize Vib-Lig were unsuccessful and obtaining the required amounts of concentrated protein for Par-Lig and Vib-Lig for further crystallization screening proved to be challenging. The available three-dimensional structures of apo Psy-Lig [166] and a homolog of DNA bound Lig E, and their high sequence similarity to Par-Lig, Vib-Lig and Vch-Lig, allowed us to build homology models of these in order to investigate the structural adaptations behind the cold-adapted behavior observed for Vib-Lig in particular. Comparison of amino acid sequences and structural models show conserved catalytic and DNA-binding regions between our cold adapted Lig Es and the mesophilic Vch-Lig but highlights interesting variations in other areas of the proteins. These are discussed in greater depth in section 4.5.3 on cold adaptation.

4.4 Roles of oligomerization and metal-binding in the nucleic acid-interacting proteins

Biomolecular oligomerization plays an important role in numerous biological processes [28]. Studying the structural oligomeric state of a protein may give insight into protein function and is of particular interest for nucleic acid binding proteins.

Among nucleic acid binding enzymes, metal ion cofactors are required for activity of nearly one-third of enzyme-mediated processes. Although various divalent metal ions can play this role *in vitro*, Mg^{2+} is most frequently involved in enzymatic catalysis, which

compared to other metals has the advantage of abundance, solubility, redox stability, relatively small size, rigid coordination geometry and unique hydration property [35,230]. Mg^{2+} forms stable complexes with phosphate-containing species, including ATP, under physiological conditions [231].

As mentioned in the introduction, the three proteins studied differ in terms of oligomerization and metal ion dependence, and this section will focus on Fur and MG Orn. The Lig Es are shown to be monomers structurally dependent on divalent cations for activity, but this was not a focus of our study on Lig E. It is however worth mentioning that rather recently, two papers gave considerable insight into the two-metal mechanism in ADLs, for both step 1 and 3 in the catalytic process [58,156] and these metal binding sites are conserved within Lig Es.

4.4.1 AsFur binds DNA in an oligomeric metal-dependent manner

The metal-responsive transcription factor Fur has been shown to act in various oligomeric states, with and without metal activation. AsFur is present as a homodimer in solution. Rationalization of the EMSA results presented in section 4.3.2 with the homology model of one dimer of AsFur in complex with the *feoAB1* operator and two dimers in complex with the *E. coli* Fur box indicates that AsFur most likely interacts as two dimers. However, it is not evident from this study whether more complex oligomeric states can possibly form during AsFur activity.

Although Fur was originally described as an iron-dependent repressor *in vivo*, studies have shown that Fur responds to and is activated by a range of divalent metals *in vitro* [51,56,67,102,232,233]. In this study, the capability of various metals to activate AsFur was evaluated by the plasmid protection assay. AsFur protected the plasmid in a strictly metal-dependent fashion the presence of the divalent metal cations Mn^{2+}, Zn^{2+}, Cu^{2+} and Co^{2+} (Fig. 6, paper I). AsFur complexed with Mn^{2+} showed considerable Fur box binding activity, although appearing weaker compared to the other metals. Since Mn^{2+} was used with success in previous studies on AsFur and other Fur homologues, and also shows similar metal coordination structurally to the preferred metal *in vivo*, Fe^{2+}, in structural studies [108], it was chosen as the preferred activating metal in subsequent studies.

In vivo, it is believed that only Fe^{2+} is present in sufficient quantities for Fur activation [67], however it is argued that elevated levels of other metals intracellularly could potentially interfere with gene regulation by Fur and even expand the DNA target repertoire [234]. Interestingly, in *A. salmonicida*, the study of a *fur* null mutant indicated up-regulation of several transport systems involved in homeostasis of other metals than iron, including a multidrug efflux pump and nickel and zinc transporter genes [235].

Thus, this study reports metal-dependent binding of AsFur to the *E. coli* Fur box consensus in the presence of numerous divalent metal ions.

4.4.2 MG Orn disulfide bond and homodimer formation

In line with most Orn proteins and other exonucleases characterized so far, MG Orn forms a homodimer in solution [68,145], which is implicated as important for the apparent processive oligoribonucleotide cleaving mechanism. The formation of the dimeric interfaces varies immensely among exonucleases, even among Orn from different species. We ought to study the importance of the disulfide bond in dimer formation and thus the function of MG Orn.

The crystal structure of MG Orn shows that an intermolecular disulfide bond connects the two monomers at the homodimer interface, in addition to other interactions as hydrophobic interactions, salt-bridges and hydrogen bonds (Supplementary Fig. 1, paper II). Compared to other homodimer interfaces in known Orn crystal structures (table 2, paper II), not all Orn homologues form an intermolecular disulfide bond between the two monomers, but it is found in *X. campestris* Orn (Xc Orn), *C. burnetii* Orn and *A. baumannii* Orn. Orn proteins that lack this disulfide bond contain an Ala or Gly residue instead of the cysteine residue at position 110. Further, compared to other Orn, MG Orn has the smallest buried surface area and number of hydrophobic interactions, and the disulfide bond may add major stability to the MG Orn homodimer. The importance of this intermolecular disulfide bond in both overall stability and hydrolytic activity has not been demonstrated in Orn so far.

The cysteine residue involved (Cys110-Cys110') in dimer formation in MG Orn was mutated to Ala and Gly to investigate the function in homodimer formation and nuclease activity. Compared to MG Orn, the mutants show considerable (more than 80 %) and complete loss of ability to degrade the dinucleotide substrate mimic *p*NP-TMP

and 7mer/10mer RNA substrates (fig. 11 and 12, paper II). Thus, loss of the intermolecular disulfide bond severely affects MG Orn's ability to act as an exoribonuclease, demonstrating the importance of residue C110 in homodimer formation and consequently, catalytic ability of MG Orn to degrade oligoribonucleotides of various lengths.

The importance of the intermolecular disulfide bond in homodimer formation and exonuclease activity was further investigated by adding the reducing agent DTT to MG Orn. Interestingly, only excessive amounts of DTT were able to slightly reduce exoribonuclease activity. These results indicate that even if the intermolecular disulfide bond is essential for homodimer formation, once the dimer is formed, other interactions as hydrophobic interactions, salt-bridges and hydrogen bonds bring sufficient strength to stabilize the homodimer interface.

4.4.3 Metal dependence of MG Orn

The metal dependence of MG Orn was evaluated using the *p*NP-TMP activity assay, which showed an absolute requirement for a divalent metal-ion, preferably Mn^{2+} over Mg^{2+} (Fig. 3A and B, paper II). Typical negative effects observed by replacing Mg^{2+} with other divalent metal ions *in vitro* are reduction of enzyme efficiency, instability or poorer substrate specificity. Interestingly, replacing Mg^{2+} with Mn^{2+} in our study on MG Orn led to better purification yields and catalytic activity rates, but with the cost of enzyme stability. Destabilization by Mn^{2+} might be due to the instability of the metal under slightly basic conditions and the tendency to be oxidized, or less ideal coordination geometry in the metal binding site. Activity studies on Ec Orn and Cps Orn show similar metal dependence; catalytic activity towards *p*NP-TMP in the presence of Mg^{2+} or Mn^{2+}, with 20% more efficient activity bound to the latter [45,142]. Other tested divalent and monovalent cations had no effect on activity.

The variations of metal ion:enzyme stoichiometry among nucleic acid-processing enzymes led to the proposal of one, two or three metal ion-mediated mechanisms. Coordination by two Mg^{2+} may form the basis for catalytic specificity, facilitating phosphoryl transfer at appropriate distance. However, one- and two-metal catalysis share a common catalytic metal ion [231].

Most of the dimeric 3´ - 5' exonucleases of the DnaQ superfamily have the DEDDh motif with one- or two-metal binding. The determined crystal structure of MG Orn gave further insight into its metal-binding and suggested a possible one-metal mechanism. Similar to Cps Orn, each monomer of MG Orn contains an Mn^{2+} ion in the active site, probably originating from the protein solution buffer. Unliganded and substrate-bound Cps Orn crystal structures revealed a homodimer with two separate active sites, residing one divalent cation each [229]. Although the structural data could not completely rule out the possibility of a second metal binding, a one-metal-dependent reaction mechanism was suggested for Cps Orn [231].

The MG Orn structure revealed that a Mn^{2+} ion is coordinated by Asp12, Glu14 and Asp163, similar to Cb Orn and Cps Orn. The *p*NP-TMP complexed D163A mutant Cps Orn crystal structure and *p*NP-TMP activity assays demonstrated the importance of the conserved Asp163 in metal coordination and RNA hydrolysis. Other solved Orn structures contain one Mg^{2+} ion in each monomer, e.g Xc Orn and Cp Orn. These metals are also coordinated by Asp12 and Glu14 as the Mn^{2+} in MG Orn, but the Mg^{2+} bound Asp112 is not involved in binding of the Mn^{2+} ion, thus binding site differs slightly.

4.5 Low-temperature adaptation

The nucleic acid binding proteins included in this study origin from organisms in Arctic marine environments and we expect these to be cold adapted. Therefore, it is interesting to explore possible cold adapted features for all three proteins, although only the Lig Es were extensively studied for this exact purpose. Relevant properties discussed for one or more of our proteins in the following sections include thermal stability, temperature optimum, amino acid sequence, and structure of the proteins.

4.5.1 Thermal stability of cold adapted proteins

Recombinant protein expression and purification of functional proteins for *in vitro* studies can be challenging, especially when working with proteins adapted to low temperatures. Cold adapted enzymes typically show intrinsic molecular instability due to increased molecular flexibility as an adaptation strategy to compensate for higher

activity at lower temperatures. Thus, these proteins are often more temperature labile and show general stability issues in experiments, even by bacterial cultivation at reduced temperatures and purification at 4 °C [223].

In line with this, our marine Arctic-derived AsFur showed major stability issues after purification (discussed in 4.1.4), which has not been reported for other Fur homologues. Difficulties to reproduce AsFur co-crystals may also be a result of the intrinsic instability observed for psychrophilic proteins.

Further, considerable loss of MG Orn activity was monitored during ultracentrifugation and long-term storage, as well as observed loss before and during TEV cleavage, resulting in high deviations in replicate measurements as well as in individual assays after storage (data not shown). These stability issues may relate to the generally low unfolding barrier caused by high flexibility typical for cold adapted enzymes.

Finally, although the psychrophilic Vib-Lig showed better overall stability during experiments compared to AsFur and MG Orn, unsuccessful crystallization trials with multiple commercial and in-house screens could be linked to intrinsic instability.

Protein stability is often analyzed by various thermal denaturation-based methods. In this study, a combination of methods was used to characterize the three nucleic acid binding proteins biophysically; Thermofluor as a low-precision high-throughput method also suitable for screening, and DSC as a high-precision and reliable low-throughput method that requires larger quantities of protein.

Thermal stability of AsFur and MG Orn

The various thermal shift assays performed in order to improve AsFur buffer compositions, described in section 4.1.4, indicate a T_m of 40 - 44 °C for AsFur. In general, knowledge about stability and thermodynamics of Fur is limited in the literature. However, the mesophilic FurA from Anabaena was studied by various biophysical techniques and found to be highly thermally stable with a T_m of around 79 ± 5 °C [236]. This was slightly unexpected as the conformational stability was only moderate, but was explained by a small heat capacity possibly due to the presence of buried polar residues in the homology model of monomeric FurA. We did not study the amino acid sequence and model of AsFur in this context, however, the low thermal denaturation temperature indicates poor intrinsic stability of AsFur.

Also relevant to cold adaptation is the lower inactivation temperature of MG Orn in pNP-TMP assays (Fig. 4, paper II), with a half-life of about 15 min at 50 °C, compared to

the mesophilic homologous enzyme from *E. coli* that shows about 50 % of residual activity after 15 min at 70 °C [45]. Even at 10 min incubation at 100 °C residual activity of *E. coli* Orn could be detected [45]. In agreement with the temperature stability profiles, Thermofluor assays indicate a melting point around 50 °C (supplementary fig. 4, paper II), showing considerably greater heat lability of MG Orn. The psychrophilic Cps ORN exhibits similar features; the activity drops sharply at temperatures over 50 °C. Additionally, CD experiments show a melting temperature already at 36 °C and denaturation of the protein at approximately 50 °C [229]. Thus, in terms of activity and thermal stability, the low-temperature sourced MG Orn shows possible cold adapted features compared to the mesophilic *E.coli* Orn.

Note however, that although these data indicate that MG Orn is cold adapted, the moderately low thermal stability and activity profiles show that similar to Psy-Lig and Par-Lig (paper III), MG Orn is better described as psychrotolerant than psychrophilic. Despite the possible psychrophilic origin in the Arctic littoral zone, these psychrotolerant enzymes are able to function sufficiently at low temperatures in their native environment, especially those who are not expressed continuously. As discussed in more detail for Lig Es in the next section, the thermal stabilities of the enzymes are not necessarily correlated to the optimal growth temperature of the host.

Detailed investigation of Lig E unfolding

More extensive protein stability characterizations, by both DSC and Thermofluor, were carried out in the study of Lig Es in paper III, focusing on cold adapted properties.

Biophysical determination of protein heat lability was performed by DSC experiments. The unfolding temperature (T_m) of the psychrophilic Vib-Lig occurred at considerably lower temperatures compared to Par-Lig and Psy-Lig (Figure 2a, paper III). Thermal stability data for the mesophilic homologue Vch-Lig are not available for comparison, but considering the nature of its origin in the host adapted to around 37 °C, it would be expected to show an even higher T_m. Thermofluor assays confirmed the results from DSC regarding unfolding temperatures (Figure 2b, paper III).

Our data are consistent with the paradigm that psychrophilic enzymes are typically characterized by a lower T_m than mesophilic homologues due to weaker and fewer interactions involved in maintaining the structure (e.g hydrophobic interactions, salt bridges and H-bonds) that can be disrupted at low temperature with less energy. Thus, thermal unfolding reflects the flexibility of proteins, effecting the ability to undergo fast

and efficient conformational changes during catalysis. However, T_m alone is not necessarily a good descriptor of cold adaptation, as a protein may be inactive even if it remains folded. Often, decreased temperature for activity of psychrophile-derived enzymes is attributed to an increase in flexibility with heat lability as a consequence [192,194,237,238]. However, for some psychrophiles the activity has been shown to drop at temperatures lower than the unfolding temperature (T_m) denaturing the structure, and an estimation of the overall stability would be uncertain. Comparison of a psychrophilic, mesophilic and thermophilic NDL [239] show a similar link between activity and thermal adaptation as we observed for our cold adapted Vib-Lig; a decrease in activity above T_{opt} is observed well before it denatures thermally. Thus, relating thermal unfolding with optimal activity gives a better understanding of cold adaptation.

4.5.2 Temperature optima of cold adapted proteins

Temperature optima of psychrophile-derived Lig Es

The growth temperatures of the host organisms are similar for the ADLs studied, showing temperature ranges typical for psychrophilic organisms; 1-22 °C [120]. Despite the psychrophilic origin of Vib-Lig, Psy-Lig and Par-Lig, only Vib-Lig exhibits classical cold adapted features with optimal activity at 20°C. In contrast, Psy-Lig and Par-Lig have pronounced higher activity optimums at well above growth temperatures of the host organism (Table 1, paper III).

Taken together with the thermal denaturation data, Vib-Lig thus exhibits classical features of cold adaptation including a low temperature optimum of activity and high thermolability compared to homologous enzymes, whereas Psy-Lig and Par-Lig have activities and thermal stabilities more similar to mesophiles.

The optimal growth temperature of the host organism is not necessarily directly correlated to the thermal stability or optimal activity of a protein. A mesophilic organism can host both mesophilic and thermophilic enzymes, while psychrophilic organisms can in principle host enzymes with thermal stability and optimal activities at temperatures well above that of their physiological conditions, as recorded for Psy-Lig and Par-Lig. This has been observed for other psychrophile-derived enzymes in the literature, such as L-haloacid dehalogenase from *Psychromonas ingrahamii*, alcohol dehydrogenase of *Flavobacterium frigidimaris* and KUC-12-keto acid decarboxylases derived from *Psychrobacter*. Although the temperature optimum is relatively high, the

low activity at lower temperatures is probably sufficient for ligase activity. It could also be argued that bacterial ADL enzymes have not evolved entirely into its cold environment after recent horizontal gene transfer, an indication of incomplete evolutionary adaptation [184].

4.5.3 Low-temperature adaptation involves surface residues in Lig E

In paper III, amino acid composition and three-dimensional structural models of Vib-Lig, Psy-Lig, Par-Lig and their mesophilic homologue Vch-Lig were compared to identify amino acids and structural features possibly involved in cold adaptation of Lig E.

Vib-Lig show typical cold adapted traits, e.g an overall decrease in hydrogen bonds [208,240,241] and increased positive electrostatic potential near the active site and on the binding site of the OB-domain for better binding of the negatively charged DNA substrate [206,215,242-247]. However, what stood out from sequence comparison and homology modelling of our Lig Es, was the differing Arginine counts and their distribution in the protein. Arg has capability to form hydrogen bonds and salt bridges, possibly adding structural stability to proteins and can contribute in more interactions with surrounding amino acids than lysine [207]. Sequence analysis showed that the mesophilic Vch-Lig has significantly increased number of Arg compared to Vib-Lig, Psy-Lig and Par-Lig, as well as higher Arg/(Lys+Arg) ratio per residue (table 2, paper III). This agrees well with comparison of three structurally homologous NAD^+-dependent DNA ligases (NDLs) adapted to different temperatures, which showed a decreased number of arginine residues for the psychrophilic *P. haloplanktis* Lig (Ph Lig) [239]. Furthermore, lysine to arginine substitutions in the psychrophilic α-amylase from *P. haloplanktis* profoundly enhanced the protein stability, demonstrating the influence of arginine content in cold adaptation [248]. This strongly indicates that the Arg count plays a relevant role in cold adaptation of Vib-Lig, Psy-Lig and Par-Lig.

Examination of the structural models confirmed the similarities in overall folds and active/catalytic sites, indicating that subtle structural adjustments in other areas of the structure are important for achieving higher catalytic activity and heat lability. Interestingly, the arginine substitutions to hydrophobic or uncharged residues that were identified by sequence analysis are generally located on the surface away from DNA-binding surface exposed regions, thus introducing unique and local hydrophobic surface patches in Vib-Lig (fig 4, paper III). This substitution of polar residues with

hydrophobic residues on the surface may also be reflected by the lower percentage of hydrophobic residues in Vib-Lig compared to Vch-Lig. Fig. 4 in paper II further shows how an increasing number of Arg substitutions to hydrophobic ones appears to be correlated with decreasing Tm.

It has been argued that a higher proportion of non-polar groups exposed to the solvent decreases the protein stability due to the "ordering of water molecules", indicating an entropy-driven destabilization of the protein structure, as observed for the psychrophilic α-amylase [249], citrate synthase [186] salmon trypsin [206] and Ph Lig [239]. The study of NAD^+-dependent DNA ligases (NDLs) also connected the decreased arginine count in psychrophiles to a significant increase of exposed hydrophobic residues to solvent, which disrupt surface hydrogen binding networks and alter the protein water surface interactions, in contrast to a more hydrophilic and charged surface area in thermophiles [250].

The lower arginine count and increased exposure to solvent of hydrophobic residues on the Vib-Lig surface, combined with the unique local distribution of hydrophobic surface patches may thus contribute to local flexibility and cold adapted behavior of Vib-Lig.

The uneven distribution of the substitutions in Vib-Lig between the two domains raises interesting questions regarding the purpose. A possible explanation may be that it tunes the dynamics of the discrete domains in a way that optimizes substrate binding and product release, rather than affecting the catalysis directly. Computer simulations and Arrhenius plots of other enzymes have also indicated that local surface flexibility outside the catalytic region affects the enthalpy/entropy balance and the overall enzyme dynamics [198,251].

Although structural adjustments of residues in the active site and catalytic regions have been reported for many psychrophilic enzymes [183,184] the number and nature of residues involved in substrate binding, flexible linker region and enzymatic activity are conserved among the homologous Lig Es adapted to different temperatures (Fig 3, supplementary figure S1a and b, paper III). This has been observed with other extremophiles and supports the rationale discussed above that residues in other parts of the enzyme influence flexibility indirectly.

4.6 Biological roles and applications

4.6.1 Fur and virulence

The vital role of Fur in host/parasite interactions through their response to environmental signals during infections is now well documented. For successful host infection, the bacteria are dependent on access to free iron, while iron scavenging is part of the host organism's first line of defense, known as the battle of iron between hosts and pathogens. As a consequence, the bacteria express genes involved in iron-uptake, mainly controlled by Fur. In addition to iron homeostasis, Fur has been implicated in the regulation of virulence factors, motility, oxidative stress, quorum sensing and bacterial fitness, all being important for host-pathogen interactions in many bacteria [81,87,104,252-255]. The importance of Fur for pathogenesis and virulence has been described in a variety of bacterial species by infection studies [254,256-259], also in fish models [70,260].

Although the understanding of different virulence mechanisms in the pathogenesis of *A. salmonicida* is limited, studies with bacteria grown at different temperatures and iron restricted conditions have pointed to the involvement of one or more iron acquisition systems [261]. Furthermore, the importance of Fur in *A. salmonicida* iron homeostasis was demonstrated by a biological study of a *fur* null mutant strain that showed reduced growth rates, loss of fitness, oxidative stress and reduced response to low iron conditions, compared to the wild-type strain. Other genes affected indirectly are involved in chemotaxis, motility and heat shock [235].

Genes controlled by AsFur are expected to be important for disease development in Atlantic salmon and cod, however the *in vivo* role of Fur in *A. salmonicida* pathogenesis has not been experimentally verified, thus it would be interesting to test knock out mutants targeting *fur* or interesting genes controlled by AsFur, for pathogenesis in fish models.

Does AsFur have a potential as a drug target?

Bacterial resistance to current antibiotics is rapidly increasing and there is an urgent demand for new strategies to fight pathogens. An approach targeting iron regulation in pathogens could be attractive for the industry. The fact that Fur homologues are absent in eukaryotes makes Fur interesting in the design of novel antimicrobial agents,

and one approach is to target the balance between iron deficiency and overload in pathogens by anti-Fur peptide inhibition.

Various *in silico*, *in vitro* and *in vivo* approaches have been utilized to understand the mechanism of inhibition behind interactions of peptides with Fur [262-264]. Michaud-Soret *et al* describes two novel small peptide inhibitors, of which one interferes with the dimerization and the other peptide-Fur complexes inhibits in the DNA-binding groove between the two Fur subunits. The optimized anti-Fur peptides show promising results and are able to decrease pathogenic *E. coli* strain virulence in a fly infection model [265].

Since the importance of iron uptake mechanisms in AsFur is established [235], uncovering the molecular basis behind DNA recognition by Fur is crucial. The deeper understanding of AsFur-DNA interactions obtained throughout this study has therefore revealed potential inhibitor binding sites in the DNA-binding domain by peptide aptamers. The residues demonstrated to be involved in AsFur-DNA interactions are mostly conserved among Fur homologues. By the plasmid protection assay, four peptide aptamers targeting EcFur (kindly donated by Dr. Isabelle Michaud-Soret) were not able to inhibit binding of the *E. coli* fur box by AsFur (data not shown). These were specifically targeting the conserved resides Tyr56 and Arg57, shown to be crucial for AsFur activity, and up to seven other residues, where only one diverges from AsFur in sequence comparisons. The lack of any observable inhibitory effect on DNA-binding could indicate a slightly different binding mode by AsFur, however as only one experiment was conducted, we cannot strictly conclude whether the peptide inhibitors have effect. It would be interesting to evaluate their ability to interfere with AsFur binding to a plasmid with a Vibrio fur box incorporated. In the future, potential novel small molecule inhibitors that specifically target AsFur can be identified by chemical library screening and *in silico* design.

In summary, new insight into the AsFur-DNA interaction opens potential for using AsFur-aptamers as antibacterial drugs in cold water vibriosis.

4.6.2 Enzymes and bioprospecting – discovering novel and valuable biological resources

Bioprospecting focuses on discovery and commercialization of new products based on biological resources. Of special value is cold-adapted enzymes possessing high activity at low temperatures and thermolability. Such enzymes from Arctic marine microorganisms have been of particular interest in our research group for the purpose of acquiring new enzymatic potency and effectiveness.

The reported Vib-Lig and MG Orn show potential to provide novel cold-active enzymes with unique properties to molecular biology research and biotechnological industry.

Applications of cold active Orn

MG Orn is an RNase that degrades small RNA to mononucleotides in order to avoid accumulation of oligoribonucleotides in cells and promote continued synthesis of new RNA transcripts. In general, microbial RNases have shown promising potential for numerous biotechnological, pharmaceutical and industrial purposes, such as removal of RNA from single-cell protein, purification of plasmid DNA for DNA vaccines and gene therapy, and preparation of specific oligonucleotides [266].

The unique features of cold active MG Orn offer an advantage over homologues counterparts in molecular biology research applications. Its potential lies in the broad specificity and relatively low inactivation temperature, useful for complete digestion of small RNA molecules. Additionally, the observed low temperature activity allows for elimination of nucleic acids from thermosensitive processes. Since MG Orn act on small RNAs, it could be utilized in combination of other RNAses to assure complete digestion of RNA.

Cold active Lig Es in molecular biology research

The potential advantage of low-temperature-active DNA ligases in general lies in molecular biology and biotechnological applications, such as genetic engineering and DNA sequencing. Vib-Lig, characterized as a genuinely cold-adapted enzyme with improved enzymatic performance in this study, may serve as a minimized model for DNA-ligase interactions.

Currently, commercially available DNA ligases are mainly mesophilic bacteriophage-derived and operate at temperatures levels where residual nucleases may interfere with the activity. The ability of Vib-Lig to act with high specificity at low temperatures therefore offer a major advantage with respect to the mesophilic ligases. Another advantage of Vib-Lig is the heat lability which allows for rapid inactivation by moderate heat treatment. There are currently no cold adapted DNA ligases commercially available.

Our comparative study of the three minimal Lig E-type ADLs originating from obligate psychrophilic bacteria and a mesophilic homologue revealed new insight into structural and biophysical determinants for low-temperature activity of DNA ligases. The identification of specific structural cold adapted features of Vib-Lig provides potential for the discovery and design of commercially optimized cold-active ligases in biotechnology, for example by random mutagenesis and structure-guided engineering. High-throughput sequencing, where nucleotides are sequenced by hybridization and ligation of base-pair mismatches, could also potentially benefit from cold active ligases.

4.7 Conclusions

This study provides new insight into molecular mechanisms and structure of three nucleic acid binding proteins with different functions in the bacterial cell, originating from marine and low-temperature habitats.

Following recombinant production, the activity and stability of AsFur, MG Orn and Lig E were characterized by a combination of biochemical, biophysical and structural techniques.

The transcription factor As-Fur, which is essential for iron homeostasis in *V. salmonicida*, bound strongly the previously predicted Vibrio Fur box consensus sequence in EMSA assays. Assays with mutated Vibrio Fur box consensus sequences verified the importance of the nucleotides T12, T13, A14 and C16 for productive binding by AsFur, as well as AT-rich regions. Additionally, as a global regulator able to recognize degenerate sequences, AsFur uses additional shape-readout modes for unspecific DNA recognition. Combined with structural homology modelling, conserved amino acid residues involved in the interaction were confirmed. For future studies of the importance of specific residues in AsFur-DNA interaction, we have identified interesting targets for site-directed mutagenesis studies. Furthermore, resolving the molecular mechanisms behind AsFur regulation presents its potential as a target in the development of small molecule inhibitors as novel antibacterial drugs for *A. salmonicida* pathogenesis.

MG Orn was a characterized as a metal-dependent 3' - 5' exonuclease that act on short RNA oligonucleotide substrates up to 10mer of length. Further, the three-dimensional structure of MG Orn determined to 3.5 Å revealed a typical homodimer formation connected by an intermolecular disulfide bond as well as other interactions. The importance of this homodimer for catalytic function was demonstrated by mutation of the cysteines involved. Structural modelling of MG Orn with ExoI in complex with ssDNA showed the structural basis behind the ability of MG Orn to accommodate and hydrolyze RNA substrates longer than a 2mer.

The Arctic marine derived Lig Es Psy-Lig, Vib-Lig and Par-Lig showed variable activity optimums and thermal stabilities, confirming Vib-Lig as a ligase with genuine cold-adapted features. Amino acid sequence comparisons and homology modelling of the three Lig Es suggested that the main feature behind cold adaptation is local surface-exposed patches with greater hydrophobicity, due to a decreased number of arginines and a general elevated number of hydrophobic residues in non-DNA-binding surface

exposed regions, compared to the mesophilic homologue. This information could be used to understand the biochemical basis for low temperature activity in other DNA ligases, and eventually used to optimize enzymes for biotechnological purposes.

Works cited

1. Scofield, M. (2007). *xPharm: The Comprehensive Pharmacology Reference*. Elsevier.
2. Watson, J.D. and Crick, F.H. (1953). *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid.* Nature, **171**(4356), 737-738.
3. Berg, J.M.e.a. (2002). *Biochemistry*. Ed 5th ed. WH Freeman.
4. Guttman, B. (2013). *DNA structure.* Brenner's Encyclopedia of Genetics. Academic Press.
5. Rutherford, K.e.a. (2013). *DNA sequence recognition.* . Encyclopedia of Respiratory Medicine. Academic Press.
6. Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B. and Mann, R.S. (2010). *Origins of specificity in protein-DNA recognition.* Annual review of biochemistry, **79**, 233-269.
7. Stormo, G.D. (2013). *Introduction to Protein-DNA Interactions: Structure, Thermodynamics, and Bioinformatics.* Cold Spring Harbor Laboratory Press.
8. Ehrenhofer-Murray, A. *DNA. Structure and function*. 2006, Encyclopedia of Respiratory Medicine.: Academic Press, Oxford. p. 37-44.
9. Jorgensen, N.O., Kroer, N. and Coffin, R.B. (1994). *Utilization of dissolved nitrogen by heterotrophic bacterioplankton: effect of substrate c/n ratio.* Appl Environ Microbiol, **60**(11), 4124-4133.
10. Pinchuk, G.E., Ammons, C., Culley, D.E., Li, S.M., McLean, J.S., Romine, M.F., Nealson, K.H., Fredrickson, J.K. and Beliaev, A.S. (2008). *Utilization of DNA as a sole source of phosphorus, carbon, and energy by Shewanella spp.: ecological and physiological implications for dissimilatory metal reduction.* Appl Environ Microbiol, **74**(4), 1198-1208.
11. Finkel, S.E. and Kolter, R. (2001). *DNA as a nutrient: novel role for bacterial competence gene homologs.* J Bacteriol, **183**(21), 6288-6293.
12. Driffield, K., Miller, K., Bostock, J.M., O'Neill, A.J. and Chopra, I. (2008). *Increased mutability of Pseudomonas aeruginosa in biofilms.* J Antimicrob Chemother, **61**(5), 1053-1056.
13. Coulocheri, S.A., Pigis, D.G., Papavassiliou, K.A. and Papavassiliou, A.G. (2007). *Hydrogen bonds in protein-DNA complexes: where geometry meets plasticity.* Biochimie, **89**(11), 1291-1303.
14. Zhang, J.-R., Marinus, M.G. and Deng, H. *Methylation and other Modifications of Nucleic Acids and Proteins*. 2015.
15. Anderson, W.F., Ohlendorf, D.H., Takeda, Y. and Matthews, B.W. (1981). *Structure of the cro repressor from bacteriophage lambda and its interaction with DNA.* Nature, **290**(5809), 754-758.

16. McKay, D.B. and Steitz, T.A. (1981). *Structure of catabolite gene activator protein at 2.9 A resolution suggests binding to left-handed B-DNA.* Nature, **290**(5809), 744-749.
17. Pabo, C.O. and Lewis, M. (1982). *The operator-binding domain of lambda repressor: structure and DNA recognition.* Nature, **298**(5873), 443-447.
18. Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M. and Iyer, L.M. (2005). *The many faces of the helix-turn-helix domain: transcription regulation and beyond.* FEMS Microbiol Rev, **29**(2), 231-262.
19. Kolkhof, P., Teichmann, D., Kisters-Woike, B., von Wilcken-Bergmann, B. and Müller-Hill, B. (1992). *Lac repressor with the helix-turn-helix motif of lambda cro binds to lac operator.* The EMBO journal, **11**(8), 3031-3038.
20. Matthews, B.W., Ohlendorf, D.H., Anderson, W.F. and Takeda, Y. (1982). *Structure of the DNA-binding region of lac repressor inferred from its homology with cro repressor.* Proceedings of the National Academy of Sciences of the United States of America, **79**(5), 1428-1432.
21. Luscombe, N.M., Austin, S.E., Berman, H.M. and Thornton, J.M. (2000). *An overview of the structures of protein-DNA complexes.* Genome biology, **1**(1), Reviews001.
22. Laity, J.H., Lee, B.M. and Wright, P.E. (2001). *Zinc finger proteins: new insights into structural and functional diversity.* Curr Opin Struct Biol, **11**(1), 39-46.
23. Lee, J.Y., Chang, C., Song, H.K., Moon, J., Yang, J.K., Kim, H.K., Kwon, S.T. and Suh, S.W. (2000). *Crystal structure of NAD(+)-dependent DNA ligase: modular architecture and functional implications.* The EMBO journal, **19**(5), 1119-1129.
24. Tateno, M., Yamasaki, K., Amano, N., Kakinuma, J., Koike, H., Allen, M.D. and Suzuki, M. (1997). *DNA recognition by beta-sheets.* Biopolymers, **44**(4), 335-359.
25. Alberts, B., Johnson, A. and Lewis, J. (2002). *Molecular Biology of the cell.* 4th ed. Garland Science.
26. Murzin, A.G. (1993). *OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences.* EMBO J, **12**(3), 861-867.
27. Theobald, D.L., Mitton-Fry, R.M. and Wuttke, D.S. (2003). *Nucleic acid recognition by OB-fold proteins.* Annual review of biophysics and biomolecular structure, **32**, 115-133.
28. Garratt, R.C., Valadares, N.F. and Bachega, J.F.R. *Oligomeric Proteins*, in *Encyclopedia of Biophysics*, G.C.K. Roberts, Editor. 2013, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 1781-1789.
29. Traut, T.W. (1994). *Dissociation of enzyme oligomers: a mechanism for allosteric regulation.* Crit Rev Biochem Mol Biol, **29**(2), 125-163.
30. Jaenicke, R. (1999). *Stability and folding of domain proteins.* Prog Biophys Mol Biol, **71**(2), 155-241.

31. Torshin, I.Y. (2002). *Functional maps of the junctions between interglobular contacts and active sites in glycolytic enzymes -- a comparative analysis of the biochemical and structural data.* Med Sci Monit, **8**(4), BR123-135.
32. Jones, S. and Thornton, J.M. (1996). *Principles of protein-protein interactions.* Proc Natl Acad Sci U S A, **93**(1), 13-20.
33. Jones, S. and Thornton, J.M. (1995). *Protein-protein interactions: a review of protein dimer structures.* Prog Biophys Mol Biol, **63**(1), 31-65.
34. Yang, W. (2011). *Nucleases: diversity of structure, function and mechanism.* Quarterly reviews of biophysics, **44**(1), 1-93.
35. Cowan, J.A. (2002). *Structural and catalytic chemistry of magnesium-dependent enzymes.* Biometals, **15**(3), 225-235.
36. Kane, C. (2001). *Encyclopedia of Genetics*. Academic Press.
37. Cooper, G. (2000). *The Cell: A Molecular Approach*. 2nd edition ed. Sinauer Associates.
38. Gilbert, W. and Müller-Hill, B. (1966). *Isolation of the Lac repressor.* Proceedings of the National Academy of Sciences, **56**(6), 1891.
39. Slattery, M., Zhou, T., Yang, L., Dantas Machado, A.C., Gordan, R. and Rohs, R. (2014). *Absence of a simple code: how transcription factors read the genome.* Trends Biochem Sci, **39**(9), 381-399.
40. Georges, A.B., Benayoun, B.A., Caburet, S. and Veitia, R.A. (2010). *Generic binding sites, generic DNA-binding domains: where does specific promoter recognition come from?* FASEB J, **24**(2), 346-356.
41. Browning, D.F. and Busby, S.J. (2004). *The regulation of bacterial transcription initiation.* Nat Rev Microbiol, **2**(1), 57-65.
42. Stojiljkovic, I., Baumler, A.J. and Hantke, K. (1994). *Fur regulon in gram-negative bacteria. Identification and characterization of new iron-regulated Escherichia coli genes by a fur titration assay.* J Mol Biol, **236**(2), 531-545.
43. Tsolis, R.M., Baumler, A.J., Stojiljkovic, I. and Heffron, F. (1995). *Fur regulon of Salmonella typhimurium: identification of new iron-regulated genes.* J Bacteriol, **177**(16), 4628-4637.
44. Ochsner, U.A. and Vasil, M.L. (1996). *Gene repression by the ferric uptake regulator in Pseudomonas aeruginosa: cycle selection of iron-regulated genes.* Proc Natl Acad Sci U S A, **93**(9), 4409-4414.
45. Niyogi, S.K. and Datta, A.K. (1975). *A novel oligoribonuclease of Escherichia coli. I. Isolation and properties.* J Biol Chem, **250**(18), 7307-7312.
46. Ghosh, S. and Deutscher, M.P. (1999). *Oligoribonuclease is an essential component of the mRNA decay pathway.* Proc Natl Acad Sci U S A, **96**(8), 4372-4377.
47. Lehman, I.R. (1974). *DNA ligase: structure, mechanism, and function.* Science, **186**(4166), 790-797.
48. Magnet, S. and Blanchard, J.S. (2004). *Mechanistic and kinetic study of the ATP-dependent DNA ligase of Neisseria meningitidis.* Biochemistry, **43**(3), 710-717.

49. Williamson, A. and Pedersen, H. (2014). *Recombinant expression and purification of an ATP-dependent DNA ligase from Aliivibrio salmonicida.* Protein Expr Purif, **97**, 29-36.
50. Williamson, A., Hjerde, E. and Kahlke, T. (2016). *Analysis of the distribution and evolution of the ATP-dependent DNA ligases of bacteria delineates a distinct phylogenetic group 'Lig E'.* Mol Microbiol, **99**(2), 274-290.
51. de Lorenzo, V., Wee, S., Herrero, M. and Neilands, J.B. (1987). *Operator sequences of the aerobactin operon of plasmid ColV-K30 binding the ferric uptake regulation (fur) repressor.* J Bacteriol, **169**(6), 2624-2630.
52. Tardat, B. and Touati, D. (1993). *Iron and oxygen regulation of Escherichia coli MnSOD expression: competition between the global regulators Fur and ArcA for binding to DNA.* Mol Microbiol, **9**(1), 53-63.
53. Escolar, L., Perez-Martin, J. and de Lorenzo, V. (2000). *Evidence of an unusually long operator for the fur repressor in the aerobactin promoter of Escherichia coli.* J Biol Chem, **275**(32), 24709-24714.
54. Nandal, A., Huggins, C.C., Woodhall, M.R., McHugh, J., Rodriguez-Quinones, F., Quail, M.A., Guest, J.R. and Andrews, S.C. (2010). *Induction of the ferritin gene (ftnA) of Escherichia coli by Fe(2+)-Fur is mediated by reversal of H-NS silencing and is RyhB independent.* Mol Microbiol, **75**(3), 637-657.
55. Teixido, L., Carrasco, B., Alonso, J.C., Barbe, J. and Campoy, S. (2011). *Fur activates the expression of Salmonella enterica pathogenicity island 1 by directly interacting with the hilD operator in vivo and in vitro.* PLoS One, **6**(5), e19711.
56. Deng, Z. *et al.* (2015). *Mechanistic insights into metal ion activation and operator recognition by the ferric uptake regulator.* Nat Commun, **6**, 7642.
57. Lovett, S.T. (2011). *The DNA Exonucleases of Escherichia coli.* EcoSal Plus, **4**(2), 10.1128/ecosalplus.1124.1124.1127.
58. Williamson, A., Grgic, M. and Leiros, H.S. (2018). *DNA binding with a minimal scaffold: structure-function analysis of Lig E DNA ligases.* Nucleic Acids Res, **46**(16), 8616-8629.
59. Lavrrar, J.L. and McIntosh, M.A. (2003). *Architecture of a fur binding site: a comparative analysis.* J Bacteriol, **185**(7), 2194-2202.
60. Agriesti, F., Roncarati, D., Musiani, F., Del Campo, C., Iurlaro, M., Sparla, F., Ciurli, S., Danielli, A. and Scarlato, V. (2014). *FeON-FeOFF: the Helicobacter pylori Fur regulator commutates iron-responsive transcription by discriminative readout of opposed DNA grooves.* Nucleic Acids Res, **42**(5), 3138-3151.
61. Holmes, K., Mulholland, F., Pearson, B.M., Pin, C., McNicholl-Kennedy, J., Ketley, J.M. and Wells, J.M. (2005). *Campylobacter jejuni gene expression in response to iron limitation and the role of Fur.* Microbiology, **151**(Pt 1), 243-257.

62. Bender, K.S. *et al.* (2007). *Analysis of a ferric uptake regulator (Fur) mutant of Desulfovibrio vulgaris Hildenborough.* Applied and environmental microbiology, **73**(17), 5389-5400.
63. Mechold, U., Ogryzko, V., Ngo, S. and Danchin, A. (2006). *Oligoribonuclease is a common downstream target of lithium-induced pAp accumulation in Escherichia coli and human cells.* Nucleic Acids Res, **34**(8), 2364-2373.
64. Doherty, A.J. and Suh, S.W. (2000). *Structural and mechanistic conservation in DNA ligases.* Nucleic Acids Res, **28**(21), 4051-4058.
65. Shuman, S. and Schwer, B. (1995). *RNA capping enzyme and DNA ligase: a superfamily of covalent nucleotidyl transferases.* Mol Microbiol, **17**(3), 405-410.
66. Carpenter, B.M., Whitmire, J.M. and Merrell, D.S. (2009). *This is not your mother's repressor: the complex role of fur in pathogenesis.* Infect Immun, **77**(7), 2590-2601.
67. Mills, S.A. and Marletta, M.A. (2005). *Metal binding characteristics and role of iron oxidation in the ferric uptake regulator from Escherichia coli.* Biochemistry, **44**(41), 13553-13559.
68. Park, A.Y., Elvin, C.M., Hamdan, S.M., Wood, R.J., Liyou, N.E., Hamwood, T.E., Jennings, P.A. and Dixon, N.E. (2008). *Hydrolysis of the 5 '-p-nitrophenyl ester of TMP by oligoribonucleases (ORN) from Escherichia coli, Mycobacterium smegmatis, and human.* Protein Expression and Purification, **57**(2), 180-187.
69. Pohl, E., Holmes, R.K. and Hol, W.G. (1999). *Crystal structure of a cobalt-activated diphtheria toxin repressor-DNA complex reveals a metal-binding SH3-like domain.* J Mol Biol, **292**(3), 653-667.
70. Ebanks, R.O., Goguen, M., Knickle, L., Dacanay, A., Leslie, A., Ross, N.W. and Pinto, D.M. (2013). *Analysis of a ferric uptake regulator (Fur) knockout mutant in Aeromonas salmonicida subsp. salmonicida.* Vet Microbiol, **162**(2-4), 831-841.
71. Berges, M. *et al.* (2018). *Iron Regulation in Clostridioides difficile.* Front Microbiol, **9**, 3183.
72. Pasqua, M., Visaggio, D., Lo Sciuto, A., Genah, S., Banin, E., Visca, P. and Imperi, F. (2017). *Ferric Uptake Regulator Fur Is Conditionally Essential in Pseudomonas aeruginosa.* J Bacteriol, **199**(22).
73. Arraiano, C.M. *et al.* (2010). *The critical role of RNA processing and degradation in the control of gene expression.* FEMS Microbiol Rev, **34**(5), 883-923.
74. Ohnishi, Y., Nishiyama, Y., Sato, R., Kameyama, S. and Horinouchi, S. (2000). *An oligoribonuclease gene in Streptomyces griseus.* J Bacteriol, **182**(16), 4647-4653.
75. Mechold, U., Fang, G., Ngo, S., Ogryzko, V. and Danchin, A. (2007). *YtqI from Bacillus subtilis has both oligoribonuclease and pAp-phosphatase activity.* Nucleic Acids Res, **35**(13), 4552-4561.

76. Braun, V. and Hantke, K. (2011). *Recent insights into iron import by bacteria.* Curr Opin Chem Biol, **15**(2), 328-334.
77. Huang, W. and Wilks, A. (2017). *Extracellular Heme Uptake and the Challenge of Bacterial Cell Membranes.* Annu Rev Biochem, **86**, 799-823.
78. Khan, A., Singh, P. and Srivastava, A. (2018). *Synthesis, nature and utility of universal iron chelator - Siderophore: A review.* Microbiol Res, **212-213**, 103-111.
79. Imlay, J., Chin, S. and Linn, S. (1988). *Toxic DNA damage by hydrogen peroxide through the Fenton reaction in vivo and in vitro.* Science, **240**(4852), 640-642.
80. Lloyd, R.V., Hanna, P.M. and Mason, R.P. (1997). *The Origin of the Hydroxyl Radical Oxygen in the Fenton Reaction.* Free Radical Biology and Medicine, **22**(5), 885-888.
81. Troxell, B. and Hassan, H.M. (2013). *Transcriptional regulation by Ferric Uptake Regulator (Fur) in pathogenic bacteria.* Front Cell Infect Microbiol, **3**, 59.
82. Butcher, J., Sarvan, S., Brunzelle, J.S., Couture, J.F. and Stintzi, A. (2012). *Structure and regulon of Campylobacter jejuni ferric uptake regulator Fur define apo-Fur regulation.* Proc Natl Acad Sci U S A, **109**(25), 10047-10052.
83. Palyada, K., Threadgill, D. and Stintzi, A. (2004). *Iron acquisition and regulation in Campylobacter jejuni.* J Bacteriol, **186**(14), 4714-4729.
84. Neumann, W., Gulati, A. and Nolan, E.M. (2017). *Metal homeostasis in infectious disease: recent advances in bacterial metallophores and the human metal-withholding response.* Curr Opin Chem Biol, **37**, 10-18.
85. Okebe, J.U., Yahav, D., Shbita, R. and Paul, M. (2011). *Oral iron supplements for children in malaria-endemic areas.* Cochrane Database Syst Rev, (10), CD006589.
86. Cornelis, P., Matthijs, S. and Van Oeffelen, L. (2009). *Iron uptake regulation in Pseudomonas aeruginosa.* Biometals, **22**(1), 15-22.
87. Vasil, M.L. and Ochsner, U.A. (1999). *The response of Pseudomonas aeruginosa to iron: genetics, biochemistry and virulence.* Mol Microbiol, **34**(3), 399-413.
88. Balasubramanian, D., Schneper, L., Kumari, H. and Mathee, K. (2013). *A dynamic and intricate regulatory network determines Pseudomonas aeruginosa virulence.* Nucleic Acids Res, **41**(1), 1-20.
89. Gao, H., Zhou, D., Li, Y., Guo, Z., Han, Y., Song, Y., Zhai, J., Du, Z., Wang, X., Lu, J. and Yang, R. (2008). *The iron-responsive Fur regulon in Yersinia pestis.* J Bacteriol, **190**(8), 3063-3075.
90. Zhou, D. *et al.* (2006). *Global analysis of iron assimilation and fur regulation in Yersinia pestis.* FEMS Microbiol Lett, **258**(1), 9-17.
91. Olakanmi, O., Gunn, J.S., Su, S., Soni, S., Hassett, D.J. and Britigan, B.E. (2010). *Gallium disrupts iron uptake by intracellular and extracellular Francisella strains and exhibits therapeutic efficacy in a murine pulmonary infection model.* Antimicrob Agents Chemother, **54**(1), 244-253.

92. Huja, S., Oren, Y., Biran, D., Meyer, S., Dobrindt, U., Bernhard, J., Becher, D., Hecker, M., Sorek, R. and Ron, E.Z. (2014). *Fur is the master regulator of the extraintestinal pathogenic Escherichia coli response to serum.* MBio, **5**(4).
93. Horsburgh, M.J., Ingham, E. and Foster, S.J. (2001). *In Staphylococcus aureus, fur is an interactive regulator with PerR, contributes to virulence, and Is necessary for oxidative stress resistance through positive regulation of catalase and iron homeostasis.* J Bacteriol, **183**(2), 468-475.
94. Pich, O.Q. and Merrell, D.S. (2013). *The ferric uptake regulator of Helicobacter pylori: a critical player in the battle for iron and colonization of the stomach.* Future Microbiol, **8**(6), 725-738.
95. Rea, R.B., Gahan, C.G. and Hill, C. (2004). *Disruption of putative regulatory loci in Listeria monocytogenes demonstrates a significant role for Fur and PerR in virulence.* Infect Immun, **72**(2), 717-727.
96. Mey, A.R., Wyckoff, E.E., Kanukurthy, V., Fisher, C.R. and Payne, S.M. (2005). *Iron and fur regulation in Vibrio cholerae and the role of fur in virulence.* Infect Immun, **73**(12), 8167-8178.
97. Yu, C. and Genco, C.A. (2012). *Fur-mediated global regulatory circuits in pathogenic Neisseria species.* J Bacteriol, **194**(23), 6372-6381.
98. Porcheron, G., Habib, R., Houle, S., Caza, M., Lepine, F., Daigle, F., Masse, E. and Dozois, C.M. (2014). *The small RNA RyhB contributes to siderophore production and virulence of uropathogenic Escherichia coli.* Infect Immun, **82**(12), 5056-5068.
99. Zhu, C., Ngeleka, M., Potter, A.A. and Allan, B.J. (2002). *Effect of fur mutation on acid-tolerance response and in vivo virulence of avian septicemic Escherichia coli.* Can J Microbiol, **48**(5), 458-462.
100. Grifantini, R., Sebastian, S., Frigimelica, E., Draghi, M., Bartolini, E., Muzzi, A., Rappuoli, R., Grandi, G. and Genco, C.A. (2003). *Identification of iron-activated and -repressed Fur-dependent genes by transcriptome analysis of Neisseria meningitidis group B.* Proc Natl Acad Sci U S A, **100**(16), 9542-9547.
101. Delany, I., Spohn, G., Rappuoli, R. and Scarlato, V. (2001). *The Fur repressor controls transcription of iron-activated and -repressed genes in Helicobacter pylori.* Mol Microbiol, **42**(5), 1297-1309.
102. Bagg, A. and Neilands, J.B. (1987). *Ferric uptake regulation protein acts as a repressor, employing iron (II) as a cofactor to bind the operator of an iron transport operon in Escherichia coli.* Biochemistry, **26**(17), 5471-5477.
103. Ochsner, U.A., Vasil, A.I. and Vasil, M.L. (1995). *Role of the ferric uptake regulator of Pseudomonas aeruginosa in the regulation of siderophores and exotoxin A expression: purification and activity on iron-regulated promoters.* J Bacteriol, **177**(24), 7194-7201.
104. Pohl, E., Haller, J.C., Mijovilovich, A., Meyer-Klaucke, W., Garman, E. and Vasil, M.L. (2003). *Architecture of a protein central to iron homeostasis: crystal structure and spectroscopic analysis of the ferric uptake regulator.* Mol Microbiol, **47**(4), 903-915.

105. Dian, C., Vitale, S., Leonard, G.A., Bahlawane, C., Fauquant, C., Leduc, D., Muller, C., de Reuse, H., Michaud-Soret, I. and Terradot, L. (2011). *The structure of the Helicobacter pylori ferric uptake regulator Fur reveals three functional metal binding sites.* Mol Microbiol, **79**(5), 1260-1275.
106. Sarvan, S., Charih, F., Askoura, M., Butcher, J., Brunzelle, J.S., Stintzi, A. and Couture, J.F. (2018). *Functional insights into the interplay between DNA interaction and metal coordination in ferric uptake regulators.* Sci Rep, **8**(1), 7140.
107. Sheikh, M.A. and Taylor, G.L. (2009). *Crystal structure of the Vibrio cholerae ferric uptake regulator (Fur) reveals insights into metal co-ordination.* Mol Microbiol, **72**(5), 1208-1220.
108. Perard, J. *et al.* (2018). *Structural and functional studies of the metalloregulator Fur identify a promoter-binding mechanism and its role in Francisella tularensis virulence.* Commun Biol, **1**, 93.
109. Pecqueur, L., D'Autreaux, B., Dupuy, J., Nicolet, Y., Jacquamet, L., Brutscher, B., Michaud-Soret, I. and Bersch, B. (2006). *Structural changes of Escherichia coli ferric uptake regulator during metal-dependent dimerization and activation explored by NMR and X-ray crystallography.* J Biol Chem, **281**(30), 21286-21295.
110. Andrews, S.C., Robinson, A.K. and Rodriguez-Quinones, F. (2003). *Bacterial iron homeostasis.* FEMS Microbiol Rev, **27**(2-3), 215-237.
111. Escolar, L., Perez-Martin, J. and de Lorenzo, V. (1999). *Opening the iron box: transcriptional metalloregulation by the Fur protein.* J Bacteriol, **181**(20), 6223-6229.
112. Baichoo, N., Wang, T., Ye, R. and Helmann, J.D. (2002). *Global analysis of the Bacillus subtilis Fur regulon and the iron starvation stimulon.* Mol Microbiol, **45**(6), 1613-1629.
113. Pich, O.Q., Carpenter, B.M., Gilbreath, J.J. and Merrell, D.S. (2012). *Detailed analysis of Helicobacter pylori Fur-regulated promoters reveals a Fur box core sequence and novel Fur-regulated genes.* Mol Microbiol, **84**(5), 921-941.
114. Escolar, L., Perez-Martin, J. and de Lorenzo, V. (1998). *Binding of the fur (ferric uptake regulator) repressor of Escherichia coli to arrays of the GATAAT sequence.* J Mol Biol, **283**(3), 537-547.
115. Lavrrar, J.L., Christoffersen, C.A. and McIntosh, M.A. (2002). *Fur-DNA interactions at the bidirectional fepDGC-entS promoter region in Escherichia coli.* J Mol Biol, **322**(5), 983-995.
116. Baichoo, N. and Helmann, J.D. (2002). *Recognition of DNA by Fur: a reinterpretation of the Fur box consensus sequence.* J Bacteriol, **184**(21), 5826-5832.
117. White, A., Ding, X., vanderSpek, J.C., Murphy, J.R. and Ringe, D. (1998). *Structure of the metal-ion-activated diphtheria toxin repressor/tox operator complex.* Nature, **394**(6692), 502-506.

118. Thompson, D.K. *et al.* (2002). *Transcriptional and proteomic analysis of a ferric uptake regulator (fur) mutant of Shewanella oneidensis: possible involvement of fur in energy metabolism, transcriptional regulation, and oxidative stress.* Appl Environ Microbiol, **68**(2), 881-892.
119. Sebastian, S., Agarwal, S., Murphy, J.R. and Genco, C.A. (2002). *The gonococcal fur regulon: identification of additional genes involved in major catabolic, recombination, and secretory pathways.* J Bacteriol, **184**(14), 3965-3974.
120. Egidius, E., Wiik, R., Andersen, K., Hoff, K.A. and Hjeltnes, B. (1986). *Vibrio salmonicida sp. nov., a New Fish Pathogen.* International Journal of Systematic and Evolutionary Microbiology, **36**(4), 518-520.
121. Pedersen, H.L., Ahmad, R., Riise, E.K., Leiros, H.K., Hauglid, S., Espelid, S., Brandsdal, B.O., Leiros, I., Willassen, N.P. and Haugen, P. (2010). *Experimental and computational characterization of the ferric uptake regulator from Aliivibrio salmonicida (Vibrio salmonicida).* J Microbiol, **48**(2), 174-183.
122. Ahmad, R., Brandsdal, B.O., Michaud-Soret, I. and Willassen, N.P. (2009). *Ferric uptake regulator protein: binding free energy calculations and per-residue free energy decomposition.* Proteins, **75**(2), 373-386.
123. Ahmad, R., Hjerde, E., Hansen, G.A., Haugen, P. and Willassen, N.P. (2009). *Prediction and experimental testing of ferric uptake regulator regulons in vibrios.* J Mol Microbiol Biotechnol, **16**(3-4), 159-168.
124. Laalami, S. and Putzer, H. (2011). *mRNA degradation and maturation in prokaryotes: the global players.* Biomol Concepts, **2**(6), 491-506.
125. Mohanty, B.K. and Kushner, S.R. (2016). *Regulation of mRNA Decay in Bacteria.* Annual Review of Microbiology, **70**(1), 25-44.
126. Kushner, S.R. (2004). *mRNA decay in prokaryotes and eukaryotes: different approaches to a similar problem.* IUBMB Life, **56**(10), 585-594.
127. Zuo, Y. and Deutscher, M.P. (2001). *Exoribonuclease superfamilies: structural analysis and phylogenetic distribution.* Nucleic Acids Res, **29**(5), 1017-1026.
128. Apirion, D. (1973). *Degradation of RNA in Escherichia coli.* Molecular and General Genetics MGG, **122**(4), 313-322.
129. Lehnik-Habrink, M., Lewis, R.J., Mader, U. and Stulke, J. (2012). *RNA degradation in Bacillus subtilis: an interplay of essential endo- and exoribonucleases.* Mol Microbiol, **84**(6), 1005-1017.
130. Celesnik, H., Deana, A. and Belasco, J.G. (2007). *Initiation of RNA decay in Escherichia coli by 5' pyrophosphate removal.* Molecular cell, **27**(1), 79-90.
131. Condon, C. (2007). *Maturation and degradation of RNA in bacteria.* Curr Opin Microbiol, **10**(3), 271-278.
132. Fang, M., Zeisberg, W.M., Condon, C., Ogryzko, V., Danchin, A. and Mechold, U. (2009). *Degradation of nanoRNA is performed by multiple redundant RNases in Bacillus subtilis.* Nucleic Acids Res, **37**(15), 5114-5125.
133. Liao, H., Liu, M. and Guo, X. (2018). *The special existences: nanoRNA and nanoRNase.* Microbiol Res, **207**, 134-139.

134. Hanekamp, T. and Thorsness, P.E. (1999). *YNT20, a bypass suppressor of yme1 yme2, encodes a putative 3'-5' exonuclease localized in mitochondria of Saccharomyces cerevisiae.* Current Genetics, **34**(6), 438-448.
135. Cohen, D., Mechold, U., Nevenzal, H., Yarmiyhu, Y., Randall, T.E., Bay, D.C., Rich, J.D., Parsek, M.R., Kaever, V., Harrison, J.J. and Banin, E. (2015). *Oligoribonuclease is a central feature of cyclic diguanylate signaling in Pseudomonas aeruginosa.* Proc Natl Acad Sci U S A, **112**(36), 11359-11364.
136. Orr, M.W., Donaldson, G.P., Severin, G.B., Wang, J., Sintim, H.O., Waters, C.M. and Lee, V.T. (2015). *Oligoribonuclease is the primary degradative enzyme for pGpG in Pseudomonas aeruginosa that is required for cyclic-di-GMP turnover.* Proc Natl Acad Sci U S A, **112**(36), E5048-5057.
137. Romling, U., Galperin, M.Y. and Gomelsky, M. (2013). *Cyclic di-GMP: the first 25 years of a universal bacterial second messenger.* Microbiol Mol Biol Rev, **77**(1), 1-52.
138. Chen, G., Zhao, Q., Zhu, F., Chen, R., Jin, Y., Liu, C., Pan, X., Jin, S., Wu, W. and Cheng, Z. (2016). *Oligoribonuclease is required for the type III secretion system and pathogenesis of Pseudomonas aeruginosa.* Microbiol Res, **188-189**, 90-96.
139. Diaz, M.H. and Hauser, A.R. (2010). *Pseudomonas aeruginosa cytotoxin ExoU is injected into phagocytic cells during acute pneumonia.* Infect Immun, **78**(4), 1447-1456.
140. Goldman, S.R., Sharp, J.S., Vvedenskaya, I.O., Livny, J., Dove, S.L. and Nickels, B.E. (2011). *NanoRNAs prime transcription initiation in vivo.* Mol Cell, **42**(6), 817-825.
141. Diaz, M.H., Shaver, C.M., King, J.D., Musunuri, S., Kazzaz, J.A. and Hauser, A.R. (2008). *Pseudomonas aeruginosa induces localized immunosuppression during pneumonia.* Infect Immun, **76**(10), 4414-4421.
142. Datta, A.K. and Niyogi, K. (1975). *A novel oligoribonuclease of Escherichia coli. II. Mechanism of action.* J Biol Chem, **250**(18), 7313-7319.
143. Nguyen, L.H., Erzberger, J.P., Root, J. and Wilson, D.M., 3rd (2000). *The human homolog of Escherichia coli Orn degrades small single-stranded RNA and DNA oligomers.* J Biol Chem, **275**(34), 25900-25906.
144. Steitz, T.A. and Steitz, J.A. (1993). *A general two-metal-ion mechanism for catalytic RNA.* Proc Natl Acad Sci U S A, **90**(14), 6498-6502.
145. Zhang, X., Zhu, L. and Deutscher, M.P. (1998). *Oligoribonuclease is encoded by a highly conserved gene in the 3'-5' exonuclease superfamily.* J Bacteriol, **180**(10), 2779-2781.
146. Chin, K.H., Yang, C.Y., Chou, C.C., Wang, A.H. and Chou, S.H. (2006). *The crystal structure of XC847 from Xanthomonas campestris: a 3'-5' oligoribonuclease of DnaQ fold family with a novel opposingly shifted helix.* Proteins, **65**(4), 1036-1040.
147. Franklin, M.C., Cheung, J., Rudolph, M.J., Burshteyn, F., Cassidy, M., Gary, E., Hillerich, B., Yao, Z.K., Carlier, P.R., Totrov, M. and Love, J.D. (2015). *Structural*

genomics for drug design against the pathogen Coxiella burnetii. Proteins, **83**(12), 2124-2136.
148. Fiedler, T.J., Vincent, H.A., Zuo, Y., Gavrialov, O. and Malhotra, A. (2004). *Purification and crystallization of Escherichia coli oligoribonuclease.* Acta Crystallogr D Biol Crystallogr, **60**(Pt 4), 736-739.
149. Wilkinson, A., Day, J. and Bowater, R. (2001). *Bacterial DNA ligases.* Molecular Microbiology, **40**(6), 1241-1248.
150. Dwivedi, N., Dube, D., Pandey, J., Singh, B., Kukshal, V., Ramachandran, R. and Tripathi, R.P. (2008). *NAD(+)-dependent DNA ligase: a novel target waiting for the right inhibitor.* Med Res Rev, **28**(4), 545-568.
151. Sriskanda, V., Moyer, R.W. and Shuman, S. (2001). *NAD+-dependent DNA ligase encoded by a eukaryotic virus.* J Biol Chem, **276**(39), 36100-36109.
152. Pitcher, R.S., Green, A.J., Brzostek, A., Korycka-Machala, M., Dziadek, J. and Doherty, A.J. (2007). *NHEJ protects Mycobacteria in stationary phase against the harmful effects of desiccation.* DNA Repair, **6**(9), 1271-1276.
153. Cheng, C. and Shuman, S. (1997). *Characterization of an ATP-dependent DNA ligase encoded by Haemophilus influenzae.* Nucleic Acids Res, **25**(7), 1369-1374.
154. Ellenberger, T. and Tomkinson, A.E. *Eukaryotic DNA ligases: Structural and functional insights*, in *Annual Review of Biochemistry*. 2008. p. 313-338.
155. Tomkinson, A.E., Totty, N.F., Ginsburg, M. and Lindahl, T. (1991). *Location of the active site for enzyme-adenylate formation in DNA ligases.* Proc Natl Acad Sci U S A, **88**(2), 400-404.
156. Unciuleac, M.-C., Goldgur, Y. and Shuman, S. (2019). *Structures of ATP-bound DNA ligase D in a closed domain conformation reveal a network of amino acid and metal contacts to the ATP phosphates.* Journal of Biological Chemistry, **294**, jbc.RA119.007445.
157. Unciuleac, M.C., Goldgur, Y. and Shuman, S. (2017). *Two-metal versus one-metal mechanisms of lysine adenylylation by ATP-dependent and NAD(+)-dependent polynucleotide ligases.* Proc Natl Acad Sci U S A, **114**(10), 2592-2597.
158. Williamson, A. and Leiros, H.S. (2019). *Structural intermediates of a DNA-ligase complex illuminate the role of the catalytic metal ion and mechanism of phosphodiester bond formation.* Nucleic Acids Res, **47**(14), 7147-7162.
159. Subramanya, H.S., Doherty, A.J., Ashford, S.R. and Wigley, D.B. (1996). *Crystal structure of an ATP-dependent DNA ligase from bacteriophage T7.* Cell, **85**(4), 607-615.
160. Nishida, H., Kiyonari, S., Ishino, Y. and Morikawa, K. (2006). *The closed structure of an archaeal DNA ligase from Pyrococcus furiosus.* J Mol Biol, **360**(5), 956-967.
161. Pascal, J.M., Tsodikov, O.V., Hura, G.L., Song, W., Cotner, E.A., Classen, S., Tomkinson, A.E., Tainer, J.A. and Ellenberger, T. (2006). *A Flexible Interface*

between DNA Ligase and PCNA Supports Conformational Switching and Efficient Ligation of DNA. Molecular Cell, **24**(2), 279-291.

162. Kim, D.J., Kim, O., Kim, H.W., Kim, H.S., Lee, S.J. and Suh, S.W. (2009). *ATP-dependent DNA ligase from Archaeoglobus fulgidus displays a tightly closed conformation.* Acta Crystallographica Section F-Structural Biology and Crystallization Communications, **65**, 544-550.
163. Supangat, S., An, Y.J., Sun, Y., Kwon, S.T. and Cha, S.S. (2010). *Purification, crystallization and preliminary crystallographic analysis of a multiple cofactor-dependent DNA ligase from Sulfophobococcus zilligii.* Acta Crystallogr Sect F Struct Biol Cryst Commun, **66**(Pt 12), 1583-1585.
164. Petrova, T. *et al.* (2012). *ATP-dependent DNA ligase from Thermococcus sp. 1519 displays a new arrangement of the OB-fold domain.* Acta Crystallogr Sect F Struct Biol Cryst Commun, **68**(Pt 12), 1440-1447.
165. Akey, D., Martins, A., Aniukwu, J., Glickman, M.S., Shuman, S. and Berger, J.M. (2006). *Crystal structure and nonhomologous end-joining function of the ligase component of Mycobacterium DNA ligase D.* Journal Of Biological Chemistry, **281**(19), 13412-13423.
166. Williamson, A., Rothweiler, U. and Schroder Leiros, H.-K. (2014). *Enzyme-adenylate structure of a bacterial ATP-dependent DNA ligase with a minimized DNA-binding surface.* Acta Crystallographica Section D, **70**(11), 3043-3056.
167. Pascal, J.M., O'Brien, P.J., Tomkinson, A.E. and Ellenberger, T. (2004). *Human DNA ligase I completely encircles and partially unwinds nicked DNA.* Nature, **432**(7016), 473-478.
168. Doherty, A.J. and Wigley, D.B. (1999). *Functional domains of an ATP-dependent DNA ligase.* J Mol Biol, **285**(1), 63-71.
169. Tanabe, M., Ishino, Y. and Nishida, H. (2015). *From Structure-Function Analyses to Protein Engineering for Practical Applications of DNA Ligase.* Archaea, **2015**, 267570.
170. Shuman, S. and Lima, C.D. (2004). *The polynucleotide ligase and RNA capping enzyme superfamily of covalent nucleotidyltransferases.* Curr Opin Struct Biol, **14**(6), 757-764.
171. Shuman, S. and Glickman, M.S. (2007). *Bacterial DNA repair by non-homologous end joining.* Nat Rev Microbiol, **5**(11), 852-861.
172. Pitcher, R.S., Brissett, N.C. and Doherty, A.J. (2007). *Nonhomologous end-joining in bacteria: a microbial perspective.* Annu Rev Microbiol, **61**, 259-282.
173. Petersen, T.N., Brunak, S., von Heijne, G. and Nielsen, H. (2011). *SignalP 4.0: discriminating signal peptides from transmembrane regions.* Nature Methods, **8**(10), 785-786.
174. Chambers, C.R. and Patrick, W.M. (2015). *Archaeal Nucleic Acid Ligases and Their Potential in Biotechnology.* Archaea, **2015**, 10.
175. Shuman, S. (2009). *DNA ligases: progress and prospects.* J Biol Chem, **284**(26), 17365-17369.

176. Gómez, F. *Extreme Environment*, in *Encyclopedia of Astrobiology*, M. Gargaud, et al., Editors. 2011, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 570-572.
177. M.T Madigan, B.L.M. (1997). *Extremophiles.* Sci Am, **246**(4), 82-87.
178. Rampelotto, P.H. (2013). *Extremophiles and extreme environments.* Life (Basel, Switzerland), **3**(3), 482-485.
179. J.A Baross, R.Y.M. (1978). *Microbial life at low temperatures: ecological aspects.* Microbial Life in Extreme Environments, 9-71.
180. Morita, R.Y. (1975). *Psychrophilic bacteria.* Bacteriol Rev, **39**(2), 144-167.
181. Atlas, R.e.a. (1997). *Microbial ecology. Fundamentals and applications.* . 4th ed. Benjamin/Cummings Science Publishing.
182. Forster, J. (1887). *ueber einige Eigenschaften leuchtender Bakterien.* Centr. Bakteroil. Parasitenk. 2, 337-340.
183. Feller, G. and Gerday, C. (2003). *Psychrophilic enzymes: hot topics in cold adaptation.* Nat Rev Microbiol, **1**(3), 200-208.
184. Georlette, D., Blaise, V., Collins, T., D'Amico, S., Gratia, E., Hoyoux, A., Marx, J.C., Sonan, G., Feller, G. and Gerday, C. (2004). *Some like it cold: biocatalysis at low temperatures.* FEMS Microbiol Rev, **28**(1), 25-42.
185. Rodrigues, D.F. and Tiedje, J.M. (2008). *Coping with our cold planet.* Appl Environ Microbiol, **74**(6), 1677-1686.
186. Russell, R.J., Gerike, U., Danson, M.J., Hough, D.W. and Taylor, G.L. (1998). *Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium.* Structure, **6**(3), 351-361.
187. Crawford, D.L. and Powers, D.A. (1992). *Evolutionary adaptation to different thermal environments via transcriptional regulation.* Mol Biol Evol, **9**(5), 806-813.
188. Baldwin, J. and Hochachka, P.W. (1970). *Functional significance of isoenzymes in thermal acclimatization. Acetylcholinesterase from trout brain.* Biochem J, **116**(5), 883-887.
189. Somero, G.N. (1995). *Proteins and temperature.* Annu Rev Physiol, **57**, 43-68.
190. Jagdale, G.B. and Gordon, R. (1997). *Effect of temperature on the activities of glucose-6-phosphate dehydrogenase and hexokinase in entomopathogenic nematodes (Nematoda: Steinernematidae).* Comp Biochem Physiol A Physiol, **118**(4), 1151-1156.
191. Smalas, A.O., Leiros, H.K., Os, V. and Willassen, N.P. (2000). *Cold adapted enzymes.* Biotechnol Annu Rev, **6**, 1-57.
192. Feller, G. (2010). *Protein stability and enzyme activity at extreme biological temperatures.* J Phys Condens Matter, **22**(32), 323101.
193. Struvay, C., Negro, S., Matagne, A. and Feller, G. (2013). *Energetics of protein stability at extreme environmental temperatures in bacterial trigger factors.* Biochemistry, **52**(17), 2982-2990.

194. D'Amico, S., Claverie, P., Collins, T., Georlette, D., Gratia, E., Hoyoux, A., Meuwis, M.A., Feller, G. and Gerday, C. (2002). *Molecular basis of cold adaptation.* Philos Trans R Soc Lond B Biol Sci, **357**(1423), 917-925.
195. Bisswanger, H. (2002). *Enzyme Kinetics: Principles and Methods*. WILEY-VCH Verlag GmbH.
196. Siddiqui, K.S. and Cavicchioli, R. (2006). *Cold-adapted enzymes.* Annu Rev Biochem, **75**, 403-433.
197. Santiago, M., Ramirez-Sarmiento, C.A., Zamora, R.A. and Parra, L.P. (2016). *Discovery, Molecular Mechanisms, and Industrial Applications of Cold-Active Enzymes.* Front Microbiol, **7**, 1408.
198. Isaksen, G.V., Aqvist, J. and Brandsdal, B.O. (2016). *Enzyme surface rigidity tunes the temperature dependence of catalytic rates.* Proc Natl Acad Sci U S A, **113**(28), 7822-7827.
199. Feller, G. (2018). *Protein folding at extreme temperatures: Current issues.* Seminars in Cell & Developmental Biology, **84**, 129-137.
200. Chiuri, R., Maiorano, G., Rizzello, A., del Mercato, L.L., Cingolani, R., Rinaldi, R., Maffia, M. and Pompa, P.P. (2009). *Exploring local flexibility/rigidity in psychrophilic and mesophilic carbonic anhydrases.* Biophys J, **96**(4), 1586-1596.
201. D'Amico, S., Marx, J.C., Gerday, C. and Feller, G. (2003). *Activity-stability relationships in extremophilic enzymes.* J Biol Chem, **278**(10), 7891-7896.
202. Bjelic, S., Brandsdal, B.O. and Aqvist, J. (2008). *Cold adaptation of enzyme reaction rates.* Biochemistry, **47**(38), 10049-10057.
203. Davail, S., Feller, G., Narinx, E. and Gerday, C. (1994). *Cold adaptation of proteins. Purification, characterization, and sequence of the heat-labile subtilisin from the antarctic psychrophile Bacillus TA41.* J Biol Chem, **269**(26), 17448-17453.
204. Feller, G. and Gerday, C. (1997). *Psychrophilic enzymes: molecular basis of cold adaptation.* Cell Mol Life Sci, **53**(10), 830-841.
205. Russell, N.J. (2000). *Toward a molecular understanding of cold activity of enzymes from psychrophiles.* Extremophiles, **4**(2), 83-90.
206. Smalås, A.O., Heimstad, E.S., Hordvik, A., Willassen, N.P. and Male, R. (1994). *Cold adaption of enzymes: Structural comparison between salmon and bovine trypsins.* Proteins: Structure, Function, and Bioinformatics, **20**(2), 149-166.
207. Aittaleb, M., Hubner, R., Lamotte-Brasseur, J. and Gerday, C. (1997). *Cold adaptation parameters derived from cDNA sequencing and molecular modelling of elastase from Antarctic fish Notothenia neglecta.* Protein Eng, **10**(5), 475-477.
208. Gianese, G., Bossa, F. and Pascarella, S. (2002). *Comparative structural analysis of psychrophilic and meso- and thermophilic enzymes.* Proteins, **47**(2), 236-249.
209. Papaleo, E., Tiberti, M., Invernizzi, G., Pasi, M. and Ranzani, V. (2011). *Molecular determinants of enzyme cold adaptation: comparative structural*

and computational studies of cold- and warm-adapted enzymes. Curr Protein Pept Sci, **12**(7), 657-683.

210. Arnold, F.H., Wintrode, P.L., Miyazaki, K. and Gershenson, A. (2001). *How enzymes adapt: lessons from directed evolution.* Trends Biochem Sci, **26**(2), 100-106.
211. Cavicchioli, R., Charlton, T., Ertan, H., Mohd Omar, S., Siddiqui, K.S. and Williams, T.J. (2011). *Biotechnological uses of enzymes from psychrophiles.* Microb Biotechnol, **4**(4), 449-460.
212. Gerday, C. *et al.* (2000). *Cold-adapted enzymes: from fundamentals to biotechnology.* Trends Biotechnol, **18**(3), 103-107.
213. Margesin, R. and Feller, G. (2010). *Biotechnological applications of psychrophiles.* Environ Technol, **31**(8-9), 835-844.
214. Chakraborty, S., Minda, R., Salaye, L., Bhattacharjee, S.K. and Rao, B.J. (2011). *Active site detection by spatial conformity and electrostatic analysis--unravelling a proteolytic function in shrimp alkaline phosphatase.* PLoS One, **6**(12), e28470.
215. de Backer, M., McSweeney, S., Rasmussen, H.B., Riise, B.W., Lindley, P. and Hough, E. (2002). *The 1.9 A crystal structure of heat-labile shrimp alkaline phosphatase.* J Mol Biol, **318**(5), 1265-1274.
216. Nilsen, I.W., Overbo, K. and Olsen, R.L. (2001). *Thermolabile alkaline phosphatase from Northern shrimp (Pandalus borealis): protein and cDNA sequence analyses.* Comp Biochem Physiol B Biochem Mol Biol, **129**(4), 853-861.
217. Lanes, O., Leiros, I., Smalas, A.O. and Willassen, N.P. (2002). *Identification, cloning, and expression of uracil-DNA glycosylase from Atlantic cod (Gadus morhua): characterization and homology modeling of the cold-active catalytic domain.* Extremophiles, **6**(1), 73-86.
218. Leiros, I., Moe, E., Lanes, O., Smalas, A.O. and Willassen, N.P. (2003). *The structure of uracil-DNA glycosylase from Atlantic cod (Gadus morhua) reveals cold-adaptation features.* Acta Crystallogr D Biol Crystallogr, **59**(Pt 8), 1357-1365.
219. Longo, M.C., Berninger, M.S. and Hartley, J.L. (1990). *Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions.* Gene, **93**(1), 125-128.
220. Vester, J.K., Glaring, M.A. and Stougaard, P. (2015). *Improved cultivation and metagenomics as new tools for bioprospecting in cold environments.* Extremophiles, **19**(1), 17-29.
221. Al Khudary, R., Stosser, N.I., Qoura, F. and Antranikian, G. (2008). *Pseudoalteromonas arctica sp. nov., an aerobic, psychrotolerant, marine bacterium isolated from Spitzbergen.* Int J Syst Evol Microbiol, **58**(Pt 9), 2018-2024.
222. Sahdev, S., Khattar, S.K. and Saini, K.S. (2008). *Production of active eukaryotic proteins through bacterial expression systems: A review of the existing*

biotechnology strategies. Molecular and Cellular Biochemistry, **307**(1-2), 249-264.
223. Bjerga, G.E.K., Lale, R. and Williamson, A.K. (2016). *Engineering low-temperature expression systems for heterologous production of cold-adapted enzymes.* Bioengineered, **7**(1), 33-38.
224. Loughran, S.T. and Walls, D. (2017). *Tagging Recombinant Proteins to Enhance Solubility and Aid Purification.* Methods Mol Biol, **1485**, 131-156.
225. Korf, U. *et al.* (2005). *Large-scale protein expression for proteome research.* Proteomics, **5**(14), 3571-3580.
226. Niiranen, L., Espelid, S., Karlsen, C.R., Mustonen, M., Paulsen, S.M., Heikinheimo, P. and Willassen, N.P. (2007). *Comparative expression study to increase the solubility of cold adapted Vibrio proteins in Escherichia coli.* Protein Expr Purif, **52**(1), 210-218.
227. Tiss, A., Barre, O., Michaud-Soret, I. and Forest, E. (2005). *Characterization of the DNA-binding site in the ferric uptake regulator protein from Escherichia coli by UV crosslinking and mass spectrometry.* FEBS Letters, **579**(25), 5454-5460.
228. Kim, S.-K., Lormand, J., Weiss, C., Eger, K., Turdiev, H., Turdiev, A., Winkler, W., Sondermann, H. and Lee, V. (2019). *Oligoribonuclease functions as a diribonucleotidase to bypass a key bottleneck in RNA degradation*.
229. Lee, C.W., Park, S.H., Jeong, C.S., Cha, S.S., Park, H. and Lee, J.H. (2019). *Structural basis of small RNA hydrolysis by oligoribonuclease (CpsORN) from Colwellia psychrerythraea strain 34H.* Sci Rep, **9**(1), 2649.
230. Maguire, M.E. and Cowan, J.A. (2002). *Magnesium chemistry and biochemistry.* Biometals, **15**(3), 203-210.
231. Sissi, C. and Palumbo, M. (2009). *Effects of magnesium and related divalent metal ions in topoisomerase structure and function.* Nucleic acids research, **37**(3), 702-711.
232. Hamed, M.Y. (1993). *Binding of the ferric uptake regulation repressor protein (Fur) to Mn(II), Fe(II), Co(II), and Cu(II) ions as co-repressors: electronic absorption, equilibrium, and 57Fe Mossbauer studies.* J Inorg Biochem, **50**(3), 193-210.
233. Zheleznova, E.E., Crosa, J.H. and Brennan, R.G. (2000). *Characterization of the DNA- and Metal-Binding Properties of Vibrio anguillarum Fur Reveals Conservation of a Structural Zn²⁺ Ion.* Journal of Bacteriology, **182**(21), 6264.
234. Hantke, K. (1987). *Selection procedure for deregulated iron transport mutants (fur) in Escherichia coli K 12: fur not only affects iron metabolism.* Mol Gen Genet, **210**(1), 135-139.
235. Thode, S.K., Baekkedal, C., Soderberg, J.J., Hjerde, E., Hansen, H. and Haugen, P. (2017). *Construction of a fur null mutant and RNA-sequencing provide deeper global understanding of the Aliivibrio salmonicida Fur regulon.* PeerJ, **5**, e3461.

236. Hernández, J.A., Meier, J., Barrera, F.N., de los Paños, O.R., Hurtado-Gómez, E., Teresa Bes, M., Fillat, M.F., Luisa Peleato, M., Cavasotto, C.N. and Neira, J.L. (2005). *The Conformational Stability and Thermodynamics of Fur A (Ferric Uptake Regulator) from Anabaena sp. PCC 7119.* Biophysical Journal, **89**(6), 4188-4200.
237. Feller, G. (2003). *Molecular adaptations to cold in psychrophilic enzymes.* Cellular and Molecular Life Sciences, **60**(4), 648-662.
238. Struvay, C. and Feller, G. (2012). *Optimization to low temperature activity in psychrophilic enzymes.* Int J Mol Sci, **13**(9), 11643-11665.
239. Georlette, D., Jonsson, Z.O., Van Petegem, F., Chessa, J., Van Beeumen, J., Hubscher, U. and Gerday, C. (2000). *A DNA ligase from the psychrophile Pseudoalteromonas haloplanktis gives insights into the adaptation of proteins to low temperatures.* Eur J Biochem, **267**(12), 3502-3512.
240. Coquelle, N., Fioravanti, E., Weik, M., Vellieux, F. and Madern, D. (2007). *Activity, stability and structural studies of lactate dehydrogenases adapted to extreme thermal environments.* J Mol Biol, **374**(2), 547-562.
241. Xie, B.B., Bian, F., Chen, X.L., He, H.L., Guo, J., Gao, X., Zeng, Y.X., Chen, B., Zhou, B.C. and Zhang, Y.Z. (2009). *Cold adaptation of zinc metalloproteases in the thermolysin family from deep sea and arctic sea ice bacteria revealed by catalytic and structural properties and molecular dynamics: new insights into relationship between conformational flexibility and hydrogen bonding.* J Biol Chem, **284**(14), 9257-9269.
242. Altermark, B., Niiranen, L., Willassen, N.P., Smalas, A.O. and Moe, E. (2007). *Comparative studies of endonuclease I from cold-adapted Vibrio salmonicida and mesophilic Vibrio cholerae.* FEBS J, **274**(1), 252-263.
243. Kim, S.Y., Hwang, K.Y., Kim, S.H., Sung, H.C., Han, Y.S. and Cho, Y. (1999). *Structural basis for cold adaptation. Sequence, biochemical properties, and crystal structure of malate dehydrogenase from a psychrophile Aquaspirillium arcticum.* J Biol Chem, **274**(17), 11761-11767.
244. Asgeirsson, B. and Cekan, P. (2006). *Microscopic rate-constants for substrate binding and acylation in cold-adaptation of trypsin I from Atlantic cod.* FEBS Lett, **580**(19), 4639-4644.
245. Moe, E., Leiros, I., Riise, E.K., Olufsen, M., Lanes, O., Smalas, A. and Willassen, N.P. (2004). *Optimisation of the surface electrostatics as a strategy for cold adaptation of uracil-DNA N-glycosylase (UNG) from Atlantic cod (Gadus morhua).* J Mol Biol, **343**(5), 1221-1230.
246. Olufsen, M., Smalas, A.O. and Brandsdal, B.O. (2008). *Electrostatic interactions play an essential role in DNA repair and cold-adaptation of uracil DNA glycosylase.* J Mol Model, **14**(3), 201-213.
247. Papaleo, E., Olufsen, M., De Gioia, L. and Brandsdal, B.O. (2007). *Optimization of electrostatics as a strategy for cold-adaptation: a case study of cold- and warm-active elastases.* J Mol Graph Model, **26**(1), 93-103.

248. Siddiqui, K.S., Poljak, A., Guilhaus, M., De Francisci, D., Curmi, P.M.G., Feller, G., D'Amico, S., Gerday, C., Uversky, V.N. and Cavicchioli, R. (2006). *Role of lysine versus arginine in enzyme cold-adaptation: Modifying lysine to homo-arginine stabilizes the cold-adapted alpha-amylase from Pseudoalteramonas haloplanktis.* Proteins-Structure Function and Bioinformatics, **64**(2), 486-501.
249. Aghajari, N., Feller, G., Gerday, C. and Haser, R. (1998). *Structures of the psychrophilic Alteromonas haloplanctis alpha-amylase give insights into cold adaptation at a molecular level.* Structure, **6**(12), 1503-1516.
250. Georlette, D., Damien, B., Blaise, V., Depiereux, E., Uversky, V.N., Gerday, C. and Feller, G. (2003). *Structural and functional adaptations to extreme temperatures in psychrophilic, mesophilic, and thermophilic DNA ligases.* J Biol Chem, **278**(39), 37015-37023.
251. Åqvist, J., Isaksen, G.V. and Brandsdal, B.O. (2017). *Computation of enzyme cold adaptation.* Nature Reviews Chemistry, **1**, 0051.
252. Tanui, C.K., Shyntum, D.Y., Priem, S.L., Theron, J. and Moleleki, L.N. (2017). *Influence of the ferric uptake regulator (Fur) protein on pathogenicity in Pectobacterium carotovorum subsp. brasiliense.* PloS one, **12**(5), e0177647-e0177647.
253. Foster, J.W. (1991). *Salmonella acid shock proteins are required for the adaptive acid tolerance response.* Journal of bacteriology, **173**(21), 6896-6902.
254. Porcheron, G. and Dozois, C.M. (2015). *Interplay between iron homeostasis and virulence: Fur and RyhB as major regulators of bacterial pathogenicity.* Vet Microbiol, **179**(1-2), 2-14.
255. Prince, R.W., Storey, D.G., Vasil, A.I. and Vasil, M.L. (1991). *Regulation of toxA and regA by the Escherichia coli fur gene and identification of a Fur homologue in Pseudomonas aeruginosa PA103 and PA01.* Mol Microbiol, **5**(11), 2823-2831.
256. Ernst, F.D., Bereswill, S., Waidner, B., Stoof, J., Mader, U., Kusters, J.G., Kuipers, E.J., Kist, M., van Vliet, A.H. and Homuth, G. (2005). *Transcriptional profiling of Helicobacter pylori Fur- and iron-regulated gene expression.* Microbiology, **151**(Pt 2), 533-546.
257. Haraszthy, V.I., Jordan, S.F. and Zambon, J.J. (2006). *Identification of Fur-regulated genes in Actinobacillus actinomycetemcomitans.* Microbiology, **152**(Pt 3), 787-796.
258. Yuhara, S., Komatsu, H., Goto, H., Ohtsubo, Y., Nagata, Y. and Tsuda, M. (2008). *Pleiotropic roles of iron-responsive transcriptional regulator Fur in Burkholderia multivorans.* Microbiology, **154**(Pt 6), 1763-1774.
259. Pi, H., Patel, S.J., Arguello, J.M. and Helmann, J.D. (2016). *The Listeria monocytogenes Fur-regulated virulence protein FrvA is an Fe(II) efflux P1B4 - type ATPase.* Mol Microbiol, **100**(6), 1066-1079.

260. Guo, Y. *et al.* (2017). *The Role of the Regulator Fur in Gene Regulation and Virulence of Riemerella anatipestifer Assessed Using an Unmarked Gene Deletion System.* Front Cell Infect Microbiol, **7**, 382.
261. Colquhoun, D.J. and Sorum, H. (2001). *Temperature dependent siderophore production in Vibrio salmonicida.* Microb Pathog, **31**(5), 213-219.
262. Cisse, C., Mathieu, S.V., Abeih, M.B., Flanagan, L., Vitale, S., Catty, P., Boturyn, D., Michaud-Soret, I. and Crouzy, S. (2014). *Inhibition of the ferric uptake regulator by peptides derived from anti-FUR peptide aptamers: coupled theoretical and experimental approaches.* ACS Chem Biol, **9**(12), 2779-2786.
263. Pérard, J., Covès, J., Castellan, M., Solard, C., Savard, M., Miras, R., Galop, S., Signor, L., Crouzy, S., Michaud-Soret, I. and de Rosny, E. (2016). *Quaternary Structure of Fur Proteins, a New Subfamily of Tetrameric Proteins.* Biochemistry, **55**(10), 1503-1515.
264. Mathieu, S. *et al.* (2016). *From Peptide Aptamers to Inhibitors of FUR, Bacterial Transcriptional Regulator of Iron Homeostasis and Virulence.* ACS Chemical Biology, **11**(9), 2519-2528.
265. Baco, E., Hoegy, F., Schalk, I.J. and Mislin, G.L.A. (2014). *Diphenyl-benzo[1,3]dioxole-4-carboxylic acid pentafluorophenyl ester: a convenient catechol precursor in the synthesis of siderophore vectors suitable for antibiotic Trojan horse strategies.* Organic & Biomolecular Chemistry, **12**(5), 749-757.
266. Hameş, E.E. and Demir, T. (2015). *Microbial ribonucleases (RNases): production and application potential.* World Journal of Microbiology and Biotechnology, **31**(12), 1853-1862.

Paper I

Biochemical characterization of Ferric Uptake Regulator (Fur) from *Aliivibrio salmonicida*. Mapping the DNA sequence specificity through binding studies and structural modelling.

Kristel Berg, Hege Lynum Pedersen and Ingar Leiros*

Norwegian Structural Biology Centre (NorStruct), Department of Chemistry, Faculty of Science and Technology, UiT the Arctic University of Norway, N-9037 Tromsø, Norway.

Abstract

Iron is an essential nutrient for bacteria, however its propensity to form toxic hydroxyl radicals at high intracellular concentrations, requires its acquisition to be tightly regulated. Ferric uptake regulator (Fur) is a metal-dependent DNA-binding protein that acts as a transcriptional regulator in maintaining iron metabolism in bacteria and is a highly interesting target in the design of new antibacterial drugs. Fur mutants have been shown to exhibit decreased virulence in infection models. The protein interacts specifically with DNA at binding sites designated as 'Fur boxes'.

In the present study, we have investigated the interaction between Fur from the fish pathogen *Aliivibrio salmonicida* (AsFur) and its target DNA using a combination of biochemical and *in silico* methods. A series of target DNA oligomers were designed based on analyses of Fur boxes from other species, and affinities assessed using electrophoretic mobility shift assay (EMSA). Binding strengths were interpreted in the context of homology models of AsFur to gain molecular-level insight into binding specificity.

Introduction

Iron is an essential nutrient for all living organisms and many key biological processes are dependent on its abundance. For bacteria, iron is crucial for growth and host colonization. Iron mostly exists in the insoluble Fe^{3+} form under aerobic conditions at physiological pH and availability of the soluble reduced form, Fe^{2+}, is restricted. Due to the ability of free iron to form toxic hydroxyl radicals through the Fenton reaction [1], the essential high-affinity uptake systems of iron and iron homeostasis in bacteria must be tightly regulated, and in most bacteria, these processes are under control of the global metalloregulator Ferric uptake regulator (Fur) [2]. Fur was first described in *Escherichia coli* [3], where it controls the expression of more than 90 genes, and its chemical properties and role in homeostasis has since been studied in homologs from multiple bacteria [4-8]. Although Fur was originally described as a repressor of genes coding for components of the ferric uptake systems found in the cell membrane, it is now understood to control the expression of toxins such as hemolysin and exotoxins, as well as proteins involved in iron-scavenging and uptake systems [9,10].

The typical model of action states that when intracellular levels of iron are high, dimeric Fur will act as a repressor by complexing Fe^{2+}, binding specific Fur recognition sites in the promoter region and preventing transcription of associated genes involved in iron uptake, storage and metabolism. Similarly, when iron is limiting, the Fe-Fur complex dissociates from the promoter and allows gene expression. Recent studies however, have broadened our understanding of Fur-mediated regulation, indicating that Fur also may function as an activator and act in an iron-independent manner [11,12], for example in *Helicobacter pylori*, all four combinations of Fur regulation have been characterized: repression and activation, with or without cofactor [13,14]. Further, apo-Fur repression has been described in *Staphylococcus aureus* [15] and *Campylobacter jejuni* [16].

In addition to iron, which is the primary functional metal bound *in vivo*, DNA-binding by Fur can be activated by other divalent metals *in vitro*; Mn^{2+}, Cu^{2+}, Cd^{2+}, Ni^{2+}, Co^{2+} or Zn^{2+} [17-21]. Mn^{2+} is considered a suitable physiological mimic of Fe^{2+} for *in vitro* studies as it is bound with a similar affinity to Fe^{2+} by *E. coli* Fur and adopts the same hexacoordinated octahedral geometry using conserved residues, as seen in recent metal-bound crystal structures of Fur from *Francisella tularensis* [17,22,23]. In contrast Zn^{2+} is bound with lower affinity and in a tetrahedral geometry [22,23].

Fur enacts its biological DNA-binding function as a dimer [24], but may exist in several oligomeric states in solution, depending on protein concentration, salt concentration and pH [25]. Each Fur monomer consists of two domains; an N-terminal winged helix-shaped domain involved in DNA binding (DNA binding domain; DBD) and a C-terminal α/β domain involved in protein dimerization (Dimerization domain; DD) (Figure 1) [26-28]. Crystal structures of apo- and holo-Fur have been available for some time from several bacterial species including *Pseudomonas aeruginosa* Fur [27], *F. tularensis* Fur [23], *Vibrio cholerae* Fur [29], *H. pylori* Fur [30], *C. jejuni* Fur [11], as well as a crystal structure of the DBD of *E. coli* Fur [31]; however, only with the recent structures of *Magnetospirillum gryphiswaldense* MSR-1 Fur (MgFur) in complex with DNA have structural details of Fur-DNA interactions become clear [22]. A series of crystal structures, which include apo-Fur, holo-Fur and two different Fur-DNA complexes, gave a better understanding of issues regarding metal-binding, molecular mechanisms and structural basis of Fur-DNA interaction, at least for that organism.

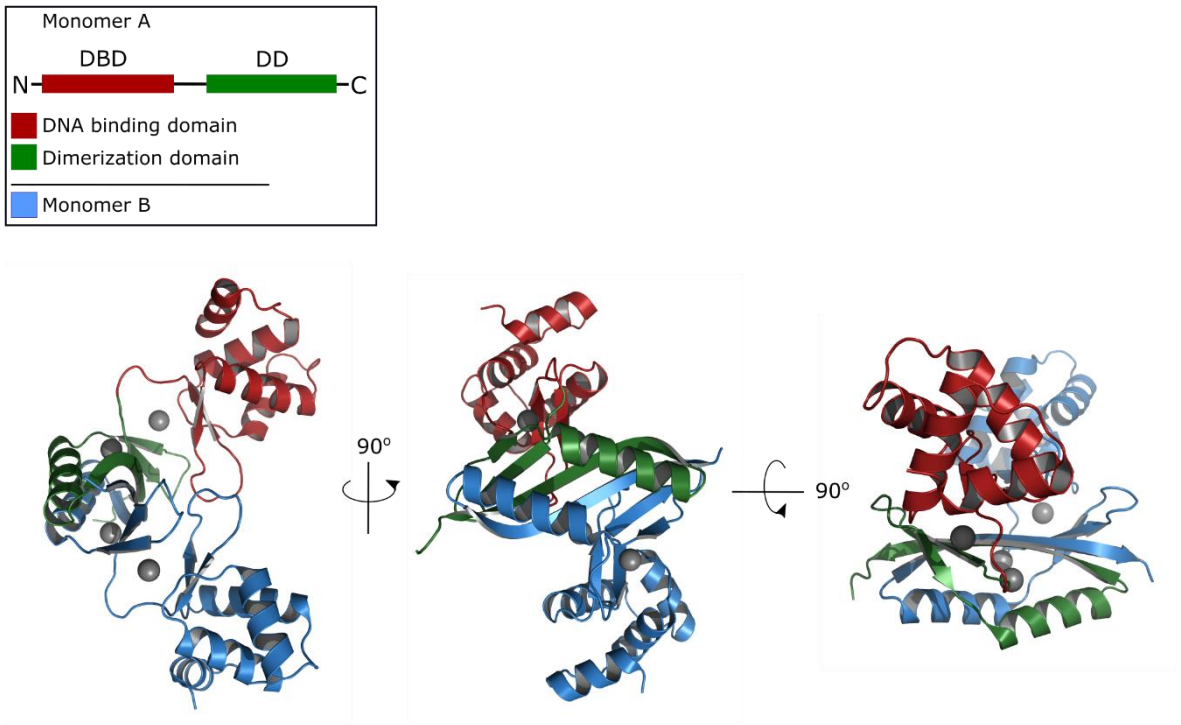


Figure 1. Cartoon representations of dimeric Fur in three different rotated views. Colour codes are as described in the inset. The figure is based on the published structure of MgFur (PDB4rb3, not showing DNA; [22]).

The Fur-DNA interaction site, generally referred to as the "Fur box", is a conserved sequence motif represented by a 19 base pair (bp) palindrome, located between the -35 and -10 sites at the promoters of Fur-regulated genes. The classical Fur box (Figure 2A) originates from DNase I protection- and footprinting-experiments on *E. coli* Fur, where a Fur dimer recognizes a 19 bp inverted repeat sequence: 5'-GATAATGATAATCATTATC-3' [32], although this exact sequence is not found in the *E. coli* genome. This inverted repeat operator site was confirmed by binding of Fur to oligonucleotides inserted into a plasmid [33]. In addition, Fur boxes from other genera have also been characterized and described [32,34-38].

Further studies showed that Fur protects a region somewhat larger than the Fur box, indicating that Fur interacts with DNA also outside the Fur box region. Several Fur dimers may bind two or more overlapping Fur boxes, even extended up to ~100 bp [37], as Fur appears to polymerize along the DNA from the initial primary site to weaker secondary sites. New insights to the binding region led to a revised model, suggesting that Fur recognizes three repeated arrays of 5'-GATAAT-3' in *E. coli* with a slight imperfection in the third array, rather than the classical palindromic 19 bp sequence [39]. In that case, dimeric Fur is suggested to interact with the AT-AT-repeat (5'-G**AT**A**AT**-3') within each hexamer, with two overlapping dimers binding. Figure 2B illustrates how the 5'-GATAAT-3' hexamer can be interpreted as three direct repeats in a tandem array F-F-F fashion, although the last hexamer could be inverted to an F-F-R arrangement or have mismatches. Alternatively, the hexamers can be arranged as two direct repeats, followed by an inverted repeat, separated by a single bp, as illustrated by the hexamer model F-F-x-R in Figure 2C. The latter organization has been shown to appear most frequently in natural Fur binding sites and also has the highest affinity for Fur [36]. A revised model of

how Fur interacts with its target was suggested by Lavrrar *et al*, who predicted the minimum Fur box as a 13 bp overlapping 6-1-6 motif (Figure 2D) with two Fur dimers interacting from opposite faces of the helix, explaining the corkscrew manner Fur wraps around the DNA duplex [40]. A similar reinterpretation of the Fur box, based on alignment of *Bacillus subtilis* Fur boxes and DNase footprinting, proposes a slightly longer Fur box core sequence consisting of two overlapping heptamer inverted repeats [34,41]. These two 7-1-7 motifs generate a slightly extended Fur consensus sequence of 21 bp (Figure 2E). Crystal structures determined for complexes of DtxR bound to its operator site show a similar binding model [42,43].

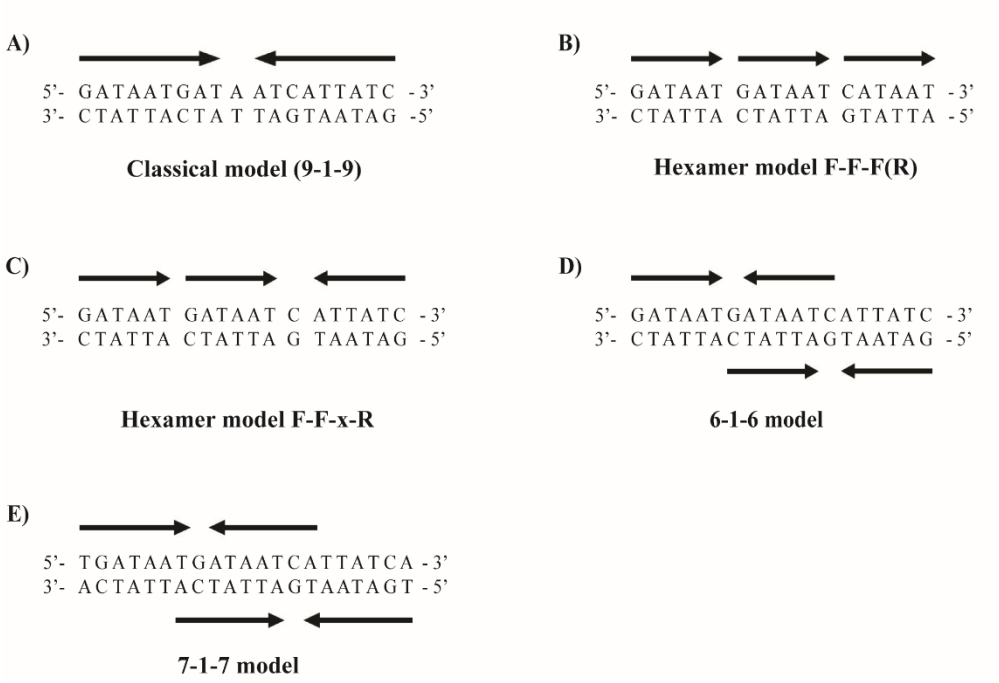


Figure 2. Alternative arrangements of the Fur box, illustrated by the *E. coli* consensus with GATAAT as the minimal recognition unit (Fur consensus NATA/TAT). Arrows mark inverted repeats/repeated arrays. **A**) The 19 bp classical model suggests two inverted repeats with an A:T basepair in between, binding a monomer each. **B**) The 18 bp hexamer model contains a minimum of three direct repeats of the hexamer GATAAT, where the AT-AT pattern within each hexamer was suggested to interact with Fur. The last hexamer may be reversed or imperfect. **C**) The 19 bp hexamer model is described as repeated arrays of three or more copies of GATAAT motifs, recognized by two hexamers in the forward direction and one hexamer at the reverse orientation, separated by one base pair. **D**) The hexamer model can be viewed as a 6-1-6 arrangement, where two overlapping hexamer inverted repeats binds Fur dimers at opposite faces of the double helix. **E**) The 21 bp 7-1-7 model defines the Fur box as two overlapping heptamer inverted repeats, also recognized by two Fur dimers at opposite faces.

In the search for the shortest recognition unit by Fur, the 7-1-7 inverted repeat was found to be the minimum in *B. subtilis* Fur [34]. Single 6-mer or 7-mer oligonucleotides showed no affinity to Fur and Fur boxes with two 6-mers bound weakly. Similar results were obtained for *E. coli* Fur, where a minimum of three repeats of the hexameric motif GATAAT was required for Fur binding [37]. Thus, in the search for a Fur box consensus, the focus is shifting towards the functional pattern within the sequence, rather than the specific sequence or length. The consensus hexamer NATA/TAT appears to be the main unit of interaction with Fur, regardless of orientation and number. In addition, Fur boxes typically have a high content of A/T bases. Experimentally and computationally determined Fur boxes in various bacteria showed

consensus sequence identity ranging from 50% to 80%. [38,41,44-46], and Fur appears to have a rather broad substrate-binding ability.

The published crystal structures of MgFur in complex with two different DNA targets demonstrate the lack of a well-defined sequence specificity, and a high degree of degeneration in the Fur box [22]. DNase I footprinting with the *feoAB1* operator showed a protected region without the typical arrays of GATAAT hexamers. However, for successful co-crystallization, the *feoAB1* operator was mutated to a near-perfect inverted repeat, which bound one dimer of MgFur with similar binding affinities to the original *feoAB1* operator. Gel shift-based assays showed that MgFur also specifically binds the *P. aeruginosa* Fur box, and furthermore, that two dimers of MgFur co-crystallized with the *P. aeruginosa* Fur box sequence (identical in sequence to the *E. coli* Fur box, which we will use throughout) [22]. These MgFur-DNA complex structures are the first to demonstrate the ability to bind DNA at different ratios.

Common for these two rather different DNA targets is the way each Fur monomer formed contacts with both DNA strands using its DNA-binding domain (DBD), which interacted with a 10-11 bp sequence containing an important G base, conserved T base and an AT-rich region characterised by a narrow minor groove. *In vivo* experiments indicated that specific Fur-DNA contacts may be directly connected to DNA shape instead of being base specific. The positively charged Lys15 in MgFur bound this narrow minor groove with enhanced negative electrostatic potential. The narrow minor groove of AT-rich sequences is highlighted as an essential feature for Fur interaction [22].

The Gram-negative Vibrionaceae family of gamma-proteobacteria include many mammalian pathogens, and the role of Fur and iron homeostasis in infection has received much attention due to its potential as a drug target [47-50]. Amino acid alignments and phylogenetic analysis shows that the Fur protein is highly conserved within the Vibrionaceae, and in the present study we have investigated Fur from the Vibrio fish pathogen *Aliivibrio salmonicida*, the causative agent of cold-water vibriosis in Atlantic salmon and cod [51]. Previously, Thode et al. demonstrated a key role of *A. salmonicida* Fur (hereafter AsFur) in iron homeostasis [52] where construction of a *fur* null mutant strain severely affected fitness and growth of the bacteria, caused oxidative stress and a general reduced ability to cope with low-iron conditions. Furthermore, evaluation of expression levels compared to the wild-type identified up-regulation of numerous genes encoding for iron uptake and storage and down-regulation of potential targets for RyhB and other sRNAs involved in iron homeostasis [52]. AsFur and its DNA target (Fur box) have previously been studied *in vitro* and *in silico*, with emphasis on identification of residues of importance for protein-DNA interactions [38,53]. A 19 bp inverted repeat Vibrio Fur box consensus, 5'-AATGATAATAATTATCATT-3', was identified by computational methods [53] and later shown to be specifically recognized and bound by AsFur *in vitro* in EMSA experiments with strong affinity [38]. Additionally, specific individual nucleotides and amino acid residues possibly interacting in the AsFur-vibrio Fur box complex have been predicted, some species-specific for AsFur. In the vibrio fur box, A14, C16 and T13 were suggested to contribute directly to the AsFur-DNA complex, on one or both strands. By homology modeling, the C16 nucleotide was predicted to be in close proximity to the amino

acids Tyr56, Arg57 and Arg70, identified by binding free energy calculations [38]. However, the base-specificity of these interactions remained elusive.

While these previous studies have mainly focused on investigating the effect of amino acid substitutions on DNA interaction, the present study aimed to elucidate the effect of nucleotide substitutions in the target DNA in an attempt to establish the binding mode of AsFur on Fur box-DNA.

In this study, we have characterized AsFur with respect to activity, thermal stability and its binding capability on a range of oligonucleotides in order to investigate the importance of key nucleotides in AsFur-DNA interaction.

Materials and methods

Cloning, expression and purification of AsFur

AsFur was overexpressed and purified with some changes from the previously described protocol [38]. Following cloning of the *fur* gene from *A. salmonicida* into the pDEST14 Gateway expression vector (Invitrogen™, USA), AsFur was overexpressed at 20ºC overnight in *E. coli* BL21-CodonPlus® (DE3)-RIL competent cells, grown in LB broth with 100 µg/ml ampicillin and 34 µg/ml chloramphenicol. Harvested cells were resuspended in lysis buffer (Buffer A; 300 mM NaCl, 50 mM Tris-HCl, 5 mM β-mercaptoethanol and 10 mM Imidazole). The histidine-rich AsFur was purified by Immobilized metal ion affinity chromatography (IMAC) on a 5 ml HisTrap HP column (GE Healthcare). Buffer A was used as the running buffer and Buffer B (300 mM NaCl, 50 mM Tris-HCL, 5 mM β-mercaptoethanol and 500 mM Imidazole) as the elution buffer. The second purification step was performed using size-exclusion chromatography (SEC) on a Superdex 200 16/60 gel filtration column (GE Healthcare) equilibrated with Buffer A without Imidazole added. AsFur purity was verified by SDS-PAGE and protein concentration was determined by NanoDrop 2000c (Thermo Scientific) using the theoretical extinction coefficient.

Thermofluor

In order to improve the stability of the purified AsFur, thermal stability in various buffer systems and salt concentrations was investigated by a Thermofluor assay [54]. Protein unfolding and its melting temperature (T_m) is monitored by using the fluoroprobe SYPRO Orange dye which emits fluorescence upon binding to exposed hydrophobic regions.

The buffer screen contained 24 buffers covering a pH range from 4.5 to 9.0. Briefly, 5 µl protein (2.5 mg/ml), 12.5 µl 2 x buffer solution (100 mM) and 7.5 µl 300 × SYPRO® Orange (Sigma Aldrich) were mixed and added to the wells of a 48-well PCR-plate (Bio-Rad). To assess the effect of various salts, 15 µl of protein (0.8 mg/ml) diluted in the appropriate buffer (Tris pH 7.5) were mixed with 7.5 µl of 300 × SYPRO® Orange (Sigma Aldrich) and 2.5 µl different

salts in concentrations ranging from 0.1-2.0 M. The plates were sealed with Microseal® 'B' Adhesive Seals (Bio-Rad) and heated in a MiniOpticon Real-Time PCR System from 20 to 80 °C in increments of 1 °C per sec. Melting curves were monitored with a charge-coupled device (CCD) camera with wavelengths for excitation and emission at 490 and 575 nm, respectively. T_m, corresponding to the midpoint of the transition curve, was determined using the supplied instrument software and monitoring the fluorescence of the HEX channel.

DNA protection assay and the effect of metals on Fur binding

The capability of purified AsFur to bind DNA in the presence of various metals was investigated using a restriction site protection assay. The aerobactin plasmid pDT10, (kindly provided by Isabelle Michaud-Soret, Grenoble, France) carries four restriction enzyme sites, with the *E. coli* Fur box incorporated into one of the HinfI sites [55]. Based on the method developed by Bagg and Neilands [18], activated Fur binds the Fur box and thereby makes the restriction site unavailable for digestion by HinfI. Fur activity is confirmed by observing digestion patterns on gel electrophoresis. Fur is active if a 1781 bp band is observed, while two bands, respectively 1530 bp and 251 bp are observed if the protein is inactive. Figure 3 (modified from [56]) summarizes the principle behind the assay.

AsFur (20 µM) was incubated with two equivalents of a range of metals (40 µM) in binding buffer (100 mM BisTrisPropane pH 7.5, 100 mM KCl, 5 mM $MgSO_4$) for 10 minutes at room temperature, followed by addition of pDT10 plasmid at 10 nM final concentration and 20 minutes additional incubation. Restriction enzyme digestion was carried out by adding 4 units per µl of HinfI to the mixture and incubating for 1h at 37°C before quenching with 0.5 mM of EDTA. The samples were run for 30 min at 100 V on 1 % agarose gel in TAE and visualized under UV light.

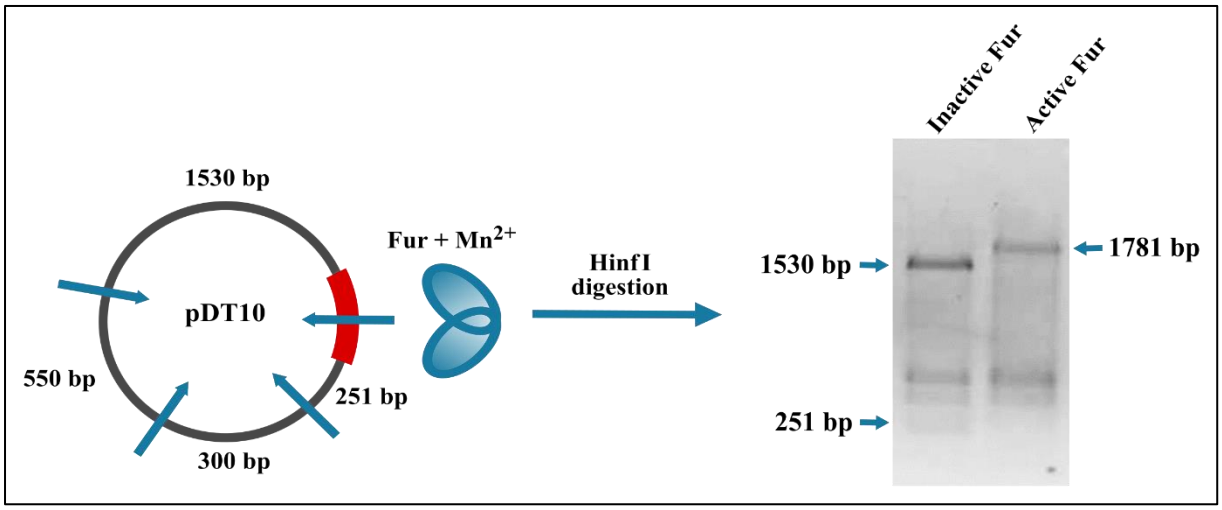


Figure 3. Plasmid protection assay. Active Fur dimer binds the incorporated *E. coli* Fur box consensus in the pDT10 plasmid and protects the specific site from HinfI digestion. The DNA protection due to Fur activity is visible as an altered migration pattern on a 1 % agarose gel. The Fur binding site is highlighted in red and HinfI restriction sites are marked by blue arrows, also including expected fragment sizes.

Design of synthetic Fur box-containing oligonucleotides

Based on proposed DNA targets (Fur box) from Fur homologs [36,41,57], numerous Fur box-containing oligonucleotides of various lengths (15-24 nt) were designed. In short, single-stranded DNA (Sigma-Aldrich) were diluted in Buffer C (50 mM HEPES pH 8, 50 mM NaCl) to 1 mM and annealed to double-stranded DNA by boiling for 5 min and cooling slowly to room temperature. Annealed oligonucleotides were separated by anion-exchange liquid-chromatography column (Mono-Q), with Buffer C as running buffer and Buffer D (50 mM HEPES pH 8, 1 M NaCl) as elution buffer, followed by dialyses in Slide-A-Lyzer Dialysis Cassettes (3.5k MWCO; Thermo Scientific) overnight back to Buffer C. DNA concentrations were measured by Nanodrop 2000c (Thermo Scientific).

Electrophoretic gel mobility shift assay (EMSA)

Unlabelled Fur box-mimicking oligonucleotides of various lengths were used as probes in EMSA assays, where DNA mobility is detected by the double stranded nucleic acid stain SybrGreen (Life Technologies), and slower mobility indicates that AsFur has complexed with the Fur box.

To complex AsFur with its DNA target, desired concentrations of purified AsFur was incubated with binding buffer (20 mM Tris acetate pH 8.0, 1 mM $MgCl_2$, 50 mM KCl, 1 mM DTT and 100 µM $MnCl_2$) at RT for 20 minutes. After addition of DNA (5 µM), the mixture was incubated for another 30 minutes before adding 10 x loading dye (30% glycerol in binding buffer). Samples were loaded on native 8 % polyacrylamide/1 x TB gels and electrophoresis was performed at 200 V for 2-2½ hours and at 6ºC with circulating buffer. Finally, the gel was incubated with SybrGreen 1:10000 in TB buffer for 20 minutes and band shifts were detected under UV light at ~254 nm. Binding strengths were examined and rated by visualization.

Analysis of AsFur compared to functional and structural homologs

Homology models of AsFur were generated using as templates the crystal structures of MgFur in complex with the *E. coli* Fur box (PDB4rb1) and the *feoAB1* operator (PDB4rb3), respectively. The modelling tools of the Swiss-model repository were utilized in default mode to obtain the homology models. The two different interactions modes and stoichiometry of MgFur interacting with dsDNA were further analysed using WinCoot [58] and visualized by PyMol (www.pymol.org). Conserved nucleotide base-protein interactions were highlighted from structure-based sequence alignments with homologous Fur crystal structures, rendered by ESPript 3.0 [59] and from the output from the NuProPlot server [60].

Results & Discussion

AsFur was purified to homogeneity through affinity- and size exclusion-chromatography

AsFur consists of 147 amino acid residues with theoretical pI and molecular weight of 5.75 and 16.6 kDa, respectively. A large-scale purification procedure of AsFur was established by Pedersen *et al* [38]. In brief, the *fur* gene from *A. salmonicida* was cloned, over-expressed in BL21-CodonPlus® (DE3)-RIL and purified to apparent homogeneity by two consecutive steps; IMAC affinity purification using HisTrap HP followed by SEC using HiLoad Superdex 200 pg. From SEC chromatography, AsFur fractions were detected at a volume corresponding to a homodimer, consistent with previous observations [38]. The SDS-PAGE analysis of purified AsFur is shown in Figure 4.

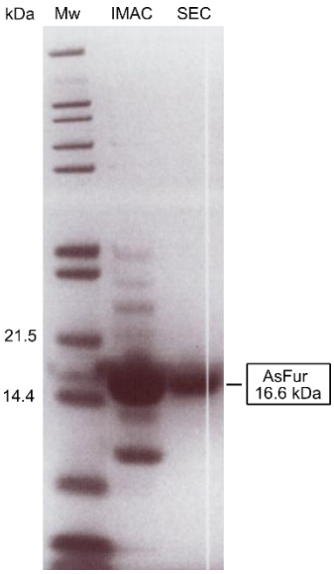


Figure 4. **Coomassie Blue-stained SDS-PAGE** showing molecular weight marker and the collected fractions from IMAC and SEC, respectively. The relevant molecular weights (Mw; kDa) are indicated in the figure.

Thermal denaturation screening on pH and salt showed a slight effect on the stabilization of Fur

Although AsFur was purified to homogeneity as seen in Figure 4, the initial protein batches showed a tendency to aggregate, with complete loss of binding activity within a week at standard storage conditions. To avoid protein aggregation and increase stability, a thermal shift assay (Thermofluor) was implemented to identify better buffer conditions. Screening of a range of buffer compositions (and pH) by Thermofluor only showed negligible effects on AsFur stability, however, activity assays indicated that a minor change in pH from 8.0 to 7.5 in Tris-buffer reduced aggregation and increased the storage stability of AsFur at 4 ˚C. Furthermore, a slight increase in NaCl concentration up to 200 mM showed a positive effect on AsFur stability (Figure 5), in comparison to $MgCl_2$ and KCl where only minor improvements could be seen. Although the storage stability of AsFur were improved by the above-mentioned changes in pH and NaCl concentrations, batch variations were still a frequent problem in the following characterization.

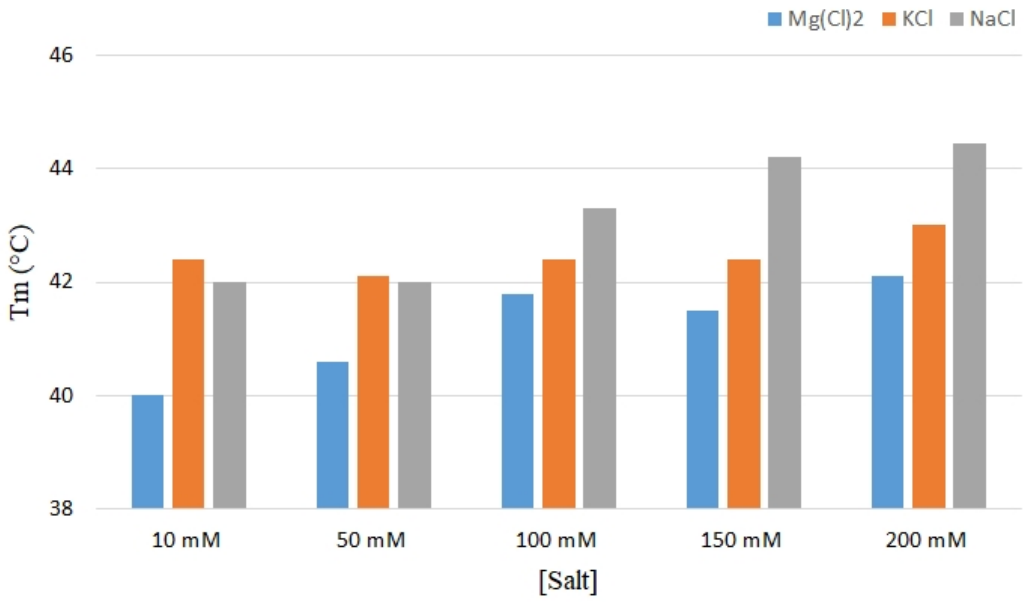


Figure 5: The effect of salt on thermostability. Buffer conditions were 30 mM Tris pH 7.5.

Presence of divalent metals alters DNA binding by AsFur

In the classical regulation pattern, iron is the primary functional metal that dimerizes and activates Fur *in vivo*. The ability of Fur to be effectively activated by a wide range of other divalent metal ions *in vitro* has prompted discussions about the true physiological metal responsible for Fur activation, although evaluation of Fur metal affinity by metal titration experiments suggests that only Fe^{2+} show sufficient affinity to activate Fur within relevant concentration ranges *in vivo* [17]. However, elevated concentrations of other metals intracellularly could have implications for the normal iron regulation and the different metal bound Fur could potentially act on different DNA targets [61].

To measure the effect of a range of metals on AsFur-DNA binding, an *in vitro* assay utilising protection of a restriction site in the aerobactin promoter was used [55]. Functional binding by Fur is envisaged by the absence of a fourth 251 bp band on the gel and an increase in size of the upper band to 1781 bp. Analysis showed that AsFur binds the aerobactin promoter in a metal-dependent fashion with Mn^{2+} present (Figure 6A). Furthermore, the results in Figure 6B show that AsFur also is able to bind the Fur box in presence of the divalent metal cations Mn^{2+}, Zn^{2+}, Cu^{2+} and Co^{2+}. Fe^{2+} is considered the most physiologically-relevant metal ion for Fur activation, however it was omitted from this panel as its rapid oxidation precludes its use under standard assay conditions. Although plasmid protection appears weaker for Mn^{2+} compared to Zn^{2+}, Cu^{2+} and in particular Co^{2+}, Mn^{2+} was still the preferred choice for subsequent AsFur-DNA interaction studies, as Mn^{2+} and Fe^{2+} have been shown to have conserved metal coordination and similar chemical behavior in structural studies [23]. The behaviour of AsFur

in the presence of Cd^{2+} could not be interpreted, as the migration pattern resembles that of untreated plasmid, suggesting that HinfI is inhibited in the presence of Cd^{2+}.

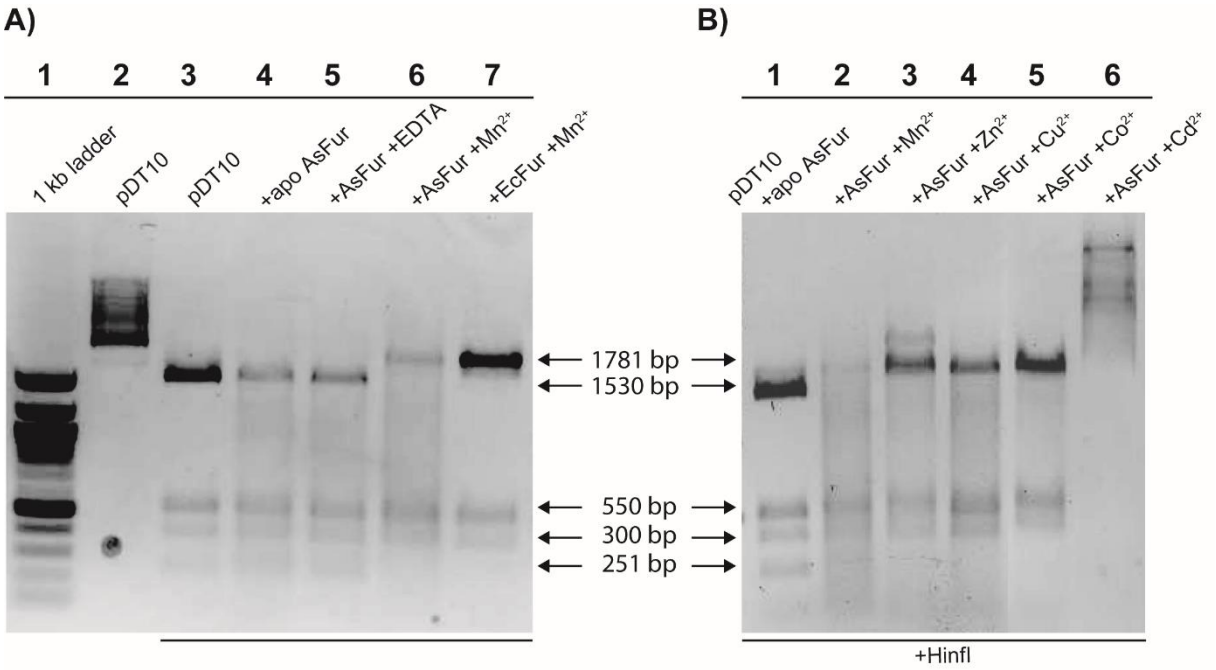


Figure 6. Plasmid protection assay verifying AsFur DNA binding in the presence of manganese A) and identifying various metals able to activate AsFur B). The plasmid pDT10 was cleaved by HinfI in the absence or presence of active Fur and the digested migration pattern was analyzed by 1 % gel electrophoresis. Active AsFur binds to the incorporated Fur box in the 1781 bp restriction fragment and protects it from being cleaved into 1530-bp and 251-bp fragments. **A**) Lane 1: 1 kb ladder; Lane 2: plasmid pDT10; Lane 3: pDT10 + HinfI; Lane 4: pDT10 + apo AsFur + HinfI; Lane 5: pDT10 + AsFur + EDTA + HinfI; Lane 6: pDT10 + AsFur + Mn^{2+} + HinfI; Lane 7: pDT10 + EcFur + Mn^{2+} + HinfI **B**) Lane 1: plasmid pDT10 + apo AsFur + HinfI; Lane 2: pDT10 + As-Fur + Mn^{2+} + HinfI; Lane 3: pDT10 + AsFur + Zn^{2+} + HinfI; Lane 4: pDT10 + AsFur + Cu^{2+} + HinfI; Lane 5: pDT10 + AsFur + Co^{2+} + HinfI; Lane 6: pDT10 + AsFur + Cd^{2+} + HinfI.

Oligonucleotides of different lengths and sequences have an effect on the Fur-DNA binding

The 19 bp inverted repeat Vibrio consensus sequence, 5'-AATGATAATAATTATCATT-3', as well as the *E. coli* Fur box, 5'-GATAATGATAATCATTATC-3', formed the templates for EMSA assays of a range of oligonucleotides (see Table 1 for oligonucleotide composition). Single bases and/or arrays of bases were mutated to examine possible important binding sites or important hexamer arrangements. Binding strengths were examined and rated.

Table 1. Oligonucleotides selected and tested for AsFur interaction by EMSA.

Identifier	Sequence	Length	B/S*	Comment
Vibrio consensus	5'-AATGATAATAATTATCATT-3' 3'-TTACTATTAATAATAGTAA-5'	19	B	Vibrio Fur box
E. coli consensus	5'-GATAATGATAATCATTATC-3' 3'-CTATTACTAATAGTAATAG-5'	19	B	*E. coli* Fur box
Vibrio anti-consensus	5'-CCGTGCGCACTCCGCAGGG-3' 3'-GGCACGCGTGAGGCGTCCC-5'	19	B	Vibrio least conserved (neg. control)
A)	5'-**GCA**GATAATAATTATCATT-3' 3'-**CGT**CTATTATTAATAGTAA-5'	19	B	Vibrio mutated 3 5'-nucleotides
B)	5'-AATGATA**C**TAATTATCATT-3' 3'-TTACTAT**G**ATTAATAGTAA-5'	19	B	Vibrio mutated A8C
C)	5'-AATGATAA**G**AATTATCATT-3' 3'-TTACTATT**C**TTAATAGTAA-5'	19	B	Vibrio mutated T9G
D)	5'-AATGATA**CG**AATTATCATT-3' 3'-TTACTAT**GC**TTAATAGTAA-5'	19	B	Vibrio mutated A8C/T9G
E)	5'-AATGATAATAA**GG**ATCATT-3' 3'-TTACTATTATT**CC**TAGTAA-5'	19	B	Vibrio mutated T12G/T13G
F)	5'-AATGATAATAATT**G**T**A**ATT-3' 3'-TTACTATTATTAA**C**A**T**TAA-5'	19	B	Vibrio mutated A14G/C16A
G)	5'-AATGATA**CG**AA**GGG**T**A**ATT-3' 3'-TTACTAT**GC**TT**CCC**A**T**TAA-5'	19	B	Vibrio mutated 6 nucleotides
H)	5'-AAATGATAATAATTAT -3' 3'- TTACTATTATTAATAT-5'	16	S	Vibrio shortened w/sticky ends
I)	5'-AATGATAATAATTAT-3' 3'-TTACTATTATTAATA-5'	15	B	Vibrio shortened
J)	5'-GATAATGATAATCATT**G**T**G**-3' 3'-CTATTACTATTAGTAA**C**A**C**-5'	19	B	*E. coli* mutated A17G/C19G
K)	5'- GATAATGATAAT**G**AT**A**ATC-3' 3'-GCTATTACTATTA**C**TA**T**TA -5'	19	S	*E. coli* 3 x GATAAT repeat w/sticky ends
L)	5'-GATAATGATAAT**G**AT**A**AT**GATAAT**-3' 3'-CTATTACTATTA**C**TA**T**TA**CTATTA**-5'	24	B	*E. coli* 4 x GATAAT repeat

* B/S refers to blunt-ended or sticky-ended oligonucleotides. Bold positions labelled in red or dark green denote modifications with respect to the Vibrio and *E. coli* consensus sequences, respectively.

The oligonucleotides were loosely grouped based on conservation and length. First, EMSA experiments were performed in order to verify interaction of AsFur with the consensus sequences from both Vibrio species and *E. coli* using the least-conserved Vibrio sequence [38] as a negative control (Figure 7). Strong interaction was observed with the Vibrio consensus sequence (Figure 7A), and moderate binding was also seen with the *E. coli* consensus sequence (Figure 7B), with essentially no interaction detected with the Vibrio least-conserved (anti-consensus) sequence even at the highest AsFur concentration (Figure 7C).

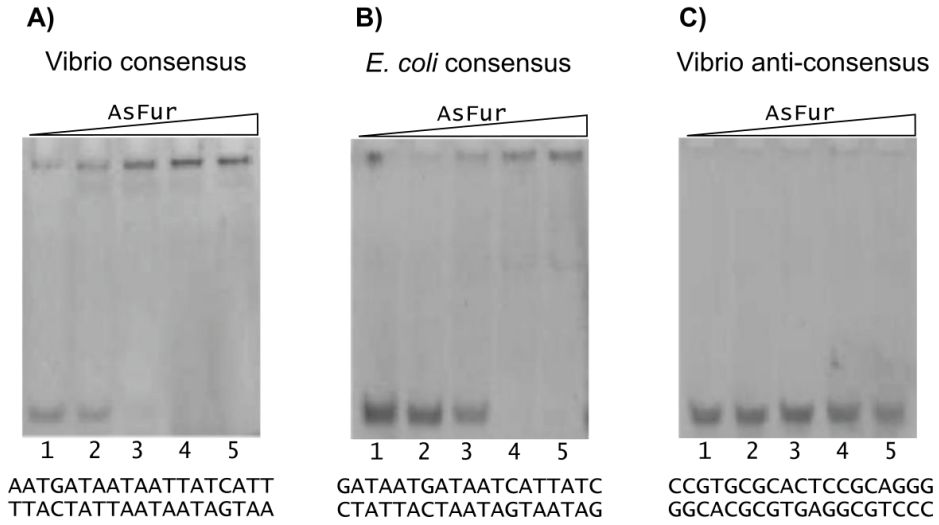


Figure 7. EMSA positive and negative controls. A) Vibrio consensus sequence. **B)** *E. coli* consensus sequence. **C)** Vibrio anti-consensus sequence. 10 μM DNA was incubated with increasing concentrations of AsFur (0, 10, 20, 40 and 80 μM) for lanes 1-5, respectively.

Subsequent EMSA experiments were run with oligonucleotides of varying content compared to the two different consensus sequences, in order to investigate the effect of specific base changes on interaction strength with AsFur. The primary targets for the experimental design were the AsFur-DNA interactions predicted by previous MD simulations and binding free energy calculations [38]. In order to further investigate the existing Fur-DNA interaction models, different oligonucleotide lengths (both shorter and longer than the 19 bp Fur box) were probed, as well as variation in the oligonucleotide termini which were either blunt or included a 1 nucleotide overhang capable of forming a 'sticky' end with adjacent DNA substrates (Table 1). As expected, the oligonucleotides showed varying interaction strength with AsFur (Figure 8). For most EMSA experiments, AsFur binding caused the substrate to be retained in the wells of the gel, which most likely reflects the tendency of AsFur to aggregate, possibly triggered by the initial DNA complex formation.

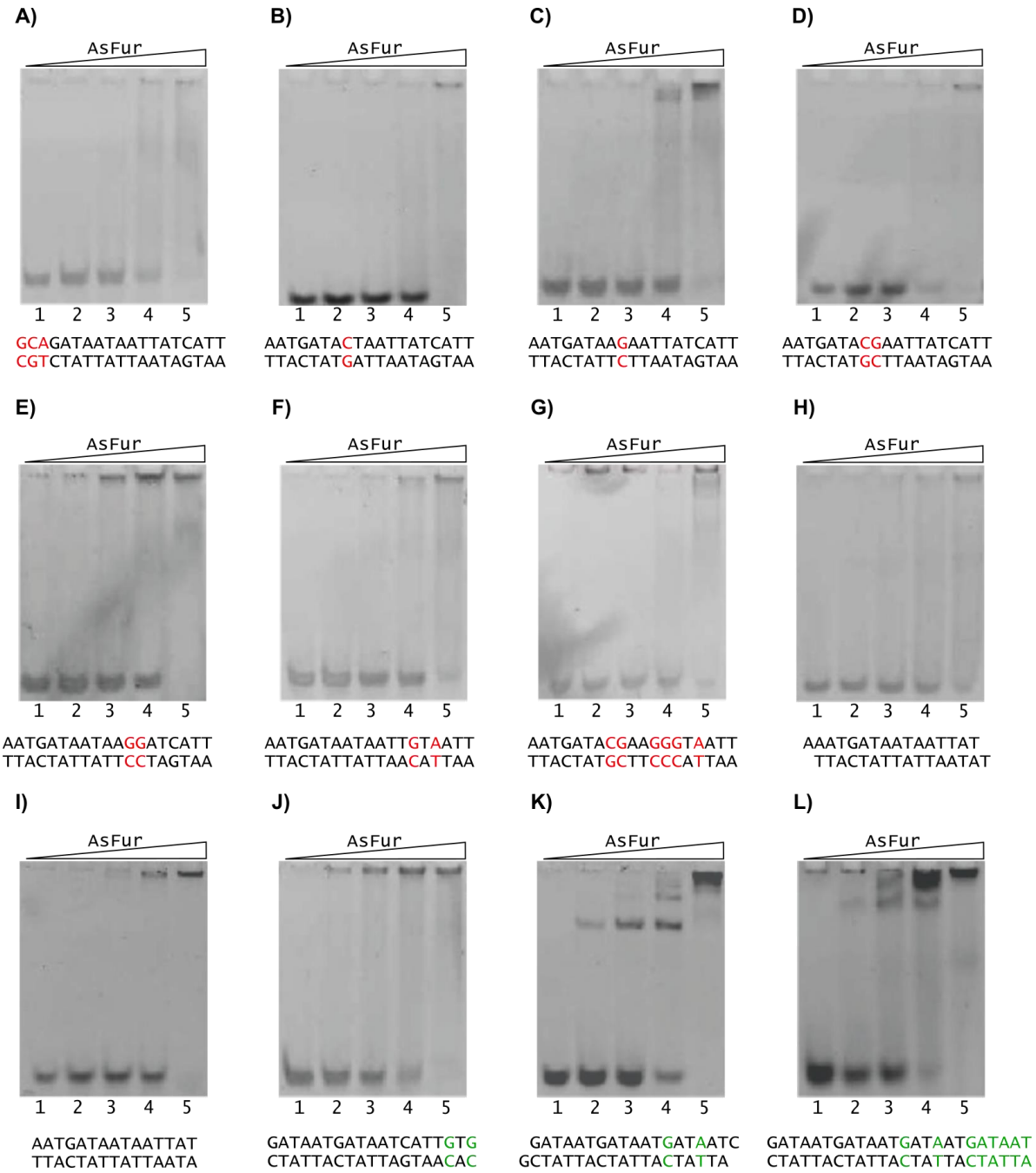


Figure 8. EMSA experiments on variants of Vibrio and *E. coli* consensus oligonucleotides modifying individual positions and/or length. For each experiment, 10 μM DNA was incubated with increasing concentrations of AsFur (0, 10, 20, 40 and 80 μM) for lanes 1-5, respectively. Positions labelled in red or dark green denote modifications with respect to the Vibrio species and *E. coli* consensus sequences, respectively.

The T12G/T13G substitution reduces interaction of AsFur with DNA considerably (Figure 8E) compared to the Vibrio consensus sequence (Figure 7A), suggesting these are key positions for interaction. The corresponding nucleotide positions in the *E. coli* Fur box are T15 and T16 which have previously been shown to interact with *E. coli* Fur by crosslinking experiments [62]. Furthermore, molecular dynamics simulations with AsFur indicated T13 as an important

contributor in protein interaction [38]. We thus present the first EMSA experiments probing these positions directly in comparison to the Vibrio consensus sequence.

When the A14G/C16A substitution (Figure 8F) is compared to the Vibrio consensus sequence (Figure 7A), much reduced interaction capacity with AsFur is observed. These nucleotide positions have previously been shown to contribute favourably to AsFur DNA binding through binding free energy simulations [38], and the EMSA results further indicate them to participate in sequence-specific interactions. It is interesting to observe that the substitutions A17G/C19G (Figure 8J) to the *E. coli* consensus sequence (Figure 7B) also has a detrimental effect, although slightly less pronounced than for the Vibrio consensus sequence. Interestingly, the results highlight the importance of both DNA strands in Fur interaction, as these nucleotide positions form part of the first **G**A**T**AAT hexamer repeat on the complementary strand of both the Vibrio consensus sequence and the *E. coli* Fur box (altered to **T**A**C**AAT and **C**A**C**AAT, respectively).

The A8C and T9G individual substitutions (Figure 8B and C) also lead to much weaker interactions when compared to the Vibrio consensus sequence (Figure 7A). AT-rich regions have previously been shown to be essential for Fur-DNA interactions [2,3,9,10]. In particular, the last T base of the GATAA**T** (T_6) unit in the hexamer repeat model described in Figure 2B and C, corresponding to the substituted T9 in the Vibrio consensus sequence, has been highlighted for its role in DNA recognition by footprinting and missing-T assays with *E. coli* Fur [36]. The matching T on the complementary strand (T_5) showed comparable effect in interactions. However, the combined substitutions A8C/T9G (Figure 8D) does not show an additive effect and has slightly stronger interaction than the individual substitutions. It is interesting that the dual removal of AT-nucleotides in the core of the Vibrio recognition sequence does not appear to further reduce binding strength.

As expected, and in a similar fashion as for the Vibrio least conserved sequence (Figure 7C) which showed almost no sign of DNA interaction with AsFur, the combined alteration of all nucleotides addressed so far (A8C/T9G/T12G/T13G/A14G/C16A; Figure 8G) produced a much-weakened interaction with AsFur, although for this EMSA gel some trace amounts of AsFur can be seen shifted to the wells of the gel throughout.

For the A1G/A2C/T3A substitutions (Figure 8A) compared to the Vibrio consensus sequence (Figure 7A), substantially reduced interaction strength can be observed. This result is interesting in view of the differences in the 5'-regions of the Vibrio and *E. coli* consensus sequences where the Vibrio consensus sequence has a three-nucleotide 'insertion' (AAT) compared to the classical *E. coli* Fur box.

The importance of a minimum length of the Vibrio consensus sequence in AsFur interactions was demonstrated by the EMSA experiments on shortened oligonucleotides compared to the Vibrio consensus sequence (Figure 8H and I), where both the 16-nucleotide sticky-end variant and the 15-nucleotide blunt-ended oligonucleotides displayed much-reduced binding strength compared to the Vibrio consensus sequence. This agrees well with previous EMSA studies on *E. coli* Fur indicating that only weak interaction is formed when oligonucleotides are considerably shorter than three hexamer repeats of the GATAAT sequence [39].

When somewhat similar experiments were performed on GATAAT hexamer repeats of the *E. coli* fur box, either with a 19-nucleotide sticky-end variant (Figure 8K) or the 24-mer quadruple repeat of the GATAAT sequence (Figure 8L), reduced interaction strengths were observed for both compared to the *E. coli* Fur box (Figure 7B). As above, these trends correspond well with EMSA experiments on EcFur with the *E. coli* Fur box, as well as a range of GATAAT repeats, where the interaction strengths were rated as Fur box > 4x GATAAT > 3x GATAAT [39]. The introduction of sticky ends to the triple GATAAT repeat in our study appears to improve binding slightly.

Analysis of AsFur compared to functional and structural homologs

To enable analysis of structural interactions contributing to specificity of binding between AsFur and variations on canonical Fur-box sequences, the sequence of AsFur was compared to structurally-determined homologs. A number of structurally-characterized homologs of AsFur were identified with sequence identities ranging from 86% to 30% (Table 2 and Figure 9). The sequence alignment between these homologs and AsFur highlights several conserved sequence patches both within the DNA-binding- and dimerization domains (DBD and DD, respectively). While the conserved positions in the DD are mainly attributed to metal-coordination, the conserved patches in the DBD are involved in interactions with the Fur box (Figure 9).

Table 2. Comparison of AsFur with known Fur structural homologs.

Abbreviation	Species	PDB	#amino acids	#identical/ #aligned	seq. id. (%)	ref.
AsFur	*Aliivibrio salmonicida*	Model	147	-	-	
EcFur	*Escherichia coli* (DBD)	2fu4	83	71/83	86	[31]
VcFur	*Vibrio cholerae*	2w57	150	125/146	86	[29]
PaFur	*Pseudomonas aeruginosa*	1mzb	136	60/126	75	[27]
FtFur	*Francisella tularensis*	5nbc	140	52/133	39	[23]
CjFur	*Campylobacter jejuni*	4ets	162	53/133	40	[11]
MgFur	*Magnetospirillum Gryphiswaldense Msr-1*	4raz/ 4rb1/ 4rb3	145	51/135	37	[22]
HpFur	*Helicobacter pylori*	2xig	150	41/137	30	[30]

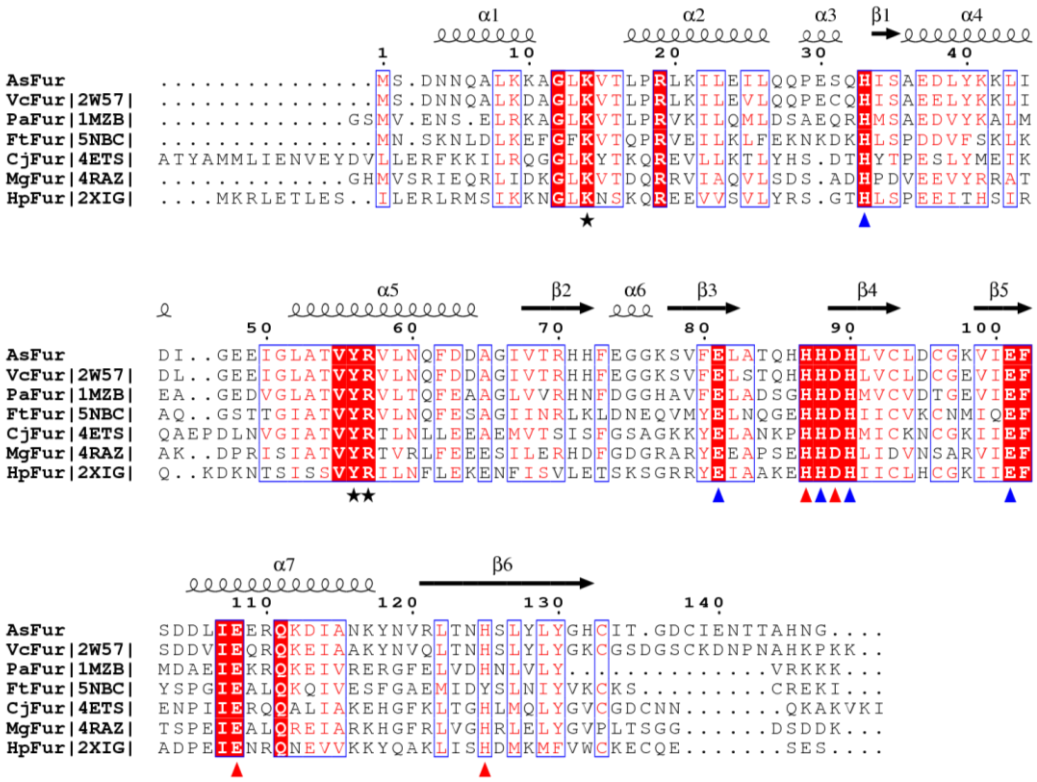


Figure 9. Structure-based sequence alignment of AsFur with known Fur structural homologs. Abbreviations are as defined in Table 2. EcFur was not included as the structure only represents the DBD. PDB identifiers are indicated between vertical lines. Secondary structure elements are shown above the alignment with spirals and arrows indicating α-helices and β-strands, respectively. Identical residues are shown in white on red background, while conserved residues are shown in red. Metal-coordinating residues are indicated with triangles (coloured blue and red for site 1 and site 2, respectively), while residues forming base contacts are indicated with a black asterisk.

For homology modelling, the recently-reported structures of MgFur were selected as templates despite the relatively low sequence similarity with AsFur, as to date these are the only structures of Fur in complex with target DNA [22]. Sequence alignment between AsFur and MgFur revealed 37% identity for the 135 residues that could be structurally aligned and enabled reliable modelling of the entire protein including the N-terminal DNA-binding domain (DBD), which is highly flexible in the un-bound form [22,63]. To compare different possible binding modes of AsFur, two models were constructed: the first based on the MgFur dimer bound to a *feoAB1* operator as a 9-1-9 inverted repeat (PDB4rb3; Figure 2A) and the second based on the two MgFur dimers bound to an *E. coli* Fur box as a 7-1-7 inverted repeat offset by 6 nucleotides (PDB4rb1; Figure 2E). Comparison of these models reveals a conservation in amino acids in the interacting regions of the two proteins (Figure 10).

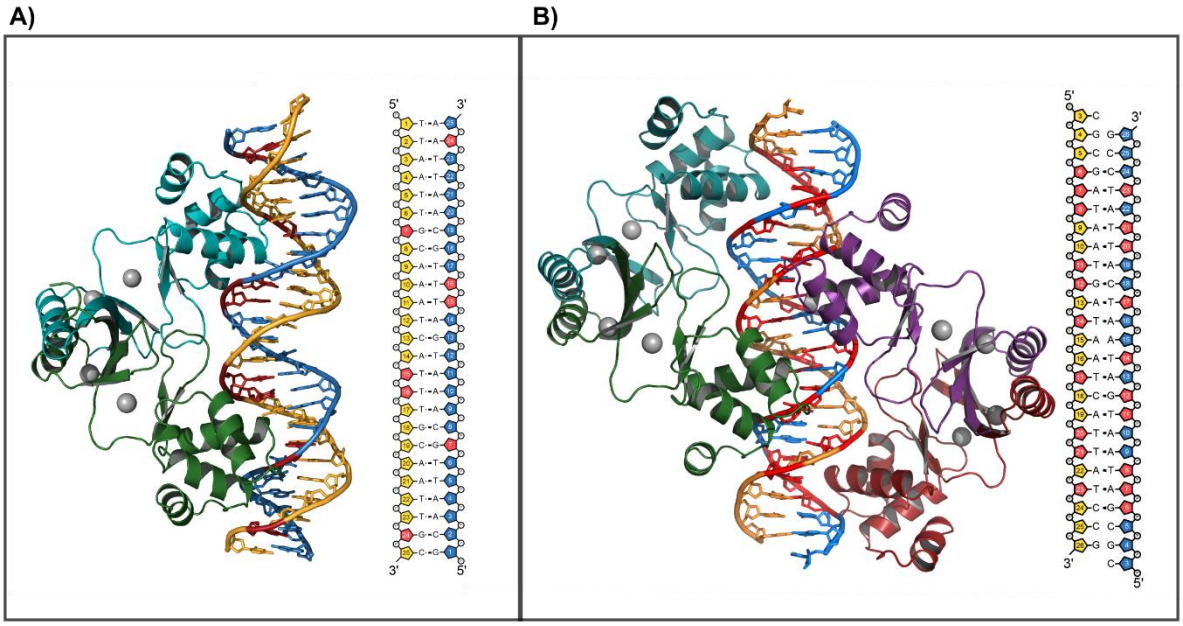


Figure 10. Homology models of AsFur in the two Fur-DNA interaction modes observed for MgFur and reported by Deng *et al* [22]. A) A dimer of AsFur interacting with the *feoAB1* operator. **B)** Two AsFur dimers interacting with the *E. coli* Fur box. Each AsFur monomer is coloured individually and the DNA strands are coloured in dark yellow and blue for the primary and complementary strands, respectively. Nucleotides coloured in red indicate base contacts with AsFur.

Analysis of structural determinants of AsFur-DNA interaction

The homology models generated based on MgFur were analysed to structurally rationalize the variations in interaction strengths from EMSA.

These strongly indicate that AsFur Tyr56 forms base-specific major groove interactions through hydrophobic interactions with the methyl groups of both T12 and T13 (Figure 11B), explaining the observed decrease in binding affinity in the T12G/T13G substitution. Interestingly, this interaction is conserved in both structural models, *i.e.* both in the forms of a 9-1-9 inverted repeat, as well as in the 7-1-7 inverted repeat offset by 6 nucleotides, highlighting the role of Tyr56 in interactions with Fur box-containing DNA.

The homology model of AsFur interacting with the *E. coli* Fur box (Figure 10B) also highlights the importance of both DNA strands in this interaction, providing a rationale for the impact of the A17G/C19G substitution (Figure 8J) with nucleotide base-contacts formed in the 5'-end of the complementary strand. The detailed view of the interactions shown in Figure 11A and C, illustrates that the nucleotide positions in the first hexamer repeat of the complementary strand (G1' and T3') form minor- and major-groove nucleotide base interactions with the conserved residues Lys14 and Arg57, respectively. Arg57 appears to form bidentate base-specific major-groove interactions, while the minor-groove interactions formed by Lys14 are base-unspecific. The corresponding Lys residue in MgFur has previously been shown to interact with the DNA target through a shape readout mechanism, where the AT-rich region in each hexamer repeat results in a narrow minor groove with enhanced negative electrostatic potential [22]. Previous

studies have indicated that most DNA-binding proteins use interplay between the base- and shape-readout modes to recognize their DNA binding sites [64]. This in turn allows for alterations in the specific nucleotide succession, and for the specific case of Fur, thus rationalises the degree of degeneracy found among Fur recognition sequences.

While Lys14, Tyr56 and Arg57 are also found to interact with nucleotide bases in the homology model of AsFur in complex with the *feoAB1* operator (Figure 10A), these interactions can not to the same extent justify a structural rationalisation of the above-mentioned effects from our EMSA experiments, making it less likely that AsFur interacts with the Vibrio consensus sequence and the *E. coli* Fur box as one dimer in the form of an 9-1-9 inverted repeat.

The AsFur homology model with the *E. coli* Fur box does not show direct contacts in the 5'-region of the Vibrio consensus sequence (upstream of the first GATAAT repeat), which is equivalent to the position of the substitution A1G/A2C/T3A; however, it is likely that Lys14 from the DBD of monomer B may undergo conformational changes in order to be involved in minor groove interactions. In fact, previous studies have suggested that the Fur box should be extended in the 5'-end, where Baichoo *et al* [34] suggested an additional T extension in the *B. subtilis* Fur box and Chen *et al* [65] that the *E. coli* Fur box should include the sequence AAT, i.e. identical to the Vibrio consensus sequence.

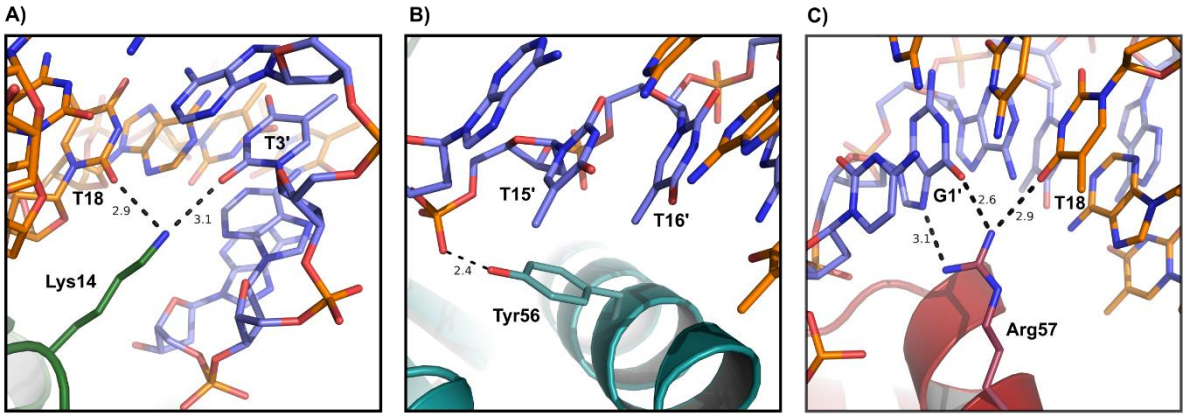


Figure 11. AsFur-nucleotide base interactions. Nucleotide base-interactions observed in the AsFur homology model based on the crystal structure of two dimers of MgFur in complex with the *E. coli* Fur box (PDB4rb1). Nucleotide numbering follows the numbering scheme used for the *E. coli* consensus sequence in Table 1. **A)** Lys14 in monomer A interacts in the minor groove with T18 on the primary strand and T3' on the complementary strand. **B)** Tyr56 in monomer B forms hydrophobic interactions in the major groove with T15' and T16' on the complementary strand (identical interactions are formed between Tyr56 in monomer D, generated through a crystallographic symmetry operation, and T15/T16 on the primary strand). **C)** Arg57 in monomer D interacts in the major groove with T18 on the primary strand and G1' on the complementary strand.

Conclusion

Fur has an important role in iron homeostasis and regulation of virulence mechanisms in many pathogenic bacteria. In an attempt to better understand the molecular basis behind DNA-recognition by AsFur, we have examined its DNA interaction with the combined use of

interaction assays and structural modeling, which allowed for a structure/function interpretation of the biochemical results obtained.

AsFur was found to be a dimer during purification conditions. Due to protein instability issues, it was difficult to further investigate the stoichiometric rates on its interaction with different consensus sequences. However, the combined output of the homology modelling and the EMSA investigations indicate that AsFur will be able to interact in the form of two dimers.

The combined results of the EMSA experiments and homology models indicate that AsFur binding strength to DNA is stronger for longer oligonucleotides than shorter, and we observed a small increase in binding strength when sticky ends were introduced to the same oligo sequence. The results further showed that no single base mutations were crucial, and that only anti-consensus depleted binding completely. However, nucleotide positions T12 and T13 (T15 and T16 in *E. coli*) and A14 and C16 (A17 and T19 in *E. coli*) previously suggested to be in direct contact with Fur, lead to a markedly reduced binding strength between AsFur and DNA when mutated. This indicated that these bases were important for AsFur-DNA specific interaction. In addition, mutations of individual and dual AT bases in the core of the vibrio consensus sequence highlighted the importance of AT-rich regions for interaction with AsFur.

The interplay between base- and shape-readout modes, allowing degeneracy between Fur consensus sequences within and between bacteria, was also for AsFur important in binding site recognition. Similarity in Fur-DNA interaction mode between bacteria through base readout by conserved Tyrosine and Arginine residues and shape readout by conserved Lysine residue.

In summary, biochemical assays combined with structural modeling has provided further insight into the AsFur-DNA interaction mode.

Acknowledgements:

We would like to acknowledge Dr. Isabelle Michaud-Soret for donating the pDT10 plasmid. We are grateful to Dr. Marcin Pierechod for his assistance in biophysical characterisation attempts and Dr. Adele Williamson for proofreading the manuscript.

Author contributions:

IL had the original project idea. IL and HLP have been involved in the overall planning of the study. KB has been primarily responsible in planning and performing the experiments, such as protein purification, plasmid protection assay, design of oligonucleotides, EMSA assays and thermal stability screenings. KB, IL and HLP have contributed to analysis of biochemical data and writing the paper. IL has been responsible for homology modelling and structural analysis thereof.

References

1. Guerinot, M.L. (1994). *Microbial iron transport.* Annu Rev Microbiol, **48**, 743-772.
2. Hantke, K. (2001). *Iron and metal regulation in bacteria.* Curr Opin Microbiol, **4**(2), 172-177.
3. Hantke, K. (1981). *Regulation of ferric iron transport in Escherichia coli K12: isolation of a constitutive mutant.* Mol Gen Genet, **182**(2), 288-292.
4. Gao, C.H., Wei, W.P., Tao, H.L., Cai, L.K., Jia, W.Z., Hu, L. and Yang, M. (2019). *Cross-talk between the three furA orthologs in Mycobacterium smegmatis and the contribution to isoniazid resistance.* J Biochem, **166**(3), 237-243.
5. Liu, J., Tian, Y., Zhao, Y., Zeng, R., Chen, B., Baishi, H. and Walcott, R.R.P. (2019). *Ferric Uptake Regulator (FurA) is required for Acidovorax citrulli virulence on watermelon.* Phytopathology.
6. Sarvan, S., Yeung, A., Charih, F., Stintzi, A. and Couture, J.F. (2019). *Purification and characterization of Campylobacter jejuni ferric uptake regulator.* Biometals, **32**(3), 491-500.
7. Smiga, M., Bielecki, M., Olczak, M. and Olczak, T. (2019). *Porphyromonas gingivalis PgFur Is a Member of a Novel Fur Subfamily With Non-canonical Function.* Front Cell Infect Microbiol, **9**, 233.
8. Wang, S. *et al.* (2019). *The Ferric Uptake Regulator Represses Type VI Secretion System Function by Binding Directly to the clpV Promoter in Salmonella enterica Serovar Typhimurium.* Infect Immun, **87**(10).
9. Prince, R.W., Storey, D.G., Vasil, A.I. and Vasil, M.L. (1991). *Regulation of toxA and regA by the Escherichia coli fur gene and identification of a Fur homologue in Pseudomonas aeruginosa PA103 and PA01.* Mol Microbiol, **5**(11), 2823-2831.
10. Vasil, M.L. and Ochsner, U.A. (1999). *The response of Pseudomonas aeruginosa to iron: genetics, biochemistry and virulence.* Mol Microbiol, **34**(3), 399-413.
11. Butcher, J., Sarvan, S., Brunzelle, J.S., Couture, J.F. and Stintzi, A. (2012). *Structure and regulon of Campylobacter jejuni ferric uptake regulator Fur define apo-Fur regulation.* Proc Natl Acad Sci U S A, **109**(25), 10047-10052.
12. Miles, S., Carpenter, B.M., Gancz, H. and Merrell, D.S. (2010). *Helicobacter pylori apo-Fur regulation appears unconserved across species.* J Microbiol, **48**(3), 378-386.
13. Danielli, A. and Scarlato, V. (2010). *Regulatory circuits in Helicobacter pylori : network motifs and regulators involved in metal-dependent responses.* FEMS Microbiol Rev, **34**(5), 738-752.
14. Carpenter, B.M., Whitmire, J.M. and Merrell, D.S. (2009). *This is not your mother's repressor: the complex role of fur in pathogenesis.* Infect Immun, **77**(7), 2590-2601.
15. Deng, X., Sun, F., Ji, Q., Liang, H., Missiakas, D., Lan, L. and He, C. (2012). *Expression of multidrug resistance efflux pump gene norA is iron responsive in Staphylococcus aureus.* J Bacteriol, **194**(7), 1753-1762.
16. Grabowska, A.D., Wandel, M.P., Lasica, A.M., Nesteruk, M., Roszczenko, P., Wyszynska, A., Godlewska, R. and Jagusztyn-Krynicka, E.K. (2011). *Campylobacter jejuni dsb gene expression is regulated by iron in a Fur-dependent manner and by a translational coupling mechanism.* BMC Microbiol, **11**, 166.
17. Mills, S.A. and Marletta, M.A. (2005). *Metal binding characteristics and role of iron oxidation in the ferric uptake regulator from Escherichia coli.* Biochemistry, **44**(41), 13553-13559.
18. Bagg, A. and Neilands, J.B. (1987). *Ferric uptake regulation protein acts as a repressor, employing iron (II) as a cofactor to bind the operator of an iron transport operon in Escherichia coli.* Biochemistry, **26**(17), 5471-5477.
19. de Lorenzo, V., Wee, S., Herrero, M. and Neilands, J.B. (1987). *Operator sequences of the aerobactin operon of plasmid ColV-K30 binding the ferric uptake regulation (fur) repressor.* J Bacteriol, **169**(6), 2624-2630.
20. Gao, H., Zhou, D., Li, Y., Guo, Z., Han, Y., Song, Y., Zhai, J., Du, Z., Wang, X., Lu, J. and Yang, R. (2008). *The iron-responsive Fur regulon in Yersinia pestis.* J Bacteriol, **190**(8), 3063-3075.

21. Ochsner, U.A., Vasil, A.I. and Vasil, M.L. (1995). *Role of the ferric uptake regulator of Pseudomonas aeruginosa in the regulation of siderophores and exotoxin A expression: purification and activity on iron-regulated promoters.* J Bacteriol, **177**(24), 7194-7201.
22. Deng, Z. *et al.* (2015). *Mechanistic insights into metal ion activation and operator recognition by the ferric uptake regulator.* Nat Commun, **6**, 7642.
23. Perard, J. *et al.* (2018). *Structural and functional studies of the metalloregulator Fur identify a promoter-binding mechanism and its role in Francisella tularensis virulence.* Commun Biol, **1**, 93.
24. Michaud-Soret, I., Adrait, A., Jaquinod, M., Forest, E., Touati, D. and Latour, J.M. (1997). *Electrospray ionization mass spectrometry analysis of the apo- and metal-substituted forms of the Fur protein.* FEBS Lett, **413**(3), 473-476.
25. D'Autreaux, B., Pecqueur, L., Gonzalez de Peredo, A., Diederix, R.E., Caux-Thang, C., Tabet, L., Bersch, B., Forest, E. and Michaud-Soret, I. (2007). *Reversible redox- and zinc-dependent dimerization of the Escherichia coli fur protein.* Biochemistry, **46**(5), 1329-1342.
26. Hernandez, J.A., Meier, J., Barrera, F.N., de los Panos, O.R., Hurtado-Gomez, E., Bes, M.T., Fillat, M.F., Peleato, M.L., Cavasotto, C.N. and Neira, J.L. (2005). *The conformational stability and thermodynamics of Fur A (ferric uptake regulator) from Anabaena sp. PCC 7119.* Biophys J, **89**(6), 4188-4200.
27. Pohl, E., Haller, J.C., Mijovilovich, A., Meyer-Klaucke, W., Garman, E. and Vasil, M.L. (2003). *Architecture of a protein central to iron homeostasis: crystal structure and spectroscopic analysis of the ferric uptake regulator.* Mol Microbiol, **47**(4), 903-915.
28. Stojiljkovic, I. and Hantke, K. (1995). *Functional domains of the Escherichia coli ferric uptake regulator protein (Fur).* Mol Gen Genet, **247**(2), 199-205.
29. Sheikh, M.A. and Taylor, G.L. (2009). *Crystal structure of the Vibrio cholerae ferric uptake regulator (Fur) reveals insights into metal co-ordination.* Mol Microbiol, **72**(5), 1208-1220.
30. Dian, C., Vitale, S., Leonard, G.A., Bahlawane, C., Fauquant, C., Leduc, D., Muller, C., de Reuse, H., Michaud-Soret, I. and Terradot, L. (2011). *The structure of the Helicobacter pylori ferric uptake regulator Fur reveals three functional metal binding sites.* Mol Microbiol, **79**(5), 1260-1275.
31. Pecqueur, L., D'Autreaux, B., Dupuy, J., Nicolet, Y., Jacquamet, L., Brutscher, B., Michaud-Soret, I. and Bersch, B. (2006). *Structural changes of Escherichia coli ferric uptake regulator during metal-dependent dimerization and activation explored by NMR and X-ray crystallography.* J Biol Chem, **281**(30), 21286-21295.
32. Escolar, L., Perez-Martin, J. and de Lorenzo, V. (1999). *Opening the iron box: transcriptional metalloregulation by the Fur protein.* J Bacteriol, **181**(20), 6223-6229.
33. Calderwood, S.B. and Mekalanos, J.J. (1988). *Confirmation of the Fur operator site by insertion of a synthetic oligonucleotide into an operon fusion plasmid.* J Bacteriol, **170**(2), 1015-1017.
34. Baichoo, N. and Helmann, J.D. (2002). *Recognition of DNA by Fur: a reinterpretation of the Fur box consensus sequence.* J Bacteriol, **184**(21), 5826-5832.
35. Pich, O.Q., Carpenter, B.M., Gilbreath, J.J. and Merrell, D.S. (2012). *Detailed analysis of Helicobacter pylori Fur-regulated promoters reveals a Fur box core sequence and novel Fur-regulated genes.* Mol Microbiol, **84**(5), 921-941.
36. Escolar, L., Perez-Martin, J. and de Lorenzo, V. (1998). *Binding of the fur (ferric uptake regulator) repressor of Escherichia coli to arrays of the GATAAT sequence.* J Mol Biol, **283**(3), 537-547.
37. Escolar, L., Perez-Martin, J. and de Lorenzo, V. (2000). *Evidence of an unusually long operator for the fur repressor in the aerobactin promoter of Escherichia coli.* J Biol Chem, **275**(32), 24709-24714.
38. Pedersen, H.L., Ahmad, R., Riise, E.K., Leiros, H.K., Hauglid, S., Espelid, S., Brandsdal, B.O., Leiros, I., Willassen, N.P. and Haugen, P. (2010). *Experimental and computational characterization of the ferric uptake regulator from Aliivibrio salmonicida (Vibrio salmonicida).* J Microbiol, **48**(2), 174-183.

39. Lavrrar, J.L. and McIntosh, M.A. (2003). *Architecture of a fur binding site: a comparative analysis.* J Bacteriol, **185**(7), 2194-2202.
40. Lavrrar, J.L., Christoffersen, C.A. and McIntosh, M.A. (2002). *Fur-DNA interactions at the bidirectional fepDGC-entS promoter region in Escherichia coli.* J Mol Biol, **322**(5), 983-995.
41. Baichoo, N., Wang, T., Ye, R. and Helmann, J.D. (2002). *Global analysis of the Bacillus subtilis Fur regulon and the iron starvation stimulon.* Mol Microbiol, **45**(6), 1613-1629.
42. Pohl, E., Holmes, R.K. and Hol, W.G. (1999). *Crystal structure of a cobalt-activated diphtheria toxin repressor-DNA complex reveals a metal-binding SH3-like domain.* J Mol Biol, **292**(3), 653-667.
43. White, A., Ding, X., vanderSpek, J.C., Murphy, J.R. and Ringe, D. (1998). *Structure of the metal-ion-activated diphtheria toxin repressor/tox operator complex.* Nature, **394**(6692), 502-506.
44. Thompson, D.K. *et al.* (2002). *Transcriptional and proteomic analysis of a ferric uptake regulator (fur) mutant of Shewanella oneidensis: possible involvement of fur in energy metabolism, transcriptional regulation, and oxidative stress.* Appl Environ Microbiol, **68**(2), 881-892.
45. Sebastian, S., Agarwal, S., Murphy, J.R. and Genco, C.A. (2002). *The gonococcal fur regulon: identification of additional genes involved in major catabolic, recombination, and secretory pathways.* J Bacteriol, **184**(14), 3965-3974.
46. Ahmad, R., Brandsdal, B.O., Michaud-Soret, I. and Willassen, N.P. (2009). *Ferric uptake regulator protein: binding free energy calculations and per-residue free energy decomposition*. Proteins, **75**(2), 373-386.
47. Mey, A.R., Wyckoff, E.E., Kanukurthy, V., Fisher, C.R. and Payne, S.M. (2005). *Iron and fur regulation in Vibrio cholerae and the role of fur in virulence.* Infect Immun, **73**(12), 8167-8178.
48. Jones, M.K. and Oliver, J.D. (2009). *Vibrio vulnificus: disease and pathogenesis.* Infection and immunity, **77**(5), 1723-1733.
49. Wright, A.C., Simpson, L.M. and Oliver, J.D. (1981). *Role of iron in the pathogenesis of Vibrio vulnificus infections.* Infection and immunity, **34**(2), 503-507.
50. León-Sicairos, N., Angulo-Zamudio, U., de la Garza, M., Velazquez-Roman, J., Flores-Villaseñor, H. and Canizalez-Roman, A. (2015). *Strategies of Vibrio parahaemolyticus to acquire nutritional iron during host colonization.* Frontiers in Microbiology, **6**(702).
51. Egidius, E., Wiik, R., Andersen, K., Hoff, K.A. and Hjeltnes, B. (1986). *Vibrio salmonicida sp. nov., a New Fish Pathogen.* International Journal of Systematic and Evolutionary Microbiology, **36**(4), 518-520.
52. Thode, S.K., Baekkedal, C., Soderberg, J.J., Hjerde, E., Hansen, H. and Haugen, P. (2017). *Construction of a fur null mutant and RNA-sequencing provide deeper global understanding of the Aliivibrio salmonicida Fur regulon.* PeerJ, **5**, e3461.
53. Ahmad, R., Hjerde, E., Hansen, G.A., Haugen, P. and Willassen, N.P. (2009). *Prediction and experimental testing of ferric uptake regulator regulons in vibrios.* J Mol Microbiol Biotechnol, **16**(3-4), 159-168.
54. Ericsson, U.B., Hallberg, B.M., Detitta, G.T., Dekker, N. and Nordlund, P. (2006). *Thermofluor-based high-throughput stability optimization of proteins for structural studies.* Anal Biochem, **357**(2), 289-298.
55. D'Autreaux, B., Touati, D., Bersch, B., Latour, J.M. and Michaud-Soret, I. (2002). *Direct inhibition by nitric oxide of the transcriptional ferric uptake regulation protein via nitrosylation of the iron.* Proc Natl Acad Sci U S A, **99**(26), 16619-16624.
56. Cisse, C., Mathieu, S.V., Abeih, M.B., Flanagan, L., Vitale, S., Catty, P., Boturyn, D., Michaud-Soret, I. and Crouzy, S. (2014). *Inhibition of the ferric uptake regulator by peptides derived from anti-FUR peptide aptamers: coupled theoretical and experimental approaches.* ACS Chem Biol, **9**(12), 2779-2786.
57. Davies, B.W., Bogard, R.W. and Mekalanos, J.J. (2011). *Mapping the regulon of Vibrio cholerae ferric uptake regulator expands its known network of gene regulation.* Proc Natl Acad Sci U S A, **108**(30), 12467-12472.

58. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010). *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, **66**(Pt 4), 486-501.
59. Robert, X. and Gouet, P. (2014). *Deciphering key features in protein structures with the new ENDscript server.* Nucleic Acids Res, **42**(Web Server issue), W320-324.
60. Pradhan, L. and Nam, H.J. (2015). *NuProPlot: nucleic acid and protein interaction analysis and plotting program.* Acta Crystallogr D Biol Crystallogr, **71**(Pt 3), 667-674.
61. Hantke, K. (1987). *Selection procedure for deregulated iron transport mutants (fur) in Escherichia coli K 12: fur not only affects iron metabolism.* Mol Gen Genet, **210**(1), 135-139.
62. Tiss, A., Barre, O., Michaud-Soret, I. and Forest, E. (2005). *Characterization of the DNA-binding site in the ferric uptake regulator protein from Escherichia coli by UV crosslinking and mass spectrometry.* FEBS Lett, **579**(25), 5454-5460.
63. Sarvan, S., Butcher, J., Stintzi, A. and Couture, J.F. (2018). *Variation on a theme: investigating the structural repertoires used by ferric uptake regulators to control gene expression.* Biometals, **31**(5), 681-704.
64. Slattery, M., Zhou, T., Yang, L., Dantas Machado, A.C., Gordan, R. and Rohs, R. (2014). *Absence of a simple code: how transcription factors read the genome.* Trends Biochem Sci, **39**(9), 381-399.
65. Chen, Z., Lewis, K.A., Shultzaberger, R.K., Lyakhov, I.G., Zheng, M., Doan, B., Storz, G. and Schneider, T.D. (2007). *Discovery of Fur binding site clusters in Escherichia coli by information theory models.* Nucleic Acids Res, **35**(20), 6762-6777.

Paper II

FEBS openbio

FEBS PRESS science publishing by scientists

Characterization of an intertidal zone metagenome oligoribonuclease and the role of the intermolecular disulfide bond for homodimer formation and nuclease activity

Yvonne Piotrowski, Kristel Berg, David Paul Klebl[†], Ingar Leiros and Atle Noralf Larsen

Department of Chemistry, Faculty of Science and Technology, SIVA Innovation Centre, UiT – The Arctic University of Norway, Tromsø, Norway

Keywords
crystal structure; homodimer; metagenome; nuclease activity; oligoribonuclease; RNA

Correspondence
A. N. Larsen and Y. Piotrowski, Department of Chemistry, Faculty of Science and Technology, SIVA Innovation Centre, UiT – The Arctic University of Norway, Sykehusvegen 23, Tromsø 9037, Norway
Tel: +47 77 64 44 78 (ANL); +47 77 62 33 58 (YP)
E-mails: atle.larsen@uit.no (ANL); yvonne.piotrowski@uit.no (YP)

[†]Present address
School of Biomedical Sciences, Faculty of Biological Sciences, Astbury Centre for Structural and Molecular Biology, University of Leeds, Leeds, LS2 9JT, UK

(Received 6 August 2019, accepted 15 August 2019)

doi:10.1002/2211-5463.12720

The gene encoding MG Orn has been identified from a metagenomic library created from the intertidal zone in Svalbard and encodes a protein of 184 amino acid residues. The *mg orn* gene has been cloned, recombinantly expressed in *Escherichia coli*, and purified to homogeneity. Biochemical characterization of the enzyme showed that it efficiently degrades short RNA oligonucleotide substrates of 2mer to 10mer of length and has an absolute requirement for divalent cations for optimal activity. The enzyme is more heat-labile than its counterpart from *E. coli* and exists as a homodimer in solution. The crystal structure of the enzyme has been determined to a resolution of 3.15 Å, indicating an important role of a disulfide bridge for the homodimer formation and as such for the function of MG Orn. Substitution of the Cys110 residue with either Gly or Ala hampered the dimer formation and severely affected the enzyme's ability to act on RNA. A conserved loop containing His128-Tyr129-Arg130 in the neighboring monomer is probably involved in efficient binding and processing of longer RNA substrates than diribonucleotides.

Oligoribonuclease (Orn) is assumed to originate from eukaryota and is present in almost all eukaryotes [1,2]. In bacteria, and based on sequenced bacterial genomes, *orn* is mainly distributed in beta- and gammaproteobacteria and firmicutes [1]. Orn is a processive 3′–5′ exonuclease that converts small oligoribonucleotides to monoribonucleotides and is important for mRNA decay in cells [3]. Studies in *Escherichia coli* show that Orn is essential for the viability of the bacteria [3], while *Pseudomonas aeruginosa* cells remain viable in the absence of Orn [4]. Interestingly, the human Orn homologue is able to degrade both small single-stranded RNA and DNA molecules *in vitro* and the authors suggest a role of human Orn in cellular nucleotide recycling [5].

In recent years, several studies of Orn in *P. aeruginosa* have broadened the view on the role(s) of Orn in

Abbreviations
CMP, cytidine 5′-monophosphate; MBP, maltose-binding protein; OD600, optical density at 600 nm; PAA, polyacrylamide; PDB, Protein Data Bank; pNP-TMP, *p*-nitrophenyl ester of thymidine 5′-monophosphate; TEV, tobacco etch virus.

bacteria. Depletion of Orn leads to accumulation of small RNA molecules in cells, and these can serve as primers for transcription initiation and lead to global alterations in gene expression [6]. Orn is also demonstrated to play a central role in intracellular turnover of the bacterial second messenger cyclic-di-GMP with implications for bacterial motility, virulence, and biofilm formation [7,8]. A recent study showed that an *orn* mutant of *P. aeruginosa* displayed reduced cytotoxicity mainly by affecting the type III secretion system, further indicating an important role of Orn in bacterial pathogenesis [9]. Furthermore, Chen and coworkers showed that a Δ*orn* mutant became highly susceptible to the antibiotic ciprofloxacin, indicating a novel role in antibacterial drug resistance [10].

Orn is a member of the DEDDh superfamily of exoribonucleases and contains four sequence motifs unique to oligoribonucleases [2]. It is a small protein of approximately 20 kDa and requires divalent cations for nuclease activity, preferably Mn^{2+} [5,11,12]. The *E. coli* enzyme exists as a homodimer in solution [12,13]. Through gel filtration experiments, the human homologue of *E. coli* Orn is in one study shown to be a tetramer in solution [5], whereas another study indicates the enzyme to be a homodimer [12]. The *E. coli* enzyme is characterized as heat-stable, has a half-life of 60 min at 65 °C, and still has residual activity after 10-min incubation at 100 °C [11]. The human enzyme is also quite thermostable and has a temperature optimum for nuclease activity around 50 °C [5]. Datta and Niyogi [14] showed that *E. coli* Orn has a higher affinity for longer chain substrates than smaller substrates, but the reaction rate was inversely proportional to the length of the chain. The nuclease activity of the human Orn homologue is also inversely proportional to the length of the single-stranded substrate [5]. Analysis of the kinetic data of human Orn indicates similar K_m values for short single-stranded RNA and DNA but degrades short RNA about fourfold more efficiently than ssDNA [5].

Crystal structures of Orn show that they are closely related and topologically arranged into an α + β fold containing 5–6 β-strands and 9–10 α-helices (PDB 2GBZ: *Xanthomonas campestris*, PDB 1J9A: *Haemophilus influenzae*, PDB 2IGI: *E. coli*, PDB 3TR8: *Coxiella burnetii*, PDB 5CY4: *Acinetobacter baumannii*). Despite several deposited Orn structures, it is yet unclear how Orn achieves the apparent processive oligoribonucleotide cleaving mechanism, but formation of a stable homodimer is indicated to be important [15]. In the *X. campestris* Orn structure (PDB 2GBZ), it is shown that Orn forms a dimer in the crystal through crystallographic symmetry. From the structural analysis, it was shown that hydrophobic interactions as well as several hydrogen bonds (H-bonds), salt bridges, and a disulfide bond contribute to the formation of a stable homodimer. A very recent publication showing among others an Orn with two uridine molecules bound in the RNA substrate binding site also provides further evidence that hydrophobic interactions, salt bridges, and H-bonds are important for dimer formation [16].

In this study, we have recombinantly produced, characterized, and determined the three-dimensional crystal structure of an arctic marine oligoribonuclease, named MG Orn. We further wanted to investigate the role of the intramolecular disulfide bond connecting the two MG Orn monomers, and our results suggest that this disulfide bond is essential for the formation of a functional homodimer and therefore also the ability of the enzyme to degrade small oligoribonucleotides. We also report the ability of MG Orn to act on longer RNA molecules. Finally, we indicate the involvement of a conserved His-Tyr-Arg loop in the neighboring monomer in binding of these longer (up to 10mer) RNA substrates.

Results

The metagenomic oligoribonuclease (MG Orn) described in this paper consists of 184 amino acid residues. The protein has been recombinantly produced with an N-terminal His_6-MBP-tag followed by a cleavage site for the tobacco etch virus (TEV) protease. After hydrolytic removal of the N-terminal tag, four amino acid residues (Gly-Ser-Phe-Thr) remain at the N terminus of MG Orn due to the recognition site of the protease. Numbering of the amino acid residues within this paper will be according to the protein sequence of MG Orn, that is, excluding the additional amino acid residues of the tag-removal reaction.

Phylogenetic analysis/sequence analysis

A phylogenetic tree based on the maximum likelihood method places MG Orn and the close homologue *Arenicella xantha* Orn in a distinct clade from the other Orn homologues (Fig. 1), as expected from the high sequence identity compared to other Orn homologues (94% versus 50–60%). These two homologues branch out early, just after the shared common gammaproteobacteria ancestor, but their origin is a rather recent event. The statistical bootstrap support value of 100 strongly indicates that MG Orn originates from a species within the *Arenicellales* order, possibly *Arenicellas* or another close relative.

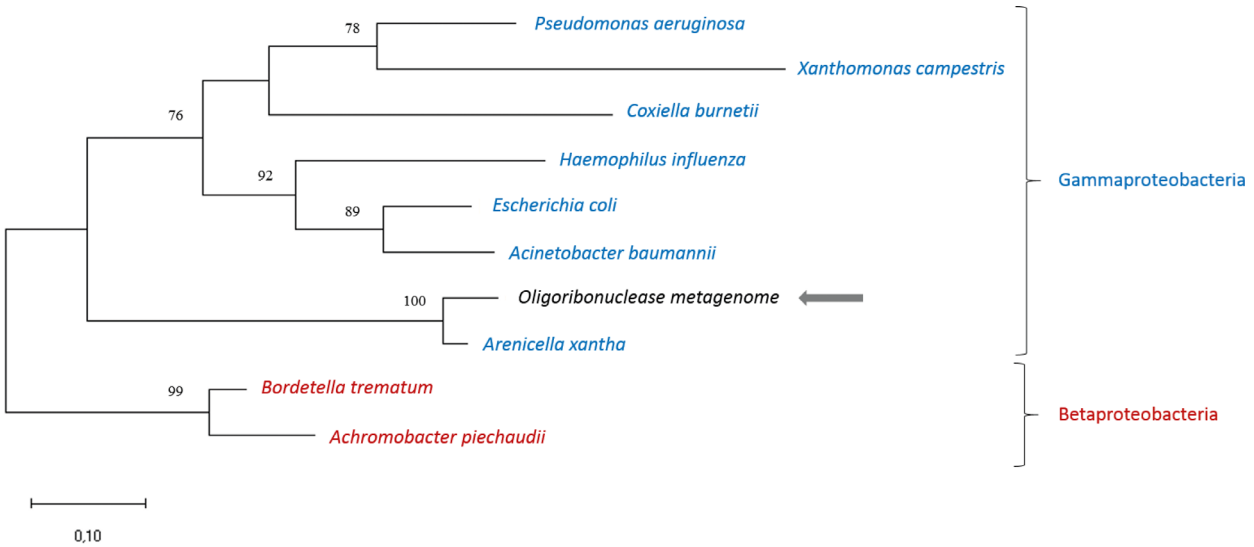


Fig. 1. Phylogenetic relationship of MG Orn protein with selected Orn homologues from gamma- and betaproteobacteria. Node numbers indicate bootstrap support values, with only values above 50 shown. The investigated metagenome sequence is marked with a gray arrow. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The sequences were obtained from GenBank with the following accession numbers: WP_113955167.1 (*Arenicella xantha*), WP_011037314.1 (*Xanthomonas campestris*), RQB22498.1 (*Pseudomonas aeruginosa*), WP_005770781.1 (*Coxiella burnetii*), SST03775.1 (*Acinetobacter baumannii*), WP_021035403.1 (*Haemophilus influenzae*), WP_042004351.1 (*Escherichia coli*), WP_025512385.1 (*Bordetella trematum*), and WP_006218241.1 (*Achromobacter piechaudii*).

Biochemical/biophysical characterization

Size-exclusion experiments were performed to investigate whether MG Orn was monomeric or dimeric in solution. MG Orn eluted as a single peak corresponding to a protein with a molecular weight of 41 kDa (Fig. 2), clearly indicating that MG Orn existed as a dimer in solution. The effect of divalent cations (Mg^{2+}/Mn^{2+}) and pH on the nuclease activity of MG Orn has been determined using the *p*NP-TMP activity assay (see Methods). MG Orn showed an absolute requirement for a divalent metal ion, with Mn^{2+} being clearly preferred over Mg^{2+} (Fig. 3A,B). Furthermore, MG Orn possessed a quite narrow pH range for optimal activity of pH 8–9 with an apparent optimum at pH 8.5 (Fig. 3C).

The effect of various salt concentrations on the nuclease activity of MG Orn was analyzed using the gel-based nuclease activity assay. MG Orn showed significant salt (NaCl) tolerance using 7mer RNA as substrate, and robust hydrolytic activity was observed in the presence of up to 250 mM NaCl. There is still some residual activity observed in the presence of 500 mM NaCl with activity gradually declining up to 2 M NaCl (Fig. 3D). Using the dinucleotide analogue *p*NP-TMP as substrate (*p*NP-TMP assay), MG Orn possessed even higher salt tolerance and little effect on nuclease activity was observed even at 2 M NaCl (results not shown).

To assess the thermal stability of MG Orn, the enzyme was preincubated at different temperatures for 15 min and residual activity was measured using the *p*NP-TMP activity assay. MG Orn was rapidly inactivated at temperatures above 48 °C, with a half-life of about 15 min at 50 °C (Fig. 4).

The 7mer RNA substrate 7mer-62OMe (5′-[FAM] CCCCC[mC]C-3′) was used to investigate the directionality of MG Orn. The substrate contains a methyl group at the 2′ hydroxyl of the ribose at C^6. This 2′-O-methylation blocks ribonuclease function. Nuclease activity proceeding in 3′–5′ direction will result in one 6mer RNA with the fluorophore FAM linked to the 5′ end and one unlabeled cytidine 5′-monophosphate (CMP). If the nuclease proceeds in 5′–3′ direction, the substrate will be cleaved into one FAM-labeled CMP, four unlabeled CMPs, and one unlabeled CDP. MG Orn proceeds in 3′–5′ direction as in all reactions a band just below the RNA substrate (7mer with 2′-O-Me) can be detected, indicating a FAM-labeled 6mer RNA (Fig. 5A). Using 5′ FAM-labeled 5mer RNA as substrate, MG Orn effectively degraded the substrate to monoribonucleotide products, further proving its 3′–5′ directionality (Fig. 5B). MG Orn was also able to degrade short single-stranded DNA (5mer and 10mer) although with much lower efficacy than with RNA (results not shown).

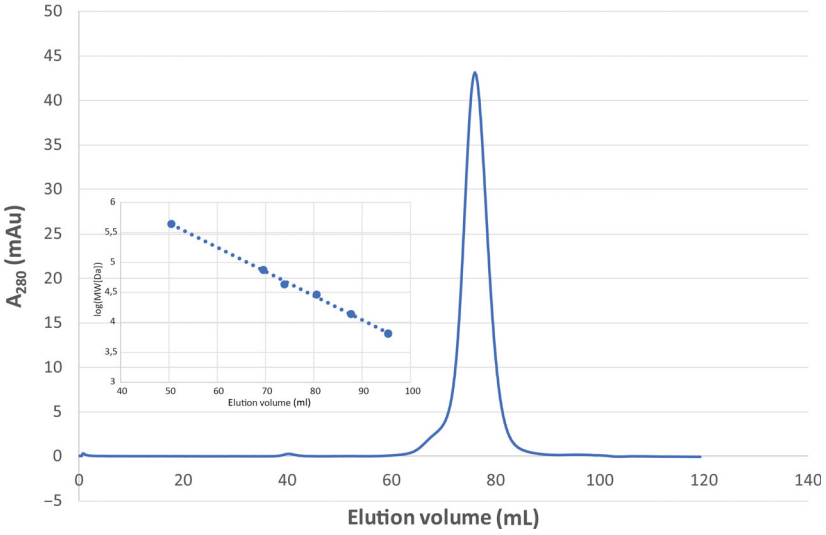


Fig. 2. Size-exclusion chromatography of MG Orn. The inset shows the calibration curve established with Ferritin (440 kDa), Conalbumin (75 kDa), Ovalbumin (43 kDa), Carbonic Anhydrase (29 kDa), RNase A (14 kDa), and Aprotinin (6.5 kDa). The R^2 value of the regression line is 0.997. Based on the calibration curve and the elution volume, the estimated size of MG Orn is 41 kDa.

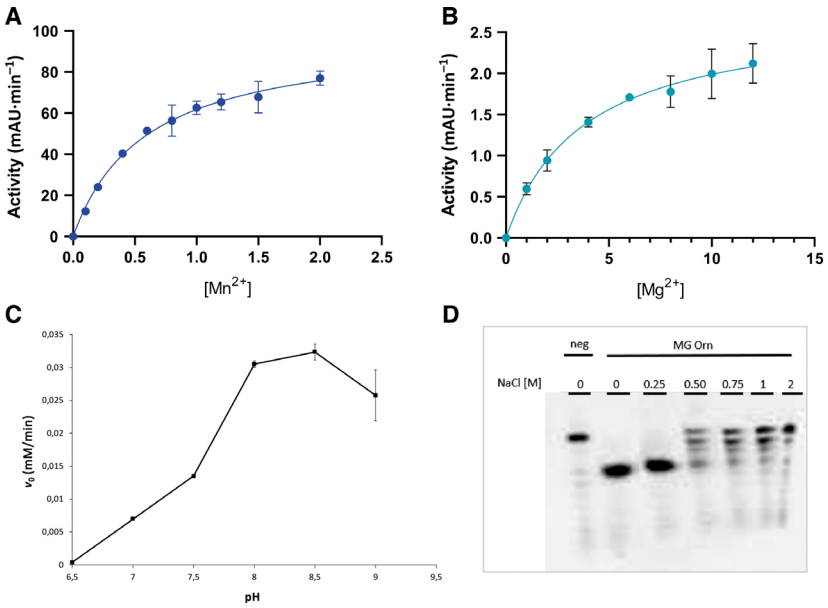


Fig. 3. Effect of Mn^{2+} (A) and Mg^{2+} (B), pH (C), and NaCl (D) on the nuclease activity of MG Orn. The effect of the metal ions and pH on the enzyme activity has been determined using the time-resolved *p*NP-TMP activity assay at 25 °C as described in Methods with 1.1 μg Orn and varying amounts of Mn^{2+} and Mg^{2+} as well as 1.5 μg MG Orn and pH values from 6.5 to 9 using MES (pH 6.5), HEPES (pH 7–7.5), and Tris (pH 8–9) as indicated in A–C. The rate of hydrolysis of *p*NP-TMP at the varying pH values was calculated according to Hamdan *et al.* [34]. Error bars indicate the standard deviation of the measurements. The effect of NaCl in a range of 0–2 M is shown in (D) and has been tested at 25 °C with the gel-based nuclease activity assay with 0.05 μM 7mer RNA substrate and 0.8 μg MG Orn in 50 mM Tris-HCl pH 8.0, 0.2 mg·mL^{-1} BSA, 2% glycerol and 1 mM $MnCl_2$ for 15 min. Samples were analyzed on 20% denaturing PAA gels (8 × 8 cm). Reaction buffer was used as negative control (Neg) instead of protein solution.

Structural analysis

The crystal structure of MG Orn was determined at 3.15 Å resolution, by the molecular replacement method, using the oligoribonuclease from *X. campestris* (PDB: 2GBZ) as a template. A summary of the data collection, refinement, and validation statistics is given in Table 1. The crystal structure of MG Orn contains three monomers in the asymmetric unit. For all three chains, a continuous polypeptide comprising amino acid residues 5–181 (chains A and B) and 5–182

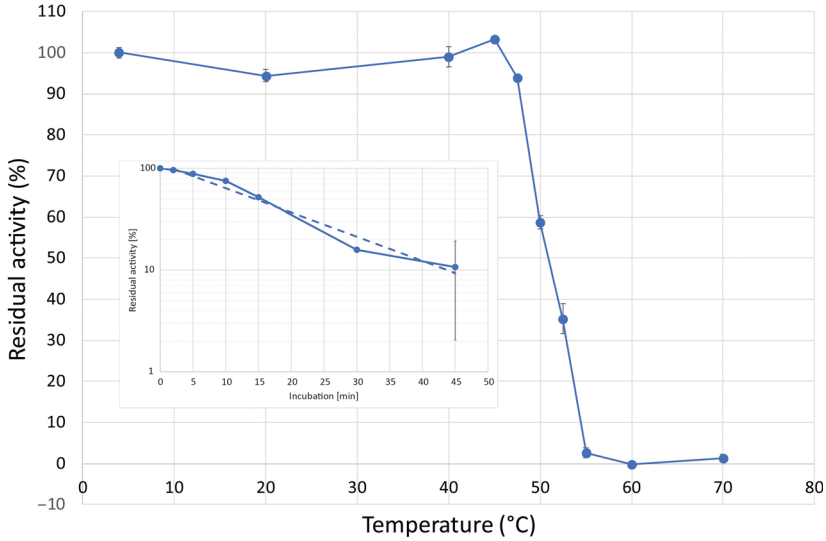


Fig. 4. Temperature stability profile of MG Orn. The enzyme was preincubated at the respective temperature for 15 min and subsequently tested with the time-resolved *p*NP-TMP assay at 25 °C in 50 mM Tris pH 8.0, 200 mM NaCl, 1 mM $MnCl_2$, 1.5 mM *p*NP-TMP, and 1.5 μg MG Orn. Activity measured of the sample preincubated at 4 °C was set as 100% residual activity. Error bars indicate the standard deviation of the measurements. The inset shows the graph determining the half-life of the enzyme at 50 °C.

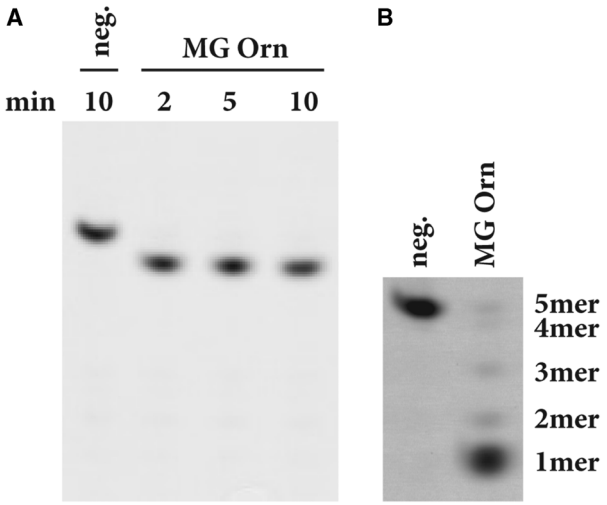


Fig. 5. (A) Directionality of MG Orn. Reactions have been performed with the enzyme assay for determination of directionality at 25 °C in 50 mM Tris pH 8.0, 150 mM NaCl, 1 mM $MnCl_2$, 1 mM DTT, 0.2 mg·mL^{-1} BSA, and 2% glycerol with 25 nM 7mer-62OMe RNA substrate and 0.74 μg MG Orn. Samples have been taken at several points in time as indicated. (B) Degradation of 5′ FAM-labeled 5mer RNA substrate at 25 °C. Reactions have been performed with the gel-based nuclease activity assay with 100 nM substrate and 40 ng MG Orn in 50 mM Tris pH 8.0, 200 mM NaCl, 1 mM $MnCl_2$. After 5 min, the samples were collected and analyzed on a 20% denaturing PAA gel (40 × 20 cm). 'Neg.' indicates the 5′ FAM-labeled 5mer RNA substrate with no enzyme added to the reaction.

(chain C), respectively, was visible in electron density and included in the final model. The three monomers were tightly restrained in refinement and are thus virtually identical with an r.m.s. deviation of 0.001 Å. Each of the three molecules in the asymmetric unit consists of nine α-helices and five β-strands in the order β1-β2-β3-α1-α2-α3-α4-β4-α5-α6-β5-α7-α8-α9. The five-stranded β-sheet forms the core of the protein. It is aligned in the order β3-β2-β1-β4-β5. Strand β2 is antiparallel to the rest. The β-sheet is flanked by the α-helices (Fig. 6A,B). Each monomer of MG Orn contains an Mn^{2+} ion in the active site, most probably resulting from buffers used for protein purification and storage of MG Orn. The metal ion is coordinated by Asp12, Glu14, both located on β1, and Asp163, located on α8 (Fig. 6B, Fig. S1).

A surface representation clearly shows a cavity within the protein comprising of the amino acid residues forming the active site and known as the DEDDh motif, that is, Asp12, Glu14, Asp112, Asp163, and His158 (Figs S1 and S2).

The three-dimensional structure of MG Orn clearly indicates that monomer A and monomer C form a homodimer within the asymmetric unit, while monomer B forms a homodimer with a crystallographic copy of itself. The A–C homodimer is illustrated and focused on in the following discussion (Fig. 7A). The homodimers are connected through an intermolecular disulfide bond (representative electron density shown in Fig. S3). The dimer is further stabilized through salt bridges, H-bonds, and hydrophobic interactions. The active site of each monomer, including coordination of the Mn^{2+} ion, is still accessible and exposed to the solvent in the homodimer.

A structural analysis through PDBePISA [17] highlighted the disulfide bond (C:Cys110 – A:Cys110), two salt bridges (C:Arg130 [NH1] – A:Glu139 [OE1] and C:Glu139 [OE1] – A:Arg130 [NH1]), and a total of 17 H-bonds as important contributors to the formation

Table 1. Data collection, processing, and structure refinement statistics. Values in parentheses are for the outermost shell

Diffraction source	BESSY II, BL 14.1
Wavelength (Å)	0.91841
Temperature (K)	100
Detector	PILATUS
Crystal-to-detector distance (mm)	647.31
Rotation range pr. image (°)	0.1
Total rotation range (°)	94
Space group	$P3_121$
a, *b*, *c* (Å)	108.32, 108.32, 101.33
α, β, γ (°)	90, 90, 120
Mosaicity (°)	0.16
Resolution range (Å)	42.57–3.15 (3.37–3.15)
Total no. of reflections	63 171 (11 289)
No. of unique reflections	12 264 (2200)
Completeness (%)	99.9 (100.0)
Multiplicity	5.2 (5.1)
$<I/\sigma(I)>$	8.2 (1.6)
$R_{p.i.m.}$	0.076 (0.551)
Overall *B* factor from Wilson plot (Å^2)	65.44
σ cutoff	None
Final R_{cryst}	0.2454
Final R_{free}	0.2655
Rotamer outliers (%)	5.34
Clashscore	5.26
No. of non-H atoms	
Protein	4376
Mn	3
Total	4379
R.m.s. deviations	
Bonds (Å)	0.003
Angles (°)	0.70
Average *B* factors (Å^2)	
Overall	82.53
Protein	82.51
Mn	103.38
Ramachandran plot (%)	
Preferred	98.29
Allowed	1.71

of the stable homodimer. Furthermore, there are numerous hydrophobic interactions to stabilize the dimer interface, as a total of 20 amino acid residues from each monomer have more than 50% of their total area toward the interface. This accounts mainly to residues in the β5-α7 region (MG Orn residues 130–145) as well as both termini.

However, several Orn sequences do not contain a Cys residue but an Ala and Gly residue at position 110, respectively (Fig. S2). Table 2 displays putative interactions adding to the formation of the dimer for the different Orn macromolecules. Of the compared structures, MG Orn has the smallest buried surface area and number of hydrophobic contributors upon dimer formation, while differences in number of H-bonds and salt bridges are less pronounced. Interestingly, only some Orn proteins form an intermolecular disulfide bridge (Cys110-Cys110′) crosslinking the monomers in the homodimer believed to be a major contributor to the overall stability of the homodimer. The effect of mutating the Cys residue in MG Orn was thus further investigated.

Structural aspects of RNA substrate binding in MG Orn

In order to investigate the structural basis for the observed *in vitro* processing of longer substrates (5mer, 7mer, and 10mer RNA) shown for MG Orn, we used complexed *E. coli* exonuclease I (ExoI) as a model. ExoI is a three-domain protein, where the N-terminal domain has homology to the DnaQ superfamily. The crystal structure of ExoI in complex with ssDNA was superpositioned to the MG Orn structure (sequence identity of 13.2% for 151 aligned amino acid residues) with an r.m.s. deviation of 2.33 Å. The overlaid structure formed the template for manual fitting of a 5mer RNA molecule into the substrate binding cleft and active site of MG Orn. The modeled 5mer RNA molecule was visualized onto both the electrostatic potential and ConSurf molecular surfaces of the functional dimer of MG Orn (Fig. 7B,C). Nucleotides in the 5′ end of this model appear to be in tight interaction with a conserved sequence patch (His128′-Tyr129′-Arg130′) in the second monomer in the functional homodimer of MG Orn (Fig. 8). This sequence patch is most likely of importance for interaction with longer substrates.

Role of the intermolecular disulfide bond for homodimer formation and nuclease function

The importance of the intermolecular disulfide bond connecting two Orn monomers was demonstrated by comparing the biochemical properties of MG Orn and the two variants OrnC110G and OrnC110A. Following the same procedure as for MG Orn production, OrnC110A and OrnC110G were recombinantly produced in *E. coli* and purified to homogeneity (Fig. S4).

Thermal stability of MG Orn, OrnC110A, and OrnC110G was evaluated by ThermoFluor assay, monitoring changes in hydrophobic fluorescent dye binding upon protein unfolding. All proteins followed the expected shapes of a thermal denaturation profile, displaying an observable melting transition between folded and unfolded states. MG Orn showed a broader thermal unfolding profile compared to the narrower profile of OrnC110A and OrnC110G, possibly

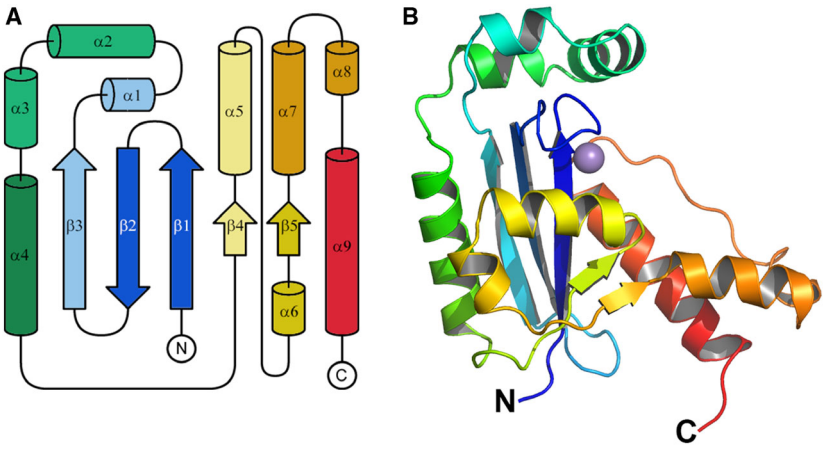


Fig. 6. Topology diagram and monomeric structure of MG Orn. (A) Topology diagram displaying the order of the secondary structure elements. (B) Cartoon representation of a monomer of MG Orn. In both figures, the N and C termini are indicated, and secondary structure elements are colored in rainbow colors ranging from blue to red. The bound Mn^{2+} ion is shown as purple sphere.

reflecting two transitions of the dimeric MG Orn (Fig. S4). The calculated melting temperature (T_m) for the three enzyme variants was similar being 55 °C for MG Orn and 55.6 °C and 54.7 °C for OrnC110A and OrnC110G, respectively (Fig. S4).

Two different assay setups were used to investigate the effect of mutating the intermolecular disulfide bridge in the homodimer. The first assay was monitoring the nuclease activity using *p*NP-TMP, a dinucleotide mimic of a natural nucleic acid, as substrate. The C110G mutation abolished approximately 80% of enzymatic activity on *p*NP-TMP, compared to MG Orn, whereas the C110A mutation almost completely inactivated the enzyme (Fig. 9).

The second assay employed utilized RNA molecules of different length as substrate. While MG Orn displayed a robust exoribonuclease activity on both 7mer and 10mer RNA, OrnC110A showed complete loss of activity on these RNA substrates. The degradation pattern indicates that OrnC110G may have a miniscule capacity to act on both 7mer and 10mer (Fig. 10). Similar results were obtained by increasing the reaction temperature to 37 °C (Fig. S5).

In order to investigate the role of the disulfide bridge after forming the functional dimer, we added increasing amount of the reducing agent DTT to MG Orn at three different temperatures (Fig. 11A,B). MG Orn showed activity in the presence of up to 10 mM DTT at 25 °C and 37 °C. Partial inhibition of activity could be detected only in the presence of 10 mM DTT at 45 °C.

Discussion

Phylogenetic analysis of the Svalbard metagenome Orn (MG Orn) described here strongly indicates that MG Orn originates from a species within the *Arenicellales* order, possibly *Arenicellas* or another close relative *Ar. xantha*. Although the dataset is limited (175 residues), the close relation of MG Orn to *Ar. xantha* Orn and its identification as a gammaproteobacterium is strongly supported by this result.

MG Orn is shown to exist as a dimer in solution in correspondence with other described Orn enzymes such as *E. coli* Orn [12,13]. It has a 3′–5′ directionality and rapidly degrades small oligoribonucleotides to monoribonucleotides. MG Orn prefers Mn^{2+} over Mg^{2+} as divalent cation at pH 8.5 for optimal nuclease activity and possesses a quite broad salt tolerance. This broad salt tolerance, with maximum approximately between 250 and 500 mM NaCl and detectable activity up to 2 M, may arise due to its marine origin and the variable salt tolerance in the littoral zone [18]. As expected, originating from a cold marine habitat MG Orn shows significantly lower thermal stability compared to its mesophilic *E. coli* counterpart. MG Orn showed a $t_{1/2}$ of 15 min at 50 °C, while *E. coli* Orn previously has been shown to still retain 50% residual activity after 60 min at 65 °C [11].

The three-dimensional structure of MG Orn indicates that it functions as a homodimer, where two monomers are connected to each other through an intermolecular disulfide bond (Fig. S3). In addition, several other interactions also contribute to the dimerization interface including hydrophobic interactions, salt bridges, and H-bonds.

In this study, we wanted to investigate the functional role of the intermolecular disulfide bridge (Cys110-Cys110′) connecting the two monomers, and mutated Cys110 to Gly and Ala, amino acid residues naturally occurring at the respective position in other deposited Orn structures. The thermal stability of MG Orn, OrnC110A, and OrnC110G was investigated using a ThermoFluor assay. The ThermoFluor data show a broader thermal unfolding profile for MG Orn

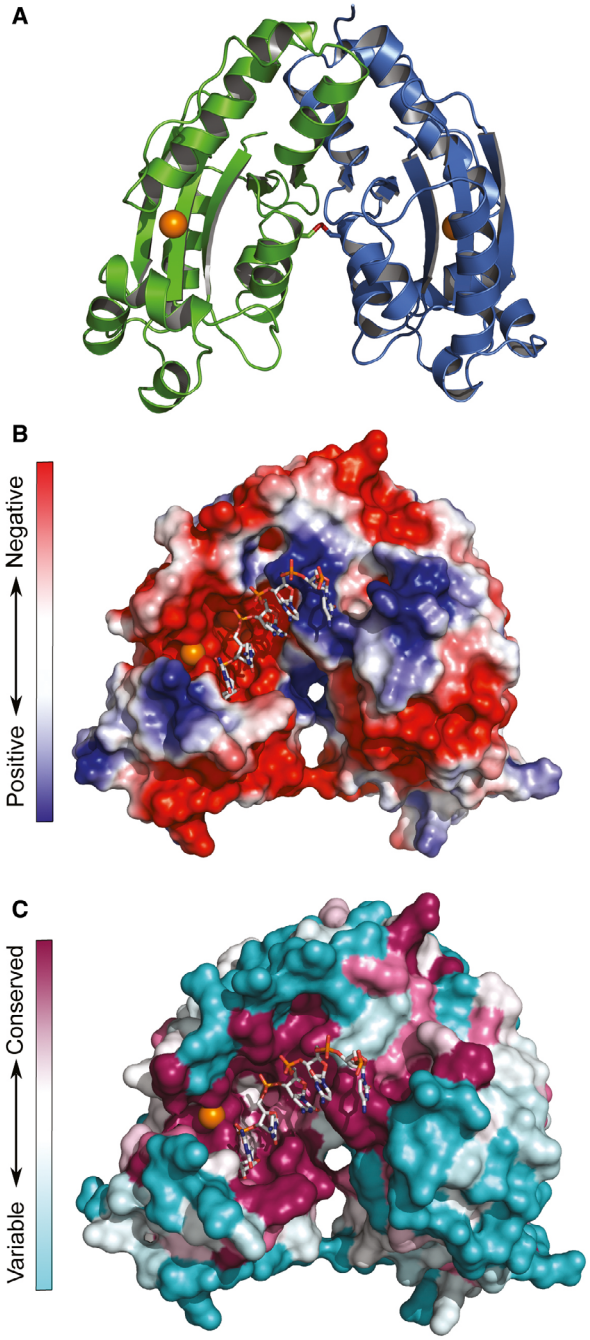


Fig. 7. The functional dimer of MG Orn. (A) Cartoon representation of the functional dimer of MG Orn with the two monomers colored individually. The Mn^{2+} ion is indicated as an orange sphere, and the intermolecular disulfide bond is shown as sticks with sulfur atoms colored orange. (B) Electrostatic potential mapped onto the molecular surface of the MG Orn dimer. The colors range from red (negative potential) to blue (positive potential). (C) Sequence conservation from the ConSurf analysis mapped onto the molecular surface of the MG Orn dimer. The colors range from deep purple (conserved residues) to mint (variable residues). In B and C, the modeled 5mer RNA is shown as a stick model.

compared to OrnC110A and OrnC110G, possibly reflecting two transitions of the dimeric MG Orn. This broader profile is probably due to dissociation of the dimer ahead of monomer unfolding. Dimer disruption allows access of the fluorescence dye to the revealed hydrophobic areas of the interface, leading to an earlier increase in the recorded fluorescence intensity. Thus, the thermal unfolding temperature of MG Orn and its variants is around 55 °C, indicating that mutation of the cysteine involved in dimer formation does not influence the thermal stability of Orn.

Using the dinucleotide substrate mimic *p*NP-TMP, as well as oligoribonucleotides of different lengths, we could show that the mutations severely affected MG Orn's ability to act as an exoribonuclease. These results indicate that residue C110 and its intermolecular disulfide bond are essential for homodimer formation and catalytic function of MG Orn. However, other interactions must also be important for maintaining the dimer formation once it is formed and was further proven by adding the reducing agent DTT to MG Orn. MG Orn showed surprising resilience toward DTT, and inhibition of exoribonuclease activity could only be detected using 10 mM DTT at 45 °C.

Coordination of the Mn^{2+} ion by Asp12 and Glu14, located on β1, and Asp163, located on α8, is also seen by the corresponding amino acid residues in *Cox. burnetii* Orn (PDB code 3TR8). *X. campestris* Orn and *Colwellia psychrerythraea* Orn (PDB code 2GBZ and 6A4A, respectively), on the other hand, each contain one Mg^{2+} ion. The site for binding of Mg^{2+} differs slightly from the binding site for Mn^{2+}. However, in both protein structures the Mg^{2+} ion is coordinated by Asp12 and Glu14 as is the Mn^{2+} in MG Orn and *Cox. burnetii* Orn. Whereas Glu14 shows the same orientation in all four proteins, the orientation of Asp12 depends on the nature of the metal-ion bound that is tilted by 35°. Binding of Mg^{2+} in *X. campestris* Orn and *Col. psychrerythraea* Orn is further supported by Asp112, residing on α5. The respective Asp residue in MG Orn and *Cox. burnetii* Orn is not involved in binding of the Mn^{2+} ion. Asp112, involved in Mn^{2+} binding as mentioned above, is also involved in Mg^{2+} binding in *Col. psychrerythraea*. Compared to Asp163 in MG Orn, this Asp residue is tilted toward the metal ion by ~ 40°.

Lately, a paper describing binding of U-U and *p*NP-TMP has been published [16]. In our study, we have shown that MG Orn efficiently acts on the dinucleotide analogue *p*NP-TMP as well as on 5mer, 7mer, and 10mer RNA substrates. In order to explain the structural basis for the observed *in vitro* processing of 'longer' oligoribonucleotides shown for MG Orn, complexed *E. coli* exonuclease I (ExoI) was used as a

Table 2. Interface analysis of the MG Orn homodimer and its homologues. Performed with PDBePISA [17]

Orn	PDB code	Buried area, Å^2	aa residue at position 110	Number of disulfide bridges	Number of H-bonds	Number of salt bridges	Number of hydrophobic contributors[a]
MG Orn	6RK6	1402.1	Cys	1	17	2	20
Xanthomonas campestris Orn	2GBZ	1753.7	Cys	1	19	4	28
Coxiella burnetii Orn	3TR8	1622.8	Cys	1	19	1	29
Acinetobacter baumannii Orn	5CY4	1602.0	Cys	1	18	6	24
Haemophilus influenzae Orn	1J9A	1638.4	Ala	0	21	4	27
Escherichia coli Orn	2IGI	1608.5	Gly	0	17	2	28
Colwellia psychrerythraea Orn	6A4A	1516.9	Gly	0	14	4	27

[a]Hydrophobic residues with more than 50% of their total area toward the interface according to PDBePISA.

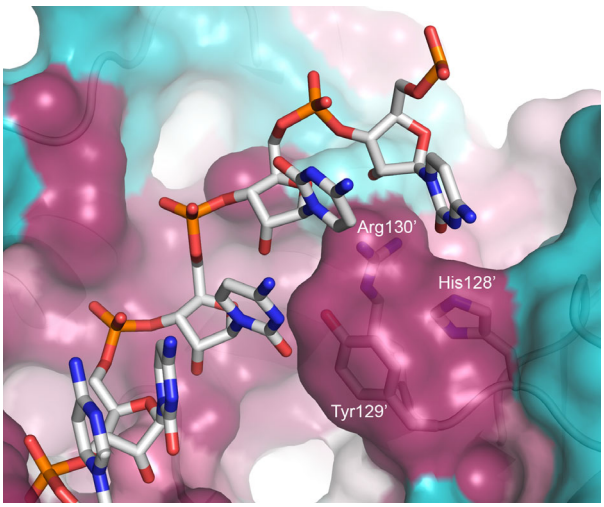


Fig. 8. The 5′ end of the modeled 5mer RNA molecule is in proximity to the conserved sequence patch His128′, Tyr129′, and Arg130′ in the neighboring monomer. The surface is colored based on the ConSurf output with colors ranging from deep purple (conserved residues) to mint (variable residues).

template for manual docking of 5mer RNA into the binding pocket. Although the sequence identity between MG Orn and ExoI is low, the structure-based alignment revealed interesting conservation in the active-site region. Notably, except for His158 (which appears to be in a somewhat flipped-out state in the MG Orn structure), all amino acid residues in the signature DEDDh cluster were structurally conserved between the two structures.

The coordination of the 3′ end of the nucleotide substrate (ssDNA in ExoI; RNA in MG Orn) into the respective active sites is at overlapping positions in MG Orn compared to ExoI. There is a marked difference in polarity between ExoI and MG Orn in the region around the 2′-position of the 3′-sugar unit of the oligonucleotide. Where this area is relatively spacious and nonpolar in ExoI, corresponding to the nature of the deoxyribose in an ssDNA substrate, it is instead rather polar in MG Orn (Thr and Ala in ExoI are replaced with His and Asn in MG Orn). A loop region (His128′, Tyr129′, and Arg130′) from the neighboring monomer in the functional dimer is in proximity to the 5′ region of the modeled RNA molecule. Tyr129′ and Arg130′ have very recently been implicated as important for binding and processing of dinucleotides. When Tyr129′ was exchanged to Ala in *Col. psychrerythraea* Orn, no significant change in hydrolytic activity against *p*NP-TMP was observed, thus indicating that Tyr129′ does not play a vital role in processing of dinucleotide substrates [16]. However, based on modeling of MG Orn with a 5mer RNA, there are clear indications that these residues indeed play an important role in stabilizing the RNA substrate when MG Orn is processing RNA molecules longer than dinucleotides. The importance of these residues is further supported by the fact that the residues in this loop are completely conserved among 150 Orn homologues.

Conclusion

This study highlights the importance of dimer formation for substrate binding and subsequent catalytic action in MG Orn. We show an important role of an intermolecular disulfide bond for the formation of the homodimer, which proves to be essential for the ability of the enzyme to degrade small oligoribonucleotides. We also show the *in vitro* ability of MG Orn to act on 'longer' RNA oligos (5–10mer), probably through the involvement of a conserved sequence loop (His128′, Tyr129′ and Arg130′) in the neighboring monomer when binding longer RNA substrates.

Methods

Bioinformatics

A maximum likelihood (ML) phylogeny (JTT model) was constructed based on a dataset containing nine Orn

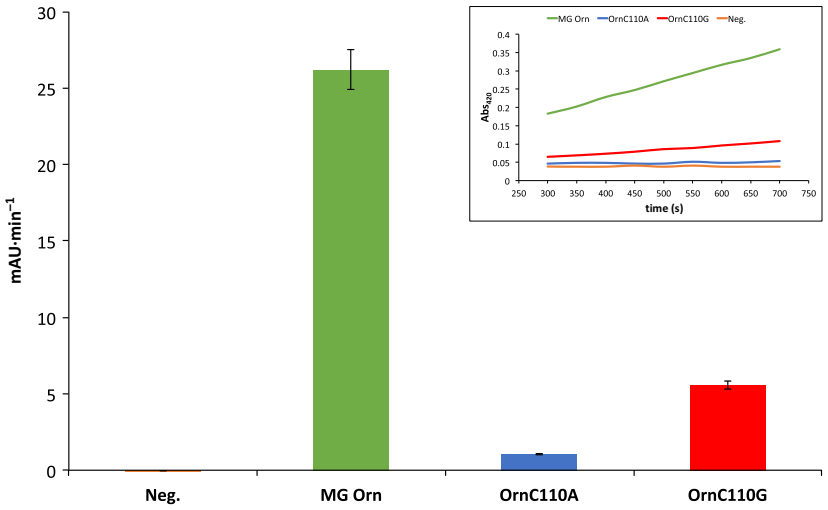


Fig. 9. Enzymatic activity of MG Orn (green), OrnC110A (blue), and OrnC110G (red). RNA degradation was tested with the time-resolved *p*NP-TMP activity assay with 1.5 mM *p*NP-TMP and 2 μg Orn in 50 mM Tris pH 8, 200 mM NaCl, 1 mM $MnCl_2$ at 25 °C. The graph shows the increase in absorbance over time for MG Orn, its mutants, and the negative control (Neg.). The calculated error bars denote the standard deviation between duplicate runs. The inset shows the absorbance (Abs_{420}) of the reaction product plotted against the time for each enzymatic reaction.

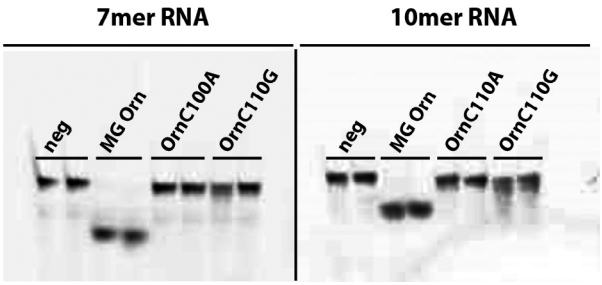


Fig. 10. Nuclease activity of MG Orn and variants on 7mer and 10mer RNA substrates. RNA degradation was investigated using the gel-based nuclease activity assay with 25 nM substrate and 0.8 μg enzyme in 50 mM Tris pH 8.0, 200 mM NaCl, 0.2 mg·mL^{-1} BSA, 2% glycerol, and 1 mM $MnCl_2$ for 15 min at 25 °C. Samples were analyzed on 20% denaturing PAA gels (8 × 8 cm). Reaction buffer replaced protein solution in the negative control (Neg).

sequences, seven from gammaproteobacteria and two from betaproteobacteria (used as outgroup), using the MEGA X software [19,20]. A bootstrap analysis was done to test the stability of nodes, using the ML method and JTT model, with 500 pseudoreplicates [19]. The protein alignment included sequences from *Ar. xantha* WP_113955167.1, *X. campestris* WP_011037314.1, *P. aeruginosa* RQB22498.1, *Cox. burnetii* WP_005770781.1, *Ac. baumannii* SST03775.1, *H. influenzae* WP_021035403.1, *E. coli* WP_042004351.1, *Bordetella trematum* WP_025512385.1, and *Achromobacter piechaudii* WP_006218241.1. Alignment files were generated using CLUSTALW [21].

Cloning of the gene encoding MG Orn

The *mg orn* gene (Fig. S6) has been cloned into the pENTR™/TEV/D-TOPO™ entry vector by Directional TOPO® Cloning from Thermo Fisher Scientific (Waltham, MA, USA) (forward primer: 5′-CACC GTG CCG CAA AAC CCA AAT GTT-3′, reverse primer: 5′- TTA GTT CAT ATC GAG CAG TAT CAG ATT GTT TCG-3′). Positive clones have been confirmed by sequencing analysis. The gene has been subsequently transferred into the destination vector pHMGWA by the LR Clonase reaction using Gateway™ LR Clonase™ II Enzyme Mix (Thermo Fisher Scientific). Positive clones have been confirmed by sequencing analysis. Due to the cloning procedure applied, the *mg orn* gene could be expressed with an N-terminal His_6-MBP-tag followed by a recognition sequence for the TEV protease (TEV protease).

Preparation of mutant constructs

Substitution of Cys110 by Ala and Gly, respectively, was performed using the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA, USA). The pHMGWA plasmid containing *mg orn* was used as a template for single substitutions with synthetic oligonucleotide primers (OrnC110A: forward primer: 5′-GCG GTA ATA GCA TTG CGC AAG ATC GCC G-3′, reverse primer: 5′-CGG CGA TCT TGC GCA ATG CTA TTA CCG C-3′; OrnC110G: forward primer: 5′-GCG GTA ATA GCA TTG GCC AAG ATC GCC G-3′, reverse primer: 5′-CGG CGA TCT TGG CCA ATG CTA TTA CCG C-3′). Both mutations were confirmed by sequencing analysis.

Recombinant expression

For recombinant expression of *mg orn* with an N-terminal His_6-MBP-tag, the plasmid has been transformed into Rosetta 2 (DE3) cells (Merck KGaA, Darmstadt, Germany). Several colonies were picked and incubated in 50 mL LB media containing 100 μg·mL^{-1} ampicillin at 37 °C, 225 r.p.m., overnight. One liter of LB/ampicillin (100 μg·mL^{-1}) medium was inoculated with 20 mL of overnight culture and grown at

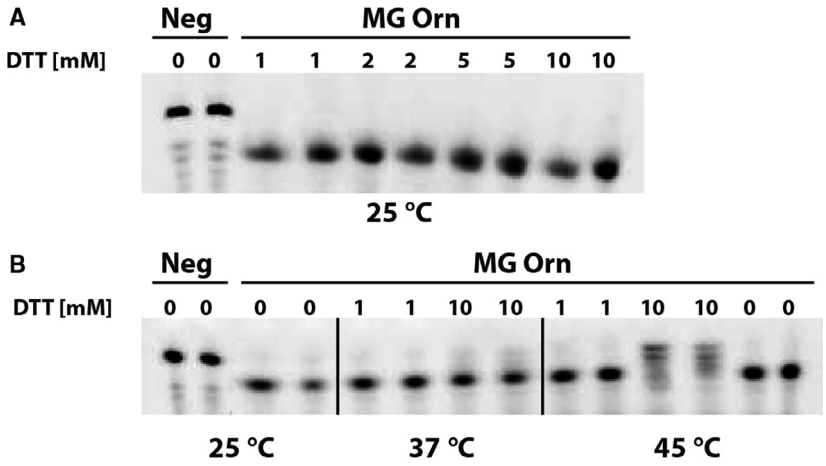


Fig. 11. (A, B) Effect of DTT and temperature on the nuclease activity of MG Orn. Reactions were performed using the gel-based nuclease activity assay with 25 nM 7mer RNA substrate and 0.8 μg MG Orn in 50 mM Tris pH 8.0, 200 mM NaCl, 0.2 mg·mL^{-1} BSA, 2% glycerol, 1 mM $MnCl_2$ with various concentrations of DTT as indicated in A and B at different temperatures (25 °C, 37 °C, and 45 °C) for 15 min and analyzed on 20% denaturing PAA gels (8 × 8 cm).

37 °C, 180 r.p.m., until cell density reached OD_{600} of 0.5. Gene expression was induced by addition of 0.5 mM IPTG, and protein production was carried out for 4 h at 20 °C and 180 r.p.m.

Protein purification of MG Orn, OrnC110A, and OrnC110G

Cell lysis and all purification steps were carried out at 4 °C. Cell pellets from a 1-L cultivation were resuspended in 30 mL lysis buffer [50 mM Tris pH 7.4 (at 25 °C), 300 mM NaCl, 1 mM $MnCl_2$], one protease inhibitor tablet (cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail; Roche, Basel, Switzerland), and 10 μg·mL^{-1} lysozyme. After incubation on ice for 30 min, cell lysis was performed by sonication with the VCX 750 from Sonics® (Newtown, CT, USA) (pulse 1.0/1.0, 20 min, amplitude 25%). Insoluble cell debris was removed by centrifugation (48 384 ***g***, 45 min, 4 °C). The supernatant was filtered through a 0.45-μM membrane and loaded onto a 1-mL HisTrap™ HP column (GE Healthcare, Wood Dale, IL, USA) in buffer A [50 mM Tris pH 7.4 (at 25 °C), 300 mM NaCl, 1 mM $MnCl_2$] and washed in three steps: (a) buffer A2 [50 mM Tris pH 7.4 (at 25 °C), 1 M NaCl, 1 mM $MnCl_2$], (b) buffer A, and (c) 5% buffer B [50 mM Tris pH 7.4 (at 25 °C), 300 mM NaCl, 1 mM $MnCl_2$, 500 mM imidazole]. His_6-MBP-tagged protein was eluted on a 20 mL gradient from 5% to 100% buffer B. Fractions containing the protein were exchanged into buffer A by using a 5-mL HiTrap™ Desalting column (GE Healthcare). TEV protease cleavage was performed overnight with 0.1 mg TEV per 1 mL of eluted fusion protein in 50 mM Tris pH 8.0 (at 25 °C), 0.5 mM EDTA, and 1 mM DTT at 4 °C (TEV protease produced in-house). Protease-treated protein was separated from remaining impurities and the His_6-MBP-tag by reverse affinity chromatography on a 1-mL HisTrap™ HP column with buffer A. Fractions from flow-through and the initial phase of the first gradient at 5% buffer B were collected, added to an equal amount of buffer A, and loaded onto a 1 mL HiTrap™ Blue HP column (GE Healthcare) in buffer A3 [50 mM HEPES pH 7.4 (at 25 °C), 50 mM NaCl, 1 mM $MnCl_2$]. The protein was eluted using 100% buffer B2 [50 mM HEPES pH 7.4 (at 25 °C), 2 M NaCl, 1 mM $MnCl_2$], and eluted peak fractions were analyzed by SDS/PAGE. Fractions containing the protein were exchanged into buffer C [20 mM HEPES pH 7.4 (at 25 °C), 150 mM NaCl, 1 mM $MnCl_2$], up-concentrated to approximately 1 mg·mL^{-1} using Amicon Ultra centrifugal filter units (MWCO 10 kDa, Merck KGaA), and stored at −20 °C for activity and stability assays with 50% (v/v) glycerol.

Size-exclusion chromatography

For crystallization trials and characterization of MG Orn stoichiometry, size-exclusion chromatography was performed. Up-concentrated Orn eluted from the HiTrap™ Blue HP column was run on a HiLoad® 16/600 Superdex® 200 pg column (GE Healthcare) at 1 mL·min^{-1} with 20 mM HEPES pH 7.5 (at 25 °C), 150 mM NaCl, 1 mM $MnCl_2$. Fractions containing MG Orn were collected and concentrated up to 3.6 mg·mL^{-1} for crystallization trials. For molecular weight determination, and thus stoichiometry, standard proteins [Ferritin (440 kDa), Conalbumin (75 kDa), Ovalbumin (43 kDa), Carbonic Anhydrase (29 kDa), RNase A (14 kDa), and Aprotinin (6.5 kDa)] have been applied onto the HiLoad® 16/600 Superdex® 200 pg column (GE Healthcare) at the same flow rate and buffer as mentioned for MG Orn. The known molecular weight and the elution volume of the individual proteins have been used to draw up a calibration curve. Based on this calibration curve and its elution volume, the molecular weight for the MG Orn macromolecule in solution has been calculated.

ThermoFluor assay

The melting temperature (T_m) of Mg Orn, OrnC110A, and OrnC110G, thus the thermal stability of the proteins, was determined by ThermoFluor experiments according to Ref.

[22]. The reactions contained 50 mM HEPES pH 7.5 (at 25 °C), 72 mM NaCl, SYPRO® Orange (Merck KGaA) in a final dilution of 6× and 4 μg of protein. All components were mixed thoroughly in a well of a thin-wall PCR plate (Bio-Rad, Hercules, CA, USA). The wells were sealed with optical-quality sealing tape (Bio-Rad). The volume of the final reaction was 25 μL. A temperature range of 10–90 °C with an increment of 0.3 °C at 3-s intervals has been scanned in the ThermoFluor experiment (excitation at 495 nm, emission at 556 nm).

Protein crystallization and X-ray data collection

Crystallization experiments were performed with a stock solution of purified MG Orn at 3.6 mg·mL^{-1} in 20 mM HEPES pH 7.5 (at 25 °C), 150 mM NaCl, 1 mM $MnCl_2$. Initial crystallization conditions were screened using the vapor diffusion method set up by a Phoenix crystallization robot (Art Robbins Instruments, Sunnyvale, CA, USA). The plates were set up with 60 μL reservoir solution and sitting drops with equal amounts of reservoir solution mixed with protein stock solution in a total drop volume of 0.5 μL. The screens were incubated at 4 °C. Diffraction-quality crystals were found after 2 weeks at a condition containing 0.1 M HEPES pH 7 (at 25 °C) and 5% PEG 8000. Crystals were harvested, transferred through a cryoprotectant solution consisting of the reservoir solution with 30% (v/v) glycerol added, and flash-cooled in liquid N_2. X-ray diffraction data were collected at BL14.1 operated by the Helmholtz-Zentrum Berlin (HZB) at the BESSY II electron-storage ring (Berlin-Adlershof, Germany; Ref.[23]). The data were indexed and integrated by XDS/XSCALE [24], before being merged and scaled by programs in the CCP4 program suite [25]. Data collection and processing statistics are presented in Table 1.

Structure determination, refinement, and analysis

The crystal structure was determined by molecular replacement using PHASER [26] in the PHENIX program package [27] with one monomer of Orn from *X. campestris* (PDB 2GBZ; Ref.[15]) as initial search model. AUTOBUILD [28] traced the full length of one of the three monomers in the asymmetric unit, and this was subsequently fed back into PHASER, resulting in improved map quality. The manual model building was done in COOT [29] interspersed by cycles of refinement using PHENIX.REFINE [30] and converged at final R_{cryst}/R_{free} values of 24.54/26.55. A summary of the refinement statistics is shown in Table 1. The atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank (www.rcsb.org) with the accession code 6RK6. Figures presented in Results section were generated using PYMOL (pymol.org).

The crystal structure of *E. coli* Exonuclease I (ExoI) in complex with ssDNA (PDB code 4JRP) was superpositioned onto the crystal structure of MG Orn using WINCOOT [29]. Based on the corresponding bound ssDNA fragment in ExoI, a 5mer RNA fragment (CCCCC) in the A-form was generated in WinCoot and manually fitted to the MG Orn active site. Shape and charge complementarities were taken into account in the adjustment. Electrostatic surface potentials were calculated through the APBS [31] plugin in PYMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC). Mapping of conserved amino acids onto the molecular surface of a dimer of MG Orn was performed through the ConSurf server [32,33]. Default parameters were used, selecting 150 sequences with sequence identities in the 35–95% range. The output was visualized using PYMOL.

pNP-TMP activity assay

The time-resolved *p*NP-TMP activity assay was performed according to Hamdan *et al.* [34]. In Falcon® 96-well assay plates, up to 2 μg Orn was mixed with reaction buffer [50 mM Tris pH 8 (at 25 °C), 200 mM NaCl, 1 mM $MnCl_2$] and 1.5 mM thymidine 5′-monophosphate *p*-nitrophenyl ester sodium salt (*p*NP-TMP; Merck KGaA) in a total volume of 100 μL. The average change in absorption (420 nm) at 25 °C was measured by monitoring the initial 100 s of each reaction using a SpectraMax® M2^e Microplate Reader (Molecular Devices, San Jose, CA, USA).

Gel-based nuclease activity assay

If not mentioned otherwise, ten microliter reactions contained 25 nM substrate (Table 3), 50 mM Tris pH 8.0 (at 25 °C), 200 mM NaCl, and 1 mM $MnCl_2$. The enzyme was added in varying amounts as indicated with each figure. The enzymatic reaction took place at various incubation times and temperatures as indicated with each figure. Addition of 2.5 μL of denaturing gel loading buffer (95% formamide, 10 mM EDTA, 0.1% xylene cyanol) terminated the reaction after the desired incubation period. Samples were heated at 95 °C for 2 min. Six microliter of each sample was loaded onto denaturing polyacrylamide gels [12% or 20% polyacrylamide/7 M urea (denaturing PAA)], and gel electrophoresis

Table 3. Sequences of RNA substrates employed in the gel-based nuclease activity assay. [FAM], derivative of the fluorophore Fluorescein

RNA substrate	Sequence (5′–3′)	Investigation of
5mer	[FAM]CCCCC	Directionality
7mer	[FAM]CCCCCCC	Effect of salt
		Effect of mutation at position 110
		Effect of reducing agent and temperature
10mer	[FAM]CCCCCCCCCC	Effect of mutation at position 110

was performed in 1× TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM EDTA) at 50 W (40 × 20 cm PAA gels) or 180 V (8 × 8 cm PAA gels) for 1 h 15 min to 1 h 30 min.

Distribution of the degradation products of the endpoint activity assay was monitored by scanning the gels for FAM fluorescence (excitation at 495 nm, emission at 517 nm) in a PharosFX Plus Imager (Bio-Rad). Analysis of the gels was performed with QUANTITY ONE 1-D Analysis Software (Bio-Rad).

Enzyme assay for determination of directionality

Ten microliters of reaction contained 25 nM RNA substrate 7mer-62OMe (5′-[FAM]CCCCC[mC]C-3′), 50 mM Tris pH 8.0 (at 25 °C), 150 mM NaCl, 1 mM $MnCl_2$, 1 mM DTT, 0.2 mg·mL^{-1} BSA, and 2% glycerol. The reaction was started by addition of 0.74 μg protein and incubated at 25 °C for 2, 5, and 10 min. Reactions were stopped by addition of 2.5 μL denaturing gel loading buffer (95% formamide, 10 mM EDTA, 0.1% xylene cyanol) and incubation at 95 °C for 5 min. For the denaturing polyacrylamide gel electrophoresis (12% polyacrylamide/7 M urea, 40 × 20 cm) a sample volume of 6 μL was loaded onto the gel. Gel electrophoresis was performed in 0.5× TBE buffer (44.5 mM Tris, 44.5 mM boric acid, 1 mM EDTA) at 50 W for 1 h 15 min, and the gel was subsequently scanned for FAM-fluorescence (excitation at 495 nm, emission at 517 nm) with the PharosFX Plus Imager (Bio-Rad).

Acknowledgements

This project was funded by the Research Council of Norway (NRC), under Grant No. 174885.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

YP has been primarily responsible in planning the experiments, performed experiments such as testing the directionality of Mg Orn, analyzed data, and contributed to writing the paper. KB has been involved in planning and performed experiments such as mutagenesis of Mg Orn, protein production of OrnC110A and OrnC110G, and comparison studies of the mutants to the wild-type enzyme. She contributed to writing the paper. DPK has been involved in planning and performed experiments such as cloning of Mg Orn, protein production, purification and crystallization of Mg Orn, and basic characterization thereof. IL has been responsible for three-dimensional structure determination of Mg Orn, analysis thereof, and writing the paper. ANL had the original project idea, has been involved in planning experiments and analyzing data, and was primarily responsible for writing the paper.

Data accessibility

Oligoribonuclease (EC 3.1.13.3), PDB code 6RK6.

References

1 Liao H, Liu M and Guo X (2018) The special existences: nanoRNA and nanoRNase. *Microbiol Res* **207**, 134–139.
2 Zuo Y and Deutscher MP (2001) Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Res* **29**, 1017–1026.
3 Ghosh S and Deutscher MP (1999) Oligoribonuclease is an essential component of the mRNA decay pathway. *Proc Natl Acad Sci USA* **96**, 4372–4377.
4 Jacobs MA, Alwood A, Thaipisuttikul I, Spencer D, Haugen E, Ernst S, Will O, Kaul R, Raymond C, Levy R *et al.* (2003) Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA* **100**, 14339–14344.
5 Nguyen LH, Erzberger JP, Root J and Wilson DM III (2000) The human homolog of *Escherichia coli* Orn degrades small single-stranded RNA and DNA oligomers. *J Biol Chem* **275**, 25900–25906.
6 Goldman SR, Sharp JS, Vvedenskaya IO, Livny J, Dove SL and Nickels BE (2011) NanoRNAs prime transcription initiation in vivo. *Mol Cell* **42**, 817–825.
7 Cohen D, Mechold U, Nevenzal H, Yarmiyhu Y, Randall TE, Bay DC, Rich JD, Parsek MR, Kaever V, Harrison JJ *et al.* (2015) Oligoribonuclease is a central feature of cyclic diguanylate signaling in *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA* **112**, 11359–11364.
8 Orr MW, Donaldson GP, Severin GB, Wang J, Sintim HO, Waters CM and Lee VT (2015) Oligoribonuclease is the primary degradative enzyme for pGpG in *Pseudomonas aeruginosa* that is required for cyclic-di-GMP turnover. *Proc Natl Acad Sci USA* **112**, E5048–E5057.
9 Chen G, Zhao Q, Zhu F, Chen R, Jin Y, Liu C, Pan X, Jin S, Wu W and Cheng Z (2016) Oligoribonuclease is required for the type III secretion system and pathogenesis of *Pseudomonas aeruginosa*. *Microbiol Res* **188–189**, 90–96.
10 Chen F, Chen G, Liu Y, Jin Y, Cheng Z, Liu Y, Yang L, Jin S and Wu W (2017) *Pseudomonas aeruginosa* oligoribonuclease contributes to tolerance to ciprofloxacin by regulating pyocin biosynthesis. *Antimicrob Agents Chemother* **61**(3), 1–11.
11 Niyogi SK and Datta AK (1975) A novel oligoribonuclease of *Escherichia coli*. I. Isolation and properties. *J Biol Chem* **250**, 7307–7312.

12 Park AY, Elvin CM, Hamdan SM, Wood RJ, Liyou NE, Hamwood TE, Jennings PA and Dixon NE (2008) Hydrolysis of the 5 '-p-nitrophenyl ester of TMP by oligoribonucleases (ORN) from *Escherichia coli*, *Mycobacterium smegmatis*, and human. *Protein Expr Purif* **57**, 180–187.
13 Zhang X, Zhu L and Deutscher MP (1998) Oligoribonuclease is encoded by a highly conserved gene in the 3'-5' exonuclease superfamily. *J Bacteriol* **180**, 2779–2781.
14 Datta AK and Niyogi K (1975) A novel oligoribonuclease of *Escherichia coli*. II. Mechanism of action. *J Biol Chem* **250**, 7313–7319.
15 Chin KH, Yang CY, Chou CC, Wang AH and Chou SH (2006) The crystal structure of XC847 from *Xanthomonas campestris*: a 3'-5' oligoribonuclease of DnaQ fold family with a novel opposingly shifted helix. *Proteins* **65**, 1036–1040.
16 Lee CW, Park SH, Jeong CS, Cha SS, Park H and Lee JH (2019) Structural basis of small RNA hydrolysis by oligoribonuclease (CpsORN) from *Colwellia psychrerythraea* strain 34H. *Sci Rep* **9**, 2649.
17 Krissinel E and Henrick K (2005) Detection of protein assemblies in crystals. *Lect Notes Comput Sci* **3695**, 163–174.
18 Weslawski JM, Wiktor J, Zajaczkowski M and Swerpel S (1993) Intertidal zone of Svalbard. 1. Macroorganism distribution and biomass. *Polar Biol* **13**, 73–79.
19 Jones DT, Taylor WR and Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* **8**, 275–282.
20 Kumar S, Stecher G, Li M, Knyaz C and Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* **35**, 1547–1549.
21 Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948.
22 Ericsson UB, Hallberg BM, Detitta GT, Dekker N and Nordlund P (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem* **357**, 289–298.
23 Mueller U, Forster R, Hellmig M, Huschmann FU, Kastner A, Malecki P, Puhringer S, Rower M, Sparta K, Steffien M *et al.* (2015) The macromolecular crystallography beamlines at BESSY II of the Helmholtz-Zentrum Berlin: current status and perspectives. *Eur Phys J Plus* **130**(7), 141.
24 Kabsch W (2010) Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125–132.
25 Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235–242.
26 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC and Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674.
27 Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–221.
28 Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung LW, Read RJ and Adams PD (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr* **64**, 61–69.
29 Emsley P, Lohkamp B, Scott WG and Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501.
30 Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH and Adams PD (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**, 352–367.
31 Baker NA, Sept D, Joseph S, Holst MJ and McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* **98**, 10037–10041.
32 Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T and Ben-Tal N (2016) ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* **44**, W344–W350.
33 Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T and Ben-Tal N (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res* **33**, W299–W302.
34 Hamdan S, Bulloch EM, Thompson PR, Beck JL, Yang JY, Crowther JA, Lilley PE, Carr PD, Ollis DL, Brown SE *et al.* (2002) Hydrolysis of the 5'-p-nitrophenyl ester of TMP by the proofreading exonuclease (epsilon) subunit of Escherichia coli DNA polymerase III. *Biochemistry* **41**, 5266–5275.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Representative view of the coordination of the modelled Mn^{2+}-ion in each of the MG Orn monomers. Highlighted amino acids are shown as sticks in atom colours, while the rest of the protein is shown as a cartoon with colouring scheme as for Figure 6. Indicated distances are given in Å.

Fig. S2. Structure-based sequence alignment of MG Orn with other determined structures of Orn homologs. The secondary structure elements of MG Orn are displayed in the top rows, where spirals and arrows depict α-helix and β-strands, respectively. Identical residues are shown in white on red background, while highly conserved residues are shown in red. Cys110 is indicated by a black asterisk, while residues in the conserved DEDDh motif are indicated by blue triangles. PDB identifiers: 2GBZ: *X. campestris* Orn; 3TR8: *C. burnetii* Orn; 5CY4: *A. baumannii* Orn; 1J9A: *H. influenzae* Orn; 2IGI: *E. coli* Orn; 6A4A: *C. psychrerythraea* Orn.

Fig. S3. Representative electron density displaying the region around the intermolecular disulphide bond connecting two MG Orn monomers. The electron density map is displayed at 1.3 times the r.m.s. deviation.

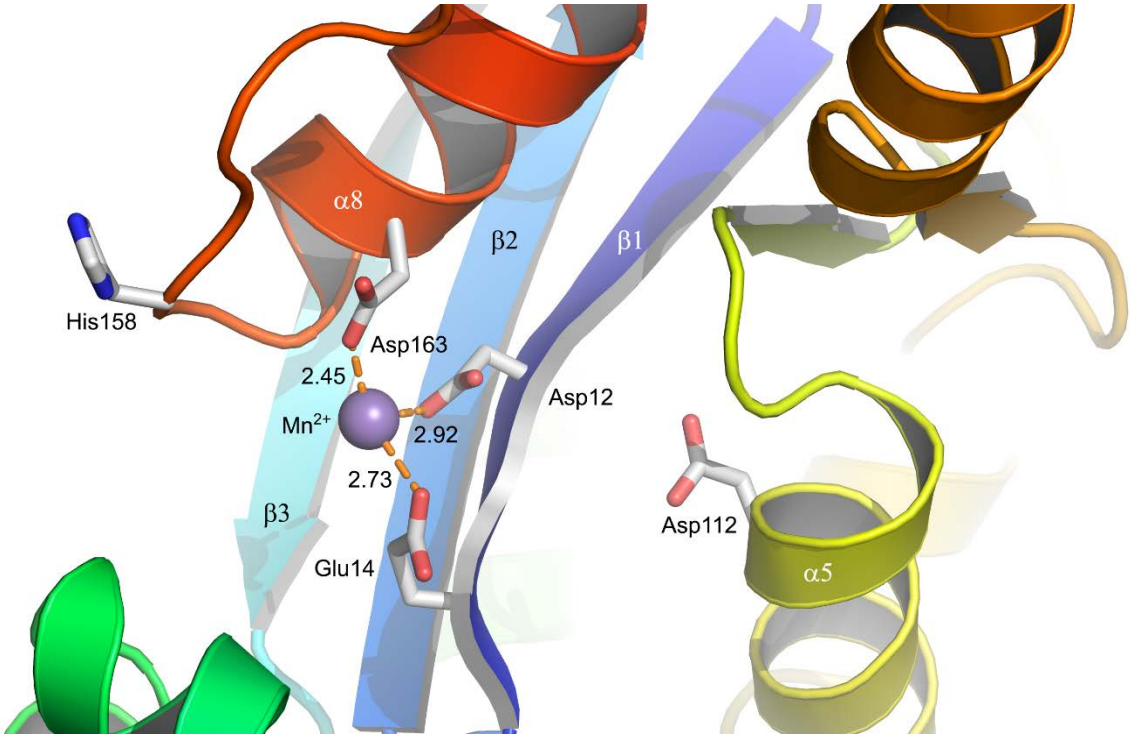
Fig. S4. Purification and thermal stability of MG Orn and its variants OrnC110A and OrnC110G. (A) SDS-PAGE gel showing purified proteins after the final purification step (standard marker Novex Mark 12, Thermo Fisher Scientific). The grey arrow marks the position of the proteins at approximately 21.5 kDa. (B) Thermofluor experiments showing the melting curves of MG Orn (green), OrnC110A (blue) and OrnC110G (red) in 50 mM HEPES pH 7.5. The thermal unfolding was recorded from 10 °C to 90 °C, in increments of 0.3 °C per sec, and the fluorescence signal was plotted as a function of temperature. The table inset sums up the measured *T*m for MG Orn and its variants.

Fig. S5. Nuclease activity of MG Orn and mutants on RNA 7mer and 10mer. RNA degradation was carried out in reaction buffer (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mg/ml BSA, 5 mM DTT, 10% glycerol, 1 mM MnCl2) for 15 minutes reaction at 25 °C and 37 °C and analyzed on 20% PAA gels (8 x 8 cm). Substrate concentration was 0.05 μM and enzyme concentration was 1.16 μM. Control reactions were run without Orn.

Fig. S6. Nucleotide sequence encoding MG Orn.

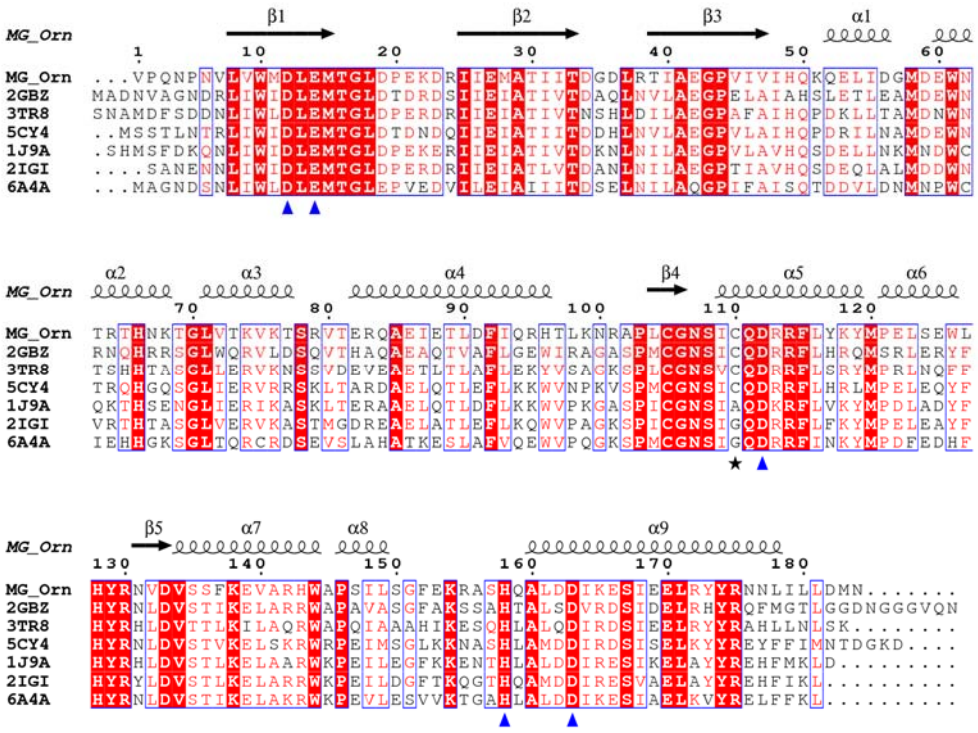
Supporting information

Supplementary Fig. 1



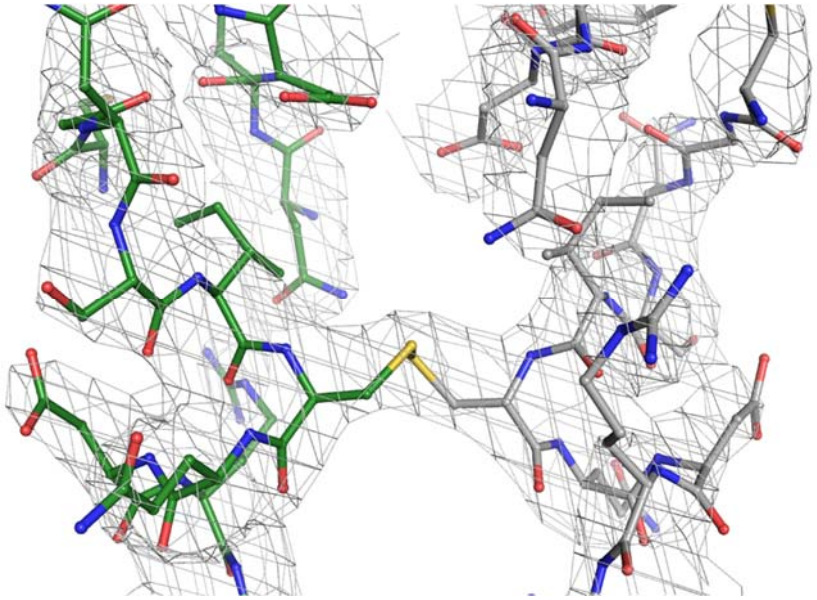
Supplementary Fig. 1. Representative view of the coordination of the modelled Mn^{2+}-ion in each of the MG Orn monomers. Highlighted amino acids are shown as sticks in atom colours, while the rest of the protein is shown as a cartoon with colouring scheme as for Figure 6. Indicated distances are given in Å.

Supplementary Fig. 2



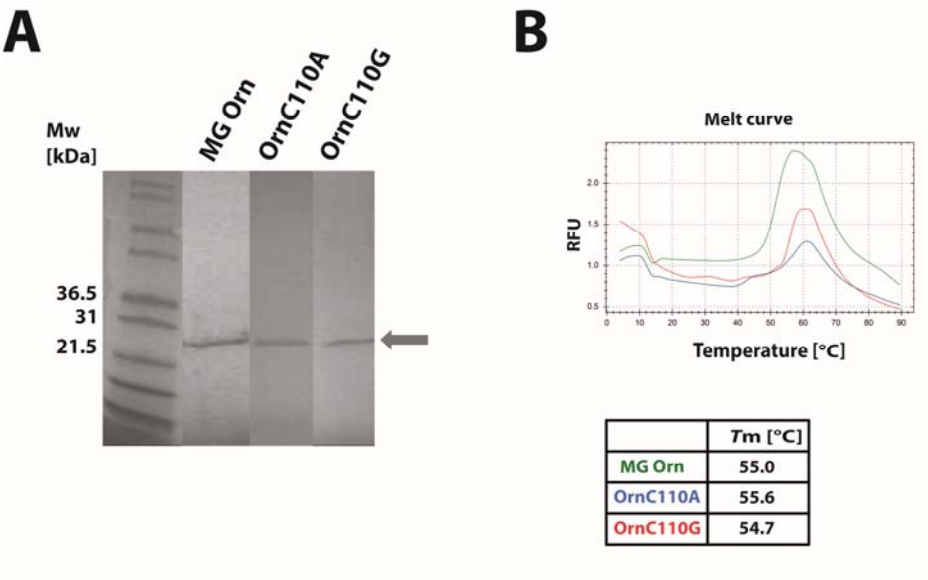
Supplementary Fig. 2. Structure-based sequence alignment of MG Orn with other determined structures of Orn homologs. The secondary structure elements of MG Orn are displayed in the top rows, where spirals and arrows depict α-helix and β-strands, respectively. Identical residues are shown in white on red background, while highly conserved residues are shown in red. Cys110 is indicated by a black asterisk, while residues in the conserved DEDDh motif are indicated by blue triangles. PDB identifiers: 2GBZ: *X. campestris* Orn; 3TR8: *C. burnetii* Orn; 5CY4: *A. baumannii* Orn; 1J9A: *H. influenzae* Orn; 2IGI: *E. coli* Orn; 6A4A: *C. psychrerythraea* Orn.

Supplementary Fig. 3



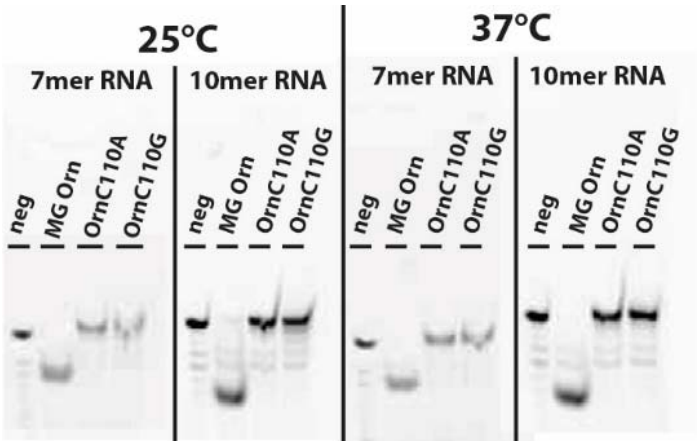
Supplementary Fig. 3. Representative electron density displaying the region around the intermolecular disulphide bond connecting two MG Orn monomers. The electron density map is displayed at 1.3 times the r.m.s. deviation.

Supplementary Fig. 4



Supplementary Fig. 4. Purification and thermal stability of MG Orn and its variants OrnC110A and OrnC110G. (**A**) SDS-PAGE gel showing purified proteins after the final purification step (standard marker Novex Mark 12, Thermo Fisher Scientific). The grey arrow marks the position of the proteins at approximately 21.5 kDa. (**B**) Thermofluor experiments showing the melting curves of MG Orn (green), OrnC110A (blue) and OrnC110G (red) in 50 mM HEPES pH 7.5. The thermal unfolding was recorded from 10 ˚C to 90 ˚C, in increments of 0.3 ˚C per sec, and the fluorescence signal was plotted as a function of temperature. The table inset sums up the measured T_m for MG Orn and its variants.

Supplementary Fig. 5



Supplementary Fig. 5. Nuclease activity of MG Orn and mutants on RNA 7mer and 10mer. RNA degradation was carried out in reaction buffer (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 1 mg/ml BSA, 5 mM DTT, 10 % glycerol, 1 mM $MnCl_2$) for 15 minutes reaction at 25 ˚C and 37 ˚C and analyzed on 20 % PAA gels (8 x 8 cm). Substrate concentration was 0.05 µM and enzyme concentration was 1.16 µM. Control reactions were run without Orn.

Supplementary Fig. 6

GTGCCGCAAAACCCAAATGTTCTCGTCTGGATGGATCTCGAAATGACGGGGCTAG
ACCCAGAAAAAGACCGCATTATTGAAATGGCAACCATCATTACCGACGGCGATTT
GCGCACTATTGCTGAGGGGCCTGTGATTGTGATTCATCAAAAGCAAGAGCTCATC
GATGGAATGGACGAATGGAATACGCGTACCCATAACAAAACAGGTTTAGTCACT
AAGGTAAAAACCAGTCGTGTGACCGAACGTCAGGCCGAAATTGAAACTCTGGAT
TTTATTCAACGGCACACGCTCAAAAATCGCGCACCACTTTGCGGTAATAGCATTT
GCCAAGATCGCCGTTTTTTATACAAGTACATGCCTGAATTAAGCGAATGGCTGCA
TTATCGCAACGTAGATGTAAGCTCGTTTAAAGAAGTGGCCAGACATTGGGCGCCT
AGCATTCTCTCAGGTTTCGAAAAACGCGCATCGCATCAAGCTTTGGACGACATCA
AAGAATCTATTGAAGAGCTGCGTTACTACCGAAACAATCTGATACTGCTCGATAT
GAACTAA

Supplementary Fig. 6. Nucleotide sequence encoding MG Orn.

Paper III

Extremophiles (2019) 23:305–317
https://doi.org/10.1007/s00792-019-01082-y

ORIGINAL PAPER

Temperature adaptation of DNA ligases from psychrophilic organisms

Kristel Berg[1] · **Ingar Leiros**[1] · **Adele Williamson**[1]

Received: 4 December 2018 / Accepted: 15 February 2019 / Published online: 2 March 2019

Abstract
DNA ligases operating at low temperatures have potential advantages for use in biotechnological applications. For this reason, we have characterized the temperature optima and thermal stabilities of three minimal Lig E-type ATP-dependent DNA ligase originating from Gram-negative obligate psychrophilic bacteria. The three ligases, denoted Vib-Lig, Psy-Lig, and Par-Lig, show a remarkable range of thermal stabilities and optima, with the first bearing all the hallmarks of a genuinely cold-adapted enzyme, while the latter two have activity and stability profiles more typical of mesophilic proteins. A comparative approach based on sequence comparison and homology modeling indicates that the cold-adapted features of Vib-Lig may be ascribed to differences in surface charge rather than increased local or global flexibility which is consistent with the contemporary emerging paradigm of the physical basis of cold adaptation of enzymes.

Keywords ATP-dependent DNA ligase · Psychrophile · Enzyme activity · Temperature optima

Introduction

DNA ligases are DNA-joining enzymes essential for survival of all organisms, due to their critical roles in DNA replication and repair. Using ATP or NAD^+ as a cofactor, DNA ligases catalyze the formation of a phosphodiester bond between the 5′ phosphate of one DNA strand and the hydroxyl group at the 3′ end of the other DNA strand, producing an intact sugar–phosphate backbone. The enzymatic reaction mechanism can be divided into three nucleotidyl-transfer steps (Ellenberger and Tomkinson 2008); the first involves the activation of the enzyme through a nucleophilic attack by a lysine residue to the adenosine cofactor ATP or NAD^+, releasing nicotinamide mononucleotide for NAD-dependent ligases (NDLs) or di-phosphate in the case of ATP-dependent ligases (ADLs). Next, the nucleophilic 5′- phosphate of the DNA attacks the phosphoramide bond to form an adenylated-DNA intermediate. The final step involves attack of the 3′-nucleophilic hydroxyl group on the new pyrophosphate bond, forming a phosphodiester bond between the 5′ and 3′ positions of the DNA and releasing the AMP. All the three chemical steps depend on a divalent cation, which is usually Mg^{2+} or in some cases Mn^{2+}.

DNA ligases are divided into two main classes based on the cofactor required in step 1 of the enzymatic reaction. The ADLs use ATP and are found in all phylogenetic kingdoms, with eukaryotes, archaea, and many viruses possessing at least one ADL that is essential for DNA replication (by joining Okazaki fragments), and some encode multiple forms with dedicated roles in DNA repair (Ellenberger and Tomkinson 2008; Plocinski et al. 2017; Shuman and Glickman 2007). NDLs, meanwhile, are found almost exclusively in bacteria, where they function in both replication and repair (Dwivedi et al. 2008; Wilkinson et al. 2001). In the cases, where accessory ADLs are identified in bacteria, it is always in addition to the essential NDLs (Pitcher et al. 2007b).

Since the first X-ray crystal structure of an ADL was solved two decades ago from bacteriophage T7 (Subramanya et al. 1996), numerous structural analyses of bacterial, archaeal, and eukaryotic ADLs have followed (Nishida et al. 2006; Pascal et al. 2004, 2006; Kim et al. 2009; Petrova et al. 2012; Akey et al. 2006; Kaminski et al. 2018; Shi et al. 2018; Williamson et al. 2014, 2018), and the wide variety

Communicated by L. Huang.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00792-019-01082-y) contains supplementary material, which is available to authorized users.

✉ Adele Williamson
adele.k.williamson@uit.no

[1] Present Address: Department of Chemistry, The University of Tromsø- The Arctic University of Norway, 9019 Tromsø, Norway

of domains and gene arrangements between the different classes of ligases has become evident. Crystallographic studies of bacteriophage T7 (Doherty and Wigley 1999; Subramanya et al. 1996) revealed a common core architecture of two essential catalytic core domains: the adenylation domain (AD) directly involved in catalysis and the site of step 1 enzyme adenylation and the smaller oligonucleotide/oligosaccharide binding domain (OB) that is also required for activity (Doherty and Suh 2000; Doherty and Wigley 1999). These core catalytic domains include six conserved motifs (I, III, IIIa, IV, V, and VI) which are involved in one or more steps of the ligation pathway (Shuman 2009). The AD and OB domains are connected by a flexible linker that allows them to reorient during DNA binding. An additional N-terminal DNA-binding domain has been described in the larger ADLs active in DNA replication in Eukarya and Archaea, and additional enzymatic domains with end-repair functions are appended to the large LigD enzymes involved in bacterial non-homologous end joining (Pitcher et al. 2007a). The Lig E group of ADLs, found predominantly in Gammaproteobacteria, do not use appending domains or unstructured loops for DNA binding or possess additional domains with independent enzymatic function; thus, these small ligases serve as a model for the minimal functional unit of the ATP-dependent ligases. The ADL from the marine psychrophile *Psychromonas* sp. strain SP041 (Psy-Lig) is the smallest DNA ligase that has been structurally studied, being 41 residues shorter than the minimal ChlV-Lig protein (Williamson et al. 2014). Recent structure–function analysis of Psy-Lig and the closely related Ame-Lig demonstrated a novel mode of ligase engagement with its DNA substrate that relies on well-ordered side-chain contacts on the surface of the conserved domains, rather than re-ordering of flexible loop regions to achieve encirclement of the DNA duplex as was previously observed for minimal viral ligases (Nair et al. 2007; Williamson et al. 2018). All Lig E-type ADLs have strong predictions for N-terminal leader sequences which are expected to direct them to the periplasm, and the demonstrated increase in activity and solubility when this predicted leader was not included in recombinantly produced *Aliivibrio salmonicida* (hereafter referred to as Vib-Lig) supports such signal processing (Williamson and Pedersen 2014). Although the biological role of such putatively periplasmic ligases and the source of ATP outside the cell remain to be determined, functions in competence and uptake of extracellular DNA have been proposed (Magnet and Blanchard 2004).

In the present study, we have characterized the temperature optima and thermal stability of Psy-Lig and Vib-Lig, both of which originate from obligate psychrophiles, along with a third homolog from *Pseudoalteromonas artica* (hereafter Par-Lig), isolated from sandy beach sediment on the Arctic island of Svalbard (Al Khudary et al. 2008). This builds on the previous work by Georlette et al. who conducted biophysical analyses and biochemical comparisons of larger, more complex NDLs from species spanning a range of growth temperature optima (Georlette et al. 2000; Georlette et al. 2003).

Living and thriving at low temperatures require that both enzyme kinetics and protein stability are adapted accordingly. It is now widely accepted that structural differences between cold-active enzymes and their mesophilic counterparts enable high specific activity at low temperatures, with a lower energy cost (D'Amico et al. 2002; Feller 2003; Struvay and Feller 2012). The physical origin of decreased temperature optima imparted by these structural changes is an active area of contemporary investigation (Åqvist et al. 2017; Arcus et al. 2016; Isaksen et al. 2016; Saavedra et al. 2018; van der Kamp et al. 2018), but it is generally observed that improved catalytic efficiency is accompanied by a reduced thermal stability and weaker substrate affinity, compared to thermophiles and mesophiles (Struvay and Feller 2012). For this reason, we have also carried out in silico comparisons of these Arctic-derived ADLs with mesophile-derived counterparts from human pathogens.

DNA ligases adapted to low temperatures offer novel potential advantages for use of these enzymes in biotechnological applications. Recently, the thermolability of a cold-adapted DNA ligase was used to develop a novel temperature-sensitive vaccine for tularemia (Duplantis et al. 2011), showing great potential in the biomedical sciences and other applications, where bacterial growth control is crucial. Furthermore, the enzymatic activity performed by DNA ligases in DNA replication and repair makes them useful tools in molecular biology and biotechnology applications, such as genetic engineering and next-generation DNA-sequencing technologies (Chambers and Patrick 2015; Shuman 2009; Tanabe et al. 2015). Cold-adapted enzymes have a potential advantage over mesophilic homologs by increasing yields of product at low temperatures, while suppressing contaminating nuclease activity. Finally, should the cold-active ligases be highly active, protocols may be carried out with smaller amounts of enzyme, due to better activity rates. In particular, short base-pair overhangs, i.e., 'sticky ends' generated by many restriction enzymes, will be stabilized due to the low melting temperature of short tracts of base-pairing involved. For these reasons, improving our understanding of temperature adaptation and identification of psychrophilic traits that could be used directly, or reverse-engineered into commercial ligase scaffolds has important biotechnological applications.

Methods

Protein expression and purification

ADLs from *Psychromonas* spp. strain SP041 (Psy-Lig) and *Aliivibrio salmonicida* (Vib-Lig) were expressed and

purified as described previously (Williamson and Pedersen 2014; Williamson et al. 2014). The gene encoding the Lig E-type ADL from *Pseudoalteromonas artica* (WP_010555135; Par-Lig), without the leader peptide, was synthesized by *Life Technologies* as the mature His-tagged, TEV-cleavable form with codon optimization for *E. coli* and supplied in the donor vector pDONR221. Transfer to the pHMGWA vector was done using Gateway® cloning (Thermo Fisher), and all steps including expression of the MBP fusion, purification, and tag removal were carried out as described for Psy-Lig and Vib-Lig.

Enzyme assays

Gel-based endpoint assays were carried out as described previously using 20 nt + 20 nt oligomers to form 40 nt product (Williamson et al. 2014, 2018). Details of substrate preparation are given in Table S1. Reactions contained 80 nM substrate, 1 mM ATP, 10 mM $MgCl_2$, 10 mM DTT, 50 mM NaCl, 50 mM Tris–HCl pH 8.0. Final enzyme concentrations of 2.5 µM and 100 µM were used for nicked and cohesive substrates, respectively. Enzymatic activity was detected by conversion of the FAM-labeled 20 nt substrate oligonucleotide into a 40 nt product, resolved by denaturing electrophoresis, detected by fluorescence on a Pharos FX Plus imager (Bio-Rad), and quantified by band intensity using the software Image J (Schneider et al. 2012). The extent of ligation activity was calculated from the ratio of these bands and expressed as a percentage. The temperature dependence of ligase activity was investigated by assaying for 15 min at temperatures between 5–55 °C for nicked substrates and 5–35 °C for cohesive substrates. Reactions were allowed to equilibrate for 1 min to the assay temperature, and then, the assay was started by addition of the enzyme.

Differential scanning calorimetry

Differential scanning calorimetry (DSC) experiments were carried out using an N-DSC III differential scanning calorimeter (Calorimetry Sciences Corporation). Purified ligases with concentrations of 1–2 mg ml^{-1} were extensively dialyzed against 50 mM HEPES pH 8.0, 100 mM NaCl to ensure complete equilibration. The enzymes were filtered through a 0.2 µm syringe filter (Millipore, Billerica, USA) and degassed for approximately 15 min before being loaded into the sample cell. The dialysis buffer was used as reference for baseline subtraction. Data analysis was performed using the program NanoAnalyse 2.4 (TA instruments). For each protein sample scanned, the corresponding buffer baseline was subtracted, and the data were normalized to the molar protein concentration calculated from the absorbance at 280 nm after dialysis and filtration. The calorimetric enthalpy was determined directly from the experimental data, and a theoretical two-state model was fitted using the routines provided in the program for determination of the van't Hoff enthalpy.

Thermofluor

Thermal denaturation of the purified ADLs with different buffers was examined by the thermofluor assay as described previously (Ericsson et al. 2006). Briefly, 5 µl of protein (1.0–1.5 mg ml^{-1}) was mixed with 1 µl of 300 × Sypro-Orange, 12.5 µl of 50 mM HEPES pH 8.0, 200 mM NaCl, added to the wells of a 96-well PCR plate (Bio-Rad) and sealed with Microseal® 'B' Adhesive Seals from Bio-Rad. Melting curves were recorded from 20 to 90 °C in increments of 0.3 °C per s using a MiniOpticon Real-Time PCR System with both FAM and HEX dye channels selected. T_m was determined using the supplied instrument software and monitoring the fluoresce of the HEX channel.

Sequence comparison

The amino acid sequences of Par-Lig, Psy-Lig, and Vib-Lig were aligned with the Lig E sequence from *Vibrio cholera* (Vch-Lig; gil147674166). N-terminal leader sequences were predicted using SignalP 4.1 and omitted from further analyses (http://www.cbs.dtu.dk/services/SignalP/) (Petersen et al. 2011). The ClustalW alignment tool in BioEdit was used to determine sequence identities and similarities. Conserved domains were analyzed by Pfam protein families' database at EMBL-EBI (http://pfam.xfam.org).

Homology modeling and analysis

Homology models of Vib-Lig, Par-Lig, and Vch-Lig were built based on the deposited crystal structure of Psy-Lig (4D05; (Williamson et al. 2014)). The sequences were uploaded to the Swiss-Model homology modeling server (Biasini et al. 2014). The A-chain of the deposited structure of Psy-Lig was selected as a modeling template for all modeled structure, as it has overall superior quality than the B-chain (with lower overall B-factor and amino acid residues generally better defined in electron density).

HBPLUS Hydrogen Bond Calculator v 3.2 (McDonald and Thornton 1994) was used to calculate hydrogen bonds in all PDB files. The hydrogen bonds included were those fulfilling the criteria for parameters donor (D), acceptor (A), acceptor antecedents (AA), and calculated hydrogen (H): maximum distance for D–A, 3.5 Å and H–A, 2.5 Å and minimum angle for D–H–A, D–A–AA, and H–A–AA of 90°. Ion-pair-interactions were investigated using the WHAT IF Web Interface (http://swift.cmbi.ru.nl/servers/html/index.html) (Vriend 1990), where interatomic distances between the side chains of the negatively charged Asp and Glu, and

the positively charges Arg, Lys, and His were tabulated with respect to being < 4 Å and < 6 Å. The APBS plugin in Pymol was used to estimate electrostatic surface potentials (Dolinsky et al. 2007).

Results

Temperature optimum and thermal stability

The aim of this study was to understand the determinants of low-temperature adaptation among DNA ligases. We chose to investigate the temperature optimum and thermal stability of Lig E ADLs from *Psychromonas* spp. strain SP041, *Aliivibrio salmonicida*, and *Pseudoalteromonas artica*, delineated Psy-Lig, Vib-Lig, and Par-Lig, respectively, as these represent psychrophilic species of bacteria isolated from a consistently low-temperature environments (Al Khudary et al. 2008; Egidius et al. 1986) (Table).

To analyze the temperature optima for ligase activity, gel-based endpoint assays were performed, both with single-nicked and overhanging substrates. Nick-sealing activity was measured by temperature intervals of 5 °C, ranging from 5 to 60 °C, or until the activity was abolished. For ligation of single-nicked substrates (Fig. 1a), there is a sharp peak of more than 50% ligation activity at around 20 °C for Vib-Lig, quickly declining to 10% activity at 30 °C, whereas the activity of Psy-Lig and Par-Lig increases with temperature from 15 °C up to an optimum of 35–40 °C, above which a sharp decline is observed. Although all ligases were cloned from psychrophilic organisms with similar growth temperatures, T_{opt} of their ligases for nicked substrates are different (Table 1).

The characterized ligases Psy-Lig, Par-Lig, and Vib-lig show a similar and relatively broad temperature optimum on the overhang substrate tested, with approximately 60–80% ligation activity from 5 to 25–30 °C, followed by a sharp decline at higher temperatures. As they all show better activity on overhang breaks at lower temperatures, we suggest that substrate stability rather than enzyme activity is the driving feature here. However, the enzymatic reaction will work very slowly at the low temperature, requiring a longer incubation time.

DSC experiments were performed to obtain a complete thermodynamic profile of the protein unfolding process of Psy-Lig, Par-Lig, and Vib-Lig. The melting temperature (T_m) was estimated to be significantly lower for Vib-Lig, 30.7 °C, compared to Psy-Lig and Par-Lig with T_m of 46.0 °C and 53.7 °C, respectively (Fig. 2a). All three ligases measured show a ratio > 1 between the van't Hoff enthalpy derived from fitting a two-state model, and the calorimetric enthalpy derived by integration of the area under the excess heat capacity curves. Such temperature profile indicates that

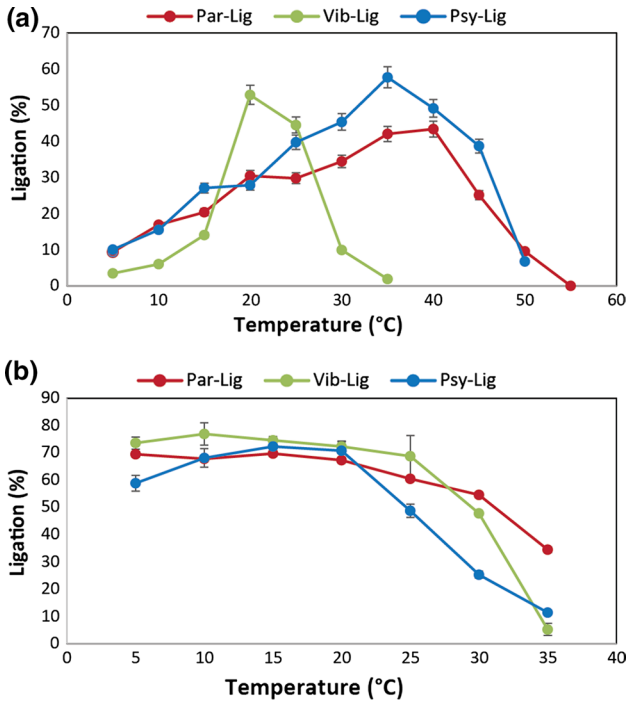


Fig. 1 Temperature optimum of Psy-Lig, Par-Lig, and Vib-lig by ligase activity assay. **a** Percentage of ligated single-nicked substrate. **b** Percentage of ligated cohesive substrate. Ligase activity was quenched after 15 min at various temperatures and quantified as percentage ligation by the intensity of the upper band relative to the sum of the two bands on the TBE-UREA gel. Ligase concentration was 2.5 µM for the nicked substrate and 100 µM for the cohesive substrate

Table 1 Literature and experimental data showing host optimal growth temperature, ligase temperature optimum for nick sealing and melting temperature

Ligase	Species of origin	Optimal growth (°C)	T_{opt} (°C)	T_m (°C)
Psy-Lig	*Psychromonas* spp. strain SP041	15[b]	35[a]	46[a]
Par-Lig	*Pseudoalteromonas artica*	10–15[c]	35–40[a]	53[a]
Vib-Lig	*Aliivibrio salmonicida*	15[d]	20[a]	30[a]

[a]This study

[b]Groudieva et al. (2003)

[c]Al Khudary et al. (2008)

[d]Egidius et al. (1986)

unfolding proceeds as a higher order oligomer; however, the irreversibility of the unfolding transition precluded detailed thermodynamic analysis.

The thermal stability in various buffer systems was measured by a thermofluor assay to confirm the DSC results and exclude the possibility that low thermal stability of observed for Vib-Lig is caused by non-ideal buffer conditions, as it

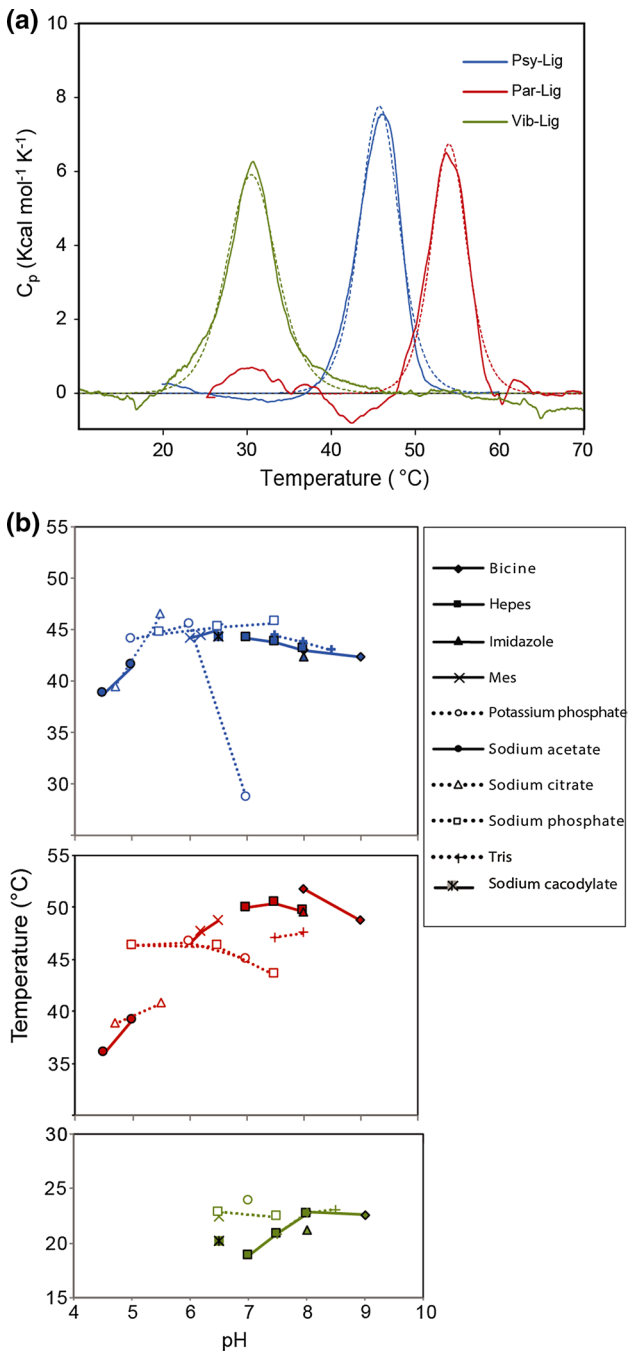


Fig. 2 Biophysical data. **a** Thermal unfolding monitored by DSC. **b** Thermal stability measured by thermofluor. Thermal unfolding parameters are given in Supplementary Table S2

has a significantly lower pI (predicted to be 5.5) relative to Psy-Lig and Par-Lig (both greater than 9.0). Thermofluor data (Fig. 2b) suggest that stability of the various ligases does not vary between pHs 6.5 and 9, with the exception of Psy-Lig which is extremely unstable in phosphate buffer at pH 7.0. Otherwise, Psy-Lig shows stability up to 46 °C and Par-Lig up to 53 °C, which is in line with DSC unfolding temperature. Also consistent with DSC data, Vib-Lig shows a lower thermal stability relative to Psy-Lig and Par-Lig with a maximum at 23 °C in all buffers down to pH 6.5. Below this, no transition could be observed, indicating that Vib-Lig was already unfolded.

Sequence comparison

Cold-active enzymes may combine rigidity and stability with a high level of flexibility. To gain further insight into the activity/stability/flexibility relationship and cold adaptation, interesting sequence and structural differences were identified by sequence alignments and homology modeling.

The enzymes studied are of similar size and share all properties common to minimal ADLs, but exhibit different temperature optima and stabilities. A structure-based sequence alignment was generated (Fig. 3). Lig E from *V. cholera* (Vch-Lig) was included as this human pathogen exhibits robust growth between 20 and 45 °C and is unable to survive at 4 °C for extended periods of time (Martinez et al. 2010). Pairwise comparison of the three experimentally examined Lig Es together with Vch-Lig shows that all sequence pairs have identities in the 40–49% range. Consistent with both enzymes deriving from members of the genus Vibrio, Vib-Lig and Vch-Lig share the highest homology in terms of sequence identity (48.4%), although they are adapted to different habitats and temperatures; thus, Vch-Lig represents a phylogenetically related mesophilic homolog of Vib-Lig. All four Lig Es contain the conserved nucleotidyltransferase family motifs I–VI and align with very few insertions or deletions, giving high confidence in placement of secondary structural elements by homology modeling (described below). Furthermore, the sequence alignment revealed high conservation of amino acids involved in substrate binding, metal binding and enzymatic activity.

Several studies have indicated the increased occurrence of some residues in cold-adapted proteins and decreased frequency of others, which has been rationalized by the physical properties of their sidechains influencing the flexibility and stability of the protein. This includes fewer salt bridges, fewer hydrogen bonds, a lower content of proline residues, a reduced Arg/(Arg + Lys) ratio, lower (Leu + Ile)/(Leu + Ile + Val) ratio and increased glycine content (Aghajari et al. 1998; Collins et al. 2005; Huston et al. 2004; Metpally and Reddy 2009; Russell et al. 1998; Saavedra et al. 2018). For this reason, we compared the amino acid content of the four proteins; however, most classic sequence 'traits' of cold adaptation, including increased glycine, decreased proline and less-packed hydrophobic core, were not apparent in Vib-Lig. Instead, higher sequence conservation appeared to be with the more phylogenetically related Vch-Lig than the other psychrophile-derived ADLs. For example, a lower number of Gly residues are often

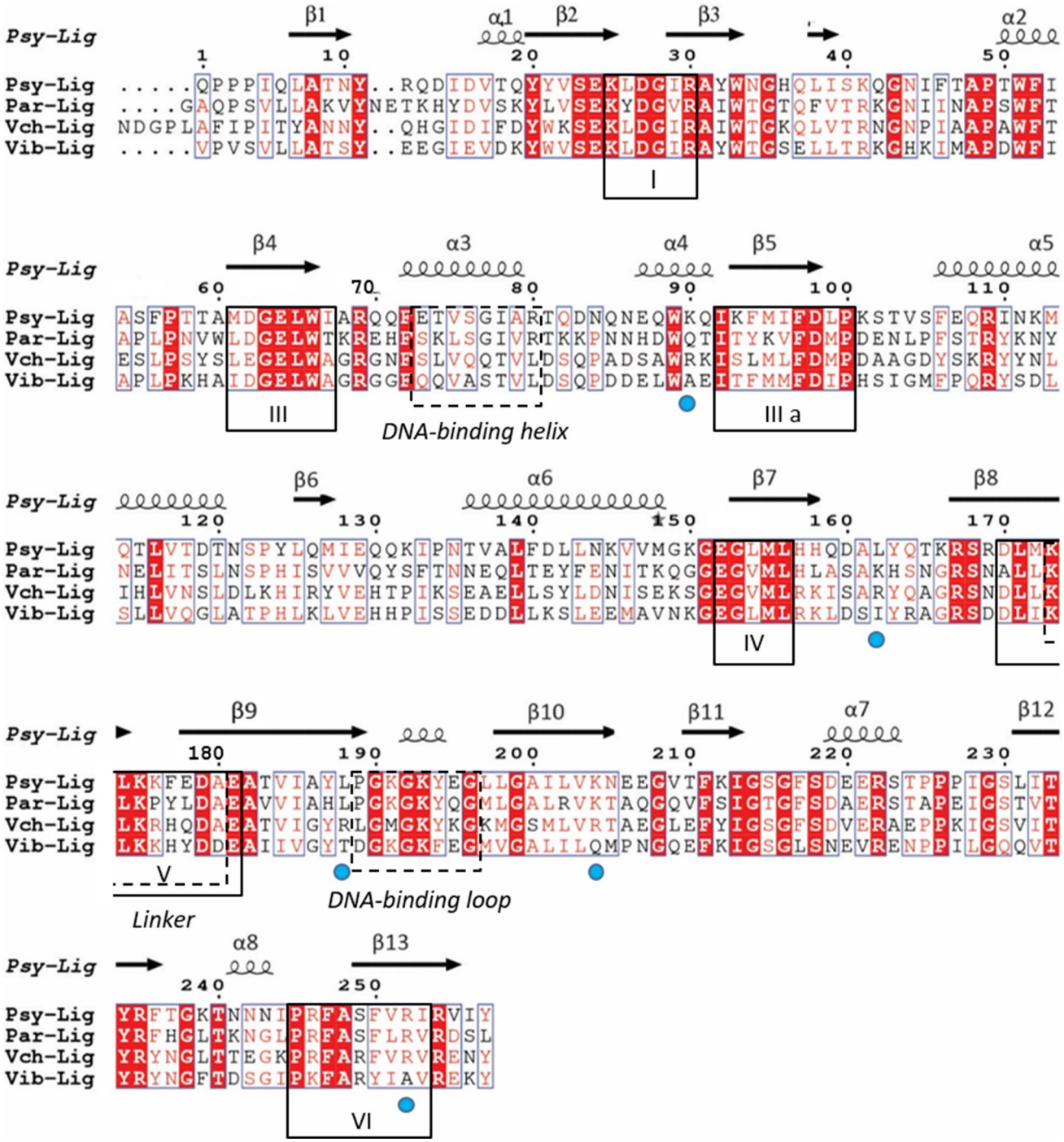


Fig. 3 Amino acid sequence alignment comparing mature ATP-dependent ligases from Psychromonas spp. strain SP041 (Psy-Lig), Aliivibrio salmonicida (Vib-Lig), Pseudoalteromonas artica (Par-Lig), and Vibrio cholera (Vch-Lig). Identical residues are shaded with red and similar residues are shown in red text. Spirals indicate α-helices and arrows indicate β-strands. Boxed amino acids represent conserved motifs of the nucleotidyltransferase enzymes. The DNA-binding elements of Lig Es are boxed with dashed lines. Surface-exposed substitutions of basic to uncharged residues in Vib-Lig are indicated by blue circles. Percentage sequence identity is given in Supplementary Table S3

pinpointed as a typical cold-adapted trait; however, this did not correlate with thermal stability of these ADLs, and most Gly residues are conserved, especially between the psychrophilic Vib-Lig and the mesophilic Vch-Lig (Table 2). Likewise, decreased Pro content has also been related to cold adaptation (Wallon et al. 1997; Zhao et al. 2010), but as Vch-Lig has fewer Pro than Vib-Lig (11 versus 13), Pro content is not an evident factor.

Another 'typical' feature of cold-adapted enzymes is a decreased number of Arg residues, which may increase stability through their capability to form hydrogen bonds and salt bridges (Aittaleb et al. 1997). In line with this, we observed the highest Arg count in the presumably mesophilic Vch-Lig (Table 2). The number of Arg residues is significantly lower for Vib-Lig (11), Psy-Lig (12), and Par-Lig (11) compared to Vch-Lig (18). This is also reflected by the ratio Arg/(Lys + Arg) per residue, which is 0.53 in Vch-Lig compared to 0.40, 0.39, and 0.39 in Psy-Lig, Vib-Lig, and Par-Lig, respectively, also supporting an overall better stability of the mesophilic molecule. Arg can contribute in more interactions with surrounding amino acids than lysine.

Table 2 Brief summary of extracted sequence features and characterization data for Psy-Lig, Vib-Lig, Par-Lig, and Vch-Lig, respectively

	Psy-Lig	Vib-Lig	Par-Lig	Vch-Lig
Sequence length	257	257	260	262
T_{opt} (°C)	35	20	35–40	–
T_{melt} (°C)	46.0	30.7	53.7	–
Calculated pI	9.1	5.3	9.5	9.0
Net charge[a]	+4	−9	+9	+4
Polar residues[b] (%)	35.8	29.2	37.3	34.0
Hydrophobic residues[c] (%)	41.3	42.8	39.2	39.7
Aromatic residues[d] (%)	10.9	9.7	11.2	11.5
Gly (number and %)	17/6.6	24/9.3	20/7.7	23/8.8
Met (number and %)	7/2.7	8/3.1	3/1.2	6/2.3
Pro (number and %)	13/5.1	13/5.1	12/4.6	11/4.2
Arg (number and %)	12/4.7	11/4.3	12/4.6	18/6.9
Arg/(Lys + Arg)	0.40	0.39	0.39	0.53
(Leu + Ile)/(Leu + Ile + Val)	0.76	0.70	0.65	0.78

[a]Residues R, K, D and E

[b]Residues G, S, T, Y, N, Q and C

[c]Residues A, V, L, I, W, F, P and M

[d]F, W and Y

However, Arg may also interact with water on the surface. Interestingly, the multiple sequence alignment (Fig. 3) shows that Arg in Vch-Lig is frequently substituted with hydrophobic residues in Vib-Lig.

Homology modeling and comparison to the crystal structure of Psy-Lig

To identify positions in the three-dimensional ligase structure, where relevant amino acid substitutions occurred, homology models of Par-Lig, Vib-Lig, and Vch-Lig were built based on the deposited structure of Psy-Lig 4D05; (Williamson et al. 2014). Increased local and/or global flexibility can be achieved by destabilization of the structure through a reduction in intramolecular forces such as salt bridges, ion-pair networks, hydrogen bonds and aromatic interaction, and increased length of loop regions (Davail et al. 1994; Feller 2003; Russell 2000). Hydrogen bond analysis shows that Vib-Lig is possibly destabilized by having fewer hydrogen bonds per residue in total, compared to Par-Lig and Psy-Lig (Table 3). In comparison, the mesophilic Vch-Lig has the highest ratio of hydrogen bonds per residue (0.816). It is interesting to note that the ratios correlate well with the measured melting temperatures Vib-Lig, Psy-Lig, and Par-Lig with low ratios giving low melting temperatures. In particular, the number of side-chain-to-main-chain hydrogen bonds is lower for the cold-adapted Vib-Lig.

Examination of the structural models also revealed that the arginine substitutions described in the preceding section are generally located on the surface, thus introducing hydrophobic surface patches in Vib-Lig (Fig. 4). Calculations by POPS (Parameter OPtimsed Surfaces (Fraternali and Cavallo 2002)) showed that the overall total area of exposed hydrophobic residues was similar among all ligases; thus, unique exposed hydrophobic patches in Vib-Lig appear to be local. Interesting Arg substitutions in Vib-Lig compared to Vch-Lig include Arg95–Ala90,

Table 3 Summary of calculated intramolecular interactions for Psy-Lig, Vib-Lig, Par-Lig, and Vch-Lig, respectively

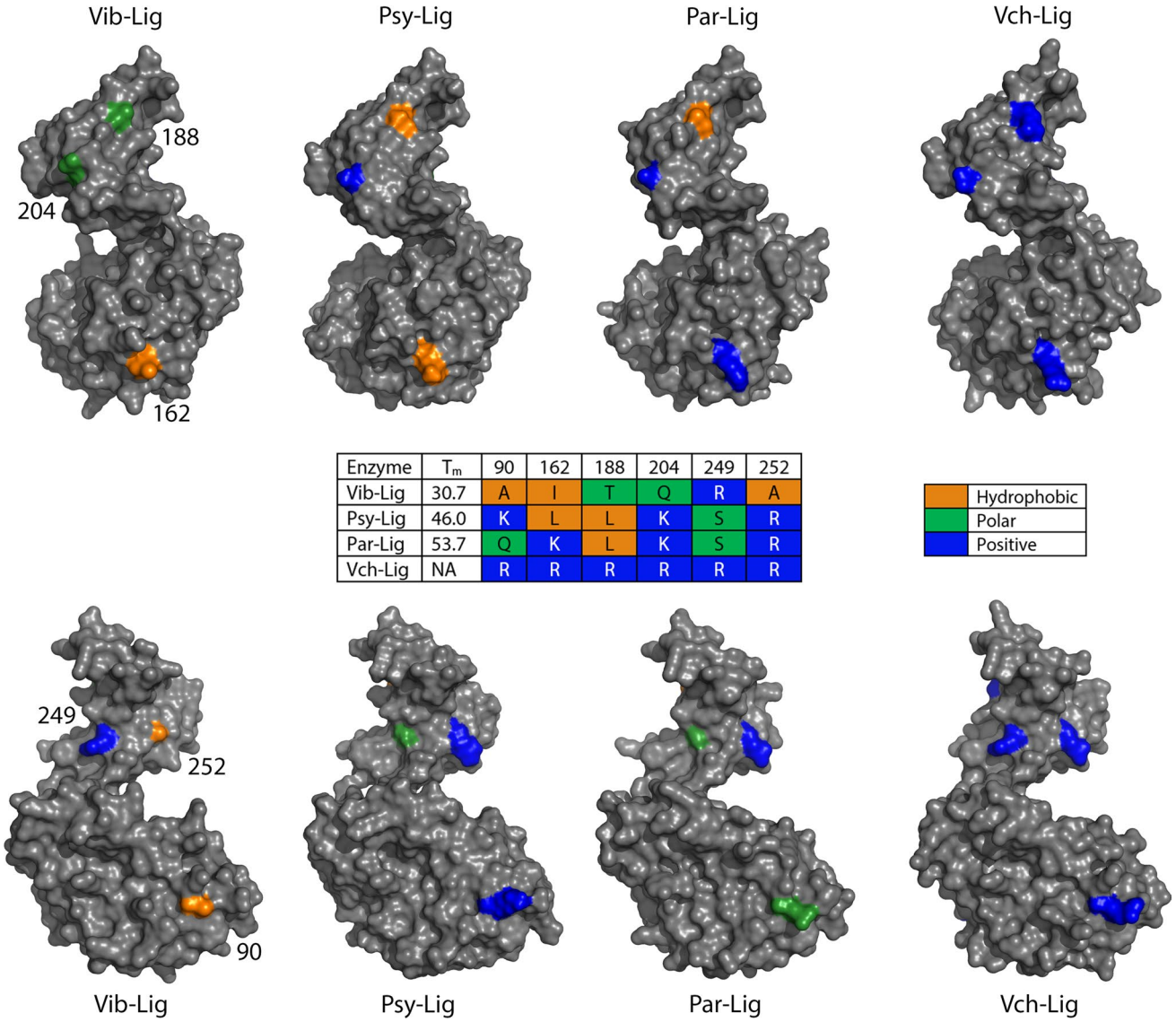
	Psy-Lig	Vib-Lig	Par-Lig	Vch-Lig
PDB ID	4d05	Model	Model	Model
Resolution	1.65 Å	–	–	–
No. of residues in PDB file	257	256	257	250
No. of hydrogen bonds per residue	0.759	0.715	0.778	0.816
No. SS hydrogen bonds per residue	0.086	0.066	0.066	0.104
No. SM hydrogen bonds per residue	0.202	0.133	0.175	0.180
No. MM hydrogen bonds per residue	0.471	0.516	0.537	0.532
No. ion pairs < 4/< 6 Å	8/19	11/18	10/16	12/23
No. 2 membered networks < 4.0 Å	6	5	6	6
No. 3 membered networks < 4.0 Å	1	3	2	3

SS side-chain-to-side-chain hydrogen bonds, *SM* side-chain-to-main-chain hydrogen bonds, *MM* main-chain-to-main-chain hydrogen bonds

Arg167–Ile162, Arg193–Thr188, Arg209–Gln204, and Arg257–Ala252 (Fig. 4). For Par-Lig and Psy-Lig, three of these Arg are substituted with Leu/Lys. The percentage of hydrophobic residues is slightly higher for Vib-Lig (42.80%) and Psy-Lig (41.25%) compared to Par-Lig (39.25%) and Vch-Lig (39.69%), possibly reflecting the substitutions of polar residues with hydrophobic residues on the surface compared to Vch-Lig. In combination, the elevated number of hydrophobic residues described above, the unique local hydrophobic surface patches and the lower number of Arg, may impart local flexibility to the Vib-Lig structure compared to its mesophilic counterpart Vch-Lig.

Electrostatic surface potential

Some cold-adapted enzymes feature an overall excess of negative charges at the surface of the protein, with a pI frequently more acidic than that of their mesophilic homologues (Feller 2003; Leiros et al. 1999; Russell 2000). Higher frequency or patches of acidic residues on the surface may increase solvent interactions and thereby lead to an overall destabilization of the enzyme by charge–charge repulsion, observed in cold-adapted trypsin and β-lactamase (Feller 2003; Leiros et al. 1999). The calculated pI of 5.3 for Vib-Lig is significantly more acidic compared to its counterparts, and also correlates with the substitution of basic



Enzyme	T_m	90	162	188	204	249	252
Vib-Lig	30.7	A	I	T	Q	R	A
Psy-Lig	46.0	K	L	L	K	S	R
Par-Lig	53.7	Q	K	L	K	S	R
Vch-Lig	NA	R	R	R	R	R	R

Fig. 4 Sequence variability mapped onto molecular surface representations of Vib-Lig, Psy-Lig, Par-Lig, and Vch-Lig. The top and bottom panels are rotated 180° views, while the middle panel shows melting temperature and substituted amino acids in selected positions for the four enzymes. Color codes: blue: positively charged residues; green: polar residues; orange: hydrophobic residues. Vib-Lig Residue numbers are included for reference between the panels

arginine residues at the surface with hydrophobic amino acids. Further examination of the charge distribution on the surface of the Vib-Lig model (Fig. 5) indicates that the DNA-binding faces of Vib-Lig remain positively charged as seen for structures of other Lig Es, while surfaces not involved in DNA binding are more positively charged compared with the more thermostable Psy-Lig. This suggests that charges in the binding surfaces of Vib-Lig are conserved and the majority of variation is located in distant areas of the protein.

Conservation of active site and DNA-binding surface

It is often suggested that low-temperature adaptation of enzymes is driven by increased local flexibility at the active site (D'Amico et al. 2002, 2006; Struvay and Feller 2012); therefore, we examined three key areas of the Vib-Lig enzyme that are essential for activity: the region surrounding the AMP-binding pocket, where the enzyme is covalently adenylated in the first step of the ligase reaction, the inter-domain linker region which undergoes significant structural changes during the catalytic cycle and the surfaces of the adenylation (AD-) and oligonucleotide-binding (OB-) domains that are in contact with double-strand DNA during nick sealing. Our comparisons reveal that the active site is strictly conserved, except for Lys 41 in Psy-Lig which is replaced by the chemically similar Arg in the other three ADLs (Fig. 3, supplementary Fig. S1a). The sequence alignment shows that the flexible linker regions connecting the two core domains are similar, preserving the hydrogen bonding pattern observed in Psy-Lig, with the exception of Par-Lig, where the equivalent of Lys 176 (Psy-Lig) is replaced by Pro (Fig. 3, supplementary Fig. S1b). Lig E-type ligases efficiently ligate DNA breaks without any additional DNA-binding domains or large flexible loop regions, instead using interactions with shorter highly structured motifs and specific charged residues found on the DNA-binding surface of the core catalytic domains (Williamson et al. 2014, 2018). In general, these motifs are well conserved between the three variants, consistent with both the equivalent positively charged DNA-binding surfaces of Vib-Lig and Psy-Lig and previous observations of consensus between Lig Es in this region (Fig. 5) (Williamson et al. 2018).

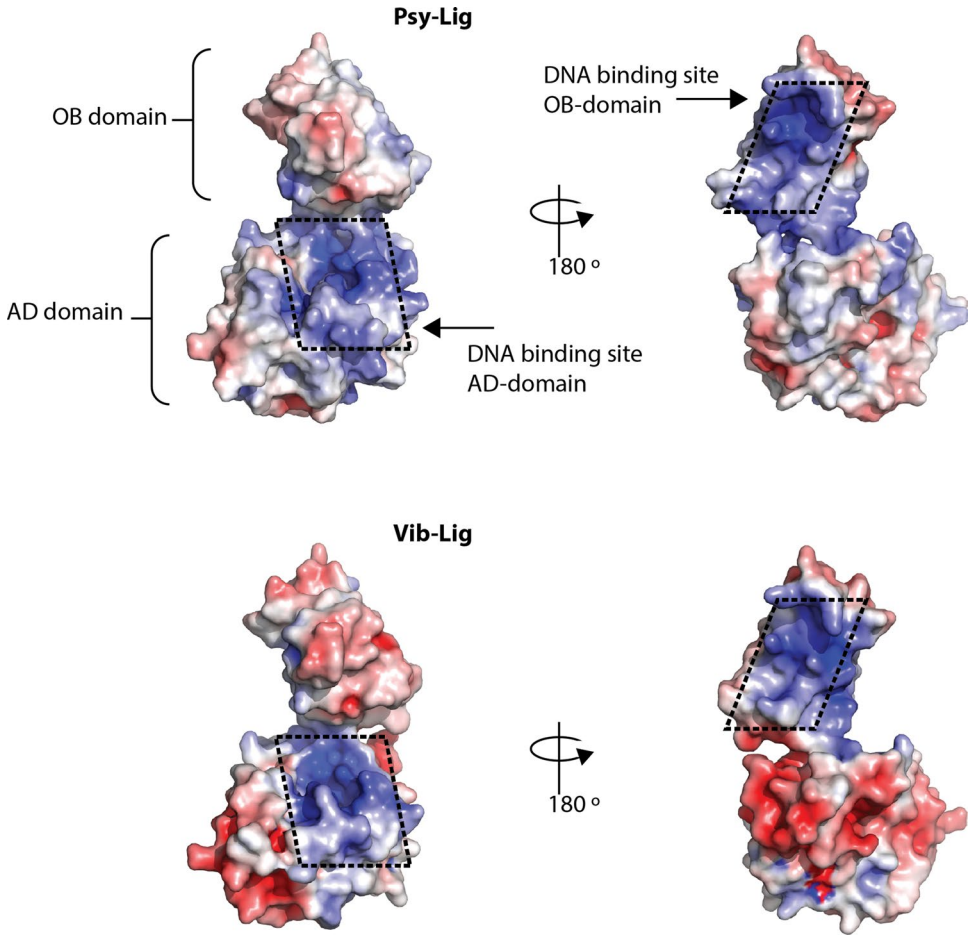


Fig. 5 Structure of Psy-Lig (top) and model of Vib-Lig (lower) colored surface charge. The surface potential was generated using APBS (Dolinsky et al. 2007), with positively charged areas shown in blue and negatively charged areas in red

Discussion

In this study, biochemical and biophysical characteristics of ATP-Dependent Ligases (ADLs) from psychrophilic organisms were analyzed in an attempt to identify typical cold-adaptation features. Vib-Lig, originating from the psychrophilic fish pathogen *Aliivibrio salmonicida* which has a growth range of 1–22 °C and an optimum of 15 °C (Egidius et al. 1986) exhibits classical features of cold adaptation including a low-temperature optimum of activity and decreased thermostability compared to homologous enzymes. In contrast, despite being derived from psychrophilic organisms, the Psy-Lig and Par-Lig enzymes are not themselves cold-adapted as both have temperature optima in the range of 35–40 °C and unfolding temperatures greater than 45 °C. This has been observed for many psychrophile-derived enzymes such as L-haloacid dehalogenase from *Psychromonas ingrahamii*, alcohol dehydrogenase of *Flavobacterium frigidimaris,* and KUC-12-keto acid decarboxylases derived from *Psychrobacter*, where the individual enzymes remain active and stable at temperatures well above the survival limit of the host organism (Kazuoka et al. 2007; Novak et al. 2013; Wei et al. 2013). The simplest rationale in the case of the ADLs is that although the temperature optimum is relatively high, the 30–40% activity recorded below 15 °C is sufficient for the biological purposes of the bacterium in its native environment, although the effects of different conditions on in vivo activity are also possible.

During ligation of double-strand breaks with cohesive ends, low temperature is an advantage to stabilize base-pairing between short stretches of complementary nucleotides at the break site. This must be balanced against decreased enzyme activity at lower temperatures. Lower temperatures allow DNA overhangs to base-pair and remain annealed long enough for the ligase to join them, at the expense of reduced ligase activity. This is directly observed in the present study during ligation of substrates with 4 nt overhangs as optimal activities are shifted to lower temperatures for all the three enzymes measured, despite of their individual T_{opt} varying when measured with a nicked substrate.

The Lig E enzymes compared in our study have moderate sequence identities (40–50%) and likely highly similar three-dimensional structures. The analyses performed indicated some sequence differences that potentially lower T_{opt} of Vib-Lig relative to homologs. One difficulty in such comparative analyses is distinguishing between substitutions imparting psychrophilicity and those that have occurred through genetic drift. To exclude possible false-positive findings based on phylogenetic resemblance, we included Lig E from *V. cholera* in our sequence comparison as previous phylogenomic studies have placed this close to Vib-Lig in evolutionary terms (Williamson et al. 2016), thus representing a genus-related but mesophilic organism. Coming from a mesophilic human pathogen, Vch-Lig is not anticipated to exhibit cold-adapted characteristics. The major differences in Vib-Lig appear to be in non-DNA-binding surface-exposed residues. Arginines are generally located on the surface of Psy-Lig, Par-Lig and Vch-Lig, and substitution of these residues introduces hydrophobic or uncharged surface patches in Vib-Lig. This is consistent with the investigation of three structurally homologous NAD^+-dependent DNA ligases (NDLs) adapted to different temperatures, where specific surface areas revealed a significant increase of exposed hydrophobic residues to solvent, in contrast to a more hydrophilic and charged surface area in thermophiles (Georlette et al. 2003), indicating an entropy-driven destabilization of the protein structure. Likewise, replacements of lysine with arginine in the psychrophilic α-amylase from *Pseudoalteromonas haloplanktis* resulted in a more stabilized enzyme with mesophilic properties, demonstrating the relevance of arginine content in cold adaptation (Siddiqui et al. 2006). It is interesting to note that these substitutions in Vib-Lig are unevenly distributed between the two domains with only two occurring in the larger catalytic adenylation domain (approximately 170 residues), and four on the smaller oligonucleotide domain (approximately 80 residues). Recent work demonstrated that substitutions increasing flexibility in different domains of adenylate kinase gave rise to different temperature effects on substrate binding and catalysis (Saavedra et al. 2018). As with DNA ligases, adenylate kinase activity involves coordinated reorientations between discrete protein domains, and it is interesting to consider whether this distribution reflects tuning of the oligonucleotide-binding domain for DNA binding/product release which are the rate limiting processes in the ligation reaction rather than the catalytic step its self (Bauer et al. 2017; Lohman et al. 2011).

Calculations of the electrostatic surface potential revealed that the cold-active Vib-Lig displays a positively charged surface near the active site and on the binding face of the OB-domain, which is important for binding of the negatively charged DNA substrate (Fig. 5), despite its overall more acidic pI. Similar results were observed for the cold-adapted uracil-DNA *N*-glycosylase (cUNG) from Atlantic cod (Leiros et al. 2003), indicating increased affinity for the negatively charged DNA compared with mesophile homologues. The number and nature of residues around the active site are conserved among the homologous ADLs adapted to different temperatures, suggesting that local cold-adapted residues are not directly involved in catalysis, but influence flexibility indirectly at some distance apart. The psychrophilic Vib-Lig is further

characterized by a decreased number of hydrogen bonds, which correlates with an increase in overall flexibility of the enzyme and affects protein-water surface interactions.

Although the decreased temperature optima of psychrophile-derived enzymes is commonly attributed to an increase in flexibility, either globally or locally, which causes a concomitant lowering of thermal stability (Smalas et al. 2000; D'Amico et al. 2002; Feller 2003; Struvay and Feller 2012), many enzymes are inactivated by temperatures below those inducing denaturation. A comparison of the NDL from the psychrophile *Pseudoalteromonas haloplanktis* with that of mesophilic NDL of *E. coli* and the thermophilic NDL of *Thermus scotoductus* indicated that structural differences imparted a temperature optimum of 18 °C, compared to 30 °C for NDL of *E. coli* and more than 60 °C for NDL of *T. scotoductus* (Georlette et al. 2000). This is accompanied by a decrease in T_m in the *P. haloplanktis* enzyme (33 °C) compared to the ones from *E. coli* (54 °C) and *T. scotoductus* (95–101 °C) (Georlette et al. 2003). The temperature optimum for activity of the *E. coli* NDL corresponds to the beginning of the thermal unfolding. *P. haloplanktis* NDL, however, shows a different link between activity and thermal adaptation; optimal activity is reached 10 °C before unfolding and the enzyme is inactivated at the beginning of the unfolding transition. A similar behaviour is observed for the activity and stability of Vib-Lig, Psy-Lig, and Par-Lig, where a decrease in activity above T_{opt} is observed in the absence of denaturation/unfolding. Recently, new paradigms have been suggested to explain this behaviour, as the classical (two-state) model is limited to enzymes, where increased catalytic activity is directly followed by thermal inactivation. These include macromolecular rate theory (MMRT), which provides a rationale for the curved temperature-rate plots observed for enzymes, independent of denaturation, and describes the temperature dependence of enzyme-catalyzed rates in the absence of denaturation by the difference in heat capacity between the enzyme–substrate complex and the enzyme transition state species (Arcus et al. 2016). The three-state equilibration model (EM) (Daniel and Danson 2013) has also been suggested to explain the temperature dependence of enzyme-catalyzed rates in the absence of denaturation. EM introduces a reversible inactivated (not denatured) form of the enzyme (E_{inact}) as an intermediate in rapid equilibrium with the active form (E_{act}), which adds a thermal buffer effect that protects the enzyme from thermal inactivation. Another explanation invokes a tuning of surface mobility through alteration of regions spatially removed from the active site which affect the overall enzyme dynamics (Åqvist et al. 2017; Isaksen et al. 2016). Computer simulations and Arrhenius plots suggest that surface rigidity/flexibility outside the catalytic region affects the enthalpy/entropy balance. Key single distant mutations may disrupt surface hydrogen bonding networks and alter the protein-water surface interactions (Isaksen et al. 2016) which may be the case with arginine substitutions in our study.

Conclusions

We have described the temperature optima and thermal denaturation profiles of three psychrophile-derived ADLs of the minimal Lig E-type. In the course of this work, we determined that two of the three, Par-Lig, and the structurally characterized Psy-Lig did not exhibit marked psychrophilic properties, while the third had typical low-temperature characteristics such as low T_{opt} and low thermal stability. Sequence comparison and homology modeling identified surface-exposed patches with greater hydrophobicity in Vib-Lig, relative to homologs, which we suggest are relevant for the experimentally observed psychrophilic properties.

Catalytic sites are often strictly conserved between homologs with different activity optima, as seen in Vib-Lig and Vch-Lig, meaning that the markedly lower T_{opt} of Vib-Lig relative to Psy-Lig and Par-Lig cannot be explained by increased active-site flexibility. We hope that future application of more sophisticated computational methods, coupled with specific mutational studies, may elucidate general principles imparting low-temperature activities. Such enhanced understanding of the molecular basis of low-temperature activity may enable us to tailor the activity optima of commercial ligases for use in biotechnological applications.

Acknowledgements This research was supported by Research Council Norway [244247, 2015]; Funding for open access charge was granted by the publication fund at the University of Tromsø.

References

Aghajari N, Feller G, Gerday C, Haser R (1998) Structures of the psychrophilic *Alteromonas haloplanctis* alpha-amylase give insights into cold adaptation at a molecular level. Structure 6:1503–1516

Aittaleb M, Hubner R, Lamotte-Brasseur J, Gerday C (1997) Cold adaptation parameters derived from cDNA sequencing and molecular modelling of elastase from Antarctic fish *Notothenia neglecta*. Protein Eng 10:475–477

Akey D, Martins A, Aniukwu J, Glickman MS, Shuman S, Berger JM (2006) Crystal structure and nonhomologous end-joining function of the ligase component of *Mycobacterium* DNA ligase D. J Biol Chem 281:13412–13423. https://doi.org/10.1074/jbc.M513550200

Al Khudary R, Stosser NI, Qoura F, Antranikian G (2008) *Pseudoalteromonas arctica* sp. nov., an aerobic, psychrotolerant, marine bacterium isolated from Spitzbergen. Int J Syst Evol Microbiol 58:2018–2024. https://doi.org/10.1099/ijs.0.64963-0

Åqvist J, Isaksen GV, Brandsdal BO (2017) Computation of enzyme cold adaptation. Nat Rev Chem 1:0051. https://doi.org/10.1038/s41570-017-0051

Arcus VL et al (2016) On the temperature dependence of enzyme-catalyzed rates. Biochemistry 55:1681–1688. https://doi.org/10.1021/acs.biochem.5b01094

Bauer RJ, Jurkiw TJ, Evans TC Jr, Lohman GJ (2017) Rapid time scale analysis of T4 DNA ligase-DNA binding. Biochemistry 56:1117–1129. https://doi.org/10.1021/acs.biochem.6b01261

Biasini M et al (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42:W252–W258. https://doi.org/10.1093/nar/gku340

Chambers CR, Patrick WM (2015) Archaeal nucleic acid ligases and their potential in biotechnology. Archaea (Vancouver, BC) 2015:170571. https://doi.org/10.1155/2015/170571

Collins T, Gerday C, Feller G (2005) Xylanases, xylanase families and extremophilic xylanases. FEMS Microbiol Rev 29:3–23. https://doi.org/10.1016/j.femsre.2004.06.005

D'Amico S et al (2002) Molecular basis of cold adaptation. Philos T Roy Soc B 357:917–924. https://doi.org/10.1098/rstb.2002.1105

D'Amico S, Collins T, Marx JC, Feller G, Gerday C (2006) Psychrophilic microorganisms: challenges for life. EMBO Rep 7:385–389. https://doi.org/10.1038/sj.embor.7400662

Daniel RM, Danson MJ (2013) Temperature and the catalytic activity of enzymes: a fresh understanding. FEBS Lett 587:2738–2743. https://doi.org/10.1016/j.febslet.2013.06.027

Davail S, Feller G, Narinx E, Gerday C (1994) Cold adaptation of proteins. Purification, characterization, and sequence of the heat-labile subtilisin from the antarctic psychrophile *Bacillus*. TA41 J Biol Chem 269:17448–17453

Doherty AJ, Suh SW (2000) Structural and mechanistic conservation in DNA ligases. Nucleic Acids Res 28:4051–4058. https://doi.org/10.1093/nar/28.21.4051

Doherty AJ, Wigley DB (1999) Functional domains of an ATP-dependent DNA ligase. J Mol Biol 285:63–71. https://doi.org/10.1006/jmbi.1998.2301

Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res 35:W522–W525. https://doi.org/10.1093/nar/gkm276

Duplantis BN, Bosio CM, Nano FE (2011) Temperature-sensitive bacterial pathogens generated by the substitution of essential genes from cold-loving bacteria: potential use as live vaccines. J Mol Med JMM 89:437–444. https://doi.org/10.1007/s00109-010-0721-3

Dwivedi N, Dube D, Pandey J, Singh B, Kukshal V, Ramachandran R, Tripathi RP (2008) NAD(+)-dependent DNA ligase: a novel target waiting for the right inhibitor. Med Res Rev 28:545–568. https://doi.org/10.1002/med.20114

Egidius E, Wiik R, Andersen K, Hoff KA, Hjeltnes B (1986) *Vibrio salmonicida* sp. nov., a new fish pathogen. Int J Syst Evol Microbiol 36:518–520. https://doi.org/10.1099/00207713-36-4-518

Ellenberger T, Tomkinson AE (2008) Eukaryotic DNA ligases: structural and functional insights. Annu Rev Biochem 77:313–338. https://doi.org/10.1146/annurev.biochem.77.061306.123941

Ericsson UB, Hallberg BM, DeTitta GT, Dekker N, Nordlund P (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. Anal Biochem 357:289–298. https://doi.org/10.1016/j.ab.2006.07.027

Feller G (2003) Molecular adaptations to cold in psychrophilic enzymes. Cell Mol Life Sci 60:648–662. https://doi.org/10.1007/s00018-003-2155-3

Fraternali F, Cavallo L (2002) Parameter optimized surfaces (POPS): analysis of key interactions and conformational changes in the ribosome. Nucleic Acids Res 30:2950–2960

Georlette D, Jonsson ZO, Van Petegem F, Chessa J, Van Beeumen J, Hubscher U, Gerday C (2000) A DNA ligase from the psychrophile *Pseudoalteromonas haloplanktis* gives insights into the adaptation of proteins to low temperatures. Eur J Biochem 267:3502–3512

Georlette D, Damien B, Blaise V, Depiereux E, Uversky VN, Gerday C, Feller G (2003) Structural and functional adaptations to extreme temperatures in psychrophilic, mesophilic, and thermophilic DNA ligases. J Biol Chem 278:37015–37023. https://doi.org/10.1074/jbc.M305142200

Groudieva T, Grote R, Antranikian G (2003) *Psychromonas arctica* sp. nov., a novel psychrotolerant, biofilm-forming bacterium isolated from Spitzbergen. Int J Syst Evol Microbiol 53:539–545. https://doi.org/10.1099/ijs.0.02182-0

Huston AL, Methe B, Deming JW (2004) Purification, characterization, and sequencing of an extracellular cold-active aminopeptidase produced by marine psychrophile *Colwellia psychrerythraea* strain 34H. Appl Environ Microbiol 70:3321–3328. https://doi.org/10.1128/aem.70.6.3321-3328.2004

Isaksen GV, Åqvist J, Brandsdal BO (2016) Enzyme surface rigidity tunes the temperature dependence of catalytic rates. Proc Natl Acad Sci 113:7822–7827. https://doi.org/10.1073/pnas.1605237113

Kaminski AM et al (2018) Structures of DNA-bound human ligase IV catalytic core reveal insights into substrate binding and catalysis. Nat Commun 9:2642. https://doi.org/10.1038/s41467-018-05024-8

Kazuoka T, Oikawa T, Muraoka I, Si Kuroda, Soda K (2007) A cold-active and thermostable alcohol dehydrogenase of a psychrotorelant from Antarctic seawater, *Flavobacterium frigidimaris* KUC-1. Extremophiles 11:257–267. https://doi.org/10.1007/s00792-006-0034-1

Kim DJ, Kim O, Kim HW, Kim HS, Lee SJ, Suh SW (2009) ATP-dependent DNA ligase from *Archaeoglobus fulgidus* displays a tightly closed conformation. Acta Crystallogr Sect F-Struct Biol Cryst Commun 65:544–550. https://doi.org/10.1107/s1744309109017485

Leiros HK, Willassen NP, Smalas AO (1999) Residue determinants and sequence analysis of cold-adapted trypsins. Extremophiles 3:205–219

Leiros I, Moe E, Lanes O, Smalas AO, Willassen NP (2003) The structure of uracil-DNA glycosylase from Atlantic cod (*Gadus morhua*) reveals cold-adaptation features. Acta Crystallogr D Biol Crystallogr 59:1357–1365

Lohman GJS, Chen LX, Evans TC (2011) Kinetic characterization of single strand break ligation in duplex DNA by T4 DNA ligase. J Biol Chem 286:44187–44196. https://doi.org/10.1074/jbc.M111.284992

Magnet S, Blanchard JS (2004) Mechanistic and kinetic study of the ATP-dependent DNA ligase of *Neisseria meningitidis*. Biochemistry 43:710–717. https://doi.org/10.1021/bi0355387

Martinez RM, Megli CJ, Taylor RK (2010) Growth and laboratory maintenance of *Vibrio cholerae*. Curr Protoc Microbiol 6:1. https://doi.org/10.1002/9780471729259.mc06a01s17

McDonald IK, Thornton JM (1994) Satisfying hydrogen bonding potential in proteins. J Mol Biol 238:777–793. https://doi.org/10.1006/jmbi.1994.1334

Metpally RP, Reddy BV (2009) Comparative proteome analysis of psychrophilic versus mesophilic bacterial species: insights into the molecular basis of cold adaptation of proteins. BMC Genom 10:11. https://doi.org/10.1186/1471-2164-10-11

Nair PA, Nandakumar J, Smith P, Odell M, Lima CD, Shuman S (2007) Structural basis for nick recognition by a minimal pluripotent DNA ligase. Nat Struct Mol Biol 14:770–778 http://www.nature.com/nsmb/journal/v14/n8/suppinfo/nsmb1266_S1.html

Nishida H, Kiyonari S, Ishino Y, Morikawa K (2006) The closed structure of an archaeal DNA ligase from Pyrococcus furiosus. J Mol Biol 360:956–967. https://doi.org/10.1016/j.jmb.2006.05.062

Novak HR, Sayer C, Panning J, Littlechild JA (2013) Characterisation of an L-haloacid dehalogenase from the marine psychrophile *Psychromonas ingrahamii* with potential industrial application. Mar Biotechnol (New York, NY) 15:695–705. https://doi.org/10.1007/s10126-013-9522-3

Pascal JM, O'Brien PJ, Tomkinson AE, Ellenberger T (2004) Human DNA ligase I completely encircles and partially unwinds nicked DNA. Nature 432:473–478. https://doi.org/10.1038/nature03082

Pascal JM et al (2006) A flexible interface between DNA ligase and PCNA supports conformational switching and efficient ligation of DNA. Mol Cell 24:279–291. https://doi.org/10.1016/j.molcel.2006.08.015

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Meth 8:785–786 http://www.nature.com/nmeth/journal/v8/n10/abs/nmeth.1701.html#supplementary-information

Petrova T et al (2012) ATP-dependent DNA ligase from *Thermococcus* sp. 1519 displays a new arrangement of the OB-fold domain. Acta Crystallogr Sect F Struct Biol Cryst Commun 68:1440–1447. https://doi.org/10.1107/s1744309112043394

Pitcher RS, Brissett NC, Doherty AJ (2007a) Nonhomologous end-joining in bacteria: a microbial perspective. Annu Rev Microbiol 61:259–282. https://doi.org/10.1146/annurev.micro.61.080706.093354

Pitcher RS, Green AJ, Brzostek A, Korycka-Machala M, Dziadek J, Doherty AJ (2007b) NHEJ protects mycobacteria in stationary phase against the harmful effects of desiccation. DNA Repair. 6:1271–1276. https://doi.org/10.1016/j.dnarep.2007.02.009

Płociński P et al (2017) DNA Ligase C and Prim-PolC participate in base excision repair in mycobacteria. Nat Commun 8:1251. https://doi.org/10.1038/s41467-017-01365-y

Russell NJ (2000) Toward a molecular understanding of cold activity of enzymes from psychrophiles. Extremophiles 4:83–90. https://doi.org/10.1007/s007920050141

Russell RJ, Gerike U, Danson MJ, Hough DW, Taylor GL (1998) Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium. Structure 6:351–361

Saavedra HG, Wrabl JO, Anderson JA, Li J, Hilser VJ (2018) Dynamic allostery can drive cold adaptation in enzymes. Nature 558:324–328. https://doi.org/10.1038/s41586-018-0183-2

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Meth 9:671–675

Shi K et al (2018) T4 DNA ligase structure reveals a prototypical ATP-dependent ligase with a unique mode of sliding clamp interaction. Nucleic Acids Res. https://doi.org/10.1093/nar/gky776

Shuman S (2009) DNA Ligases: progress and prospects. J Biol Chem 284:17365–17369. https://doi.org/10.1074/jbc.R900017200

Shuman S, Glickman MS (2007) Bacterial DNA repair by non-homologous end joining. Nat Rev Microbiol 5:852–861. https://doi.org/10.1038/nrmicro1768

Siddiqui KS et al (2006) Role of lysine versus arginine in enzyme cold-adaptation: modifying lysine to homo-arginine stabilizes the cold-adapted alpha-amylase from *Pseudoalteramonas haloplanktis*. Proteins 64:486–501

Smalas AO, Leiros HK, Os V, Willassen NP (2000) Cold adapted enzymes. Biotechnol Annu Rev 6:1–57

Struvay C, Feller G (2012) Optimization to low temperature activity in psychrophilic enzymes. Int J Mol Sci 13:11643–11665. https://doi.org/10.3390/ijms130911643

Subramanya HS, Doherty AJ, Ashford SR, Wigley DB (1996) Crystal structure of an ATP-dependent DNA ligase from bacteriophage T7. Cell 85:607–615

Tanabe M, Ishino Y, Nishida H (2015) From structure–function analyses to protein engineering for practical applications of DNA ligase. Archaea (Vancouver, BC) 2015:267570. https://doi.org/10.1155/2015/267570

van der Kamp MW, Prentice EJ, Kraakman KL, Connolly M, Mulholland AJ, Arcus VL (2018) Dynamical origins of heat capacity changes in enzyme-catalysed reactions. Nat Commun 9:1177. https://doi.org/10.1038/s41467-018-03597-y

Vriend G (1990) WHAT IF: a molecular modeling and drug design program. J Mol Graph 8(52–56):29

Wallon G et al (1997) Sequence and homology model of 3-isopropylmalate dehydrogenase from the psychrotrophic bacterium *Vibrio* sp. I5 suggest reasons for thermal instability. Protein Eng 10:665–672

Wei J, Timler JG, Knutson CM, Barney BM (2013) Branched-chain 2-keto acid decarboxylases derived from *Psychrobacter*. Fems Microbiol Lett 346:105–112. https://doi.org/10.1111/1574-6968.12208

Wilkinson A, Day J, Bowater R (2001) Bacterial DNA ligases. Mol Microbiol 40:1241–1248

Williamson A, Pedersen H (2014) Recombinant expression and purification of an ATP-dependent DNA ligase from *Aliivibrio salmonicida*. Protein Expres Purif 97:29–36. https://doi.org/10.1016/j.pep.2014.02.008

Williamson A, Rothweiler U, Schroder Leiros H-K (2014) Enzyme-adenylate structure of a bacterial ATP-dependent DNA ligase with a minimized DNA-binding surface. Acta Crystallogr Sect D 70:3043–3056. https://doi.org/10.1107/S1399004714021099

Williamson A, Hjerde E, Kahlke T (2016) Analysis of the distribution and evolution of the ATP-dependent DNA ligases of bacteria delineates a distinct phylogenetic group 'Lig E'. Mol Microbiol 99:274–290. https://doi.org/10.1111/mmi.13229

Williamson A, Grgic M, Leiros HS (2018) DNA binding with a minimal scaffold: structure–function analysis of Lig E DNA ligases. Nucleic Acids Res 46:8616–8629. https://doi.org/10.1093/nar/gky622

Zhao JS, Deng Y, Manno D, Hawari J (2010) *Shewanella* spp. genomic evolution for a cold marine lifestyle and in situ explosive biodegradation. PLoS One 5:9109. https://doi.org/10.1371/journal.pone.0009109

Supporting Information

Temperature adaptation of DNA ligases from psychrophilic organisms

Kristel Berg, Ingar Leiros, Adele Williamson*

Current address: Department of Chemistry, The Arctic University of Norway, N-9019 Tromsø, Norway

Correspondence: adele.k.williamson@uit.no

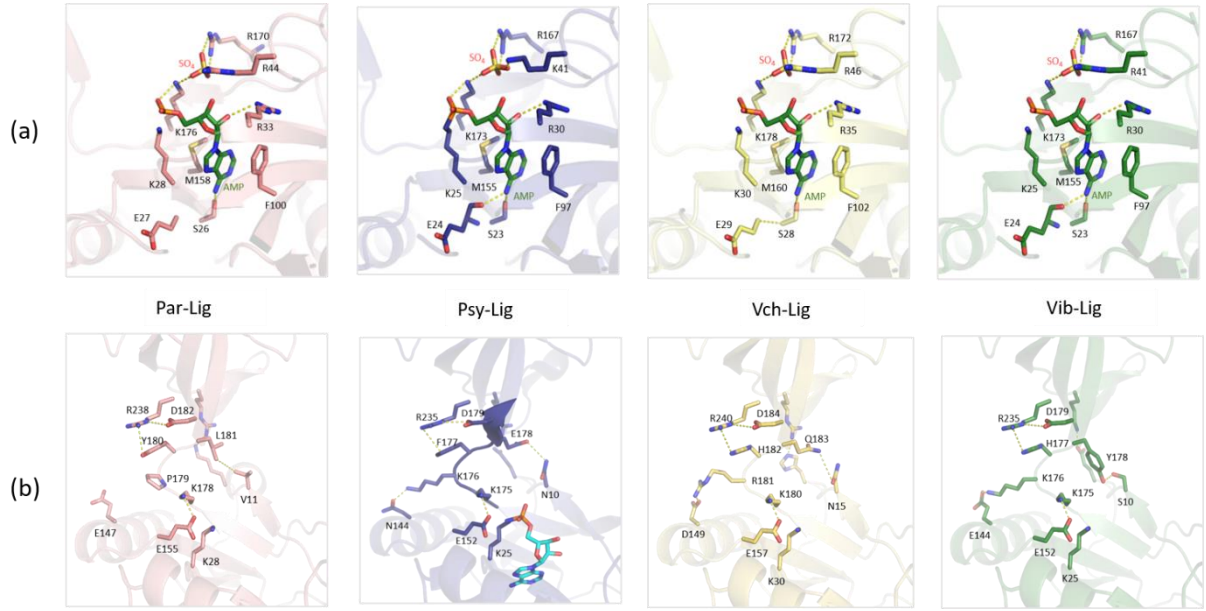


Fig S1 Structural comparison of key areas essential for activity in *Psychromonas spp*. strain SP041 (Psy-Lig), *Aliivibrio salmonicida* (Vib-Lig), *Pseudoalteromonas artica* (Par-Lig) and *Vibrio cholera* (Vch-Lig). (a) Active site showing residues making contact with the cofactor. The nucelotide cofactor is labeled in green and the sulfate ion is in yellow. (b) Inter-domain linker region between the AD and OB domains.

Table S1. Ligase substrates for gel-based assays

Oligo	Name	Sequence	Modification
1	5-prime_nick	tccgaattcgagctccgtcg	5' phosphate
2	3-prime_nick	aggccatggctgatatcgga	5' fam (6-carboxyfluorescein)
3	single nick complement	cgacggagctcgaattcggatccgatatcagccatggcct	
4	cohesive 5-prime complement	ggatccgatatcagccatggcct	5' phosphate
5	cohesive 3-prime complement	cgacggagctcgaatt	

* Nicked substrate was formed by annealing oligos 1, 2 and 3 in ratio 5:1:5. Cohesive substrate annealed oligos 1+5 and 2+4 (1:1.2 and 1.2:1.2) separately, followed by mixing at ambient temperature.

Table S2. Thermal unfolding parameters monitored by DSC.

Protein	ΔH_{cal} (kcal mol^{-1})	T_m	ΔH_{vH} (kcal mol^{-1})	Ratio $\Delta H_{vH}/\Delta H_{cal}$	ΔS_{unf} (kcal mol^{-1} K^{-1})
Psy-Lig	46.10	46.00	128.50	2.80	1.40
Vib-Lig	49.30	30.70	92.20	1.90	0.20
Par-Lig	34.28	53.65	149.24	4.40	0.11

Table S3. Sequence identity matrix for the mature ATP-dependent ligases from *Psychromonas spp*. strain SP041 (Psy-Lig), *Aliivibrio salmonicida* (Vib-Lig), *Pseudoalteromonas artica* (Par-Lig) and *Vibrio cholera* (Vch-Lig).

	Psy-Lig	Vib-Lig	Par-Lig	Vch-Lig
Psy-Lig	100	43.5	45.7	42.7
Vib-Lig		100	40.0	48.4
Par-Lig			100	46.2