Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Cotutelle internationale : Université libanaise

Présentée et soutenue par :
Mirna HAJJ

le 17 février 2021

Titre :

Etudes phylogénomiques et moléculaires des hélicases de type Lhr chez les Archaea et les Bactéries

---

École doctorale et discipline ou spécialité :
ED BSB : Biotechnologies

Unité de recherche :
LMGM - Laboratoire de Microbiologie et Génétique Moléculaires

Directeur/trice(s) de Thèse :

Béatrice CLOUET-D'ORVAL et Ziad ABDEL-RAZZAK

Jury :
M. Herve LE HIR, Rapporteur
Mme Ghislaine HENNEKE , Rapporteure
M. Jacques OBERTO, Rapporteur
M. Matthieu ARLAT, Examinateur
Mme Jamila BORJAC-NATOUR, Examinatrice
Mme Hala CHAMIEH, Examinatrice
M. Ziad ABDEL-RAZZAK, co-directeur de thèse
Mme Béatrice CLOUET-D'ORVAL, Directrice de thèse

Phylogenomics & Molecular studies of Lhr-type SF2 helicases in Archaea and Bacteria.

Helicases are proteins that use ATP energy to unwind nucleic acids and to remodel protein-nucleic acid complexes. They are involved in almost every aspect of the DNA and RNA metabolism and participate in numerous repair mechanisms maintaining cellular integrity. Helicases are classified into six superfamilies (SF1-6). The Lhr-type proteins belong to SF2 helicases that are poorly characterized to date. A phylogenomic study performed by Chamieh et al classified SF1 and SF2 helicases from archaeal sequenced genomes and showed that Lhr-type proteins are ubiquitous in Archaea (Chamieh et al. 2016). Another study conducted by Clouet D'orval et al, determined the interaction networks of proteins involved in RNA metabolism, such as the ribonuclease aRNase J and the helicase ASH-Ski2, they identified a cross-talk between the translation, RNA degradation and transcription machineries in Thermococcales a group of -hyperthermophilic Archaea- and remarkably Lhr-type helicase was found to be a partner in these networks (Phung et al. 2020). In this context, my PhD thesis aim to perform in-depth phylogenomic analyses of Lhr-type helicases in Archaea and to extend this work further to bacteria, to dissect the molecular function of aLhr2 in *Thermococcus barophilus*, a model organism used for biochemical and genetic studies in Thermococcales. Further, we aim to investigate the role of the bacterial Lhr (bLhr) helicase in *E.coli* where the gene is co-transcribed with the RNase T in proteobacteria.

The first part of our work was initiated with a bibliographic survey (published book chapter; Hajj et al, 2019)/ followed by phylogenomic studies on the Lhr-type proteins. We were able to define the Lhr-type proteins as ubiquitous enzyme in Archaea and identify five orthologous groups. Based on these analyses, we proposed an evolution route for the five archaeal and bacterial Lhr groups and hypothesize on their functions in the cell (Hajj et al, Manuscript in preparation). In a second part, we focused on the molecular study of aLhr2 from *Thermococcus barophilus* (*Tbar*). To investigate the enzymatic activities of *Tbar*-aLhr2, *alhr2* gene was cloned and the recombinant protein recombinant *Tbar*-aLhr2 protein produced using the pET expression system in *E. coli.* We demonstrated that *Tbar*-aLhr2 is an ATPase that has the same affinity for single stranded DNA and RNA and can specifically anneal and unwind DNA/RNA duplex with a 3' overhang. Finally, proteomic and transcriptomic analyses were performed to identify *Tbar*-aLhr2 protein network and to determine the effect of *lhr2* deletion (Δ*lhr2*) on gene expressions in *T. barophilus*. In light of all these results, we propose that *Tbar*-aLhr2 is involved in transcription or/and DNA repair in Thermococcales and acts on R-loops (RNA/DNA duplex) (Hajj et al, Manuscript in preparation). In the third part of the work, we initiated a functional study of the bLhr helicase of *E. coli* (*Eco*). To examine a putative interaction between *Eco*-bLhr and RNase T that are expressed as an operon, the *Eco-blhr* gene was cloned and the recombinant *Eco*-bLhr protein produced. RNase T is a ribonuclease known to be involved in DNA repair and tRNA/rRNA metabolism.

Finally, we discussed and compared the putative role(s) of Lhr helicases in Bacteria and Archaea in RNA metabolism and DNA repair in eliciting RNA as a key player in the repair of DNA damage.

Etudes phylogénomiques et moléculaires des hélicases de type Lhr chez les Archaea et les Bactéries

Les hélicases, classées en six superfamilles (SF1-6), sont des protéines qui utilisent l'énergie de l'ATP pour dérouler les acides nucléiques et pour remodeler les complexes protéines-acides nucléiques. Elles sont impliquées dans presque tous les aspects du métabolisme de l'ADN et de l'ARN en participant à de nombreux mécanismes de maintien de l'intégrité cellulaire. Les protéines de type Lhr sont des hélicases SF2 qui sont pour la plupart non caractérisées. Récemment, par des approches phylogénomiques, Dr H. Chamieh et ses collaborateurs ont classé toutes les hélicases SF1 et SF2 présentes dans les génomes d'Archaea et ont montré que les protéines de type Lhr sont ubiquitaires (Chamieh et al. 2016). De plus, en déterminant les réseaux d'interaction des protéines impliquées dans le métabolisme de l'ARN, comme la ribonucléase aRNase J et l'hélicase ASH-Ski2, Dr B. Clouet-d'Orval et ses collaborateurs ont identifié un lien entre les machines de traduction, de dégradation de l'ARN et de transcription chez les Thermococcales -archaea hyperthermophile- avec, au sein de ces réseaux, une protéine annotée comme une hélicase de type Lhr (Phung et al. 2020). Dans ce contexte, les travaux de ma thèse ont pour objectif d'effectuer des analyses phylogénomiques approfondies des hélicases de type Lhr chez les archées et les bactéries, de disséquer la fonction moléculaire de aLhr2 chez *Thermococcus barophilus*, organisme modèle pour les études biochimiques et génétiques chez les Thermococcales, et d'étudier le rôle de l'hélicase Lhr de *E. coli* où le gène *lhr* est en opéron avec le gène codant pour la RNase T.
Dans une première partie, une étude bibliographique (publication d'un chapitre de livre ; Hajj et al, 2019) et des analyses phylogénomiques ont permis de définir les protéines de type Lhr comme ubiquitaire chez les archées et d'identifier cinq groupes d'orthologues. Ces analyses permettent de proposer un chemin évolutif pour les protéines Lhr d'archées et de bactéries et d'émettre des hypothèses sur leurs fonctions dans la cellule (Hajj et al, Manuscrit en préparation). Dans une deuxième partie, nous nous sommes focalisés sur l'étude moléculaire de aLhr2 de *Thermococcus barophilus* (*Tbar*). Pour étudier les activités enzymatiques de *Tbar*-aLhr2, le gène a*lhr2* a été cloné et la protéine recombinante *Tbar*-aLhr2 exprimée à l'aide du système d'expression pET chez *E. coli*. Nous avons démontré que *Tbar*-aLhr2 est une ATPase avec une affinité similaire pour l'ADN et l'ARN simple brin qui peut spécifiquement former et dérouler des duplex ADN/ARN avec une extrémité 3' sortante. Enfin, nous avons effectué des analyses protéomiques et transcriptomiques pour identifier le réseau de protéines associées à *Tbar*-aLhr2 et pour déterminer l'impact de la délétion du gène *alhr2* (Δ*alhr2*) sur l'expression génique chez *T. barophilus*. Au regard de ces résultats, nous proposons que *Tbar*-aLhr2 est impliqué au niveau de la transcription et/ou de la réparation de l'ADN chez les Thermococcales en agissant au niveau des « R-loop » (duplex ARN/ADN) (Hajj et al, Manuscrit en préparation). Dans une troisième partie, nous avons initié une étude fonctionnelle de l'hélicase bLhr de *E. coli* (*Eco*). Pour tester l'interaction entre *Eco*-bLhr et la RNase T qui sont exprimées au sein d'un même opéron, le gène *Eco-blhr* a été cloné et la protéine recombinante *Eco*-bLhr exprimée. La RNase T est une ribonucléase connue pour être impliquée dans la réparation de l'ADN et le métabolisme des ARNt et ARNr.
Finalement, une discussion permet de comparer le(s) rôle(s) proposé(s) pour les hélicases de type Lhr chez les bactéries et les archées en dégageant l'ARN comme acteur clé dans la réparation des dommages de l'ADN.

Table of contents

INTRODUCTION 9
I-ARCHAEA 10
1-Discovery & Classification 10
1-1 Archaea domain in the tree of life 10
1-2 Taxonomy Classification 14
2-Archaea are ubiquitous micro-organisms 17
2-1 Archaea of the geochemical terrestrial cycles. 17
2-2 The archaeome 18
3-Specific archaeal characteristics 19
3-1 Archaeal S-layer 20
3-2 Archaeal lipidic membrane 20
3-3 the Archaeal motility structure: the Archaellum 21
4-Archaeal study models 23
5-Mosaic aspects of archaeal informational system 27
5-1 Replication 27
5-2 Transcription 31
5-3 Translation 34
5-4 Post-transcriptional regulation of gene expression 37
II-HELICASES 40
1-Discovery of Helicase enzymes as cellular motors 40
2-Helicase Classification 43
2-1 The six superfamilies of helicases 43
2-2- Structural characteristics and conserved motifs of SF1 & SF2 53
3-Biochemical and mechanistic activities 59
3-1 NTP binding and hydrolysis 59
3-2 Unwinding activity 59
3-3 Annealing activity 61
3-4 Translocation and protein displacement 63
4-General cellular functions of DNA & RNA helicases 64
5-Lhr helicase subfamily belonging to SF2 69
5-1 Lhr bacterial helicases 69
5-2 Lhr archaeal helicases 73
6-Publication 1 : Hajj M. et al. 2018 76
THESIS OBJECTIVES 77
MATERIALS AND METHODS 81
I-Phylogenetic analysis 82

1-Building Lhr-type dataset 82
2-Identification of Lhr-type orthologous groups of proteins 83
3-Multiple Sequence Alignments and Phylogenetic Tree Constructions 83
4-Genomic context analysis 84
5-Structural model predictions 84
II-Construction of expression vectors 87
1-Construction of expression vectors carrying aLhr2 gene (TERMP_00533). 87
2-Construction of expression vectors carrying aLhr2 Dom4 and aLhr2 ΔDom4. 87
3-Construction of expression vectors carrying aLhr2 with punctual mutations. 88
III-Production and purification of recombinant proteins 90
1-Expression and production of *Tbar*-aLhr2 recombinant proteins. 90
2-Purification of *Tbar*-aLhr2 recombinant proteins. 90
2.1 Purification of aLhr2 proteins expressed in pET11b vectors 90
2.2 Purification of aLhr2 proteins expressed in pET15b vectors 91
2.3 Quantification and storage of proteins 91
IV-Enzymatic assays 93
1-Radiolabelled substrates 93
2-ATPase hydrolysis assay. 93
3-Nucleic acid binding assay. 94
4-Annealing assay 94
5-Unwinding assay 94
V-Protein interaction networks and Transcriptomic analysis 95
1-Pull down assay 95
2-Proteomic analysis 96
3-Transcriptomic RNA seq analysis 96
VI-*In vitro* interaction preliminary assays test of *Ecol*-Lhr-856 and RNAse T 96
1-Construction of expression vectors carrying *Ecol*-Lhr-856 gene (JW1645) 96
2-Expression and production of *Ecol*-Lhr-856 and *Ecol*-RNaseT recombinant proteins. 97
3-Purification of *Ecol*-Lhr-856 recombinant proteins. 97
4-Co-purification of recombinant protein on Nickel column 97
RESULTS 99
I-Phylogenetic and structural analysis of Lhr-type helicase 100
1-Archaeal and bacterial Lhr-type sequences are collected from three complementary search approaches 100
2-Five Lhr–type orthologous groups are identified by MCL clustering. 103
3-Taxonomic occurrence identifies aLhr1 & aLhr2 proteins as ubiquitous in archaeal genomes 107
4-Specificities at the structural level are observed for each Lhr-type orthologous group 112

5-Genomic context analysis of the genes coding for Lhr–type proteins helps in predicting the function(s) of some Lhr helicases ... 116
II-Biochemical studies of *Tbar*-aLhr2 protein. ... 118
1-*Thermococcus barophilus* as model of study ... 118
2-Production and Purification of *Tbar*-aLhr2 recombinant proteins ... 120
2-1 Constructions of pET11b-*Tbar*-aLhr2 expression vectors ... 120
2-2 Optimization of expression and purification experimental conditions ... 122
3-Biochemical activities of *Tbar*-aLhr2 wild-type and variants ... 129
3-1 ATPase activity ... 131
3-2 Nucleic acid affinity ... 136
3-3 Unwinding activity ... 140
3-4 Annealing activity ... 145
4-Conclusions ... 148
III-Towards the significance of aLhr2 of *T.barophilus* ... 150
1-Identification of protein interaction network of *Tbar*-$(His)_6$-aLhr2 ... 150
1-1 The Pull-down methodology ... 150
1-2 Overview of the interaction network of *Tbar*-$(His)_6$-aLhr2 ... 154
2-Transcriptomic analysis of aLhr2 helicase-lacking strain ... 161
3-Conclusions ... 163
IV-Functional study of *Ecol*-Lhr ... 164
1-Constructions of pET21b-*Ecol*-bLhr-856 expression vectors ... 164
2-Optimization of Co-purification experimental conditions ... 168
Discussion ... 171
1-Landscape of Lhr-type proteins the three domains of life ... 172
2-Is archaeal aLhr2 enzymes acting on DNA or RNA hybrids? ... 177
3-*Tbar*-aLhr2 network at the crossroad of translation, RNA processing/decay and DNA repair pathways ... 180
4-Archaeal aLhr2 helicases acting on R-loop structures? ... 185
5- Is there a Cross-talk between bLhr-HTH and RNase T in *E.coli* ... 186
6-General conclusion ... 187
Bibliography ... 189
FIGURES ANNEXE ... 231

List of FIGURES

Figure 1 Timeline and milestones in the discovery of Archaea 11
Figure 2 Models of the tree of life 13
Figure 3 Archaeal taxonomic tree 16
Figure 4 Lipids of archaeal, bacterial and eukaryotic cell membrane 22
Figure 5 RNAP subunits in the three domains of life 30
Figure 6 Archaeal transcription machinery 32
Figure 7 Mosaic molecular machinery involved in RNA metabolism in Archaea 39
Figure 8 Discovery of helicases 41
Figure 9 SF1 and SF2 subfamilies 46
Figure 10 The catalytic core of helicase superfamilies 47
Figure 11 SF2 helicase core 54
Figure 12 Helicases unwinding activity features 58
Figure 13 Mechanistic functions of helicases 62
Figure 14 Studied bacterial and archaeal Lhr-type protein 67
Figure 15 Protein-protein interaction networks including Lhr2 75
Figure 16 PhD Objectives 80
Figure 17 The Lhr-type domains and approaches used in collecting Lhr–type protein sequences 102
Figure 18 MCL clustering for defining Lhr-type protein clusters 104
Figure 19 Phylogenetic trees of the five Lhr-type orthologous groups 106
Figure 20 Distribution of the archaeal Lhr-type groups throughout the phylogeny of Archaea 109
Figure 21 Distribution of the bacterial Lhr-type groups throughout the phylogeny of Bacteria 111
Figure 22 Weblogo sequences of the Lhr-type orthologous groups 113
Figure 23 Features of Lhr-type orthologous groups 114
Figure 24 Homology of aLhr2 of Paby and Tbar 119
Figure 25 Helicase enzymatic activity and protein variants used in this study 121
Figure 26 Experimental workflow for production and purification of recombinant proteins 124
Figure 27 Set up heterologous overexpresssion of Tbar aLhr2 WT in E.coli 125
Figure 28 Over Expression of recombinant Tbar-aLhr2 variants 127
Figure 29 Purification of recombinant aLhr2 WT and variants 128
Figure 30 ATP hydrolysis assays 130
Figure 31 aLhr2 WT and variants ATPase activity 132
Figure 32 Kinetic of ATP hydrolysis of aLhr2 WT 133
Figure 33 Kinetic of ATP hydrolysis of aLhr2 ΔDom4 and T215A 135
Figure 34 Tbar-aLhr2 affinity for nucleic acids 138
Figure 35 Binding affinity curves of Tbar-aLhr2 WT for single strand-nucleic acids 139
Figure 36 Unwinding activity assay 141
Figure 37 Unwinding activity of Tbar-aLhr2WT and Tbar-aLhr2 ΔDom4 in presence of ATP 143
Figure 38 Unwinding in presence of analogs of ATP & Initial velocity of unwinding reaction 144
Figure 39 Duplex annealing assay 146
Figure 40 Annealing activity assay of Tbar-aLhr2WT and ΔDom4. 147
Figure 41 Mechanistic activities of Tbar-aLhr2 149
Figure 42 Expression and Purification of Tbar-$(His)_6$-aLhr2 152
Figure 43 Pull-down assay 153
Figure 44 Interaction network obtained using Tbar-$(His)_6$-aLhr2 as bait protein 157
Figure 45 Changes in transcriptional profile of aLhr2 helicase-lacking strain 162
Figure 46 bLhr-HTH of E.coli 165
Figure 47 RNAse T of E.coli 167
Figure 48 Co-purification assay 169
Figure 49 Co-purification of Ecol bLhr-HTH-856 and RNaseT 170
Figure 50 Structure and gene cluster of Lhr-type families 174
Figure 51 Tbar-aLhr2 duplex unwinding and strand annealing 179
Figure 52 Common protein candidates of aLhr2 and putative protein partners 181
Figure 53 Proposition of the implication of aLhr2 in R-loops processing based on the protein interaction profile. 184

List of FIGURES ANNEX

ANNEX 1 Phylogenetic tree of the Lhr-type OGs together with protein sequences retrieved in the Asgard phylum 232
ANNEX 2 Phylogenetic tree of the Lhr-type orthologous groups with HRQ-like helicase as outgroup 233
ANNEX 3 Sequence alignment of Lhr-type families 234
ANNEX 4 Chromatogram and SDS PAGE of purification steps of aLhr2 W557A, I512A and T215A 235
ANNEX 5 Chromatogram and SDS PAGE of purification steps of aLhr2 Dom4 and ΔDom4 236
ANNEX 6 Binding affinity of aLhr2 variants with 26nt substrate 237
ANNEX 7 Binding affinity of aLhr2 Dom4 238

List of Tables

Table 1……26
Table 2……29
Table 3……36
Table 4……45
Table 5……51
Table 6……72
Table 7……85
Table 8……86
Table 9……89
Table 10……92
Table 11……123
Table 12……137
Table 13……155
Table 14……158
Table 15……176
Table 16……182

INTRODUCTION

I-ARCHAEA

1-Discovery & Classification

1-1 Archaea domain in the tree of life

Prior to the work of Carle Woese and Georges E. Fox in 1977, the domains of life were represented as a tree with only two branches corresponding to the eukaryotes and prokaryotes.Their work revolutionised evolutionary biology by proposing a tree of life consisting of three domains: Eukaryotes, Bacteria and Archaea (Woese and Fox 1977)(Olsen and Woese 1993) (**FIGURE 1 & FIGURE 2A**). Initially, the archaeal domain was referred as "Archaeobacteria" domain which included methanogenic bacteria and the bacterial domain as "Eubacteria" which included all other bacteria. The first complete genome of autotrophic archaea the one of *Methanococcus jannaschii* (Bult et al. 1996) (**FIGURE 1**). This step provided the first opportunity to compare complete genetic complements and biochemical pathways among the three domains of life.

Archaea were for a long time considered only as extremophiles (Cavicchioli 2011). Indeed, their habitats are diverse under conditions where few bacteria and eukaryotes can tolerate. They are found in swamps (methanogenic, anoxic habitats) (Knittel et al. 2005), in hypersaline waters such as the Dead Sea (Halophiles) (Zvyagintseva, I.S, Tarasov, A.L 1988), at very high temperatures such as in hydrothermal vents on the seabed or in hot springs such as in Yellowstone Park (Thermophiles) (Gugliandolo and Maugeri 2019), or, conversely, in permanently cold conditions such as the dry lakes of the Antarctic (Psychrophiles) and at extreme pH levels (Acidophiles and Alkaphiles; (DeLong 1992; Fuhrman et al. 1992).

For more than 20 years, archaeal microorganisms have been classified into two groups (clades): the Euryarchaea and the Crenarchaea (Olsen and Woese 1993) (**FIGURE 1**). This classification was based on pairs of paralogous genes, the translation factors EFTu/ EFG and two related ATPase subunits. This study showed that Crenarchaeota have the same "sulfur-dependent" thermophilic phenotype, while the Euryarchaeota are phenotypically diverse. The predominant Euryarchaeal phenotype was the methanogenic.

Since 2002, with the development of high throughput sequencing techniques and the progress of phylogenomic analyses, several new groups such as Korarchaea, Nanoarchaea and Thaumarchaea were identified **(FIGURE 1)**.

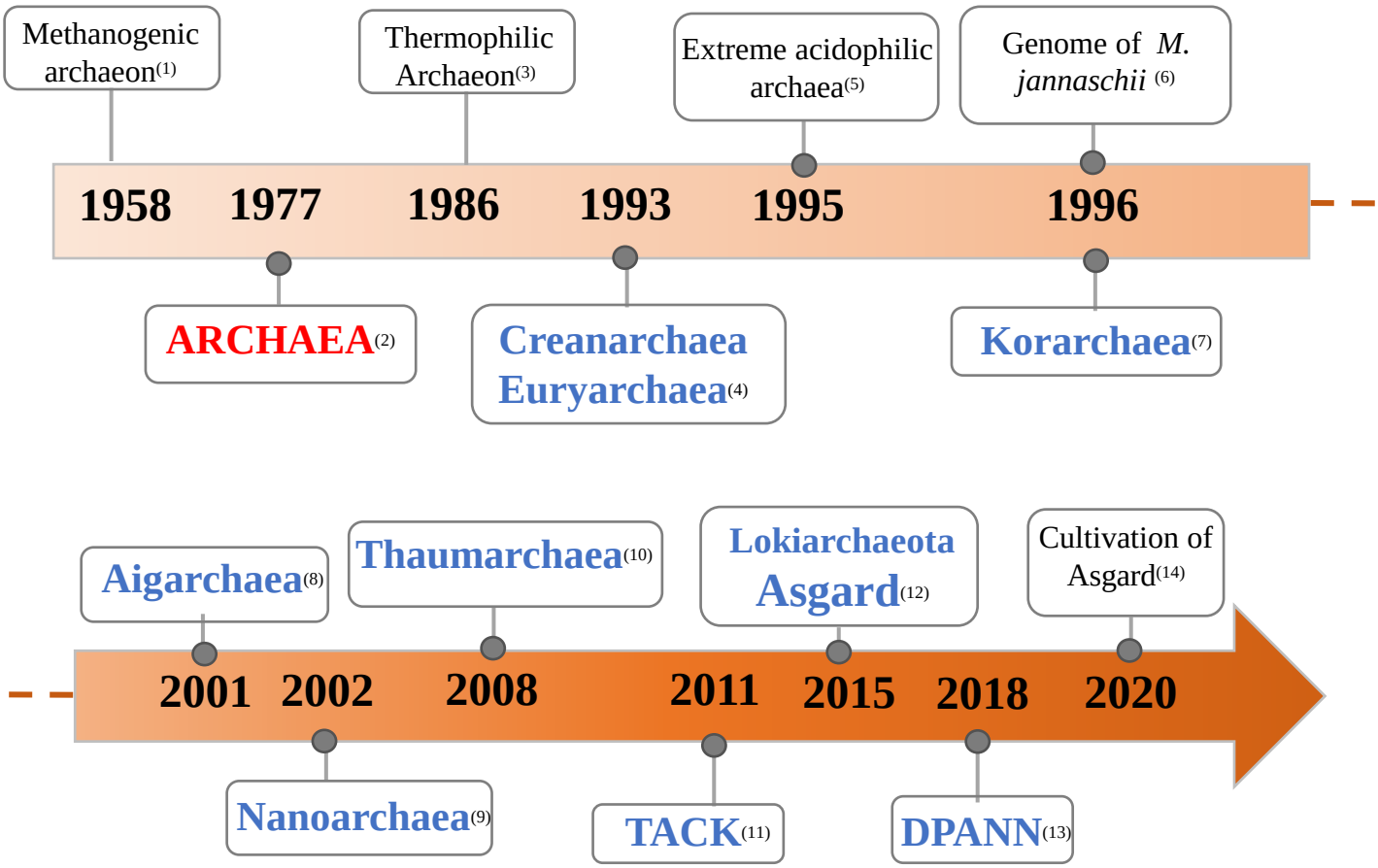


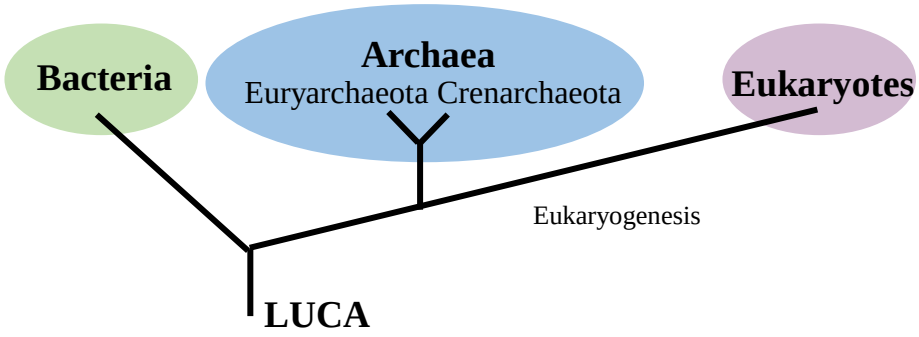
FIGURE 1. Timeline and milestones in the discovery of Archaea.
Archaea as a domain of life (in red) and of the different archaeal taxonomic groups (in blue). Key features (in black) and dates are indicated.
DPANN: Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota; TACK: Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota.
(1) Smith & Hungate, 1958 (2) Woese and Fox 1977 (3) Stetter 1986 (4) Olsen and Woese 1993 (5) Schleper et al 1995 (6) Bult et al 1996 (7) Barns et al 1996 (8) DeLong and Karl 2001 (9) Huber et al 2002 (10) Brochier-Armanet et al. 2008 (11) Guy & Ettema 2011 (12) Spang et al 2015 (13) Castelle and Banfield 2018 (14) P López-García 2020.

The Thaumarchaeota phylum was initially classified as mesophilic Crenarchaeota (Fuhrman et al. 1992; DeLong 1992). Subsequent comparative genomics based on concatenated data set of 53 ribosomal proteins common to archaea and eukarya revealed that Thaumarchaeota forms a separate and deep-branched phylum within the archaea phylogeny (Brochier-Armanet et al. 2008) **(FIGURE 1)**.

From 2011, the phylogeny of the archaea is enriched with a new classification that establishes the TACK group, which includes Thaumarchaea, Aigarachaea, Crenarchaeea and Korarchaea, and the DPANN group, which is characterised by small genomes (~0.5 to 1.5 Mbp) and made up of Diapherrites, Parvarchaea, Aenigmarchaea, Nanoarchaea and Nanohaloarchaea (Kellner et al. 2018). The Aigarchaeota phylum was proposed base on the sequencing of *Candidatus Caldiarchaeum subterraneum* genome, a thermophilic uncultivated organism phylogenetically close to Thaumarchaeota and Crenarchaeota (Nunoura et al. 2011) **(FIGURE 1)**. Interestingly, this genome encoded for eukaryotic-like proteins of the ubiquitin-proteasome system which identify it as part of a new archaeal phylum. The name Aigarchaeota, from the Greek « αση », meaning «dawn» or «aurora», these organisms represent intermediate characteristics of hyperthermophilic and mesophilic archaea (Nunoura et al. 2011). Two years later, the Geoarchaeota phylum was proposed based on the analysis of sequences of four metagenomes from microbial acidic mat samples at high temperatures and a low oxygen concentration of the Yellowstone Park (Kozubal et al. 2013).

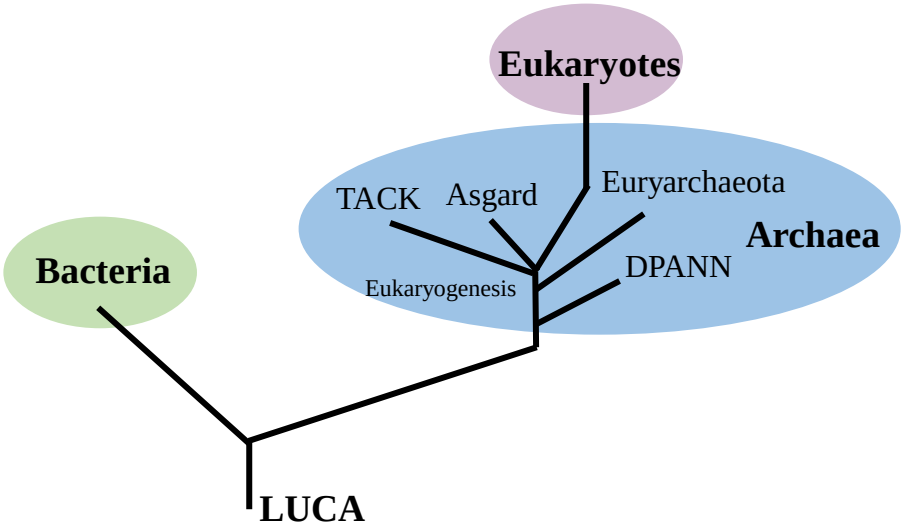
Since then, with the expansion of metagenomics and single-cell sequencing technics (Rinke et al. 2013), current data with more than 201 archaeal genomes available on the NCBI site, allow to further extend this diversity and to clarify this phylogeny with the identification of Geoarchaea, Bathyarchaea and Verstraetearchaea within the TACKs and new lines affiliated to the Euryarchaea group (Spang et al. 2015; Zaremba-Niedzwiedzka et al. 2017; Da Cunha et al. 2018; Eme et al. 2018; Spang et al. 2017). In 2011, phylogenetic trees based on Maximum-likelihood using ribosomal 16S RNA sequences allowed identifying a distinct superphylum composed of the «Thaumarchaeota, Aigarchaeota, Crenarchaeota et Korarchaeota» phyla also called «TACK» superphylum (Guy and Ettema 2011) **(FIGURE 1)**. In addition, a new superphylum so-called «DPANN» named after the acronym of its five phyla: Diapherotrites, Parvarchaeota, ARMAN, Nanohaloarchaea, Aenigmarchaeota (Kellner et al. 2018) **(FIGURE 1**).

A.



Carl Woese's
Three-domain tree

B.



New
Two-domain tree

FIGURE 2. Models of the tree of life.
A. In the 'three primary domains' (3D) scenario, the eukaryotes (purple), archaea (blue) and Bacteria (green) form three primary domains, each with a specific most recent common ancestor. **B.** In the 'two primary domains' (2D) scenario, archaea and bacteria are the two primary domains.
Branch lengths and number of lineages within each domain are arbitrary.
(ARCHAEA AND THE MEANING OF LIFE, Microbiology Society).

In 2015, with the discovery of the new branch of the Lokiarchaea found in marine sediments near Loki Castle (hydrothermal vents in the middle based on metagenomic data from the Atlantic Ocean), the phylogeny of the archaea opens new perspectives on the common ancestor of eukaryotes and archaea (Spang et al. 2015) **(FIGURE 2)**. Based on this new classification, two evolutionary scenarios are proposed according to phylogenomic approaches and are the subject of much debate (Gribaldo and Brochier-Armanet 2020). The first one, so-called 3D model, proposes three branches of life with a common ancestor preceding the diversification of archaea and eukaryotic sister lines (Zhu et al. 2019; Da Cunha et al. 2018; Forterre 2013) **(FIGURE 2A)**. The second one, so-called 2D model, proposes a two-branch tree with the eukaryotic domain emerging from the Asgard domain **(FIGURE 2B)** (Williams et al. 2020; Eme and Ettema 2018; Spang et al. 2018; Eme et al. 2018; Gribaldo and Brochier-Armanet 2020; Dacks et al. 2016; Klinger et al. 2016; Hug et al. 2016; Doolittle 2020). This scenario is based on the discovery of genes coding for eukaryotic signature proteins (ESP) present in all members of the Asgard lineage (Spang et al. 2018; Eme et al. 2018; Zaremba-Niedzwiedzka et al. 2017) and more specifically in the Heimdallarchae group (Williams et al. 2020; Zaremba-Niedzwiedzka et al. 2017; Schleper and Sousa 2020).

1-2 Taxonomy Classification

Current archaeal taxonomy includes four superphyla: the DPANN, Euryarchaea, TACK, and Asgard groups (Spang et al. 2017) **(FIGURE 3)**.

The DPANN are characterized by their small size genomes of ~0,5 à 1,5 Mbp. This superphylum includes 10 phyla: Aenigmarchaeota, Altiarchaeota, Diapherotrites, Mamarchaeota, Micrarchaeota, Nanoarchaeota, Nanohaloarchaeota, Pacearchaeota, Parvarchaeota, and Woesearchaeota (Castelle and Banfield 2018) **(FIGURE 3)**. Analyses based on phylogenetic markers suggest that the DPANN superphylum diverged early from the archaeal tree (Dombrowski et al. 2020).

The Euryarchaeota includes 12 orders for which complete genomes were sequenced and strains could be cultivated: Methanopyrales, Thermococcales, Archaeoglobales, Methanobacteriales, Thermoplasmatales, Halobacteriales, Methanomicrobiales, Methanosarcinales Methanococcales, Methalocellales, Methanoplasmatales, and Nanohaloarchaea (Sakai et al. 2008; Spang et al. 2015) **(FIGURE 3)**. Initial classification was

based on the SSU rRNA phylogeny (Mihajlovski et al. 2008; Paul et al. 2012)(Narasingarao et al. 2012). Now a large number of uncultivated organisms from environmental studies have been added. More specifically, hyperthermophiles of the Thermococcales order which has the largest number (around 300) of isolates and 45 species is composed of three genera: the Thermococcus (Zillig et al. 1983), the Pyrococcus (Fiala and Stetter 1986), and the Palaeococcus (Takai et al. 2000) Thermococcales organisms are hyperthermophilic, their temperatures of optimal growth vary between 80 and 100°C (Prieur et al. 2004).

The TACK superphylum includes the Thaumarchaeota, Aigarchaeota, Crenarchaeota, Korarchaeota, and Bathyarchaeota order (Spang et al. 2015; Inagaki et al. 2003). Bathyarchaeota is constituted of 25 subgroups (Zhou et al. 2018). Crenarchaeota phylum is divided into five orders: Desulfurococcales, Sulfolobales, Thermoproteales, Acidilobales, and Fervidicoccales (Prokofeva et al. 2009; Perevalova et al. 2010) **(FIGURE 3)**. Most Thermoproteales and Desulfurococcales are anaerobic and hyperthermophilic. Sulfolobales are mostly aerobic and their growth temperatures are lower than for the two other groups. Thaumarchaeota represents a very diverse phylum comprising 8 classes, 10 orders, 28 families, 31 genus, and 103 species (Sheridan et al. 2020). Most of the genomic divergence in this phylum is driven by gene duplication or loss and by horizontal gene transfer (Hua et al. 2018). Comparative genomics shows that Thaumarchaeota and Aigarchaeota share 1154 genes with their common ancestor. The genetic diversity of Aigarchaeota is essentially generated by horizontal gene transfer (Hua et al. 2018).

The Asgard are composed of Lokiarchaeota, Thorarchaeota, Odinarchaeota, and Heimarchaeota (Zaremba-Niedzwiedzka et al. 2017) **(FIGURE 3)**. Recently was added *Promethearchaeum symtrophicum* to Lokiarchaeota (Imachi et al. 2020) **(FIGURE 1)**. This major discovery opens many perspectives and will provide new elements insights into eukaryogenesis (Schleper and Sousa 2020). The Asgard phylum branches as a sister clade to the TACK (Guy and Ettema 2011) **(FIGURE 3)**.

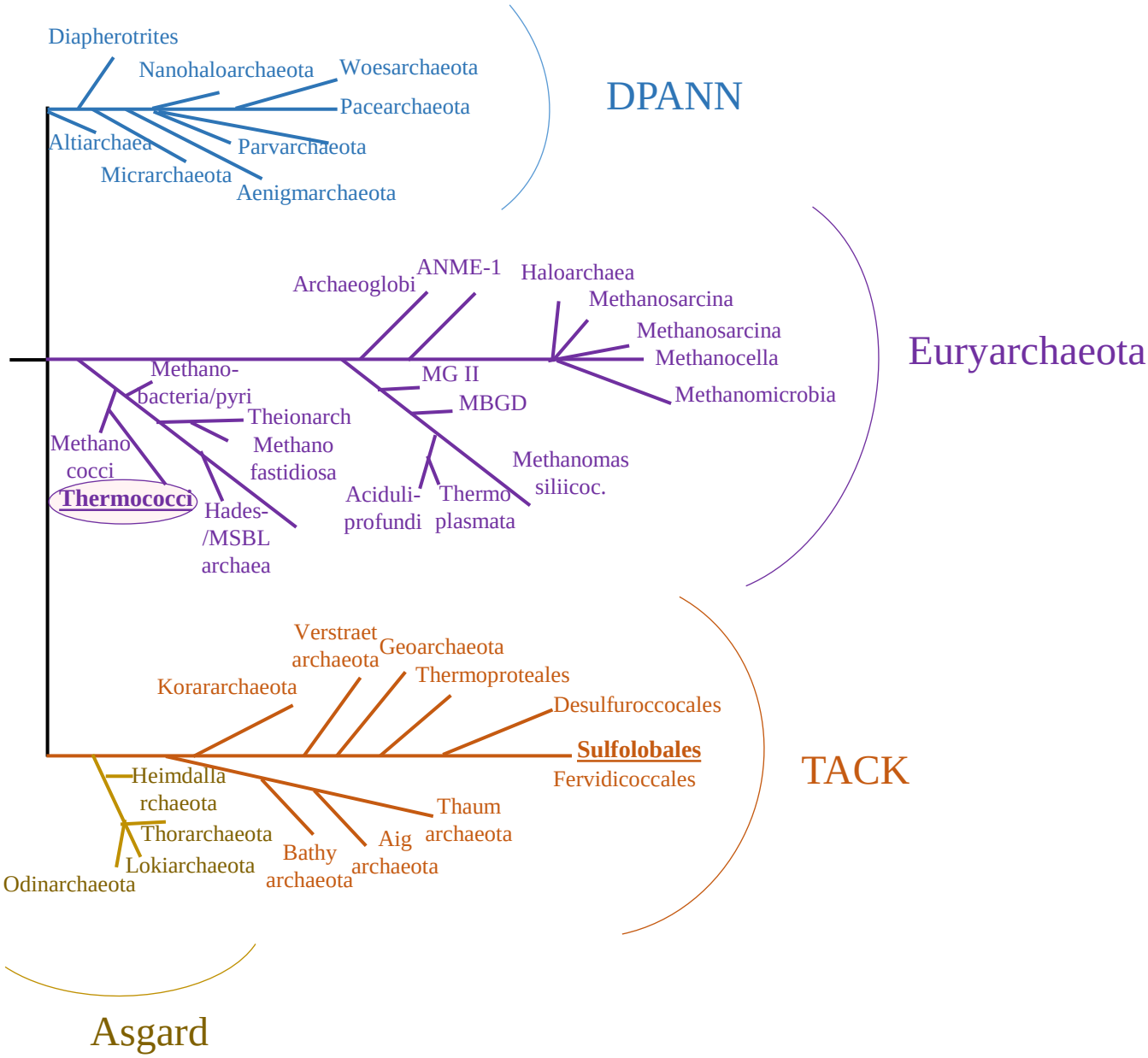


FIGURE 3. Archaeal taxonomic tree.
Schematic illustration of the archaeal phylogenetic tree. So far, archaea are divided into four major lineages: Euryarchaea (purple), TACK (orange), DPANN (blue), and Asgard (gold-yellow).
(Spang et al 2017).

2-Archaea are ubiquitous micro-organisms

At present, archaea are considered to be ubiquitous organisms present in all terrestrial environments, i.e. about 20% of the biomass of the oceans (DeLong and Pace 2001), as well as in the human microbiota, on the skin, and in the digestive and respiratory systems. However, their impact on human health remains to be determined. (Drancourt et al. 2017; Mihajlovski et al. 2010; Oxley et al. 2010; Lurie-Weinberger and Gophna 2015; Gaci et al. 2014; Brugère et al. 2014; Moissl-Eichinger et al. 2018; Koskinen et al. 2017).

The Halobacteriaceae family of archaea dominates the Vestfold Hills lake in Eastern Antarctica with a salinity of 320 $g.l^{-1}$ and temperatures between -14 and -18°C (DeMaere et al. 2013). Two orders of Euryarchaeota are particularly adapted to halophilic environments, the Halobacteriales and the Nanohaloarchaea (Zvyagintseva, I.S, Tarasov, A.L 1988). These organisms can grow at a high salt concentration of 2.5 to 5.2M of NaCl for the most extreme, and of 0.5 to 2.5M for the moderate halophiles. Psychrophilic uncultivated archaea were discovered in the oceans where the water temperature is around 4°C (DeLong 1992; Fuhrman et al. 1992). Euryarchaeota were also identified in mesophilic environments (temperature up to 50°C), such as rice fields for certain methanogens (Knittel et al. 2005).

In terrestrial or maritime hot springs environment and hydrothermal vents, hyperthermophilic archaea belonging to both Crenarchaeota and Euryarchaeata superphyla were isolated (Gugliandolo and Maugeri 2019). Non-cultivated Crenarchaeota were also found in non-thermophilic environments such as the marine water column, the gut of a deep-sea cucumber, and in lake sediments (Besseling et al. 2019)(McInerney et al. 1995)(MacGregor et al. 1997).

2-1 Archaea of the geochemical terrestrial cycles.

Euryarchaeota are mostly strict anaerobic and autotrophic, and some heterotrophs can be found (Schönheit et al. 2016). Methanogenesis is the main metabolic process in autotrophs. Methanogenic archaea gain energy by reducing C1 and C2 compounds, including CO2, formate, acetate, methanol, ethanol, methylamines, and methyl sulfides to methane (Garcia et al. 2000; Liu and Whitman 2008). Based on substrate use, traditional methanogenic Euryarchaeota lineages are characterized as hydrogenotrophic (H2 and CO2), aceticlastic (acetate), methylotrophic (X-CH3), and H2-dependent methylotrophic (H2 and X-CH3). Methanobacteriales, Methanococcales, Methanopyrales, Methanocellales, and Methanomicrobiales are strictly hydrogenotrophic (Liu and Whitman 2008). Some members

of the Methanosarcinales perform hydrogenotrophic methanogenesis (Welander and Metcalf 2005) but they are also capable of aceticlastic and/or methylotrophic methanogenesis (Fournier 2009). Furthermore, Euryarchaeota includes sulfate-reducers as Archaeoglobus (2011).

Most Crenarchaeota are anaerobic. For energy-generating metabolism, sulfur (S) is commonly used by these archaea as an electron acceptor (Leigh and Whitman 2013). Besides, many other electron acceptors are used by various species including oxygen (O2). Regarding electron donors, besides hydrogen gas (H2), aerobes may also use Sulfur as an electron donor. Many Crenarchaeal species may also use organic compounds (Leigh and Whitman 2013). Some marine Crenarchaeota are also capable of nitrification, suggesting these organisms may affect the oceanic nitrogen cycle (Stein and Klotz 2011).

Thaumarchaeota which can perform the oxidation of ammonia to nitrite represents the only archaeal taxonomic group that resides in large numbers in the global aerobic terrestrial and marine environments on Earth (Kimble et al. 2018). Archaea which are sulfur-dependent are confined mostly to hot environments. Acidophiles and anaerobic non-thermophilic methane oxidizers have a potential impact on the environment as they release metals. However, the metabolisms of a large number of archaea remain to be explored (Offre et al. 2013).

Hence, archaea are important for the cycle of organic and inorganic carbon, nitrogen, and sulfur, and influence greenhouse gas emission. Anaerobic archaea can perform important steps in the carbon cycle and exclusively methanogenesis and anaerobic methane oxidation (Offre et al. 2013).

2-2 The archaeome

It is now known that host-associated microbial communities have an important role in shaping the health and fitness of plants and animals. It was recently shown that the archaeal community, the so-called archaeome, is now an important component of host-associated microbiomes. The archaeome is composed of various lineages, Euryarchaeota dominates the archaeal species in animals (Janssen and Kirs 2008; Snelling et al. 2014; Shi et al. 2015). Methanogenic archaea have been detected in the mammalian gastrointestinal tract with a diversity limited to four orders: Methanomicrobiales, Methanosarcinales, Methanobacteriales, and Thermoplasmatales (Moissl-Eichinger et al. 2018).

Using high-throughput sequencing approaches, members of Euryarchaeota and Crenarchaeota orders were identified within the human gut microbiome (Gill et al. 2006; Oxley et al. 2010).

So far, only *Haloferax massiliensis* strain could be successfully isolated and characterized from the human gut (Khelaifia and Raoult 2016).

Archaea showed positive effects on human health. *Methanomassiliicoccales* strains might be used as probiotics against metabolic disorders associated with TMA (trimethylamine) produced by gut bacteria (Borrel et al. 2017). A recent molecular and culture-based study has also shown the presence of methanogens in brain abscesses (Drancourt et al. 2017), which means a possible involvement of archaea in the pathogenicity of this severe infection. Thaumarchaeota represents a detectable part of the human skin microbiome (Probst et al. 2013). The ammonia-oxidizing Nitrososphaera could potentially be associated with a reduction in smell and an improvement in the skin constitution. Using quantitative PCR, the abundance of archaea on human skin was shown to be influenced by human age and skin physiology (Moissl-Eichinger et al. 2017).

Although no archaeal species causing disease has been identified yet (Bang and Schmitz 2015), the methanogenic archaea in the gastro-intestinal tract contribute to methane production and have been recently shown to potentially be involved in disease-relevant processes (Borrel et al. 2020).

Archaea represent a significant component of the plant microbiome**.** Methanogens and ammonium-oxidizing archaea have been found to colonize the rhizosphere and the roots of plants at high abundances, as they can provide anoxic micro-niches (Herrmann et al. 2008). Also, recent studies on the natural vegetation of alpine bogs revealed that the plant genotype also influences colonization by archaea (Herrmann et al. 2008; Aminov 2013)

3-Specific archaeal characteristics

Archaea are unicellular microorganisms whose size varies between 0.1 and 15µm (Eun et al. 2018). Archaeal cells have a variety of shapes. Some of them have a typical round (shell) (Zinder and Dworkin 2006) rod (bacillus) or shapes (Huber et al. 1987), comparable to those observed in bacteria. Archaea could also have a disc, spiral, or filament shape (Burns et al. 2004; Pivovarova et al. 2002). However, despite their shape similarities with bacteria, most archaeal cells do not possess peptidoglycan and have membranes containing isoprene-based lipids attached to glycerol-1-phosphate through ether bonds (Langworthy et al. 1972; Tourte et al. 2020). Archaeal genomes have a high gene density and consist of a circular chromosome ranging in size from 0.49 to 5.75 Mbp, for *Nanoarchaeum equitans* and *Methanosarcina acetivorans*, respectively (Kellner et al. 2018; Wang et al. 2015; Koonin and Wolf 2008).

3-1 Archaeal S-layer

Most archaea have a rigid protein layer that surrounds the cell and that could help in maintaining shape and regulating intracellular chemical composition (Klingl 2014). In most cases, this outer layer is an S-layer for Surface layer composed of proteins and glycoproteins. The S-layer is found in almost all archaeal organisms excepting for Halococcus, Ignicoccus, Natronococcus, Methanobrevibacter, Methanosphaera, Thermoplasma, and *Thermosphaera aggregans* (Albers and Meyer 2011; Richards et al. 2018; Gambelli et al. 2019). The S-layer protein structures show an oblique, square, or hexagonal symmetry, depending on which the geometry and size of the pores between the S-layers (Albers and Meyer 2011). S-layers are between 5 and 25 nm of thickness and are rather straight on the outer surface, while the inner surface is wavy. All archaeal proteins in the S-layer are N- or O-glycosylated (Jarrell et al. 2014).

3-2 Archaeal lipidic membrane

Archaea have a membrane founding a barrier between the cell and its environment **(FIGURE 4).** This cell membrane is deeply different from that of bacteria and eukaryotes with four major differences (Jain et al. 2014)(Siliakus et al. 2017). The basic constituents of the archaeal membranes are phospholipids with glycerol with an L-chirality (L-glycerol) **(FIGURE 4)**. The presence of distinct stereoisomers implies the distinct enzymes having active sites of different topologies (Carbone et al. 2015). The nature of the bond between glycerol and the side chains of phospholipids also differs. Archaea use an ether bond to link isoprenoid hydrocarbon side chains to the *sn*-glycerol-1-phosphate backbone, while in bacteria and eukaryotes fatty acid side chains are linked via an esther bond (Corcelli et al. 2012)(Jain et al. 2014) (**FIGURE 4).** The side chains of the lipids in bacterial and eukaryal membranes are composed of fatty acids of 16 to 18 carbons, while archaea have isoprene chains of 20 carbons. The unique membrane lipids of archaea have been implicated in the survival and adaptation of the organisms to extreme environments (Jain et al. 2014).

Recently, Asgard suggests a possible clue to eukaryotic membrane complexity (Kalde et al. 2019; Zaremba-Niedzwiedzka et al. 2017). Asgard genomes encode a repertoire of eukaryotic-like proteins including, homologs essential for cytoskeleton as actin (Stairs and Ettema 2020), which is critical for membrane remodeling in eukaryotes (Pollard and Cooper 2009); exosome biosynthesis pathways as ESCRT-I, II, and III systems (Spang et al. 2015).;

Transport Protein Particle (TRAPP), essential for exocytosis, endocytosis, protein sorting, and cytokinesis (Kalde et al. 2019; Zaremba-Niedzwiedzka et al. 2017); regulatory systems as small GTPases (Dey et al. 2016), also implicated with BAR domains, and longin domains in building, maintaining, and defining many membrane-bound compartments (Rout and Field 2017); protein transporter, as the coatprotein complex II (COPII) responsible for the transport of protein cargoes from the Endoplasmic Reticulum (ER) to the Golgi apparatus (Schlacht and Dacks 2015).

3-3 the Archaeal motility structure: the Archaellum

Archaea possess a unique motility structure termed archaellum (Aminov 2013; Shahapure et al. 2014). The flagellum in bacteria and the archaellum are both considered as a rotating structure. Although their compositions and structural organizations are fundamentally different (Jarrell and Albers 2012). The direction of the rotation of the motility structure in archaea, is performed by the chemotaxis system in a manner similar to the flagellar system in bacteria (Briegel et al. 2015) except that the chemotaxis system of archaea requires an adaptor protein to allow communication with the archaellum motor (Quax et al. 2018). The archaellum allows archaea to look for favorable conditions for their growth by moving towards nutrients and oxygen (Quax et al. 2018). For example, *Halobacterium salinarum* has an archaellum at one or two poles depending on the growth state (5 to 10 filaments on one pole under optimum growing conditions and 5 to 10 filaments on each pole in deficiency conditions) (Tarasov et al. 2000).

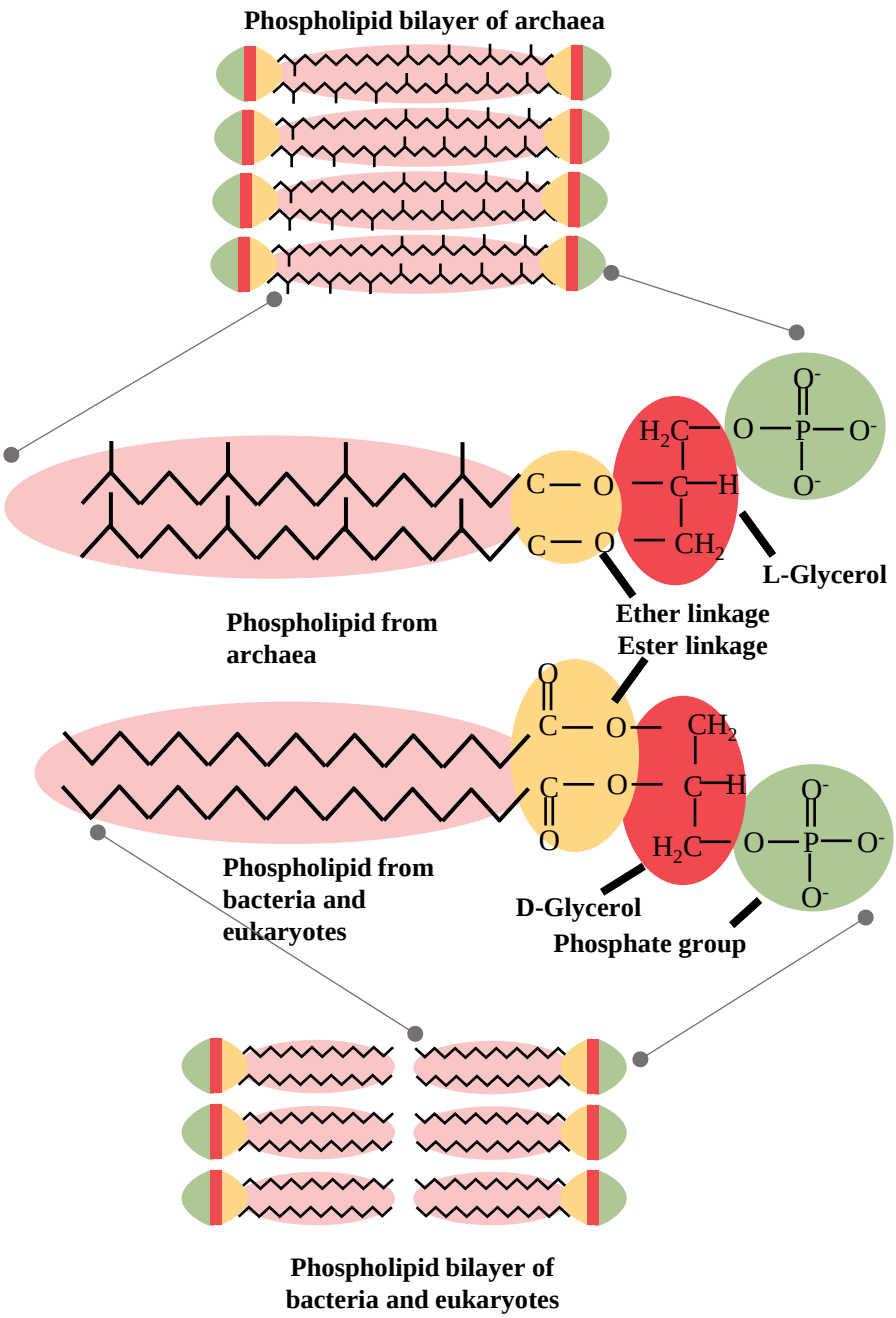


FIGURE 4. Lipids of archaeal, bacterial and eukaryotic cell membrane.
The lipids found in the archaeal membrane are fundamentally different from those found in eukaryotic and bacterial membranes. In eukaryotes and bacteria, the glycerol moiety is ester-linked to an sn-glycerol-3-phosphate backbone, whereas in archaea the isoprenoid side chains are ether-linked to an sn-glycerol-1-phosphate moiety. The figure shows the structure of monolayer-forming tetraether lipids found in some archaea, for example, the thermoacidophilic archaeon *Thermoplasma acidophilum*, in which the hydrophobic core consists of C40C40 caldarchaeol.
(Sonja-Verena Albers and Benjamin H. Meyer 2011)

4-Archaeal study models

Because Archaea and Eukarya share a related-informational system (see Part 1-**FIGURE 2**), studies on archaeal models helped to discover new functions in eukaryotes. For instance, the identification of a new family of DNA topoisomerase type II in archaea contributed to the identification of the SPO11 protein responsible of chromosome breakage during meiotic recombination in Eukarya (Bergerat et al. 1997). Importantly, several hereditary diseases or cancers are caused by mutations in genes encoding human proteins that have homologous counterparts in archaea (Shin et al. 2014). Therefore studying archaeal homolog proteins, for example, in hyperthermophilic archaea, facilitates protein structure-function analyses (Vieille and Zeikus 2001). In this case, due to the stability of thermophilic proteins, while expressed in mesophilic hosts, thermophilic or hyperthermophilic enzymes resist heating treatment and chemical denaturants (such as a solvent or guanidinium hydrochloride). Finally, enzymatic assays at high temperatures allow using higher substrate concentrations, lower viscosity, fewer risks of microbial contaminations, and often higher rates of reaction.

Nevertheless, the main challenge for using archaea as a study model lies in the difficulties in the cultivation of archaeal strains and the lack of genetic markers which have impeded the development of genetic tools in archaea (Farkas et al. 2013; Adam et al. 2017; Baker et al. 2020). So far, it is possible to transform some Crenarchaeal and Euryarchaeal strains, to delete or replace genes using the Pop-in Pop-out technique, and to express proteins carrying a tag *in vivo* (Allers and Mevarech 2005; Farkas et al. 2013; Atomi et al. 2012).

The development of a genetic system in these archaea offered a new tool for their usage as study models. Owing to their thermostable enzymes, thermophilic archaea of kingdoms of Euryarchaeota and Crenarchaeota, as mentioned, have always been of interest to biochemists and structural biologists. They offer significant potential for biotechnology, and for researchers wishing to use a multidisciplinary approach that combines genetics with biochemistry. Therefore, genetic systems have been developed for methanogens (Kohler and Metcalf 2012) and halophiles (Leigh et al. 2011) as well as thermophilic Euryarchaeota (Thiel et al. 2014), Thermococcales, and Crenarchaeota (Portnoy and Schuster 2006; Wagner et al. 2012). We will cite here some examples of archaeal strains used as study models for different purposes **(Table 1)**.

Sulfolobus acidocaldarius is part of the aerobic hyperthermophilic Sulfolobales which are the most developed Crenarchaeota models due to their potential to be easily transformed, their sensitivity to a wide range of antibiotics and their low mutation rate despite high-temperature

environment (Wagner et al. 2012)(Aagaard et al. 1996) **(Table 1)**. For example, *S.acidocaldarius* was used to study transcription machinery, replication proofreading at high temperatures (Langer et al. 1995; Pühler et al. 1989) (Grogan et al. 2001; Grogan 2003; Lundberg et al. 1991). In addition, special features of *S. acidocaldarius* including its capacity to grow synchronously in culture and intercellular chromosomal gene exchange have facilitated studying archaeal cell cycle (Aagaard et al. 1995; Grogan 1996). We can also mention that *Sulfolobus islandicus* and *Sulfolobus solfataricus*, another Sulfolobales strains, help to investigate host-virus interactions and catabolic enzymes as GAPDH (Glyceraldehyde-3-phosphate dehydrogenase), and Phosphoglycerate kinase (Bräsen et al. 2014). (Li et al. 2016; Zhang et al. 2016, 2013a)(Reno et al. 2009; Held and Whitaker 2009)(Albers and Driessen 2008). However, *S. solfataricus* genome houses several hundred mobile elements which cause significant genetic instability (Brügger et al. 2002).

Haloferax volcanii is an aerobic and mesophilic archaea with an optimum growth temperature of 45°C **(Table 1)**. It is a strict halophilic archaea adapted to salty environment (optimal concentration of 2.5 M). The genome of *H. volcanii* DS2 was completely sequenced (Hartman et al. 2010). The H26 is the strain commonly used in the laboratory. This strain has a deletion in the pyrE2 gene encoding orotate phosphoribosyltransferase, involved in the de novo synthesis of uracil, and is used as a selection and counterselection marker (Bitan-Banin et al. 2003). The easy growth conditions, the short generation time, and simple genetics have made *H. volcanii* a model organism for haloarchaeal biology with a wide number of genetic tools which allows insertion-deletion of sequences via homologous recombination (Bitan-Banin et al. 2003), protein labeling (Zatopek et al. 2018) and inactivation of specific genes expression via the CRISPR system (Maier et al. 2019). Gene expression reporters based on β-galactosidase and green fluorescent protein (GFP) are also available in *H. volcanii*. Finally, *H. volcanii* is useful for the production of biotechnologically attractive halophilic enzymes with application potentials in sustainable fine chemical synthesis, bioprocessing, bioremediation, biofuel, and bioplastic production (Haque et al. 2020)

Pyrococcus abyssi is a hyperthermophilic archaea isolated for the first time in hydrothermal vents located in northern Fiji at a depth of 2000m (Erauso et al. 1993) **(Table 1)**. Its optimum growth temperature is 96°C with a doubling time of 33min. For the cultivation of this strain at a temperature of 90°C under anaerobic conditions and at a pH of 6.4 are required. The genome of *Pyrococcus abyssi* is one of the first to be sequenced and annotated (Cohen et al.

2003; Gao and Wang 2012). Its chromosome is around 1.7 Mbp with a GC percentage of around 44%, codes for over 1800 proteins, and has a plasmid (pGT 5). Although genetic tools have not yet been developed for this strain, it is considered a model organism in the study of DNA maintenance (Pluchon et al. 2013). The stability and resistance of its thermophilic proteins make it a model organism for biochemistry.

Thermococcus barophilus is an hyperthermophilic archaea of the order of thermococcales, closely related to *P. abyssi*, which serves now as a standard genetic model which was developed by M. Jebbar's team and used in this study **(Table 1)** (Thiel et al. 2014). *T. barophilus* strain MP was the first true hyperthermophilic piezophile archaeon isolated in 1993 from a deep-sea hydrothermal vent at depth of 3550 meter on the Mid-Atlantic Ridge (Marteinsson et al. 1999). The name *Thermococcus barophilus* has Greek roots, thermo for heat, kokkos for the spherical cells, baros for weight, and philos for loving. Overall, the name means "organism with a spherical body that gravitates to heat and to the pressure of the water column." The genome of *T. barophilus* has been sequenced and annotated (Vannier et al. 2011), and strains with multiple genetic backgrounds have been constructed (Thiel et al. 2014; Marteinsson et al. 1999; Birien et al. 2018). *T.barophilus* is anaerobic and sulfur-metabolising archaeon and grows in a rich medium from 48°C to 100°C, with an optimum at 85°C, and within a pressure range of 0.1 to 85 MPa, with an optimum of 40 MPa. *T.barophilus* is an obligate piezophile for temperatures over 95°C (Thiel et al. 2014). The strain overexpresses a heat shock protein when grown at low pressure (0.3 MPa), indicating that this organism could sense low pressure as a stress (Vannier et al. 2011, 2015). The genetic manipulations established in *T.barophilus* rely on uracil auxotrophy and simvastatin resistant as selectable markers, using pop-in/pop-out method (Thiel et al. 2014). In this context, *T.barophilus* is now considered as an appropriate model to decipher the molecular mechanism in Thermoccales. In this context, in this study, we proceeded to perform biochemical, transcriptomics, and proteomics studies on aLhr2 from *T. barophilus* strain.

Table 1 Archaeal organism models.

Phylum	**CRENARCHAEA**	**EURYARCHAEA**		
Order	Sulfolobales	Halobacteriales	Thermococcales	
Strain	*Sulfolobus acidocaldarius*	*Haloferax volcanii*	*Pyrococcus abyssi*	*Thermococcus barophilus*
Habitat	Terrestrial solfataric springs	Sea Lake ocean	deep-sea hydrotherma vent	chimney hydro-thermal vent
Growth	acidophilic Aerobic 80 °C	3-4 M salt Aerobic 45°C	Anaerobic 96 °C	High-pressure Anaerobic 85°C
	Genetics Molecular Biology	Genetics	Biochemistry	Genetics Molecular Biology

Some characteristics of the main archaeal study model

5-Mosaic aspects of archaeal informational system

The archaeal informational systems i.e. molecular machines implicated in replication, transcription, translation, recombination, and DNA repair processes are closely related to those identified in eukaryotes with the addition of bacterial-like features (Brochier-Armanet et al. 2011; Lyu and Whitman 2017). In this context, archaeal informational systems feature some mosaic aspects.

5-1 Replication

Archaea have a circular chromosome, similar to that of bacteria (Raymann et al. 2014; Kelman and Kelman 2014; Yao and O'Donnell 2016), with one or more origins of replication (ORC) depending on the species (Kelman and Kelman 2018). The replication is initiated by the formation of the Cdc6-MCM protein complex at the origin of replication **(Table 2)**. Cdc6 is homologous to the eukaryotic Cdc6 protein and allows the recruitment of the MCM helicase (MiniChromosome Maintenance) which locally unwinds the DNA duplex and initiates the formation of the replication fork (Sakakibara et al. 2009) (Kelman et al. 1999; Costa and Onesti 2009; Kelman et al. 2020). This complex then recruits the GINS and GAN factors, which play an essential role in the establishment and maintenance of replication forks (Marinsek et al. 2006; MacNeill 2010). GAN (for GINS-associated nuclease) is a protein of the DHH phosphoesterase family and is homologous to the eukaryotic Cdc45 protein (Makarova et al. 2012; Oyama et al. 2016; Marinsek et al. 2006; Nagata et al. 2017). Once the double-stranded DNA (dsDNA) has been unwounded by MCM helicase, the RPA proteins (for replication protein A) bind and overlap the exposed single-stranded DNA (ssDNA) (Dickey et al. 2013). Some archaea contain one to several homologs of eukaryotic RPA, as in *Pyrococcus furiosus* and *Pyrococcus abyssi* which include a trimeric complex RPA41/ RPA14/ RPA32 (Komori and Ishino 2001) or *Methanocaldococcus jannaschii* (Kelly et al. 1998) which has only one. In other archaea the RPA proteins are similar to bacterial SSBs or contain both bacterial and eukaryotic homologs (Wadsworth and White 2001; Raymann et al. 2014). DNA synthesis in archaea is performed by two DNA polymerases, PolB (homologous in eukaryotes) and PolD (unique for archaea) (Greenough et al. 2015). As these DNA polymerases move only in the 5 'to 3' direction and the unwound DNA duplex is antiparallel, primase (p41 and p46) synthesized the discontinuous strand in the form of a series of Okazaki RNA fragments (Bocquier et al. 2001; Chemnitz Galal et al. 2012; Frick and Richardson

2001). The PCNA factor (for proliferating cell nuclear antigen) is recruited by the Replication Factor C (RFC) and combines with DNA polymerase to give it a high processivity. PCNA forms a trimeric ring that encircles dsDNA and then binds to DNA polymerase for fast and efficient DNA strand synthesis. PCNA also interacts with other factors of nucleic acid metabolism and the cell cycle, via its PIP motif and is also involved in DNA repair (Pan et al. 2011). Fen1 endonuclease and Lig1 DNA ligase are then recruited for RNA primer removal and DNA ligation, respectively (Balakrishnan and Bambara 2013; Henneke 2012). Data on the termination of archaeal replication is limited, the circular nature of their chromosomes suggests a mechanism similar to that of bacteria that terminates in a *ter* region of the chromosome using the Tus protein (Duggin et al. 2011). However, neither the presence of a *ter*-like sequence nor that of a protein homologous to Tus has been described in archaea.

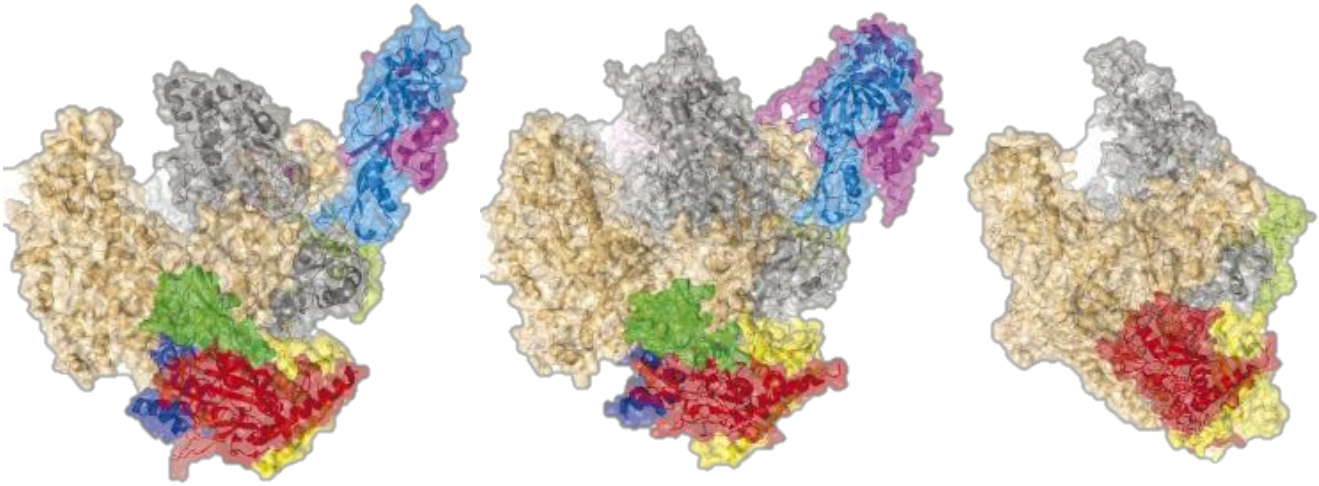
Table 2 Archaeal replication machinery

Function	Protein complexes
Origin recognition	Cdc/ORC
Replicative helicase	MCM
Helicase loader	Cdc6
Pre-IC	GINS
ssDNA-binding protein	RPA
	SSB
Elongation complex	Primase
Sliding clamp	PCNA
Clamp loader	RFC
Replicative DNA polymerase	PolB/D
Primer removal	Fen1

Proteins implicated in DNA replication in archaea. Protein complexes highlighted in green have homologues in Eukaryotes.

A.

Archaea **Eukarya** **Bacteria**



Sulfolobus solfataricus (PDB 2PMZ*)*

Saccharomyces cerevisiae (PDB 1NT9)

Thermus aquaticus (PDB 1I6V)

B.

Bacteria	Archaea	Eukaryote RNAPII
β’	Rpo1’/Rpo1’	RPB1
β	Rpo2’/2”	RPB2
α	Rpo3	RPB3
α	Rpo4	RPB11
ω	Rpo5	RPB6
	Rpo6	RPB5
	Rpo7	RPB8
	Rpo8	RPB10
	Rpo10	RPB12
	Rpo11	RPB4
	Rpo12	RPB7
	Rpo13	RPB9

FIGURE 5. RNAP subunits in the three domains of life.
A.The X-ray structures of archaeal RNAP, eukaryotic RNAPII, and bacterial RNAP. **B.** Subunits composition of RNA polymerases in the three domains of life. The coloured circles indicate corresponding subunits in the structures above.
Adapted from (Werner 2008)

5-2 Transcription

The transcription machinery in Archaea is a unique RNA Polymerase (RNAP) composed of 10 subunits in Euryarchaea and 12 subunits in Asgard and TACK (**FIGURE 5A**). In addition, crenarchaeal RNAPs have an extra-specific subunit, Rpo13 (Korkhin et al. 2009). These subunits are homologuous of the subunits of the eukaryotic RNA polymerase II (**FIGURE 5B**) (Bell et al. 1998)(Hirata et al. 2008). For a reminder, in eukaryotes, RNA polymerase I (RNAPI) transcribes rRNAs, RNA polymerase II (RNAPII) mRNAs and non-coding RNAs, and RNA polymerase III (RNAPIII) 5S rRNAs, tRNAs, and some snRNAs (Armache et al. 2013). Basal transcription factors that direct transcription initiation and elongation are similar between archaeal RNAP and RNAPII (Rowlands et al. 1994; Qureshi et al. 1997; Hausner and Thomm 2001; Gehring et al. 2016) (**FIGURE 5B**).

The transcription cycle includes three steps, initiation, elongation, and termination (Werner 2013) **(FIGURE 6)**. In archaea, three elements are required during initiation, the B-recognition-element BRE and TATA box, both are DNA sequences that are recognized by basal transcription factors, TFB «Transcription Factor B» and TBP «TATA-Binding Protein», respectively, and the transcription start site (TSS) at +1 position **(FIGURE 6)**. These elements influence subsequent steps in the promoter opening more specifically the RNAP recruitment. The pre-initiation complex composed of TBP, TFB, and RNAP, is critical and sufficient for the formation of the nascent transcription bubble and for RNA synthesis in archaea *in vitro* **(FIGURE 6)** (Werner and Weinzierl 2002). In contrast to the TBP and TFB factors, the identified BRE consensus differs strongly between archaea (Bell et al. 1999)(Lagrange et al. 1998), the TATA-box consensus shows strong conservation from archaea to eukaryotes. (Babski et al. 2016; Jäger et al. 2014; Li 2015; Wurtzel et al. 2010). The TFE factor, by its binding to the RNAP clamp, facilitates the separation of DNA strands and the loading of the DNA template strand on RNAP (i.e. Opening of RNAP clamp and formation of transcription bubble) **(FIGURE 6)**. The formation of the transcription bubble by the TFE is facilitated by the AT-rich region between the BRE element and the +1 position requiring less energy for the DNA strands unwinding (Smollett et al. 2017). This organization is close to that found in bacteria (Zuo and Steitz 2015). To begin the elongation step in archaea, the pre-initiation complex recruits Spt4/5, homologous of the bacterial NusG **(FIGURE 6)**. The substitution of TFE by Spt4/5 at the same RNAP binding site allows the transition from initiation to transcription elongation (Grohmann et al. 2011).

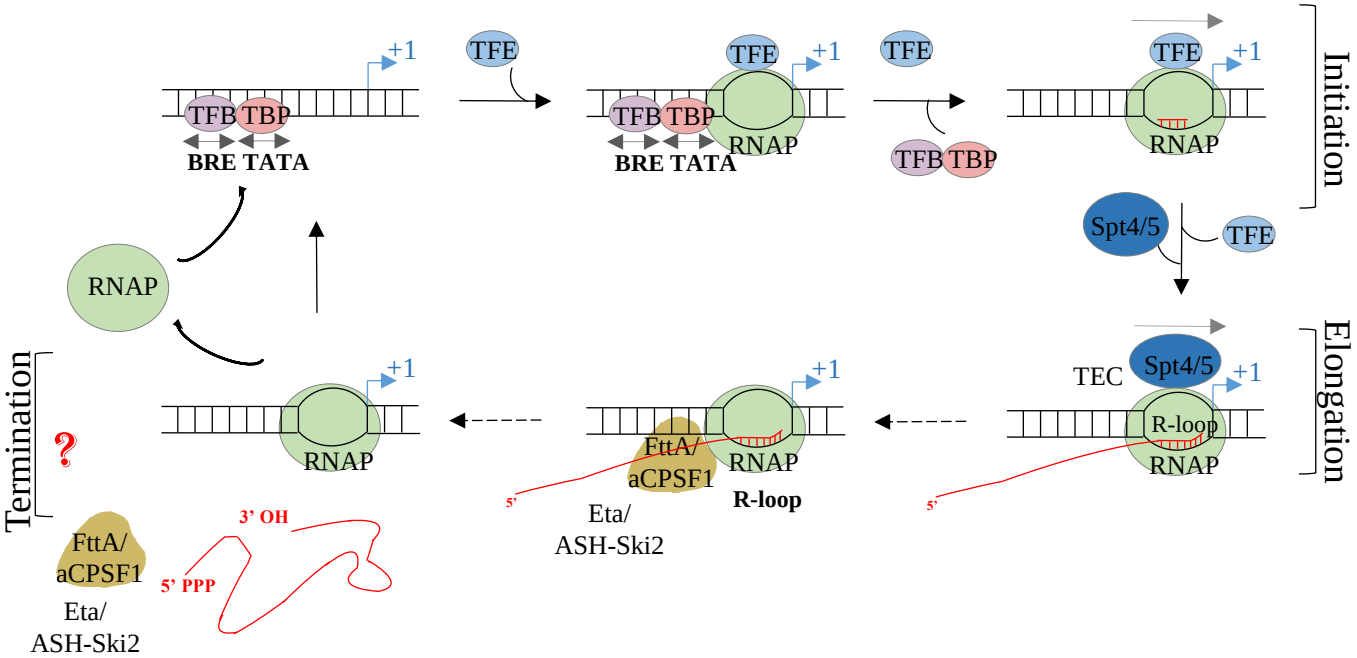


FIGURE 6. Archaeal transcription machinery.
Schematic representation of the archaeal transcription machinery including RNAP and other transcription factors during initiation, elongation, and termination steps.
TBP: TATA-binding protein; BRE: The B recognition element; TFIIB: Transcription factor II B; RNAP: RNA polymerase; TEC: transcription elongation complex.

Transcription termination remains the least understood step in RNA synthesis in archaea. In bacteria and eukaryotes, several DNA sequences and factors are able to induce the dissociation of the elongation complex (EC) triggering the termination of transcription. In archaea, the aCSPF1 protein which is the homolog of eukaryotic CPSF73 possesses an endoribonucleolytic with preference to CA dinucleotides (Phung et al. 2013). The archaeal homologous of the RPB1 and RPB2 subunits of RNAP have been found in the aCPSF1 interaction network of *Pyrococcus abyssi* (Phung Thesis, 2017). A recent study has shown that aCPSF1 (named FttA in the publication) is able to cleave and release nascent transcripts originating from the elongation complex in *Thermococcus kodakaraensis* and proposes a mechanism in which aCPSF1/FttA would be responsible for the termination of transcription (Sanders et al. 2020). However, a functional interaction between RNAP and aCPSF1/FttA remains to be shown. The mechanisms of transcription termination in archaea remain to be determined and may vary among archaeal groups.

The majority of sequenced Archaeal genomes encode multiple orthologous of general transcription factors responsible for regulating the level of gene expression in the cell (Peeters et al. 2015). Transcription factors may act by activating or repressing the transcription. For example, the binding of transcriptional repressor to the sequence overlapping the BRE element or the TATA-box prevents the access of TBP/TFB basal transcription factors (Lee and Young 2013; Lipscomb et al. 2009; Dahlke and Thomm 2002). Usually, transcriptional activators bind to a primary operator site upstream or partially overlapping the BRE element, so-called UAS (Upstream Activating Sequence), stimulating the recruitment of TBP /TFB to the promoter through protein-protein interactions (Wilkinson et al. 2010; Ochs et al. 2012).

Most of archaeal transcriptional regulators identified so far possess bacterial-like DNA binding domains, such as RHH (ribbon-helix-helix) and HTH (helix-turn-helix) motifs (Aravind and Koonin 1999; Pérez-Rueda and Janga 2010). Based on in silico analyzes, it is predicted that the Leucine Responsive-Regulatory Protein (Lrp) family, a well-known family of regulators in bacteria, is widely present in archaea (Brinkman et al. 2003). Lrp-like proteins have been annotated in all available genomes of archaea and are considered one of the most abundant families of proteins regulators of transcription (Pérez-Rueda and Janga 2010; Wurtzel et al. 2010; Charoensawan et al. 2010). Archaeal Lrp-like regulators usually bind close to the promoter region, allowing the formation of a ternary complex TBP-TFB-Lrp.

These regulators can affect the formation of the PIC in a positive and negative way (Peeters and Charlier 2010).

In eukaryote, chromosomal DNA is structured by histones, which together form nucleoprotein complexes called nucleosomes. Among archaea, almost all known Euryarchaeota genomes encode counterparts of histones. Some Crenarchaeota genomes encode for proteins capable of structuring DNA in such a way similar to histones (Mattiroli et al. 2017). In *M.jannaschii*, the UAS is located directly upstream of the rubredoxin (rb2) promoter downstream of the TATA box. Both histones and the Lrp-type regulator Ptr2 have been shown to bind this region (Ouhammouch et al. 2003; Ouhammouch and Geiduschek 2001). Ptr2 transcription factor binds to the UAS and activates transcription by recruiting TBP to the promoter, which leads to high expression levels. In the absence of Ptr2, the histone tetramers bind to the UAS resulting in transcriptional repression. The binding of TBP and TFB can overcome transcriptional repression mediated by the histones (Ouhammouch et al. 2003; Ouhammouch and Geiduschek 2001).

Other transcription factors are also widespread in archaea, such as MarR, ArsR/SmtB and TrmB (Lee et al. 2008; Lemmens et al. 2019; Itou et al. 2008; Kumarevel et al. 2009). So far, only a minor group of all existing archaeal regulators has been characterized.

5-3 Translation

Even though archaeal mRNAs have bacterial features (no introns and a 5' triphosphate at the 5'end), archaea have a translation apparatus close to those described in eukaryotes (Lecompte et al. 2002; Londei 2005). Translation of mRNAs starts with the recruitment of the small subunit of the ribosome and the initiation factors.

Ribosomes are ribonucleoprotein particles composed of conserved rRNAs that carry catalytic activity (ribozyme) and ribosomal proteins. They consist of a large and a small subunit, the 30S and 50S subunits in archaea, forming a binding site for aminoacylated tRNAs (site A), a binding site for deaminoacylated tRNA (site P), and a tRNA exit site (E site). The 30S subunit consists of 16S RNA and 25 proteins and allows the sequence of the mRNA to be decoded. The 50S subunit contains 23S and 5S RNA and 39 ribosomal proteins which structure the catalytic site. Ribosomal proteins of archaea are either universal or homologous to those of eukaryotes (Lecompte et al. 2002; Armache et al. 2013; Ban et al. 2000, 2014).

Archaeal initiation factors are aIF1, aIF1A, aIF2, and aIF5B which are homologous to eukaryotic factors **Table 3** (Kyrpides and Woese 1998; Benelli and Londei 2011; Schmitt et

al. 2019). aIF1 contributes to the identification of the initiation codon, aIF1A occupies the A site and aIF2 allows binding of the tRNAi to the P site. Once the initiator codon is recognized, the large 50S subunit is recruited using the factor aIF5B. Another factor aIF6, homologous to factor eIF6, is certainly involved in translation but its role in archaea as in eukaryotes is not clearly identified (Benelli and Londei 2009). The elongation steps are mainly controlled by two proteins with GTPase activity found in the three domains of living organisms: aEF1A and aEF2 **Table 3**. During the first step, an amino-acylated tRNA enters the A site of the ribosome and recognizes the codon of the mRNA with the help of the factor aEF1A. The second step is the transpeptidation during which the amino acid carried by the tRNA in site A is added to the nascent peptide chain carried by the tRNA in the site P. aEF1B allows the recycling of aEF1A-GDP into aEF1A-GTP. Translocation is the last step and causes the reciprocal movement of the ribosome and the mRNA which causes the shift of the codon in the A site. It is the binding of the aEF2 factor which allows the translocation of the ribosome and the release of the tRNAs from the site E. The termination of translation and the recycling of ribosomes have been little described in archaea. Termination occurs when the ribosome recognizes one of the three stop codons (UAA, UAG or UGA) at site A and requires the intervention of the release factor (RF) aRF1 which is homologous to eukaryotic eRF1 (Kobayashi et al. 2012). The dissociation of ribosome subunits and their recycling is coordinated by the ABCE1 protein which is homologous to eukaryotic ABCE1 (Barthelme et al. 2011).

Table 3 Translation factor in Archaea

	Name	Function	Eukaryotic homologues
Initiation factors	aIF1A	Unkonwn/ A site of ribosome	aIF1A
	aIF5B	Adjusts tRNAi in P site	aIF5B
	aIF1	Recognition start codon	aIF1
	aIF2	Position tRNAi/ P site	aIF2
	aIF6	Ribosomal subunits dissociation ?	aIF2
Elongation factors	aEF1A	aminoacyl-tRNA to ribosome-GTP	eEF1A
	aEF1B	Recycling of aEF1A-GDP/ aEF1A-GTP	eEF1B
	aEF2	GTP hydrolysis	eEF2
Termination factors	aRF1	Recognize of stop codons	eRF1
	ABCE1	Dissociation of ribosome subunits	ABCE1

The function and eukaryotic homologous of archaeal translation initiation and elongation factor.

5-4 Post-transcriptional regulation of gene expression

The study of archaea is very important to understand the evolutionary history of life, but also, they are used as a simplified context to the study of eukaryotes, notably for studying the fundamental processes conserved across all domains as RNA metabolism. Even after 40 years of research on archaea, numerous elements of the mechanism of post-transcriptional regulation of the gene expression are still to be determined.

The key players in RNA processing and degradation are the ribonucleases (RNases). RNases catalyze the exo- or endoribonucleolytic cleavage of phosphodiester bond and act in a complex with other enzymes such as RNA helicases, poly (A) polymerases, and pyrophosphohydrolases. To date, only ten families of RNase have been reported (Clouet-d'Orval et al. 2018). Endoribonucleolytic activities are carried by the universal RNase H, RNase P, and RNase Z endonucleases which are involved in tRNA and rRNA maturation processes. Shared with eukaryotes, the EndA and Nob1 RNase are also important in those processes.

The 3'-5' exoribonucleolytic activity i.e. degradation of RNAs from their 3 'end, is critical in all three domains of life. In archaea, this activity is provided by two 3'-5' exoribonucleases: the RNA exosome and the aRNase R. The archaeal exosome, which is homologous to the eukaryotic exosome, is found in a vast majority of archaea groups except for the Methanomicrobials, Haloferacales, Halobactériales, and Methanococci groups (Clouet-d'Orval et al. 2018). The core of the archaeal exosome is composed of a trimer of Rrp41/Rrp42-dimer which forms the barrel, the catalytic site is carried by Rrp41. The exosomal cap is a heterotrimer composed of the RNA binding proteins Rrp4 (which have an RNA binding domain S1 and KH) and /or Csl4 (have an S1 and ZN domain) (Witharana et al. 2012). aRNase R, homologous of the bacterial RNase R, is found only in Methanomicrobials, Haloferacales, and Halobacteria (Clouet-d'Orval et al. 2018).

The β-CASP family, which comprises RNA cleavage (CPSF73 and RNase J) and DNA repair (Artemis, SNM1 and PSO2) enzymes, is part of the metallo-β-lactamase superfamily, found in the three domains of life (Dominski et al. 2013). Among the β-CASP enzymes, two groups have emerged based on their cellular functions and phylogenomic, one is homologous to the eukaryotic protein CPSF73 and the other to the bacterial RNase J protein. RNase J exonuclease is found in bacteria, Euryarchaea, and plants. In bacteria, the RNase J ribonuclease has a key role in rRNA maturation and mRNA turnover (Linder et al. 2014; Durand and Condon 2018; Mathy et al. 2007)(Dominski et al. 2013). The 5'-3'

exoribonucleolytic activity of aRNase J is very processive *in vitro* and was firstly identified in Thermococcales and Methanococci (Clouet-d'Orval et al. 2010; Levy et al. 2011). Recently it was shown that aRNase J is specifically engaged with the ASH-Ski2 helicase, homologous of the eukaryotic Ski2, and the RNA exosome (Phung et al. 2020) **(FIGURE 7)**. This underlines the mosaic aspect of archaeal RNA machinery that harbor RNases and helicases with homologs in eukaryotes and bacteria. More specifically, aRNase J, ASH-Ski2 and the RNA exosome in *Pyrococcus abyssi* have an extended shared protein network (Phung et al. 2020). Interestingly in the network of *Paby*-ASH-Ski2 and the RNA exosome, a partner annotated as a helicase of the Large-related-helicase super family (Lhr2) was identified **(FIGURE 7)**. One of the purposes of this PhD study is the study of the structure-function of Lhr2.

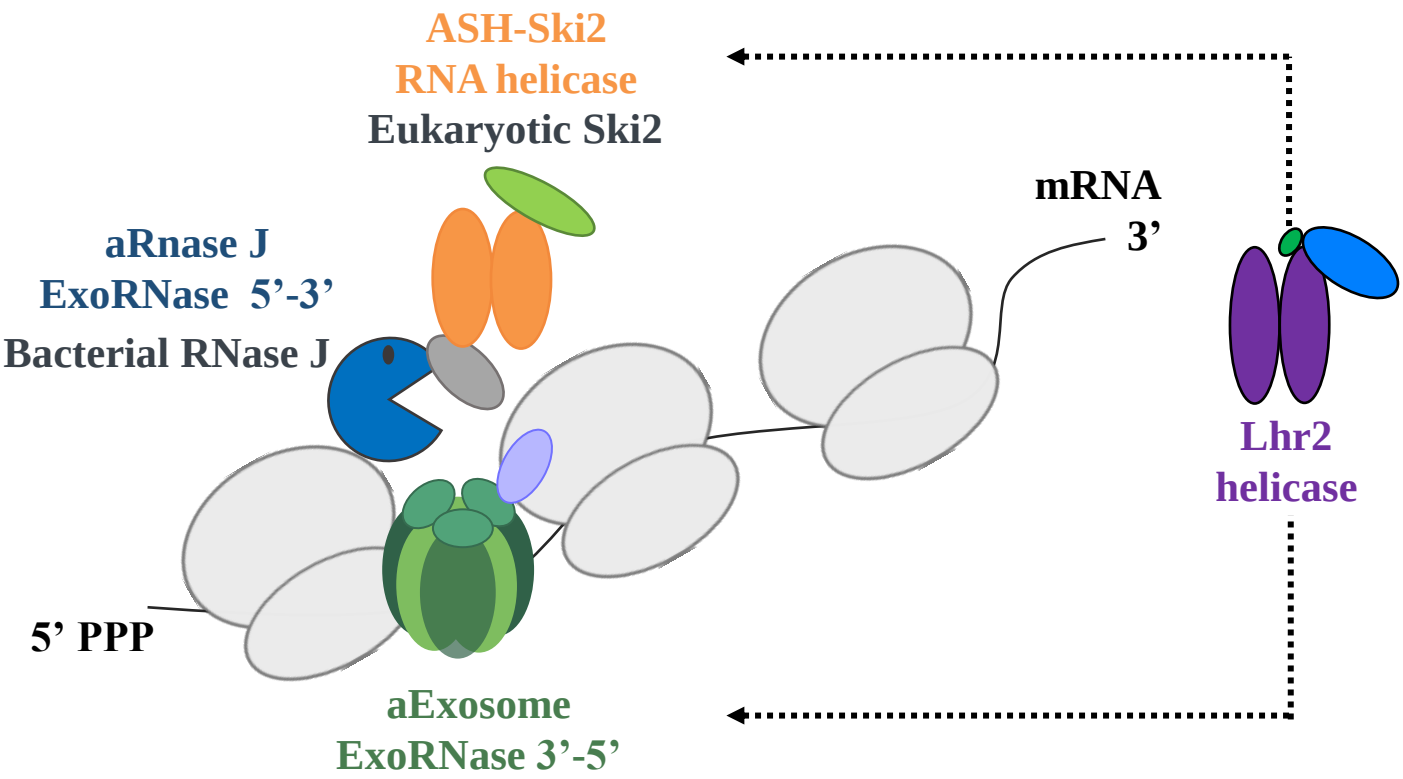


FIGURE 7. Mosaic molecular machinery involved in RNA metabolism in Archaea.
The complex formed by the exoribonuclease aRNase J and ASH-Ski2 helicase is involved with the RNA exosome in mRNA processing in Archaea. Lhr2 helicase is found in the protein interaction network of ASH-Ski2 and the aExosome.

II-HELICASES

1-Discovery of Helicase enzymes as cellular motors

The discovery of proteins capable of ATP-dependent enzymatic unwinding of duplex DNA was first reported in the late 1970s (**FIGURE 8)** (Abdel-Monem et al. 1976; Abdel-Monem and Hoffmann-Berling 1976). The Helicase I was isolated from *E. coli*, but later shown to be encoded by the TraI gene of the F episome (Abdel-Monem et al. 1983). Three years after, Hoffmann-Berling coined the term "helicase" for these proteins (Kuhn et al. 1979). Thereafter, numerous DNA helicases were discovered in the three kingdoms of life: in the eukaryotic lily plant (Hotta and Stern 1978); in bacteriophage (Venkatesan et al. 1982); in mammals (Hübscher and Stalder 1985), in viruses with SV40 protein (Stahl et al. 1986), in yeast (Sugino et al. 1986); in human (Tuteja et al. 1990); and in 1992 and 1996 respectively the first mitochondrial and chloroplast DNA helicase were identified (Hehman and Hauswirth 1992). Subsequently, the first biochemically active malaria parasite DNA helicase from *Plasmodium cynomolgi* was reported (Tuteja et al. 2002) (**FIGURE 8)**.

Remarkable progress was shown over the years in developing experimental techniques to detect and measure the unwinding activity of the helicases (Brosh and Matson 2020). From the S1 nuclease digestion assay (Abdel-Monem et al. 1976) to native polyacrylamide gel resolution of helicase reaction products used for gene 4 protein (Matson et al. 1983). A similar strand displacement assay to measure helicase activity *in vitro* was also used for the bacteriophage T4 gene 41 protein (Venkatesan et al. 1982). Subsequently, Fluorescence resonance energy transfer to monitoring in real-time the unwinding helicase activity (Bjornson et al. 1994). In 2004, the Single-molecule manipulation technique using optical tweezers allowed measuring the rate, lifetime, and processivity of the UvrD helicase as a function of ATP, and for estimating the step size (Dessinges et al. 2004). One year after, a technique combining optical tweezers and fluorescence microscopy so called ''single-molecule'' was useful to understand the relationship between the three-dimensional structure of the protein and its function (Comstock et al. 2015).

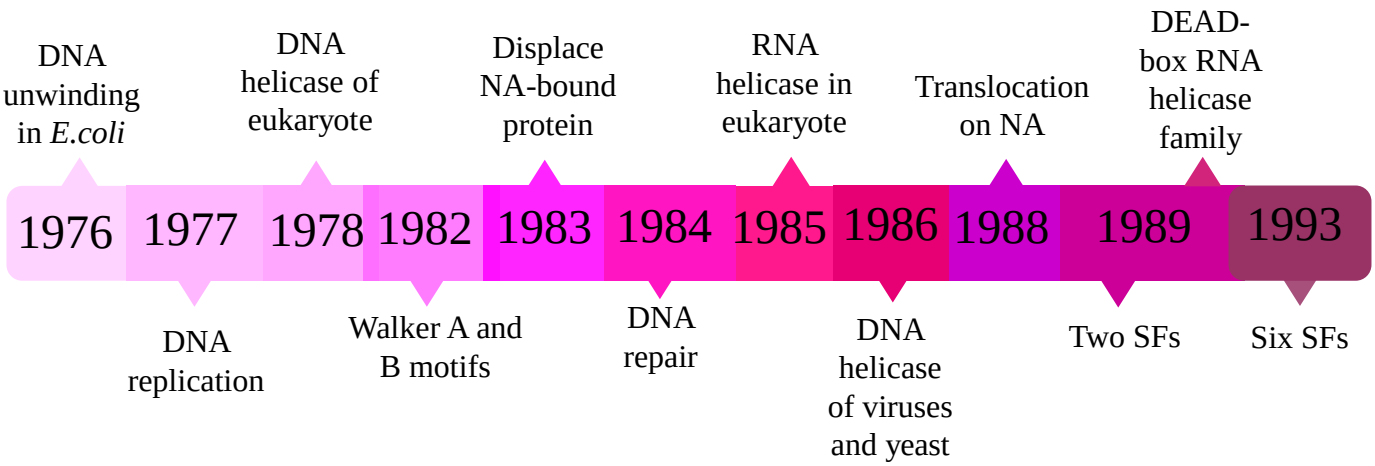


FIGURE 8. Timeline of helicases discovery.
Discovery of first bacterial DNA unwinding enzyme in **1976;** DNA helicases are involved in DNA replication in **1977;** First eukaryotic DNA helicase in **1978;** Discovery of Walker A and Walker B motifs **1982;** DNA helicases displace nucleic acid-binding protein in **1983;** DNA helicases are involved in DNA repair in **1984;** First eukaryotic RNA helicase in **1985;** First DNA helicase discovered in viruses and yeast in **1986;** Helicases translocate on nucleic acid strand in **1988;** Helicases were classified into two superfamilies in **1989;** Discovery of DEAD-box RNA helicases and classification of helicases into two superfamilies in **1989;** and the classification of helicases in six superfamilies **in 1993**.

Therefore, DNA helicases were discovered because of their ability to catalyze the separation of the complementary strands of dsDNA, this activity fulfills a function of this class of enzyme in a range of biological processes. The first studied function of helicases was in DNA replication revealed in the phage φ X174 (Scott et al. 1977) **(FIGURE 8)** and in *E.coli* where the Rep helicase was showed to use the energy of ATP to separate dsDNA at the replication fork (Aoyama and Hayashi 1986). The replication function of helicases relies also on the ability of these enzymes to displace nucleic acid-protein complex, firstly showed for the T4 Dda helicase (Bedinger et al. 1983) **(FIGURE 8)**. The translocation activity of helicases was firstly predicted for the gene 4 protein that binds and translocates on the single-stranded template (Myers and Romano 1988). The mechanism of helicases translocation on dsDNA during the replication process was demonstrated for the primosome of *E.coli* (Lee and Marians 1989). The role of DNA helicases in DNA repair was also investigated over the year. In 1984, genetic studies demonstrated the importance of *E.coli* RuvA, RuvB, and RuvC in recombinational repair (Lee and Marians 1989). Subsequently, it has been shown that the ATPase/helicase RuvB together with the DNA binding protein RuvA catalyzed HJ (Holliday junction) branch-migration in an ATP-dependent manner (Parsons et al. 1992). This discovery was significant and was the first to link human RecQ helicases to numerous diseases of premature aging and cancer (Karow et al. 2000; Constantinou et al. 2000).

From the end of the 1970s, several laboratories were trying to reconstitute eukaryotic protein synthesis *in vitro*. The first helicase shown to unwind duplexed RNA was the eukaryotic translation initiation factor, eIF4A (Ray et al. 1985) **(FIGURE 8)**. The role of this RNA helicase in translation initiation was then established in eukaryotes (Pause and Sonenberg 1993). Later, RNA helicases were shown to be involved in mRNA splicing (PRP5 and PRP28 of yeast) (Wassarman and Steitz 1991), in ribosome assembly (SrmB and DeaD of *E.coli*) (Fuller-Pace and Lane 1992), in rRNA processing (Fuller-Pace et al. 1993), and in *E.coli* RNA degradome (Py et al. 1996) suggesting that these proteins have a role in the modulation of RNA structure (Tanner and Linder 2001).

2-Helicase Classification

2-1 The six superfamilies of helicases

The first classification of helicases by Walker and his collaborators described in 1982 was based on the sequence analysis of α- and β-subunits of ATP synthase proteins, including a known bacterial ATPase, the recombination protein RecA, adenylate kinase, and myosin from both nematode and rabbit. They identified two clustered amino acid regions which shared strong similarity with other ATP-binding proteins (Walker et al. 1982) **(FIGURE 8)**. These motifs, popularly known as Walker loop or P-loop (phosphate-binding loop), occurred at positions corresponding to the proposed nucleotide-binding site in the protein crystal structure (Pai et al. 1977). The association of these sequences with actual ATP-binding appeared to be established. Thus, the motif GXXXXGKT was proposed as a common nucleoside triphosphate (NTP) binding fold present in each of the analyzed proteins. An alignment of the amino acid sequences corresponding to the Walker A and B of the bacterial RecD helicase with a viral NTP motif-containing RNA helicase was then described (Gorbalenya et al. 1988a).

In 1989 Linder and colleagues observed that several proteins (nine helicases including eIF4A) that displayed RNA-dependent ATPase activity and/or helicase activity with RNA duplexes shared conserved motifs in their amino acid sequence (Linder et al. 1989) **(FIGURE 8)**. This family named DEAD-box was then developed to include about 30 proteins in a wide range of organisms from bacteria to humans (Schmid and Linder 1991).

Basing on primary structure analyses, the presence of seven conserved amino acid motifs found in a superfamily (SF) of proteins was observed, and three superfamilies and two families of putative helicases were described (Gorbalenya et al. 1988b, 1989). the computer-aid analysis of the amino acid sequences of helicases helps classifying helicases and putative helicases: The DEAD and DEAH families were represented as two protein families within SF2. SF1 comprises a number of prokaryotic and eukaryotic putative helicases including the Senl-like family. the majority of the putative helicases of RNA viruses belong to SFl, SF2, or SF3. The SF4 helicase includes DnaB helicase and gp41 and gp4 of bacteriophage. Rho helicase is found in the SF5.

Subsequently, It has also become apparent that many nucleic acid motors are members of the AAA+ (ATPases Associated with various cellular Activities), represented as Superfamily 6 (Singleton et al. 2007). With the accumulation of the structure of many helicases families, it was possible to integrate the data from sequence, structure and mechanism to classify the

existing helicases (Singleton and Wigley 2002). As a result of their classification, Singleton and his colleagues proposed the existence of six helicase superfamilies (SF1-SF6) (Singleton et al. 2007)). In 2010, Fairman-Williams and his colleagues redefine the classification of SF1 and SF2 based on the accumulation of new structural data from several SF1 and SF2 families. This allowed to complement and extend the existing categorizations (Fairman-Williams et al. 2010) **(FIGURE 9).** The classification proposed by Fairman-William was a helpful step toward explaining common features as well as differences among specific helicases.

- **SF1 and SF2 helicases**

The two largest acknowledged superfamilies, SF1 and SF2 group the non-hexameric helicases performing diverse cellular functions. The SF1 includes three subfamilies UvrD, Pif1-like, and Upf1-like, and the SF2 comprises ten subfamilies; the RecG-like, RecQ-like, Ski2-like, T1R, Swi/Snf, RIGI-like, DEAD-box, DEAH/RHA, and NS3/NPH-II (Fairman-Williams et al. 2010) **(FIGURE 9)**. The SF1 and SF2 subfamilies include a minimum of three members from each organism (Fairman-Williams et al. 2010).

Helicases can also be subclassified according to the polarity of unwinding or translocating along with DNA/RNA strand (Fairman-Williams et al. 2010). For example, the UvrD/Rep helicases which belong to the SF1A subfamily translocate in the 3'to 5' direction (Matson 1989), while the Pif1-like and UPF1-like helicases which belong to the SF1B subfamily translocate in the 5' to 3' direction (Chakrabarti et al. 2011; Sanders 2010).

Interestingly, SF1 & SF2 subfamilies either contain exclusively DNA or RNA helicases, or both (Fairman-Williams et al. 2010) **(Table 4).** In families with both DNA and RNA helicases, individual proteins mostly function either as RNA or DNA helicase. For example, based on *in vitro* activity assays, the DEAD-box helicases bind specifically RNA substrates (Tanner and Linder 2001; Rocak and Linder 2004); while the RecQ-like, RecG-like, Swi/Snf, T1R, and Rad3/XPD bind exclusively DNA substrates (Chu and Hickson 2009; Azeroglu and Leach 2017; Dürr et al. 2005; Liu et al. 2008; Jankowsky et al. 2011). The DEAH/RHA and Ski2-like are able to target both RNA and DNA substrates (He et al. 2011; Johnson and Jackson 2013). Regarding the SF1 superfamily, only the UPF1-like helicases are able to precede both DNA and RNA substrates (Azzalin and Lingner 2006).

The hallmarks of SF1 and SF2 superfamilies is a conserved helicase core formed by the two RecA-like domains in tandem (RecA1 and RecA2) **(FIGURE 10A).** The common helicase

core of both superfamilies contains characteristic and conserved sequence motifs, where the similarity is higher within each family and the level of similarity in most motifs decreases throughout the respective superfamilies (Fairman-Williams et al. 2010). Amino acid residue motifs are conserved in at least 75% of the proteins belonging to the given superfamily of helicases. These motifs are usually clustered in a core region of 200–700 amino acids, separated by stretches of non-conserved sequences (Gorbalenya and Koonin 1993).

Table 4. Nucleic acid preference of SF1 and SF2 helicases families.

Subfamily	Superfamily	Nucleic Acid Preference	
		DNA	RNA
DEAD-box	SF2		+
DEAH/RHA		+	+
NS3/NPH-II		+	+
Ski2-like		+	+
RIG-I-like		+	+
RecQ-like		+	
RecG-like		+	
Swi/Snf		+	
T1R		+	
Rad3/XPD		+	
UvrD/Rep	SF1	+	
Pif1-like		+	
Upf1-like		+	+

Nucleic acid preference indicates the ability to unwind either DNA or RNA. (Fairman-Williams et al. 2010).

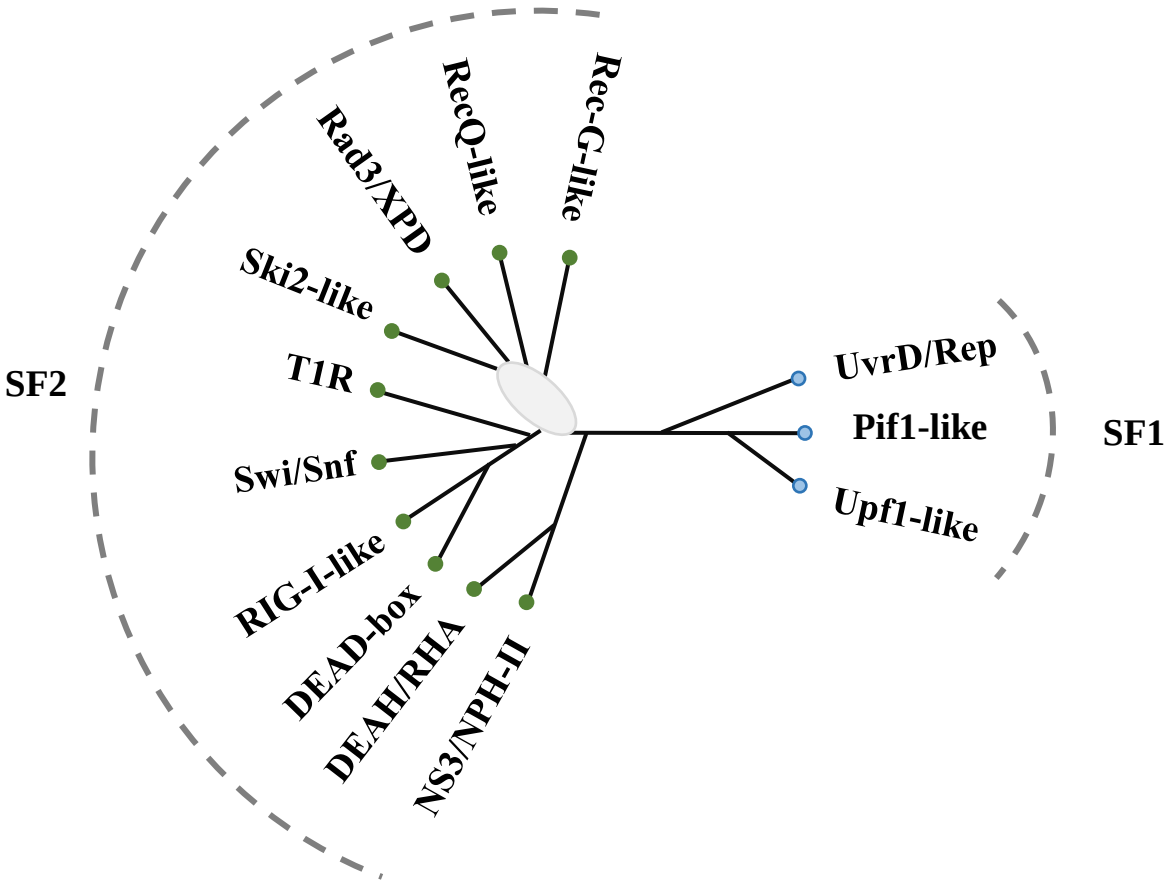


FIGURE 9. SF1 and SF2 subfamilies.
Schematic unrooted cladogram showing the three SF1 subfamilies (right), and the 10 SF2 subfamilies (left).
From (Fairman-Williams, Guenther, and Jankowsky 2010).

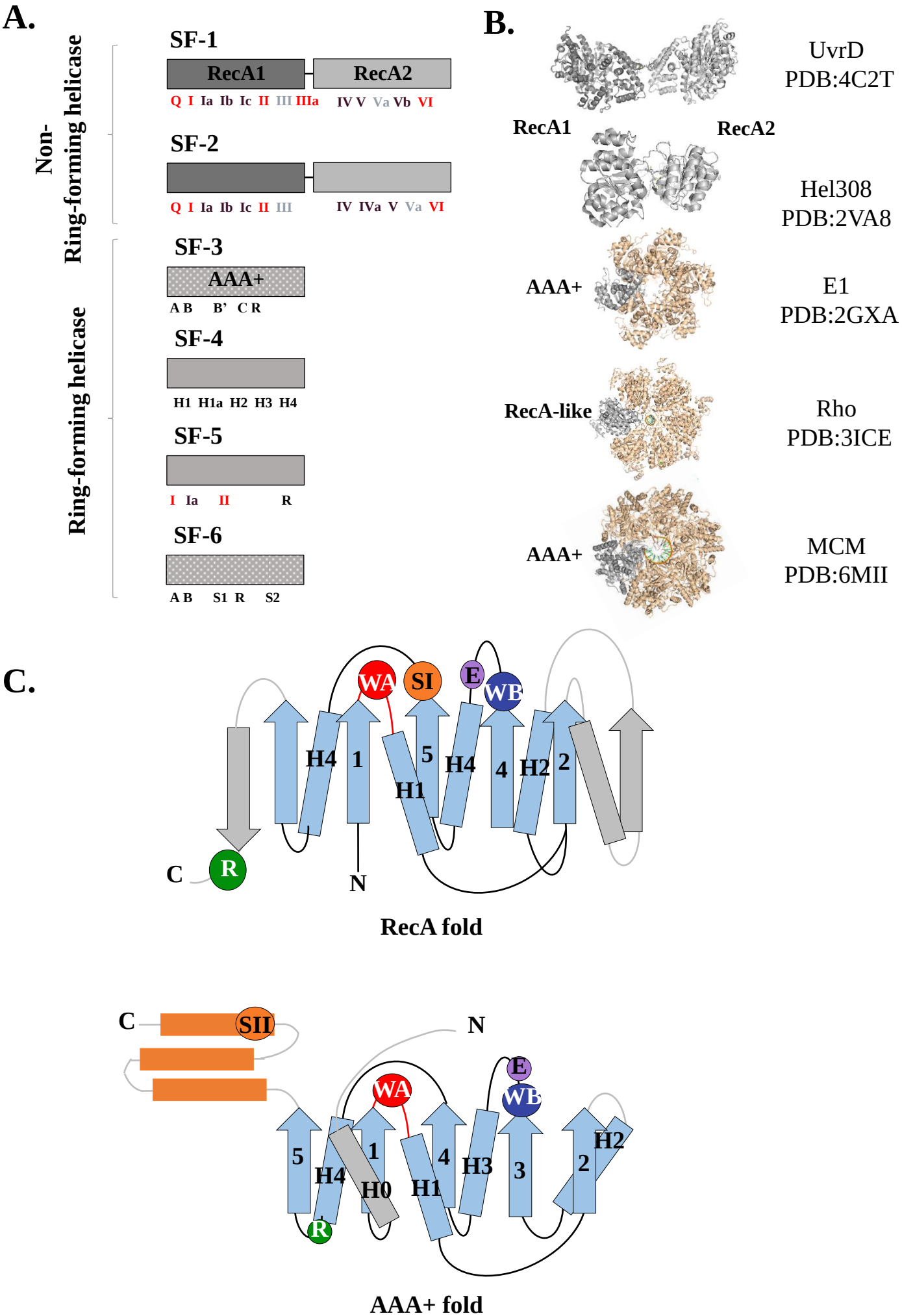

A.
Non-Ring-forming helicase
Ring-forming helicase
SF-1
RecA1
RecA2
Q I Ia Ib Ic II III IIIa IV V Va Vb VI
SF-2
Q I Ia Ib Ic II III IV IVa V Va VI
SF-3
AAA+
A B B' C R
SF-4
H1 H1a H2 H3 H4
SF-5
I Ia II R
SF-6
A B S1 R S2
B.
UvrD
PDB:4C2T
RecA1
RecA2
Hel308
PDB:2VA8
AAA+
E1
PDB:2GXA
RecA-like
Rho
PDB:3ICE
AAA+
MCM
PDB:6MII
C.
WA
SI
E
WB
H4
1
5
H1
H4
4
H2
2
C
R
N
RecA fold
C
SII
N
E
WB
WA
5
H4
1
H0
H1
4
H3
3
2
H2
R
AAA+ fold

FIGURE 10. The catalytic core of helicase superfamilies.
A. Schematic representation of the helicase core observed in each superfamily, and the distribution of the different motifs. The SF1 and SF2 possess two RecA-like domains in tandem. The SF4 and SF5 are hexameric helicases, each monomer possesses one RecA-like domain. The SF3 and SF6 helicases are also hexameric, each monomer possesses one AAA+ domain. **B.** Cartoon structural representation of monomeric and hexameric helicases with RecA-like and AAA+ domains. **C.** A comparison between the RecA and the AAA+ folds. Both RecA and AAA+ belong to the ASCE subgroup of the P loop NTPases, and comprise a 5-stranded parallel b –sheet flanked by helices both in front and behind the plane of the sheet. The secondary structure elements that are common to the ASCE proteins are shown in blue: the P loop containing the Walker A motif (WA) is shown in red, the Walker B motif in dark blue, the catalytic glutamate in violet (E), the arginine finger (R) in green and the Sensor I (SI) and Sensor II (SII) in orange. The AAA+ is characterized by a C-terminal helical domains (in orange) (Franziska Bleichert, Michael R. Botchan, and James M. Berger 2017).

Helicases of SF1 and SF2 superfamilies are generally monomeric but some candidates are able to function as monomers or dimers. Each monomeric SF1 or SF2 helicase consists of two RecA domains (Velankar et al. 1999; Maluf et al. 2003; Bujalowski et al. 1994; Lee and Yang 2006). For example, the UvrD SF1 and the SF2 Hel308 helicases form an active dimer in solution, while the T4 Dda and HCV NS3 form monomeric helicases and translocases (Stelter et al. 2013; Richards et al. 2008b)(Eoff and Raney 2006) **(FIGURE 10B)**.

The RecA-like domains belong to the ASCE subgroup of the P loop NTPases, and comprise a five-stranded parallel β-sheet, flanked by helices both in front and behind the plane of the sheet (Story and Steitz 1992; Story et al. 1992) **(FIGURE 10C)**. The Walker A motif (WA) is located between the strand β-1 and helix α-1 (Walker et al. 1982) **(FIGURE 10C)** . The Walker-B motif (WB), is located between the strand β-4 and helix α-4. An arginine finger similar to those first observed in GTPase activating proteins helps coupling ATP binding and hydrolysis to large-scale conformational changes between adjacent RecA folds **(FIGURE 10C)** (Wittinghofer et al., 1997). Another component of the active site is the conserved glutamate which primes a water molecule for the nucleophilic attack on the γ-phosphate group of ATP. The conserved glutamate occurs immediately downstream of the Walker B aspartate **(FIGURE 10C)** (Hilbert et al. 2015; Leipe et al. 2003). Finally, a polar residue at the C-terminal tip of strand β-5, sensor-II (SII), is conserved in many RecA-like ATPases and might function as a γ-phosphate sensor in order to transmit conformational changes to other parts of the protein upon ATP binding (Iyer et al., 2004b; Story and Steitz, 1992; Yoshida and Amano, 1995).

The helicase catalytic core fold is highly conserved in SF1 and SF2, but it is now clear that these enzymes perform diverse functions on nucleic acids both *in vitro* and *in vivo due to the* existence of accessory domains outside the helicase core. For example, the XPD helicase is involved in transcription and Nucleotide Excision Repair (NER) (Kuper et al. 2014) while the RecQ helicase is involved in the repair of double-strand breaks (DSBs) (Bernstein et al. 2010), even that both belong to SF2 superfamily.

- **Archaeal SF1 & SF2 helicases (book chapter Hajj et al 2018-PUBLICATION 1)**

Few studies address the role of helicases in archaea particularly at the cellular level and only a few members from the SF1 and SF2 families were characterized. In the frame of this thesis work, we review the SF1 and SF2 subfamilies of helicases from the archaeal domain in a book chapter (**PUBLICATION 1**, (Hajj et al. 2019). Phylogenomic studies from archaea sequence genomes showed the distribution and occurrence of SF1 and SF2 families in archaea (Chamieh et al. 2016). This study showed that SF1 superfamily in archaeal sequenced genomes comprises only two subfamilies, the Upf1-like and UvrD-like. No archaeal homolog of Pif1-like was identified. Similarly, no homologs of SF2 RecG-like and DEAH/RHA helicases were found in archaea, and the archaeal SF2 superfamily includes the Swi2/Snf, RIG-I, DEAD-box, Rad3/XPD, Ski2-like, Rad25/XPB, RecQ-like, UvrB, Cas3, Sfth, reverse gyrase, T1R, and Lhr helicase families. The distribution of helicases identified by genome-wide analysis by Chamieh and her colleagues in archaea, human and *E.coli* is represented in **Table 5**. Structural analysis of helicases showed the conservation of the helicase core (e.g RecA1 and RecA2) and as mentioned before the accessory domains provide substrate specificity to the helicase. Several biochemical and genetics studies were performed on archaeal helicases, in addition to determination of the three-dimensional structure of the protein. We can cite the well-studied Hef RIG-I-like helicase firstly identified from the hyperthemophilic archaeon *Pyrococcus furiosus* as a protein factor that can stimulate holiday junction resolution by the Hjc resolvase (Newman et al. 2005; Komori et al. 2002; Nishino et al. 2005). It was suggested that this helicase plays a role in processing stalled replication forks as it produces splayed duplexes from DNA forks and four-way junctions and on fork-structured DNA in vitro (Lestini et al. 2010). Hence, a book chapter was published where we reviewed the literature on the current knowledge of archaeal SF1 and SF2 helicases (Hajj et al. 2019). Many archaeal helicases remain to be studied such an SF2 specific helicase characteristic of Euryachaoeta named ASH from the ski2-like family (Manon Batista thesis, 2020) and the Lhr-type SF2 family which is the subject of this PhD work.

Table 5. Distribution of helicases in the three domains of life.

Subfamily	Superfamily	Organism		
		Archaea	Human	*E.coli*
DEAD-box	SF2	+	+	+
DEAH/RHA		+	+	
ERCC4/RIG-I/Hef/XPF		+	+	
Ski2-like		+	+	+
Cas3		+		+
RecQ		+	+	+
Rad52		+	+	
Swi/Snf2		+	+	+
T1R		+	+	+
XPD/Rad3-like/DinG		+	+	+
Lhr		+		+
Sfth		+		+
Reverse gyrase		+		
UvrD-like	SF1	+	+	+
Pif1-like			+	+
Upf1-like		+	+	+

The table represents the presence (+) or absence (-) of SF1 and SF2 helicases revealed by genome-wide analysis in archaea, human, or *Ecoli.* (Chamieh et al, 2016)

- **SF3 to SF6 helicases**

Briefly, the SF3 to SF6 superfamilies form hexameric toroid structures (ring structure) (Dillingham 2011; Lohman et al. 2008; Patel and Picha 2000) **(FIGURE 10B)**. Helicases of the SF3 comprise viral helicases for example the E1 helicase of Papillomaviruses (Enemark and Joshua-Tor 2006) (**FIGURE 10B**). The SF4 superfamily is often associated to primase enzyme during replication. This superfamily includes helicases from bacteriophage as GP4 of T7 phage, bacteria as the replicative helicase DnaB, and mitochondria (TWINKLE helicase) (Fernández-Pevida et al. 2015) **(FIGURE 10B)**. The SF5 includes the replicative and transcription termination factor Rho helicase (Thomsen and Berger 2009). The superfamily SF6 includes the replicative helicase Mini Chromosome Maintenance (MCM) of the pre-replication complex in eukaryotes and archaea and the RuvB helicase in bacteria (Meagher et al. 2019) (Byrd and Raney 2017) **(FIGURE 10B)**.

The hexameric structure of the SF3 to SF6 helicases shows six subunits (monomer), containing each one variant of ASCE. The SF4 and SF5 helicases use the RecA-like variant, while those of the SF3 and SF6 use another variant of ASCE, the AAA+ fold (Medagli and Onesti 2013) **(FIGURE 10B)**. the AAA+ domain is structurally similar to the RecA-like and possesses a distinct α/β fold (Ogura and Wilkinson 2001) **(FIGURE 10C)**. The AAA+ domain contains 200–250 amino acids with a central β-sheet in $\beta5$-$\beta1$-$\beta4$-$\beta3$-$\beta2$ strand order. The β-sheet is flanked on both sides by α-helices to form a three-tiered α-β-α sandwich. Features that distinguish AAA+ family members from other P-loop NTPases include the insertion of $\beta4$ between $\beta1$ and $\beta3$, the lack of an antiparallel β-strand adjacent to $\beta5$, and the lack of any additional strands directly adjacent to either $\beta5$ or $\beta2$ (Hanson and Whiteheart 2005)(Iyer et al. 2004b)(Iyer et al. 2004b, 2004a) **(FIGURE 10C)**. The AAA+ domain has in addition to the Sensor II, the Sensor I motif which might play a role in orienting the catalytic water molecule for the attack on the γ phosphate (Thomsen and Berger 2008).

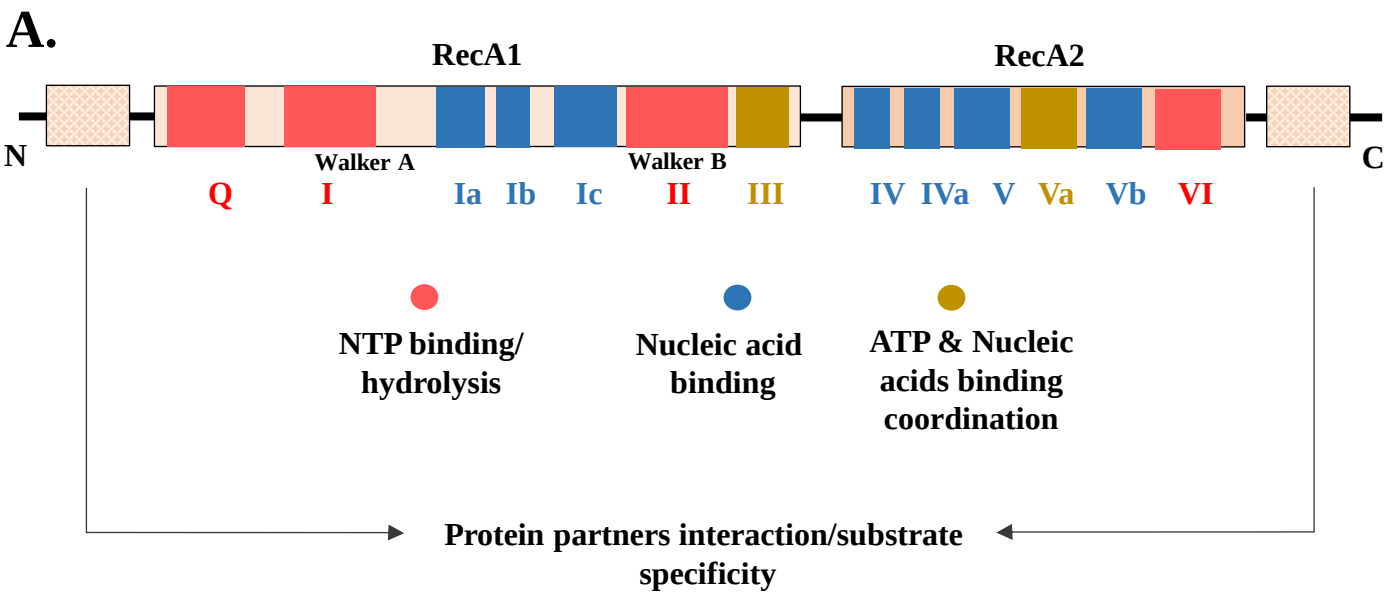
The ATP binding sites are also located at the interface between each two adjacent subunits of the hexameric helicase. The interaction between the ATP molecule and these sites brings closer the Walker A and Walker B motifs and the arginine amino acid which allow the hydrolysis of the ATP (Yoshimoto et al. 2010; Goetz et al. 1988). The ATP hydrolysis induces conformational changes between each of the two subunits which induce the movement along the nucleic acid substrate. The ring structure forms a channel that encircles

the DNA or RNA substrate which stabilizes the enzyme binding to the substrate and allows a diversity of processes (Joo et al. 2019).

Many examples exist, the unwinding activity of DNA in the 5'-3' direction, that allows the DNA polymerase to synthesize a complementary strand during the replication (DnaB in *E.coli*) (Egelman et al. 1995). The hexameric ring of RuvB, a member of the AAA+ helicase family, encircles and translocate along dsDNA and then promotes branch migration of Holliday junctions (Neuwald et al. 1999)(Miyata et al. 2000).
One example of hexameric RNA helicase is the Rho transcription termination factor. Rho binds RNA by recognizing the *rut* site (Rho utilization site), and translocate in the 5'-3' direction. Binding of Rho helicase to the transcript induces the closing of the ring to encircle nucleic acids. The interaction between nucleic acids and Rho factor induces ATP hydrolysis and translocation of the helicase (Miyata et al. 2000; Sachsenmaier and Waldsich 2013).

2-2- Structural characteristics and conserved motifs of SF1 & SF2

The conserved helicase core formed of RecA1 and RecA2 includes seven motifs located as follows: motifs Q I, Ia, II, and III in domain RecA1 and motifs IV, V, and VI in domain RecA2 **(FIGURE 11A)** (Fairman-Williams et al. 2010). The two domains are separated by a cleft, where the conformational changes induced by ATP binding and hydrolysis are expected to take place. Motif I (Walker A), with motif II (Walker B), and VI have the highest level of sequence conservation across the two SFs while the consensus sequence of the other motifs is different between SF1 and SF2 but also from one SF2 subfamily to another. These residues are located at the cleft between the two RecA-like domains (Leipe et al. 2002) **(FIGURE 11B).** Their arrangement corresponds to their functional role in ATP hydrolysis. Motif IV, Ia, Ib, Ic, IVa, V, and Vb are located at the opposite face of the ATP hydrolysis site on the two RecA-like domains and contribute to the formation of the nucleic acids binding channel (Dehghani-Tafti et al. 2019) **(FIGURE 11B)**.
The binding of ATP, in the presence of RNA or DNA substrate, results in a "closed" conformation which causes a rotation of 20° between the two RecA domains. The hydrolysis of ATP and the release of ADP leads to the "open" conformation (Jarmoskaite and Russell 2014). These changes in conformation confer to the helicases the unwinding or annealing activity (Yang and Jankowsky 2005).



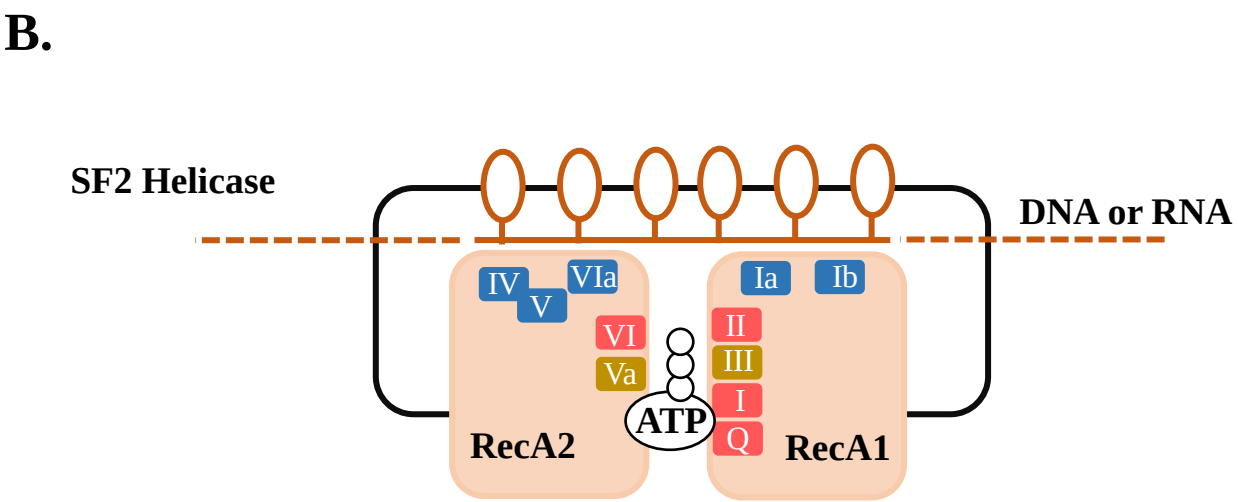


FIGURE 11. SF2 helicase core.
A. Schematic representation of the localization of the SF1 and SF2 conserved motifs on the RecA1 and RecA2 and their functions. The Q, I, II, VI are implicated in NTP binding and hydrolysis. Motifs Ia and Ib are implicated in nucleic acid binding. Motifs Ia, Ib, Ic, IV, Iva, and V are implicated in RNA substrate binding. And finally, motif III and Va are implicated in the coordination between NTP binding and hydrolysis and Nucleic acid unwinding. **B.** Schematic representations of SF2 helicase. The subdomains are as indicated, and the RecA1 and RecA2 subdomains form a motor core and adopt a "RecA"-like fold. Nucleotide cofactor is bound in a cleft between these two subdomains.

Q motif

This motif was firstly discovered in the DEAD-box helicases (Tanner et al. 2003). It consists of an amino acid sequence containing invariant glutamine (Q), hence its name. In yeast, site-specific mutagenesis studies show that the Q motif of eIF4A is involved in ATP binding and hydrolysis (Tanner et al. 2003). The glutamine amino acid of the Q motif coordinates the adenine residue of ATP, and though led to ATP hydrolysis **(FIGURE 11)**. This motif is absent in helicases with no preference for ATP such as NS3/NPH-II and DEAH/RHA (Tanner 2003). Nevertheless, in other helicases, it has been shown that this motif is involved in DNA binding and not ATP (Ding et al. 2015). Motif Q is also implicated in the affinity of DEAD-box helicases to RNA substrates and the regulation of their helicase activities (Cordin et al. 2004).

Motif I (Walker A)

In the consensus sequence (AxxGxGKT), the lysine (K) of GKT binds the β and γ phosphates of the ATP molecule, and the mutation of lysine to the uncharged amino acid asparagine abolished the binding of ATP **(FIGURE 11B)** (Pause and Sonenberg 1992). The side chain of the invariant K of motif I bound the Mg2+ ion, and the β phosphate of the bound NTP and may act to stabilize the transition state during catalysis. Mutation in this motif leads to an ATPase-deficient mutant, which also lacked helicase activity, demonstrating that ATP hydrolysis was required for the unwinding of duplex nucleic acid (Hall and Matson 1999).

Motif Ia

This motif was shown to be involved in ssDNA binding **(FIGURE 11)** (Marintcheva and Weller 2003). Mutational analysis in this motif in UL9 helicase of HSV-1 showed that mutants exhibited wild-type levels of intrinsic ATPase activity with severe defects in ssDNA-stimulated ATPase activity and ssDNA binding (Marintcheva and Weller 2003). These results indicate that residues from the Ia motif contribute to the ssDNA-binding and helicase activities of UL9 and are essential for viral growth.

Motif II (DEXX)

This motif is shared by all known and putative helicases **(FIGURE 11)**. Mutational analysis in the D of DEAD shows the implication of this motif in ATP hydrolysis, and the helicase activity of mutants was affected (Pause and Sonenberg 1992). With the growing number of identified helicases, some variations in motif II have been observed, and these are arranged in

different subgroups of SF2 such as DEAD, DEAH (Studer et al. 2020), DEXH (Jankowsky and Jankowsky 2000) and DEAH box proteins (Gilman et al. 2017).

Motif III

This motif is implicated in the coordination of ATP hydrolysis and nucleic acids binding to mechanical activity including translocation (Zhang et al. 2013b). Mutational analysis shows that mutants in this motif retained ATPase activity but lacked significant helicase activity (Graves-Woodward et al. 1997). This motif is also important for converting the energy from ATP binding and hydrolysis to RNA helicase activity and RNA binding affinity (Banroques et al. 2010).

Motif IIIa

This motif comprises a conserved tyrosine which stacks on the adenine of the ATP molecule (Fairman-Williams et al. 2010; Nishikiori et al. 2012). This motif frequently annotated as motif IV is specific for the SF1 family (Nishikiori et al. 2012)

Motif IV

This motif has a phenylalanine in its consensus sequence **(FIGURE 11)**. In DEAD-Box proteins this conserved amino acid was shown to be required for the cooperation of ATP-dependent binding of RNA substrates (Banroques et al. 2008).

Motif IVa

This motif is specific for SF2 (Fairman-Williams et al. 2010). Mutational analysis shows the implication of this motif in DNA-stimulated ATPase activity. In addition, the conserved Q residue in the QxxR motif is predicted to help in RNA binding (Dürr et al. 2005).

Motif V

This motif interacts with the sugar-phosphate backbone of the DNA. The motif V has shown to be implicated in nucleic acid affinity and in ssDNA-dependent ATP hydrolysis (Graves-Woodward and Weller 1996) **(FIGURE 11)** . This motif is also involved in the NTP hydrolysis step required for potyvirus RNA helicase activity, thus playing an essential role in virus RNA replication inside the infected cell (Fernández et al. 1997).

Motif VI

This motif is required for the nucleic acid-dependent NTP hydrolysis **(FIGURE 11)**. Mutational analysis shows that this motif has an effect on a variety of activities including

ssDNA binding, ATP hydrolysis, and ligand-induced conformational change (Sriskanda and Shuman 1998). The fact that this motif is close to both the NTP binding site and the DNA binding site suggests that it is may act to communicate between the two sites and is required for the helicase to move along the DNA substrate (Hall and Matson 1999) **(FIGURE 11B)**.

Putative helicases share these motifs but do not necessarily confirm it as an active helicase, though their activity was not yet demonstrated. For example, the SF2 NS3h translocate on ssDNA as a monomer, but the translocase activity does not correspond to comparable DNA unwinding activity or protein-displacement activity (Matlock et al. 2010).

In almost all cases, non-conserved accessory domains are attached to or inserted within the RecA-like fold and provide the specificity of action and the regulation of the enzymatic activity such as Zn-fingers, OB-folds, dsRBDs (Cui et al. 2008; He et al. 2010; Zhang and Grosse 2004) **(FIGURE 11A)**.

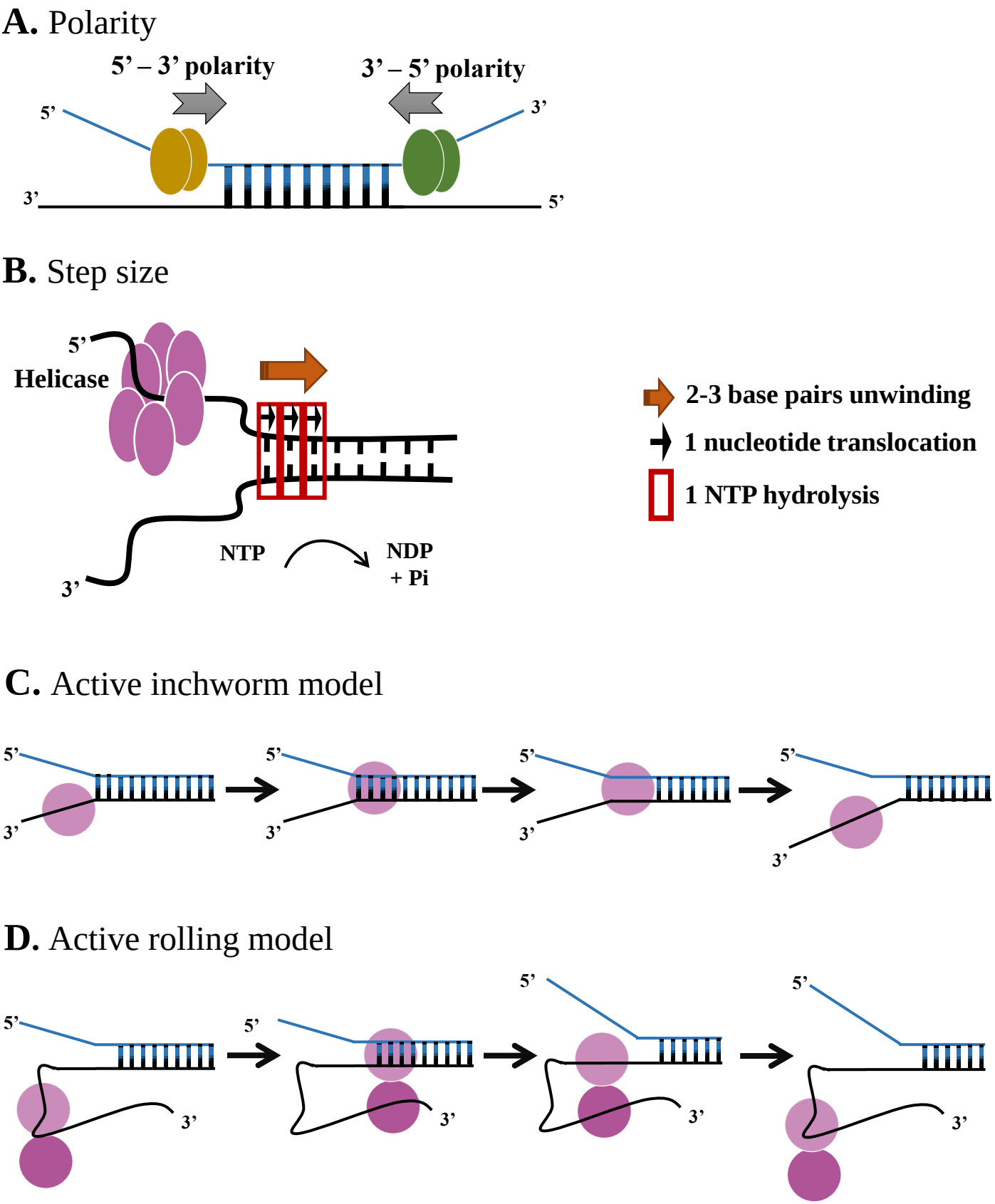


FIGURE 12. Helicases unwinding activity features.
A. Helicases (purple oval) are able to unwind duplex structure in 3' to 5' or in 5' to 3' direction. **B.** Unwinding step size depends on the number of unwounded base pair between two successive rate-limiting steps (kinetic step size) or per one hydrolyzed ATP molecule (chemical step size). Two types of active unwinding mechanism exist: The inchworm model **C.** and the rolling model **D.**

3-Biochemical and mechanistic activities

3-1 NTP binding and hydrolysis

The specificity of the hydrolyzed nucleotide varies among helicase families. Most helicase preferentially hydrolyze ATP relative to the other three nucleoside triphosphate types (UTP, GTP, and CTP) (Singleton and Wigley 2003). The conserved amino-acid motifs of helicases play a critical role in defining nucleoside triphosphate preference. For example, the Q motif of the DEAD-box helicase family confers specificity for ATP **(FIGURE 11)** (Tanner 2003). Usually, a divalent metal cation binds specially to the β and γ phosphates of the NTP (Frick et al. 2007). This metal ion is an essential component of the overall reaction of ATP binding and hydrolysis. Most helicases use magnesium as a cofactor (Jezewska et al. 2005). But other metal ions may substitute Mg2+, as for the UvrD helicases where Co2+ could support a catalytic efficiency similar to that with Mg2+ (Curti et al. 2007). Generally, the NTPase activity of helicases is stimulated by its binding to nucleic acid substrates. For example, the Prp2 and Prp3 of the DEAH/RHA helicase family have a very low NTPase activity in absence of nucleic acid substrates. Their NTPase activity is stimulated by RNA or DNA regardless of its nucleic acid sequence (Kim et al. 1992; Tanaka et al. 2007). However, some helicases display an NTPase activity independently of the presence of nucleic acid. This is the case of the replicative helicase DnaB, which is able to hydrolyze all NTPs molecules even in the absence of DNA substrates (Roychowdhury et al. 2009).

3-2 Unwinding activity

The energy from NTP hydrolysis is required for helicase binding to the substrate and subsequent activities (Singleton and Wigley 2003). The unwinding activity of a helicase is a complex process that can be characterized by numerous features such as polarity, processivity, step-size, and mechanism of action.

The polarity of unwinding is defined by the orientation of the protein translocation on nucleic-acid strand, from the 5’ to 3’ or 3’ to 5’ **(FIGURE 12A)**. Most helicases show a unidirectional polarity even though some bipolar helicases were identified (Wong et al. 2013) such as HerA, a helicase of thermophilic archaea (Constantinesco et al. 2004). *In vitro*, the unwinding activity of most of the studied helicases requires a single loading single-stranded DNA or RNA (Rajagopal and Patel 2008; Singh et al. 2019) to unwind double-stranded

nucleic acids and are unable to unwind duplexes with blunt ends. Exceptions are BLM and *E.coli* RecQ helicases, that are able to unwind blunt-end duplex DNA structure (Hishida et al. 2004; Mohaghegh et al. 2001). For blunt-ended duplex processing, some helicases were suggested to bind DNA duplex and to generate a ssDNA loading strand for ATP-dependent DNA translocation and unwinding. For example, the Lhr protein of *Methanothermobacter thermautotrophicus* is able to remodel fork DNA substrate which would yield the ssDNA needed to trigger the ATP-dependent DNA translocation, thus unwinding the fork (Buckley et al. 2020).

The processivity of helicases is defined by the number of base-pair separated before the release of the enzymes from the DNA or RNA substrate (Chakrabarti et al. 2019). Some helicases exhibit a high processivity with a separation of several kilo base-pairs per second, as for the RecB and RecD of the RecBCD complex involved in DNA repair in *E.coli* (Lohman and Fazio 2018) and the Upf1 helicase, a highly processive helicase that tightly holds onto NA, allowing long-lasting action (Kanaan et al. 2018). While other helicases are able to unwind only some base pair before the release of the enzyme from its substrate (Kanaan et al. 2018). The processivity of a helicase depends on the oligomeric state and the type of nucleic acid substrate (Patel and Donmez 2006). For example, the Pif1 helicase is able to unwind a duplex substrate in its dimeric state only, while the monomeric state of the protein is inactive on double-stranded substrate (Barranco-Medina and Galletto 2010). Also, some helicases show a higher processivity to unwind RNA/DNA duplexes than DNA/DNA duplexes substrates (Byrd and Raney 2017).

The step size of helicases is defined in two different ways **(FIGURE 12B)**. 1-The kinetic step corresponds to the number of bases or base pairs unwound between two successive rate-limiting steps that are repeated during DNA unwinding (Lucius et al. 2003). For exemple, a kinetic step size of 3bp means that the helicase proceeds through a slow conformational change that occurs every three base pairs. 2-The chemical step size refers to the number of base pairs unwound per one molecule of hydrolyzed ATP (Eoff and Raney 2010).

Mechanism of action Two different mechanisms for unwinding activity was described, active, and passive. In a passive mechanism, the helicase is not involved directly in the destabilization of the dsDNA but waits for the base pairs to open spontaneously by thermal fluctuations before it moves and binds the newly opened bases (Manosas et al. 2010). This

type of mechanism is attractive for helicases that can move and occupy one base at a time Since the terminal bp at the junction opens and closes at a very fast rate (Guéron and Leroy 1995; Patel and Donmez 2006). The active mechanism requires the binding of the helicase to both ss and duplex DNA at an unwinding junction simultaneously (Lohman and Bjornson 1996; Xie 2016). Different types of active mechanisms of unwinding were described. The «inchworm model» (Yarranton and Gefter 1979) **(FIGURE 12C)**, and the «rolling model» **(FIGURE 12D)** (Lohman and Fazio 2018; Lohman and Bjornson 1996) require the functional helicase to possess at least two DNA binding sites (Lohman 1992, 1993). These sites accommodate intermediates in which two regions of ssDNA are bound simultaneously to the helicase. These two models were proposed for Rep/PcrA helicases (Bird et al. 1998; Velankar et al. 1999). A third model, the «torsional mechanism» in which the enzyme binds simultaneously the two strands of DNA at the ssDNA/dsDNA junction to unwinds the dsDNA has been also described (Lohman 1992). Finally, regarding the unwinding of the ring-shaped helicases, it was proposed that one of the single strands passes through the center of the ring-shaped protein, while the complementary single strand passes along the outside of the ring (Patel and Picha 2000).

3-3 Annealing activity

Some helicases were shown to catalyze a strand-annealing activity (Bansbach et al. 2009; Yusufzai and Kadonaga 2010) **(FIGURE 13A)**. For example, the BLM helicase (Bloom Syndrome Helicase) possesses a DNA annealing activity suggesting a role in the processing of Okazaki fragments during DNA replication by inhibiting illegitimate recombination (Bartos et al. 2006). Moreover, the annealing activity of the p68 and p72 RNA helicases (yeast nuclear DEAD-box RNA helicases) rearrange the RNA secondary structure (Rössler et al. 2001). The annealing activity, contrary to the unwinding activity, appears to be independent of ATP hydrolysis (Chamot et al. 2005) **(FIGURE 13 A&B)**. The annealing activity of helicases is influenced by the substrate length. For example, when the duplex length of the complementary region of the nucleic acid substrate increase, the annealing activity is higher, as in the case of BLM helicase (Machwe et al. 2006). The DNA unwinding and strand annealing activities can be uncoupled. For example, the annealing activity of RecQ5β helicase is comprised in its C-terminal region (Garcia et al. 2004). Furthermore, studies of RECQ4 protein revealed that some missense mutants lose unwinding activity but still possess strand annealing activity (Xu and Liu 2009).

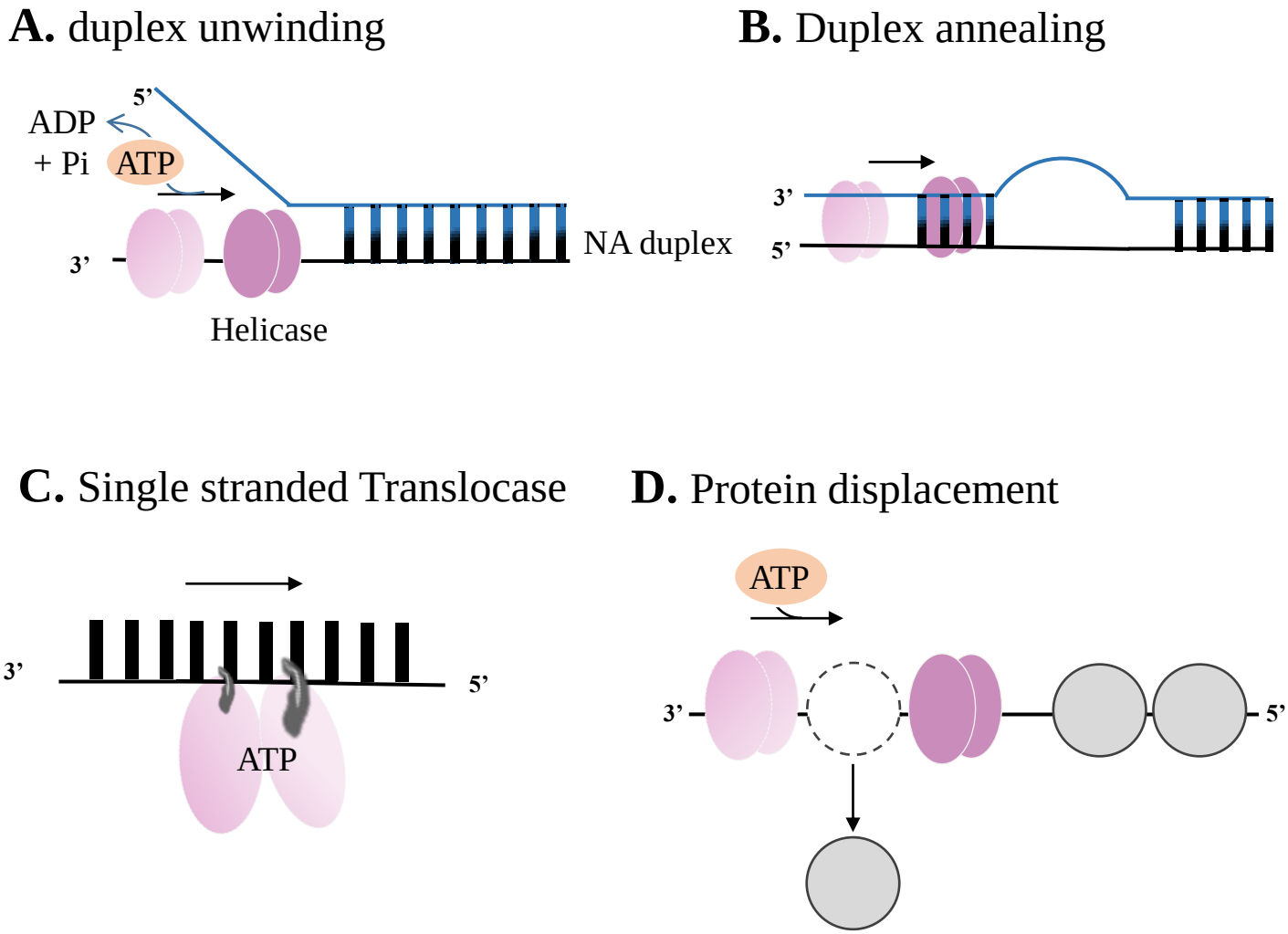


FIGURE 13. Mechanistic functions of helicases.
A. Helicases (purple oval) use ATP hydrolysis energy for duplex unwinding **B.** In the absence of ATP, some helicases are able to anneal nucleic acid duplex. **C.** Helicases use the energy of ATP hydrolysis also for translocating and protein displacement along nucleic acid substrate **D.** The inchworm model of translocation on ssDNA is represented. The alternative tight and lose interactions are indicated by the hand icons (Gao et al 2020).

3-4 Translocation and protein displacement

Monomeric SF1 and SF2 helicases used inchworm mechanism to translocate along with nucleic acid. This mechanism implicates the helicase to move by alternating affinities between domains (RecA1 and RecA2) and ssDNA **(FIGURE 13C)** (Gao and Yang 2020). Each helicase has at least two ssDNA attaching points (Lee and Yang 2006). The binding of ATP induces domain closure. The RecA1 and the associated DNA exiting channel loose contacts with the 3' half of ssDNA and release one nucleotide, while domain RecA2 still holds the 5' half of the ssDNA near the single-double stranded junction. During ADP and Pi releases, the RecA domains open up, and the RecA2 releases the ssDNA and allows it to move towards the exit by one nucleotide. The translocation is of one nucleotide per ATP hydrolysis (Saikrishnan et al. 2009).

During their translocation, many helicases share the ability to remove proteins from DNA for replication, recombination, transcription, or repair processes to occur **(FIGURE 13D)**. For example, the *E.coli* RecBCD helicase displaces nucleosomes during unwinding (Eggleston et al. 1995). The DExH/D helicases have been demonstrated to disrupt RNA–protein interactions without an unwinding activity (Fairman et al. 2004).

During replication, the MCM helicase functions as replicative helicase in archaea and eukaryotes (Shin et al. 2003). In eukaryotes, this helicase disrupts nucleosomes, probably with the aid of histone modifications and chaperones (Alabert and Groth 2012). When provided with ATP, the archaeal MCM helicase has the ability to unwind DNA bound by DNA-binding proteins notably the HMtA2 histone, the transcription repressor TrpY, and the transcription pre-initiation complex and dislodge proteins from dsDNA (Shin et al. 2007).

Head-collisions with RNAPs cause fork pausing during the replication (Merrikh et al. 2011). *in vitro,* the dsDNA translocase Mfd provides a potential backup mechanism after a head-on collision by displacing the RNAP (Pomerantz and O'Donnell 2010). In *E.coli*, Rep and UvrD helicases, can each promote movement of replisomes along with protein-bound DNA *in vitro* and *in* vivo (Guy et al. 2009; Boubakri et al. 2010).

4-General cellular functions of DNA & RNA helicases

Due to their dominant presence, it is necessary to understand helicases functional role in the biological processes and their interactions with various cofactors and proteins of different biological pathways.

By their mechanical activities, helicases were found to be implicated in almost all biological processes involving DNA (DNA helicase) or RNA (RNA helicase) (Pyle 2008; Jankowsky and Fairman 2007). RNA helicases are essential for most processes of RNA metabolism such as translation (Shen and Pelletier 2020), transcription (Epshtein et al. 2010), ribosome biogenesis (Martin et al. 2013), RNA splicing (Cordin and Beggs 2013), RNA transport (Gong et al. 2005), RNA editing (Li et al. 2011), RNA decay (Khemici and Linder 2018), and R-loops processing (Chakraborty and Grosse 2011). They also play an important role in sensing viral RNAs (Hardwick and Luisi 2013; Khemici and Linder 2018; Jarmoskaite and Russell 2014; Brüning et al. 2018; Linder and Jankowsky 2011; Skourti-Stathaki et al. 2011). Many cellular processes, such as transcription, DNA replication, recombination, and repair, involve the separation of nucleic acid strands that necessitates the use of a plethora of DNA helicase (Brosh and Matson 2020; Knoll and Puchta 2011).

In human, more than fifty helicases have been characterized biochemically, and around ten of them are associated with pathologies (Mohaghegh et al. 2001; Xu and Liu 2009; Cantor and Guillemette 2011). Among these are the dual RNA/DNA helicase, Senataxin, an orthologous of the yeast RNA helicase (Sen1p), which plays a role in the maturation of ncRNAs and R-loops processing (Dutta et al. 2020). Defects in senataxin cause the neurodegenerative disorder ataxia with oculomotor apraxia 2 (AOA-2) (Yüce and West 2013). The XPB and XPD subunits are associated with three genetic disorders: xeroderma pigmentosum (XP), cockayne syndrome (CS) and trichothiodystrophy (TTD) (Coin et al. 2007; Rouillon and White 2010; Richards et al. 2008a; Tirode et al. 1999).

In bacteria, genetic studies for helicases were performed and two helicases were shown to be essential for the survival of the cell: The chromosome replication fork helicase DnaB, and the PcrA DNA helicase (Subramanya et al. 1996; Velankar et al. 1999; Soni et al. 2003). All sequenced bacterial genomes encode an essential hexameric helicase equivalent to DnaB of *Escherichia coli* (Zhang et al. 2014; Marie et al. 2017).

Helicases function does not rely only on the conserved motifs, but also on accessory domains located around their structural cores which frequently contain specific additional functionalities such as DNA-binding, protein-binding, or oligomerization.

The eukaryotic DHX29 is a large multidomain helicase that carries accessory domains, the WH domain, the ratchet domain, and the OB domain, forming a complex regulatory network (Dhote et al. 2012). The WH domain performs the binding to the 40S subunit, while the OB domain mediates RNA-stimulated NTPase activity and inhibits basal NTPase activity in the absence of RNA. Mammalian DDX3 and its yeast ortholog Ded1 play roles in the export and translation initiation of mRNAs and act as translational repressors (Tarn and Chang 2009). The protein contains in addition to the helicase core N- and C-terminal extensions, which appear to mediate interactions with other proteins as eIF4G, PABP, eIF4E, and eIF4F that ensure the recruitment of Ded1/DDX3 to specific mRNAs (Hilliker et al. 2011; Shih et al. 2008; Soto-Rifo et al. 2012).

The bacterial SF2 RecG helicase contains beside the SF2 helicase domain a Wedge domain, which specifically binds at the junction of DNA arms (Singleton et al. 2001). Aromatic residues within this domain are essential for this substrate recognition. This domain is capable by itself of binding Holliday junctions with high affinity, but not replication forks (McGlynn and Lloyd 2001). In addition, when the wedge domain is deleted, the RecG still displays helicase activity, but with reduced substrate affinity (Briggs et al. 2005; McGlynn and Lloyd 2001). Thus, the wedge domain is seen as a substrate specificity factor and a processivity factor.

The XPD helicases possess a C-terminal FeS domain. When this domain is disrupted, XPD still binds to ssDNA and has DNA-dependent ATPase activity (Rudolf et al. 2006; Pugh et al. 2008). The FeS domain contains a wedge structure implicated in DNA duplex separation (Spies 2014).

The UvrB helicase is known for its function in DNA repair. The helicase comprises a β-hairpin accessory domain implicated in DNA binding and lesion detection (Webster et al. 2012). Mutagenesis studies have shown that residues of the β-hairpin domain are essential for the handover of DNA from UvrA to UvrB and damage recognition (Van Houten et al. 2005).

RNA helicases frequently function in multi-subunit complexes, for example, the degradosome complex in bacteria and the translation initiation and the mRNA spliceosomal complex in eukaryotes.In eukaryotes, The DEAD-box protein eIF4AIII is an essential part of the EJC multiprotein complex (Ballut et al. 2005). The EJC complex is probably involved in

nucleocytoplasmic transport of spliced mRNA, as well as the nonsense-mediated mRNA decay pathway (NMD) (Le Hir et al. 2000). eIF4AIII binds spliced mRNA in the exon junction complex (Shibuya et al. 2004). The ATP hydrolysis activity of eIF4AIII is inhibited in the presence of the MAGOH–Y14 dimer of the EJC complex. Although, its RNA binding retains a very high affinity (Noble and Song 2007). In *E.coli*, the degradosome is a multiprotein complex involved in the processing of rRNA and the degradation of mRNA (Carpousis 2002). RhlB helicase is an important functional component of the RNA degradosome, and it is activated by the protein-protein interactions with the C-terminal portion of the ribonuclease RNase E. A region of the C-terminal domain that lies at a distance from the helicase's catalytic site of RhlB is probably responsible for recognizing the RNase E peptide (Chandran et al. 2007). The interaction between RNAseE and RhlB boosts the ATPase activity and represses the RNA binding activity of the helicase. Similar degradosomes assemble in *B. subtilis* and *Staphylococcus aureus* containing the CshA DEAD-box RNA helicase (RhlB-equivalent), and RNases Y (functional equivalent of RNase E) (Lehnik-Habrink et al. 2011).

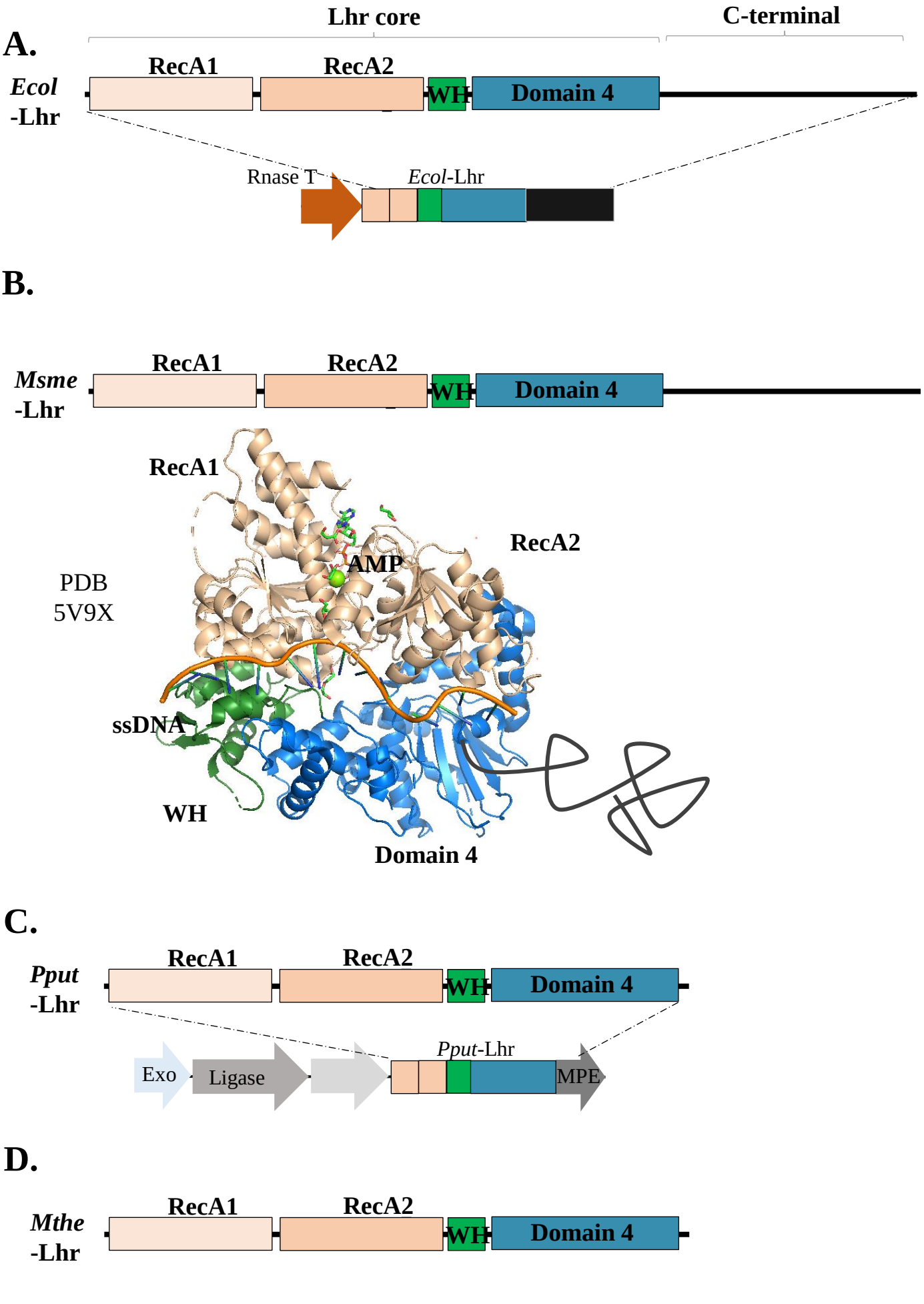

A.
Lhr core
C-terminal
RecA1
RecA2
WH
Domain 4
Ecol
-Lhr
Rnase T
Ecol-Lhr
B.
Msme
-Lhr
RecA1
RecA2
WH
Domain 4
RecA1
RecA2
AMP
PDB
5V9X
ssDNA
WH
Domain 4
C.
Pput
-Lhr
RecA1
RecA2
WH
Domain 4
Pput-Lhr
Exo
Ligase
MPE
D.
Mthe
-Lhr
RecA1
RecA2
WH
Domain 4

FIGURE 14. Studied bacterial and archaeal Lhr-type protein.
A. The primary structure of *Ecol*-Lhr protein. The gene encoding *Ecol*-Lhr occurs in a gene cluster with *Ecol*-RNAseT. **B.** The primary (up panel) and tertiary (lower panel) structure of *Msme*-Lhr PDB 5V9X revealed the organization of SF2 RecA1 and RecA2 helicase core (light pink), a WH domain (forest green) and a novel domain specific for Lhr-type named Domain4 (blue). **C.** The primary structure of *Pput*-Lhr protein. A gene cluster including *Pput*-lhr gene with an exonuclease, a ligase and a metallophosphoesterase encoding genes is represented. **D.** The primary structure of *Mthe*-Lhr2 protein.

5-Lhr helicase subfamily belonging to SF2

In this work, we are going to characterize the Lhr-type from the SF2 superfamily. Lhr for the (**l**ongest **h**elicase **r**elated) protein was firstly discovered in *E.coli* in 1995 (Reuven et al. 1995). Lhr is a large protein (1507 and 1538 amino acids residues in *Mycobacterium smegmatis* and *E. coli*, respectively) (Reuven et al. 1995; Ordonez and Shuman 2013). Lhr helicases are prokaryotic since members of this family were only found in bacteria and archaea, and it is particularly abundant in actinomycetes and proteobacteria (Ordonez and Shuman 2013). Lhr-type family belongs to the superfamily 2 of helicases. Bacterial Lhr-type protein comprises the conserved helicase core of two RecA-like domains in tandem and an additional C-terminal domain formed by a Winged-helix domain (WH) and a specific Lhr-type Domain 4 (Ejaz et al. 2018) **(FIGURE 14)**. The structure of the C-terminal extension is not yet resolved and the exact function of Lhr-type family remains unclear. Specific studies aimed to decipher Lhr-type activities and protein partners in archaea and bacteria **(Table 6).**

5-1 Lhr bacterial helicases

In bacteria, and particularly in *E.coli*, the importance of Lhr rose from its occurrence in a cluster with RNase T encoding gene **(FIGURE 14A)** (Reuven et al. 1995). This interaction has been demonstrated at the transcriptional level, as the Lhr and RNase T are co-transcribed, but this interaction at the protein level was not established neither *in vitro* nor *in vivo*. Mutational analysis showed that *Ecol*-Lhr encoding gene is a non-essential gene. Compared to the WT, *Ecol*-**Δ**lhr strains display the same rates of recovery; do not show growth defects (compared to the WT) at low (22°C) and high (44°C) temperatures; no increased sensitivity to UV and H_2O_2; and no effect on the efficiency of the phage P1-mediated transduction, which was also tested for double mutants (*lhr* with other helicase genes: *recA, recB, recD, uvrA, uvrB, uvrD, rep*, and *rho*) (Reuven et al. 1995). Nevertheless, double mutant strain **Δ***radA* and **Δ***lhr* showed lower sensitivity to AZT (antimicrobial) and Ciprofloxacin (antibiotic), affecting the DNA repair, but not to UV relative to the *radA* single mutant (Cooper et al. 2015). Thus mutations in RadA cause synergistic effects with those in the Lhr helicase, which led to propose that *Ecol*-Lhr protein may act redundantly to RadA in DNA repair or is required to remove potential lesions that require RadA for repair (Cooper et al. 2015).

In *Mycobacterium smegmatis*, the level of mRNA encoding *Msme*-Lhr was increased by > 2-fold upon damage by UV irradiation or by treatment with mitomycin C, a DNA crosslinker. The expression of Lhr in mycobacteria is upregulated in these conditions (Rand et al. 2003).
Msme-Lhr is a nucleic acid-dependent ATPase and dATPase with a preference for DNA over RNA. The optimal cofactor for the ATPase activity is Ca^{2+}. In addition, *Msme*-Lhr is a 3'-5' helicase and translocase. When provided with ATP, the protein is able to unwind 3'tailed DNA duplex, Y-forked DNA, and RNA/DNA duplex (only with DNA loading strand). No duplex dissociation activity of the protein was detected on blunt duplex DNA and 3'-tailed RNA duplex (Ordonez and Shuman 2013) **Table 6.** The helicase preference towards 3'RNA-DNA tailed hybrids questions the exact role of Lhr which might also function in RNA metabolism (Ordonez and Shuman 2013).
The crystal structure of *Msme*-Lhr (1-856) in complex with AMP-PNP and a single-stranded DNA confirm Lhr as a new SF2 helicase family in terms of domain organization (Ejaz et al. 2018) **(FIGURE 14B)**. The structure defines the conserved helicase core (RecA1 and RecA2), followed by a winged-helix (WH) domain and Domain 4 **(FIGURE 14B)**.
The WH-domain displays a similar fold to the one observed in Hjm and RecQ helicases organization (Ejaz et al. 2018). The domain widely represented in the three domains of life (Giraldo 2003), is found in different cellular pathways from transcription regulation to RNA processing (Teichmann et al. 2012). The WH of *Ecol*-RecQ have a role in DNA binding (Huber et al. 2006), and in binding to *E. coli* SSB, forming a complex with enhanced helicase activity (Gangloff et al. 1994). Although, This domain is not required for DNA unwinding (Kocsis et al. 2014). The ATPase activity of Hel308 from *Archeoglobus fulgidus* was found to be independent of the WH domain (Büttner et al. 2007).

Domain 4 is a novel structural module whose function is unknown. Mutational analysis performed by introducing punctual mutations in *Msme*-Lhr domains shows the importance of some amino acids in the protein binding to nucleic acid and mechanical activity (Ejaz et al. 2018). The Arg297 of the domain 2 (RecA2) is critical for coupling ATP hydrolysis to movement on the ssDNA loading strand. Thr145 (RecA1) and Ile528 (WH domain) are essential for duplex nucleic acid unwinding. Ser253, Trp255, and Lys410 amino acid forming the RecA2 side chain are important in the contacts of the protein to the 5′ DNA segment. Finally, the Trp597 (Domain 4) is important for coupling ATP hydrolysis to DNA translocation and therefore duplex unwinding (Ejaz et al. 2018).

Pseudomonas Putida has two copies of *lhr* gene in its genome, one encodes an Lhr with a shorter amino acid sequence than the other so-called *Pput*-Lhrcore (Ejaz and Shuman 2018) **(FIGURE 14C)**. This protein was shown to possess an enzymatic activity that matches *Msme*-Lhr, with some differences **Table 6**. *Pput*-Lhrcore has ssDNA-dependant ATPase activity more active in the presence of Manganese than Calcium as cofactor. *Pput*-Lhrcore is able in presence of ATP to unwind RNA/DNA duplex when the loading strand on the 3' extremity is a single-stranded DNA and the displaced strand is either RNA or DNA. The protein is also able to dissociate DNA/DNA duplex, but not RNA/RNA duplex. Moreover, genome context analysis shows that some bacterial and archaeal *lhr* gene including that of *P. Putida* occurs in a gene cluster with three other genes **(FIGURE 14C)**. A metallophosphodiesterase (MPE), with endonuclease activity (putative homologous of Mre11), and genes that are genetically related to a putative DNA ligase and a putative exonuclease in some bacterial genomes. The occurrence of this cluster suggests a role in a DNA repair pathway in bacteria. However, a preliminary experiment *in vitro* showed that *Pput*-Lhrcore and *Pput*-MPE do not form a complex when they are mixed (Ejaz and Shuman 2018). Altogether, these results suggest the role of bacterial Lhr helicase in DNA repair but supplemented genetic and functional analysis are required for confirming this hypothesis.

Table 6. Studied enzymatic activities of Lhr protein from archaea and bacteria.

Lhr-type		Archaea			Bacteria	
		S.sol		***M.the***	***M.sme***	***P.put***
Enzymatic activities	Structure	Monomer	Dimer	ND	Monomer	Monomer
	ATPase	DNA-Independent		ND	-ATP/dATPase -DNA -Calcium	-ATP/dATPase -Manganese
		ND		ND	K_m 105 nM k_{cat} 366 min-1	K_m 37.104 nM k_{cat} 198 min−1
	Polarity	3'-5'		3'-5'	3'-5'	3'-5'
	Translocase	ND		ND	Yes	Yes
	Nucleic acid Binding	-Fork -3′tail dsDNA	-Fork -3′-tail dsDNA	Fork DNA	ssDNA	ssDNA
	Unwinding (ATP dependent)	-fork DNA -3'tail dsDNA	No activity	HJ and forked	-3'tail DNA -fork DNA, -RNA:DNA (DNA loading strand)	-3'tail DNA -RNA:DNA (DNA loading strand)
	Annealing (ATP independent)	-Fork DNA	-Fork DNA	ND	ND	ND

This table summarizes the enzymatic activities showed by the studied archaeal (*Sulfolobus solfataricus* and *Methanothermobacter thermautotrophicus* and bacterial (*Mycobacterium smegmatis* and *Pseudomonas putida*) Lhr proteins. The indicated nucleic acid substrate or cofactor are preferred by the protein for ATPase, translocase, nucleic acid binding, unwinding, and annealing. ND: not determined.

5-2 Lhr archaeal helicases

In archaea, two Lhrs are commonly found in the sequenced genomes (Chamieh et al. 2016). They differ by the presence of a cysteine-rich region which occurs in some metal-binding proteins in the C-terminal region of Lhr2 protein (Chamieh et al. 2016).

In *Sulfolobus islandicus* the deletion of an Lhr-type encoding gene (ΔSiRe_1605) increases the mutant strain sensitivity to MMS (DNA alkylating agent, methyl methane-sulfonate) and leads to the downregulation of many genes (Song et al. 2016). Some of these genes encode for enzymes involved in nucleotide metabolism and DNA repair. The mutant also shows the silencing of over 80 genes located at a specific region. From these silenced genes we can cite: a methyltransferase of the FkbM family (*SiRe_0853*), glycosyltransferase family I, and II (*SiRe_0869* and *SiRe_0874*), IS6 family transposase (*SiRe_0856*), and glycoside hydrolase family 1 (*SiRe_ 2224*).

Biochemical analysis of Hel112 of *Sulfolobus solfataricus* of the Lhr-type family showed that it is a helicase that forms a stable monomer and dimer in solution (De Felice et al. 2007). It is a nucleic-acid dependant ATPase, ATP-dependent DNA helicase, and 3'-5' translocase. The protein binds in preference forked DNA substrate in its dimeric and monomeric form. The dimeric form is also able to bind 3′-tail dsDNA, HJ (Holliday junction) structure, and negatively supercoiled plasmid DNA **Table 6**. The dimeric protein binds ssDNA with low affinity while the monomer has no affinity for this substrate. Only the dimeric protein possesses an ATP-dependant unwinding activity with forked DNA, 3'tail dsDNA, and HJ structure. The protein is unable to dissociate bubble or blunt-ended DNA. Finally, both monomeric and dimeric forms are able to re-associate in an ATP-independent manner forked DNA. In addition, the dimeric protein can assemble Y-shaped forks and ssDNA to form HJ structure. Hence *Ssol*-Hel112 can bind and promote HJ substrate processing. Hence, it was supposed to be a homolog of the eukaryotes RecQ family helicases, despite the absence of RecQ domains in *Ssol*-Hel112 sequence (Valenti et al. 2012).

Lhr of *Methanothermobacter thermautotrophicus* is a 3'-5' translocase that exhibits binding affinity for forked DNA (Buckley et al. 2020) **(FIGURE 14D)**. The protein is able to dissociate in preference HJ and forked DNA structure but also 3'tail dsDNA, 3'tail RNA/DNA duplex with DNA loading strand. No unwinding activity was detected on blunt dsDNA in our conditions *in vitro*.

Lhr2 of *Pyrococcus abyssi* was detected in the protein interaction network of proteins implicated in DNA replication and repair, the replication protein RPA, the DNA helicase Rad25, the DNA repair AlkA protein, and the N-glycosylase/DNA lyase OGG1 protein (Pluchon et al. 2013) (**FIGURE 15)**. *Paby*-Lhr2 was also identified in the RNA helicase ASH-Ski2 and the RNA exosome interaction networks (Phung et al. 2020) (**FIGURE 15)**. No mechanistic or functional study of *Pbay*-Lhr2 or any thermococcales Lhr2 has yet been established.

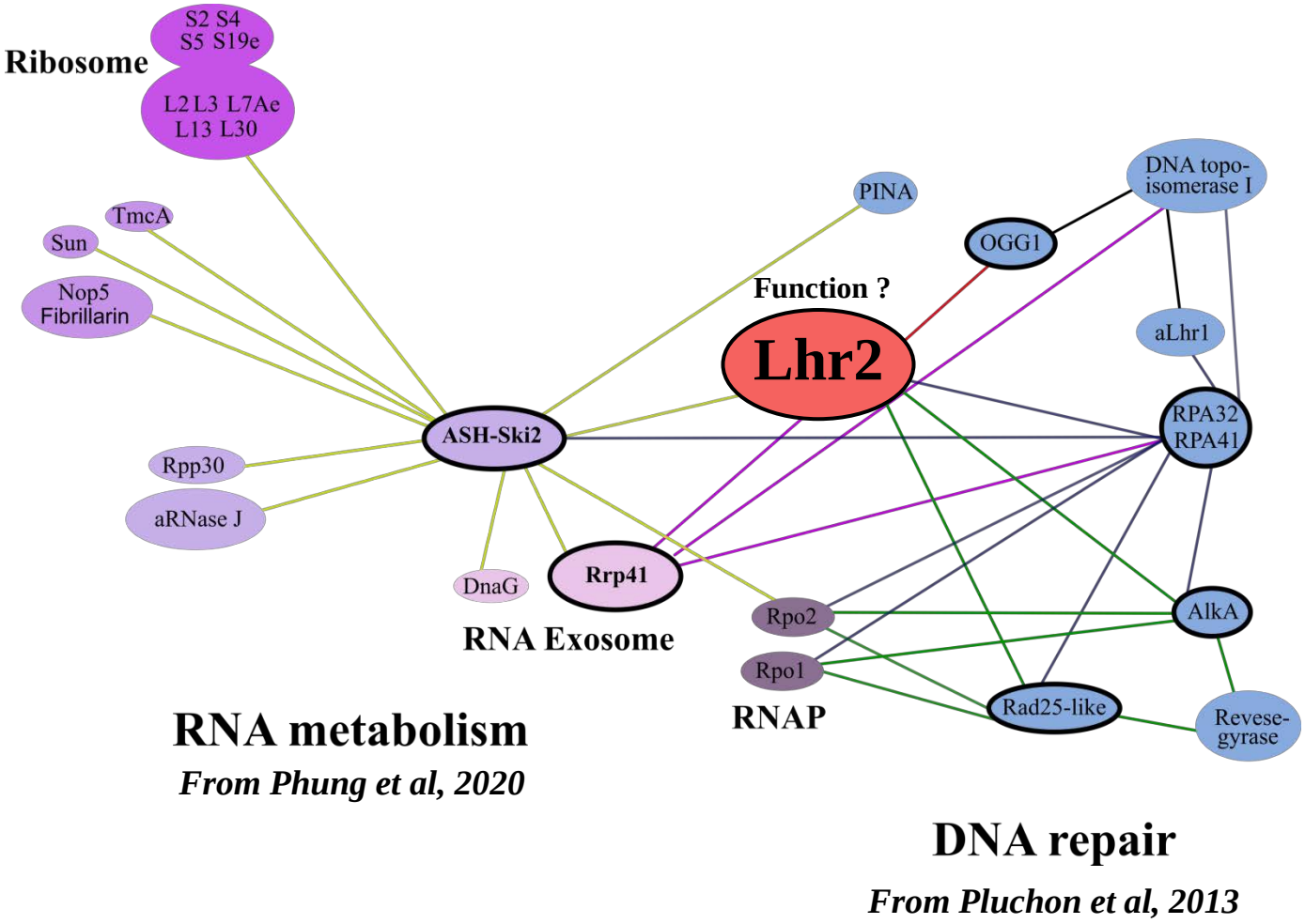


FIGURE 15. Protein-protein interaction networks including Lhr2.
Networks were obtained with the pull-down assay using (His_6)-*Paby*-ASH-Ski2 (gold line), (His_6)-*Paby*-RPA (blue line), (His_6)- Rrp41 (pink line), (His_6)- Rad25-like (green line), (His_6)-OGG1 (red line) and (His_6)-DNA-TOPO I (black line) as bait protein. Lhr2 protein was found to be common in these networks that include proteins implicated in RNA metabolism (shadow of purple) and in DNA replication and repair (shadow of blue).

6-Publication 1 : Hajj M. et al. 2018

Archaeal SF1 and SF2 helicases: Unwinding in the extreme

CHAPTER

1 ARCHAEAL SF1 AND SF2 HELICASES: UNWINDING IN THE EXTREME

Mirna Hajj[1,3], Samar El-Hamaoui[2], Manon Batista[3], Marie Bouvier[3], Ziad Abdel-Razzak[2], Béatrice Clouet d'Orval[3] and Hala Chamieh[1,2]

[1]Laboratory of Applied Biotechnology, Azm Center for Research in Biotechnology and Its Applications, Lebanese University, Tripoli, Lebanon [2]Department of Life and Earth Sciences, Faculty of Science, Lebanese University, Tripoli, Lebanon [3]Laboratoire de Microbiologie et Génétique Moléculaires, Centre de Biologie Intégrative (CBI), Centre National de la Recherche Scientifique (CNRS), Université de Toulouse, UPS, Toulouse, France

INTRODUCTION

In 1977 Carl Woese and collaborators identified Archaea as a separate domain of life. Since then, Archaea have been considered as valuable study models to understand the diversity of life styles on earth [1]. The first-discovered Archaea were distinguished by their ability to thrive in challenging habitats such as high salinity, pH, temperature, and high pressure. However, novel high-throughput sequencing methods permitted the identification that Archaea constitute a considerable fraction of the Earth's ecosystems with astonishing diversity and omnipresence. Remarkably, archaeal microorganisms are found to play pivotal roles in geochemical cycles as well as being part of human gut microbiota [2,3].

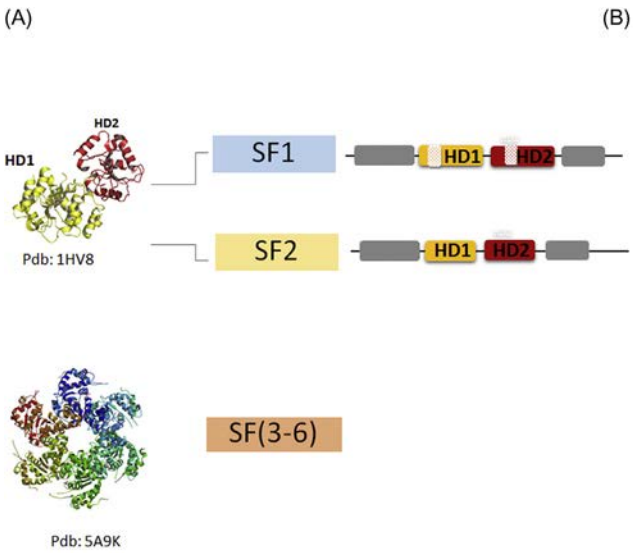
Original classification based on 16S rRNA showed that the archaeal phylogeny embraces two major phylogenetic groups, named Euryarchaeota and Crenarchaeota [4]. Subsequently, phylogenomic analyses using an increasing number of sequenced archaeal genomes led to the characterization of several phyla, including Euryarchaeota and two main "superphyla," namely the TACK superphylum (Thaumarchaeota, Aigarchaeota, Crenarchaeota, and Korarchaeota) and the DPANN superphylum (Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, and NanohaloArchaea) [5,6]. Metagenomics analyses allowed the identification of a novel archaeal clade, named Asgard, with an expanded repertoire of eukaryotic signatures. These findings provide novel hypothesis on the origin of Eukarya within the archaeal domain [7].

At first glance, archaeal cells look like bacterial cells, however unusual composition of membrane lipids and cell envelope made irrevocably clear the existence of profound differences between Archaea and Bacteria. Archaeal membranes are composed of ether lipids instead of ester lipids and the cell envelope does not contain peptidoglycan [8]. While Archaea share some bacterial essential functioning systems of energy metabolism, the informational processing system which includes DNA replication, transcription, and translation, are closely related to Eukarya [9–12]. Moreover, the archaeal genome is organized by either eukaryotic-like histone proteins or

Helicases from all Domains of Life. DOI: https://doi.org/10.1016/B978-0-12-814685-9.00001-4
© 2019 Elsevier Inc. All rights reserved.

bacterial-like nucleoid-associated proteins [13]. In this mosaic setting, we are interested in deciphering the panel of helicases existing in Archaea that would drive numerous fundamental metabolic pathways.

Helicases are molecular motors that couple the use of energy to countless biological processes. By catalyzing the separation of double stranded nucleic acids into a single stranded one and the dissociation of nucleic-acid associated proteins, helicases participate in all aspects of DNA and RNA metabolism, and help in chromatin remodeling [14–17]. Helicases are grouped into six superfamilies (SF1–6) based on amino acid sequence similarity, oligomeric state (monomeric or hexameric), activity (substrate as single-stranded DNA (ssDNA), double-stranded DNA (dsDNA), or RNA, translocating activity), and polarity (5′→3′, 3′→5′) [16] (Fig. 1.1). The two largest acknowledged superfamilies, SF1 and SF2, which group nonhexameric helicases, perform diverse cellular functions in DNA replication, repair, recombination, RNA metabolism, and protein translation. SFs3–6 form hexameric toroid structures (Fig. 1.1) [18–20]. SF3 helicases comprise viral helicases.



(B)

SF1	Archaea
UvrD-like	✓
Upf1-like	✓
Pif1-like	X

SF2	Archaea
Rec-G like	X
DEAH/RHA	X
Swi2/Snf	✓
XPF/Hef/ERCC4/RIG-I	✓
DEAD-box	✓
XPD/Rad3/DinG	✓
Ski-2 like	✓
XPB/Rad25-like	✓
RecQ-like	✓
UvrB	✓
Cas 3	✓
Sfth	✓
Lhr	✓
T1R	✓
Reverse gyrase	✓

FIGURE 1.1

(A) Domain organization of SF superfamilies and (B) occurrence of SF1 and SF2 helicases in Archaea. Helicases are classified into six superfamilies which include the hexameric helicases SF3 to SF6 and the non hexameric helicases SF1 and SF2. SF1 and SF2 helicases share a conserved helicase core composed of two RecA-like domain folds (HD1 in yellow and HD2 in red). SF1 helicases present multiple insertions within the helicase core. SF1 and SF2 helicases are divergent in their N- and C-terminus which are attributed to the diversity of helicase functions in vivo (represented by gray boxes). Crystal structure of SF2 DEAD-box helicase from *Methanococcus janaschii*, Pdb: 1HV8. Crystal structure of hexameric helicase E1 from papillomavirus, Pdb: 5A9K.

SF4 and SF5 function as replicative and transcription termination factors, respectively. SF6 include the minichromosome maintenance helicases and the RuvB helicase [20,21].

The SF1 and SF2 helicases are characterized by a conserved helicase core formed by two RecA domains which consist of nine characteristic sequence motifs named Q, I, Ia, Ib, and II to VI. These motifs slightly differ between SF1 and SF2 and among those Walker A (motif I) and Walker B (motif II) bind NTPs (Fig. 1.1) [22,23]. The specificity of action of SF1 and SF2 helicases has been mostly attributed to their accessory domains present at their N- and C-terminus in addition to their conserved helicase core. SF1 and SF2 were also classified based on their translocation polarity. Two groups emerged: the SF1/2A and SF1/2B with 3′−5′ or 5′−3′ translocation polarity, respectively [16]. Another classification was further refined based on the sequence conservation of bacterial and eukaryal-like helicases and on structural and mechanistic features allowing the identification of twelve distinct families (Fig. 1.1) [24]. The SF1 includes UvrD-like/Rep, Pif-1-like, and Upf1-like families, whereas the SF2 accounts for the Rec-G like, RecQ-like, XPD/Rad3/DinG, Ski-2 like, type 1 restriction enzyme helicase subunit (T1R), Swi2/Snf, XPF/Hef/ERCC4/RIG-I nuclease helicase, DEAD-box, and the DEAH/RHA families [18,25,26].

To overcome the gap of knowledge on archaeal helicase families, a comprehensive *in silico* analysis allowed retrieving the first exhaustive list of SF1 and SF2 helicases in Archaea. Each family was named based on knowledge of the function of their bacterial and eukaryotic counterparts (Fig. 1.1) [27]. Most of the already-known SF1 and SF2 families are represented in Archaea. Two major archaeal SF1 families could be defined, the UvrD-like and Upf1-like. Remarkably, the SF1 helicases are mostly restricted to the Euryarchaeoeta phylum [27]. In addition, none of the sequenced archaeal genomes encode the Pif-1-like helicases which have been shown in Eukarya to be involved in the maintenance of both the mitochondrial and nuclear genomes [28]. Finally, all the major SF2 families were retrieved in the archaeal classification with the exception of the bacterial Rec G-like involved in DNA replication, recombination and repair [29], and the DEAH/RHA families that functions in premessenger RNA splicing and ribosome biogenesis [30] (Fig. 1.1). Structural and mechanistic studies were reported for some members of these families [31−37]. Nevertheless, associated cellular processes have not been yet addressed.

THE Uvr AND XP HELICASE FAMILIES

Helicases play essential roles at different steps of the DNA repair pathways. Nucleotide excision repair (NER), base excision repair (BER), and DNA mismatch repair (MMR) are the main known pathways, and all act on one of the two strands of the damaged DNA. Among these pathways, NER is the most versatile one and seems to be conserved in all domains of life [38−40]. Briefly, an initial recognition of DNA damage is followed by excision of the damaged base of the ssDNA. These steps require the coordinated action of DNA repair helicases and nuclease enzymes to detect the damaged DNA. The existence of NER pathway in Archaea has not been formally demonstrated, but it is suggested based on the presence of NER pathway like-enzymes. Most archaeal genomes encode for a homologue of the eukaryotic XP (*Xeroderma pigmentosum*) helicases. In addition, few encode for a homologue of the prokaryotic Ultraviolet repair (Uvr) helicase [27,40]. It has been suggested that Archaea use a simplified version of the eukaryal NER pathway and that the bacterial NER systems come from horizontal gene transfer from mesophilic archaeal species [40].

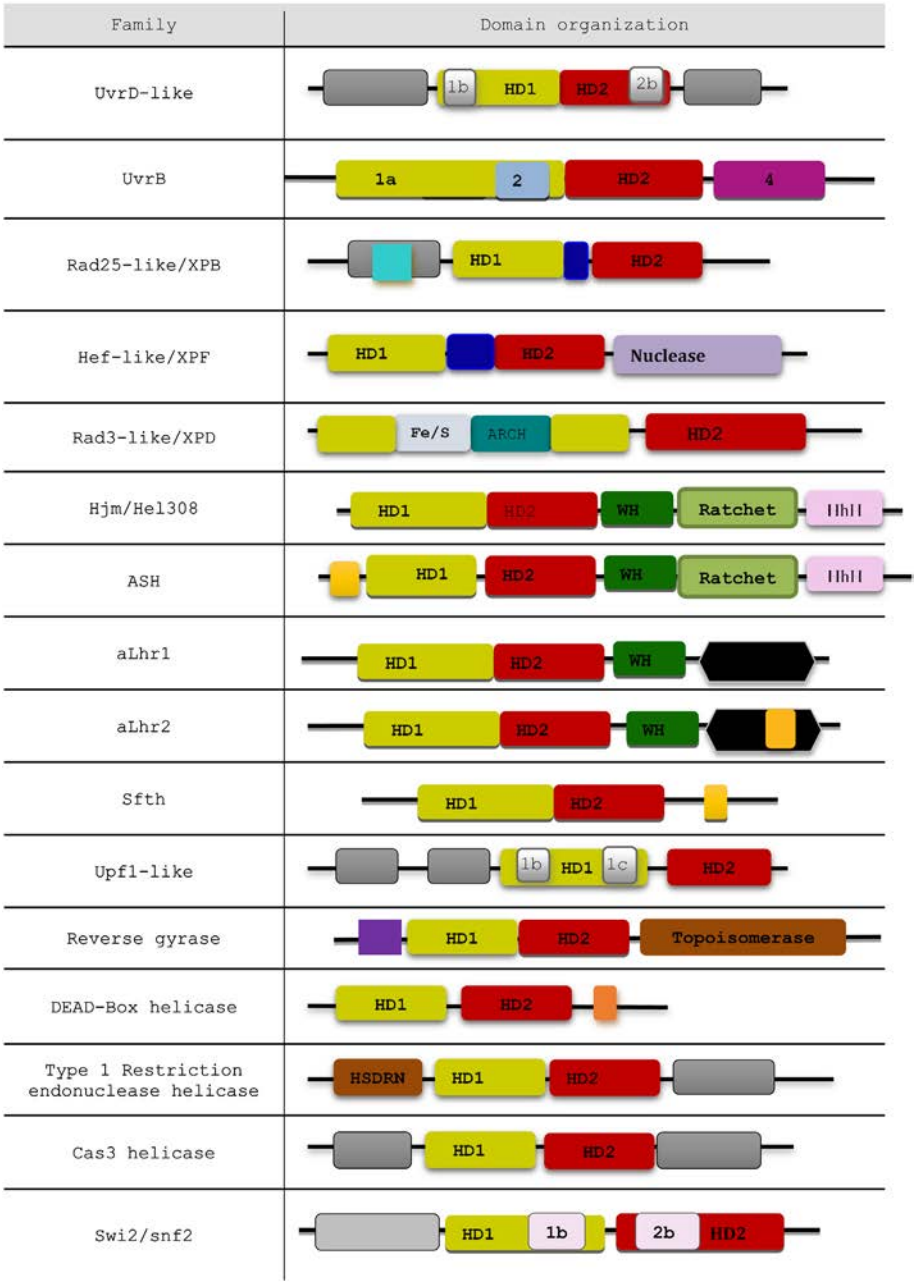


FIGURE 1.2

Structural domain organization of archaeal SF1 and SF2 families.
The Helicase domain 1 (HD1) and 2 (HD2) are respectively represented in yellow and red in all helicase families. The C- and N-terminal additional domains are colored as follows: UvrB: domain 1a corresponds to HD1 and contains two insertion domains 1b and 2 that are represented in green and blue, respectively, HD2 represents domain 3; XP helicases: the thumb motif located between HD1 and HD2, the C-terminal damage recognition domain (DRD) and the FeS cluster and ARCH domain inserted in HD1 are respectively represented in blue, cyan, blue-white and aquamarine; DEAD-box helicase: the C-terminal DbpA-domain is in orange; Type 1 Restriction

(*Continued*)

Briefly, the Uvr A/B/C/D proteins are known to be the key actors of the NER pathway in Bacteria [41]. Helicase activities are required in two steps of the NER pathway: the first one carried out by the UvrB helicase verifies the damage, whereas the second is part of the post-incision complex and involves the UvrD or PcrA activity which removes the excised fragment containing the damage.

The Archaeal UvrD-Like Helicase

The UvrD-like helicases are described as SF1 helicase with a 3′−5′ polarity. These helicases consist of two RecA helicase core domains HD (named 1a and 2a) and two inserted auxiliary domains 1b and 2b within both HD domains, respectively (Fig. 1.2). These auxiliary domains possess DNA binding motifs [42]. Genes encoding UvrD-like helicases are particularly abundant in genomes of the Euryarchaeota, in particular in the Halobacteriales and Methanosarcinales clades [27]. However, euryarchaeal UvrD-like helicases show an extreme divergence of their N- and C-terminal unstructured regions that are proposed to interact with UvrB. Intriguingly, the UvrD encoding genes do not show the same taxonomic distribution as the other Uvr-like proteins across the archaeal phylogeny. This questions the physiological role(s) of the UvrD-like helicases in Archaea [27]. In fact, UvrD from *Escherichia coli* has been shown to play other roles in DNA replication by acting on Okazaki fragments and in methyl-directed mismatch repair (MMS) [43−46].

The Archaeal UvrB-Like Helicase

The UvrB-like helicases have been described as SF2 helicase with a 3′−5′ polarity and, based on *E. coli* UvrB, a weak ATPase and unwinding activities. Further findings on *Mycobacterium* UvrB suggest that UvrB-like proteins could also display strong ATPase and helicase activities [47,48]. The crystal structure of UvrB showed four domains: domains 1 to 4. Domain 1 was further divided into 1a and 1b. Domains 1a and 3 are conserved regions with the SF2 family members and contain their characteristic motifs (Fig. 1.2), The three auxiliary domains are domain 1b (which is supposed to favor additional interactions with the DNA), domain 2 (which provides interaction with UvrA protein), and domain 4 (which interacts with UvrA and UvrC proteins) [49,50]. Archaeal UvrB-like members are abundant in Euryarchaeota and possess a conserved C-terminal domain which is supposed to interact with UvrC [27]. However, the *in silico* detection of these helicases in the archaeal phylogeny are not yet supported by *in vivo* and *in vitro* experimental evidences and the existence of archaeal UvrA/B/C/D complexes or subcomplexes have not yet been demonstrated. Genetic studies showed that, in halophilic Archaea, the deletion of the genes encoding the UvrA/B/C repair proteins render the cells hypersensitive to ultraviolet (UV) radiation (Table 1.1) [51]. Other studies performed in *Haloferax volcanii* showed that UvrA/B/C proteins work together with the PCNA factor named NerA in repairing DNA damage resulting from exposure to mitomycin [52,53].

◀ endonuclease helicase: the N-terminal nuclease domain is in brown; Swi2/Snf: the domain 1B inserted in HD1 and the domain 2B inserted in HD2 are in light pink; reverse gyrase: the additional N-terminal domain and the C-terminal topoisomerase domain are respectively in purple and in brown; Hjm and Lhr: WH domain is colored in dark green; Hjm and ASH: the ratchet is in olive green, the helix-loop-helix (HLH) is in pink; ASH, Lhr and Sfth: cysteine-rich region are in yellow. All grey boxes indicate divergent regions within protein families. Black boxes indicate regions of unidentified functions.

Table 1.1 Phenotypical Analyses of Archaeal Strains Knocked-out for Gene Encoding SF2 Helicases

Gene	ORF ID	Organism	Predicted Function	Phenotype	Ref.
XPB	XPB1: SiRe_1128	*Sulfolobus islandicus*	NER	None observed	[75]
	XPB2 SiRe_1526		NER	None observed	[75]
	TK0928	*Thermococcus kodakarensis*	NER	Negligible DNA-damage sensitivity to (UV) irradiation, methyl methanesulfonate (MMS) and mitomycin C (MMC)	[68]
Hef/XPF	TK1021	*Thermococcus kodakarensis*	Helicase-endonuclease	General sensitivity to UV irradiation and MMC. Particular hypersensitivity to MMS	[68]
	HVO_RS19270	*Haloferax volcanii*	Helicase endonuclease	Moderate sensitivity to DNA crosslinking agents	[69]
XPD	TK0784	*Thermococcus kodakarensis*	NER	Slight sensitivity to UV irradiation, MMS, and MMC	[68]
	SiRe_1685	*Sulfolobus islandicus*	NER	None observed	[75]
Hjm/ Hel308	TK1332	*Thermococcus kodakarensis*	DNA helicase	Sensitivity to MMS	[68]
	PF_RS03450	*Pyrococcus furiosus*	DNA helicase	None observed	[78]
	SiRe_0250	*Sulfolobus islandicus*	DNA helicase	Lethal	[75,81]
DEAD	TK0306	*Thermococcus kodakarensis*	RNA helicase	Decreased cell growth	[110]
UvrA/B/ C	AE004437.1	*Halobacterium sp.*	NER	Hypersensitivity to UV irradiation	[51]
Reverse Gyrase	AB117612	*Thermococccus kodokarensis*	Topoisomerase	Decreased growth	[97]
Lhr2	SiRe_1605	*Sulfolobus islandicus*	DNA/RNA helicase	Sensitivity to MMS	[90]
ASH	TK0566	*Thermococcus kodakarensis*	Unknown	Slow cell growth, sensitivity to DNA-damaging agents	[86]

Archaeal XP helicases were extensively studied at the structural level and have greatly contributed to our understanding of the eukaryal and bacterial SF2 helicases functions and mode of actions [32,36,37,40,54–58]. In eukaryotes, two types of NER pathways exist: firstly, the TCR (transcription-coupled Repair) pathway which acts to detect DNA damage and which is linked to the transcription machinery [59]; and secondly a global genome repair (GGR) pathway that involves the XPC-hr23B heterodimer to detect DNA damage [60]. In both pathways, a multi-subunit complex

composed of the Transcription Factor II-H complex and of two main helicases XPB and XPD is required to bind and extend the ssDNA around the damage site. The activity of these two helicases allows other NER factors such as the helicase-nuclease XPF-ERCC1 and XPG endonuclease to cleave the 5′ and 3′ sides of the lesion, respectively. In Archaea, a pathway similar to NER has not yet been demonstrated. The recognition of the lesion has been suggested to occur during transcription by the action of SSB proteins or the RNAP, or through an unidentified damage recognition protein [40]. Most of the archaeal organisms encode for XP-like helicases XPB and XPD and the nuclease-helicase XPF, but they do not follow the same overall distribution across archaeal members raising the question of how the NER pathway will operate in the absence of one of the XP helicase partners. An endonuclease named Bax1 (Binds archaeal XPB) was found to form a stable complex with XPB in vitro. It was suggested that Bax1 may be the archaeal counterpart of the eukaryal nuclease XPG by acting together with XPB in cleaving 3′ of the DNA damaged site [58].

XPB/rad25 Helicase

XPB/Rad25-like helicases have been described as SF2 helicases with a 3′−5′ polarity that function in eukaryotic NER pathways, but are also part of the basal transcription machinery in eukaryotes [61]. The crystal structure of XPB from *Archaeoglobus fulgidus* (AfuXPB) gave the first insight on the organization of an archaeal helicase from the SF2 family [54]. The resolved structure of AfuXPB showed, in addition to the two conserved helicase core domains (HD1 and HD2), a 100 amino acid N-terminal domain named the DRD domain (damage recognition domain) resembling the mismatch recognition domain (MRD) of the DNA repair MutS protein and is responsible for recognition of distorted and damaged DNA (Fig. 1.2). The DRD domain does not seem to be conserved across archaeal XPBs as the XPB sequences from Archaea are extremely divergent at their N- and C-termini [27]. Another domain found in XPB is the thumb motif located between HD1 and HD2 (Fig. 1.2). This motif is commonly found in DNA-dependent DNA polymerases and thought to be involved in branched DNA binding. In addition, the HD1 domain contains a RED motif that has a key role in DNA unwinding [54].

Hef-Like Helicase

Hef-like (*H*elicase-associated *E*ndonuclease for *F*ork-structured DNA) helicases have been described as SF2 helicases with a 3′−5′ polarity that in some cases retain, in addition to the helicase domain, a nuclease domain similar to the eukaryotic XPF endonuclease. The archaeal Hef was first identified from the hyperthemophilic archaeon *Pyrococcus furiosus* as a protein factor that can stimulate holiday junction resolution by the Hjc resolvase [62,63]. Most archaeal genomes, except for Thermoplasmatales, encode a Hef-like helicase. The Hef-like helicases exist in two forms: the long form consists of an N-terminal helicase domain carrying a conserved helicase core domain fused to a C-terminal nuclease and specific to the Euryarchaeota; the simpler version lacks the helicase domain in the Crenarchaeota and Thaumarchaeota (Fig. 1.2) [64]. The crystal structures from *Pyrococcus furiosus (Pfu)* and *Aeropyrum pernix* (Aep) Hef-like helicases provided the first structural information on the XPF family [65,66]. The C-terminal endonuclease domain is structurally related to the one found in the eukaryotic nucleases Mus81 and XPF. In addition to these domains, a conserved “helix-hairpin-helix” (HhH2) motif involved in protein dimerization and DNA binding was also discovered [56,62]. Evidence suggests that PfuHef plays a role in processing stalled

replication forks as it produces splayed duplexes from DNA forks and four-way junctions [35,56,67]. The recombinant *Thermococcus kodakarensis* Hef protein (TkoHef) exhibits a similar activity on fork-structured DNA in vitro [68]. Moreover, the strain knocked-out for Tko*hef* is highly sensitive to mitomycin C, suggesting that Hef is involved in numerous repair processes and is critical for DNA interstrand cross-link repair (Table 1.1) [68]. Similar observations were made on a *H. volcanii* strain that was deleted for the gene-encoding Hef. This strain exhibits end-joining defects, as well as homologous recombination and cross-link repair deficiencies. Conversely, *hef* deletion does not render the cells sensitive to UV radiation, meaning that the NER pathway is not impaired [69].

XPD/Rad3-like

The XPD/Rad3-like family is the only SF2 helicase family with a 5′−3′ polarity with a weak DNA unwinding activity in vitro [70]. Archaeal XPDs were extensively studied at the structural and biochemical levels with four crystal structures published from three different archaeal organisms: *Thermoplasma acidophilum (TaaXPD*), *Sulfolobos acidocaldarius (SsAXPD),* and *Sulfolobus tokodaii (SstXPD)* [55,71−73]. These structures showed that archaeal XPD-like proteins consist of the helicase core domain (HD1 and HD2) and two additional domains that are inserted into HD1: an iron−sulfur 4FeS domain and an Arch domain defined by its arch-shaped structure (Fig. 1.2). The 4FeS domain is essential for the helicase activity and functions as a wedge structure involved in duplex separation. The Arch domain function is not characterized for archaeal XPD-like protein, however it has been shown to be essential for interaction with the CAK—cyclin-dependent kinase (CDK)-activating kinase—in eukaryotic XPD helicase. The published data using Archaea as models provided insights into the mechanism of XPD unwinding where, in Eukaryotes, it is thought that ssDNA passes first through the groove between the HD2 and arch domains, moving then through a hole encircled by the Arch, FeS-cluster, and HD1 domains [55]. Finally, it was shown that SsaXPD can efficiently unwind in vitro 5′ overhang, Y fork, or bubble substrates [74].

Altogether, it is not yet fully understood how the NER pathway is operating in Archaea; as deletion of either XPB or XPD did not affect DNA repair in *Sulfolobus islandicus* and only slight sensitivity to DNA damaging agents was observed in the null mutants of *T. kodakarensis* (Table 1.1) [68,75]. One possible explanation is the presence of numerous XPB/Rad25 and XPD/Rad3 encoding genes. These genes could be functionally redundant, and in this case no clear phenotype would be observed [68,75]. Further genetic studies are required to fully elucidate the exact role of XP proteins in DNA repair in Archaea.

THE Ski2-LIKE FAMILY

The Ski2-like family was originally named after the Ski2 RNA helicase that works with the exosome in the eukaryotic turnover and quality control of mRNAs. These SF2 enzymes are mostly RNA helicases, however, only one member—the archaeal Hel308 member—is described as a DNA helicase [76]. In Eukarya, three major helicase groups exist, the cytoplasmic Ski2 and the nuclear Mtr4 associated with the eukaryal exosome machineries, and the Brr2 helicases involved in RNA quality control and RNA splicing, respectively. Initially, the archaeal DNA helicase homolog of the human HelQ was named Hel308 [77]. Subsequently, it appeared that Hel308 members belong to the Ski2-like family [76]. The protein Mth810 was called Hel308a since it resembled the human

Hel308 helicases implicated in DNA repair [77]. The *P. furiosus* was named Hjm (Holliday junction migration) as it was detected through its activity on model Holliday junctions in vitro [78]. Importantly, Hel308 members are commonly found through all the archaeal phylogeny with few exceptions [27]. These enzymes display a ssDNA stimulated ATPase activity and have in vitro activities on DNA forks and holiday junctions. They translocate along both DNA strands in a $3' \rightarrow 5'$ direction and a better efficiency on forked structures was observed [77]. Deciphering the biochemical activities of *S. tokodaii* Hjm identified, in contrast to other Hjm, unwinding activities in both directions [79]. Hjm was shown to interact with the endonuclease/resolvase Hjc in vitro, therefore it has been proposed that archaeal Hjm/Hel308 promotes replication fork regression through its interaction with Hjc [79,80]. However, in *T. kodakaraensis*, a knockout mutant of the encoding gene is viable in contrast to *Sulfolobus* where the Hjm helicase has been shown to be an essential protein (Table 1.1) [68,75,81].

High resolution structure of Hjm/Hel308 from *A. fulgidus* in an apo form and in complex with DNA captured the duplex-unwinding reaction and showed that initial strand separation is not dependent on ATP [82]. Hel308 consists of the two tandem RecA core domains (HD domains 1 and 2) (Fig. 1.2). The resolved structure showed a protruding region of β-hairpin secondary structure between motifs Va and VI in HD2 [31,36]. This hairpin was proposed to guide nucleic acid strand separation and thus plays a crucial role in coupling translocation to helicase activity. A third domain found in the C-terminal region is a "winged helix" domain (WH domain) identified as a sequence-specific DNA-binding domain in transcription factors, particularly those of the forkhead family (Fig. 1.2) [83]. The WH domain makes extensive contacts with branched structures and may play a role as a "hinge," permitting Hel308 to bind and then close tightly around its DNA substrate. Mutations in the WH domain abolished DNA binding and unwinding activity [83]. Domain 4 (named as "ratchet") was proposed to position nucleic acids for translocation linked to ATPase functions. At the C-terminal, a helix-loop-helix (HLH) caps the Hel308 helicase and exhibits typical DNA-binding activity with strong binding to ssDNA. The HLH domain possibly functions to limit the enzyme processivity on branched substrates [36]. Subsequently, the resolution of structures of the eukaryal Mtr4, Brr2, and Ski2 RNA helicases allowed to define the hallmarks for the Ski2-like family based on structural signature domains the helicase SF2 core and the typical C-terminal domain consisting of WH and a ratchet domain (Fig. 1.2). In addition, each family member will possess its own specific accessory domains [76].

Nevertheless, the exact role of archaeal Hel308/Hjm is yet to be uncovered. Since most Archaea lack RecQ, with the exception of Methanosarcinales, Hel308 is proposed to function in DNA recombination and repair as an alternative to the RecQ helicase commonly found in Bacteria and Eukarya [84]. When Hel308 was introduced into an *E. coli* replication-deficient strain, the RecQ-like phenotype was restored [77,84]. Hel308 was shown to physically interact with the Replication Protein A (RPA) that binds to ssDNA and helps loading Hel308 at stalled replication forks [85]

In addition to the Hel308 from the Ski2-like family, phylogenomic analysis demonstrated the presence of a helicase group specific to Archaea and mainly to Euryarchaoeta named ASH for Archaea Specific Helicase. ASH has been shown to be evolutionary related to the Ski2-like family [27]. On the sequence level and as shown by secondary structure predictions, ASH is more similar to Hel308 than the other RNA helicases of the Ski2-like family. At the N-terminal, the ASH helicase differ by the presence of a cysteine-rich region which is predicted to be a metal binding motif (Fig. 1.2) [27]. Further, ASH was proposed to act as a transcription termination factor—also named

Eta for euryarchaeal termination activity [86]. Indeed, ASH was shown to disrupt the transcription elongation complex in vivo in *T. kodokarensis*. In addition, deletion of TK0566 (encoding ASH/Eta) results in slow growth and renders cells sensitive to DNA damaging agents (Table 1.1) [86].

THE Lhr-LIKE AND Sfth-LIKE HELICASE FAMILIES

Lhr (**L**ongest **H**elicase **R**elated protein) and Sfth, two SF2 helicases with unidentified functions, are the most widespread in archaeal genomes [27]. Indeed, all archaeal genomes encode at least for one Lhr helicase and Lhr gene duplication events are commonly observed. Sfth is found in all Archaea except for the thermoplasmatales and some methanogens. The role of these two helicases have not yet been studied in spite of their large abundance across sequenced genomes.

Lhr that was first discovered in *E. coli* in 1995 is a large protein (1507 and 1538 amino acids residues in *Mycobacterium smegmatis* and in *E. coli*, respectively) [87]. Lhr is a prokaryote helicase since members of this family are only found in Bacteria and Archaea. They are particularly abundant in actinomycetes and proteobacteria [88]. Genetic studies in *E. coli* showed that the cellular growth is not impaired by the deletion of *lhr* [87]. Upon damage by UV irradiation or by treatment with mitomycin C, the expression of Lhr is upregulated in mycobacteria [88]. The first crystal structure of the first half of *M. smegmatis* Lhr (1−856) in complex with AMP-PNP and ssDNA just became available and exemplifies Lhr as a new SF2 helicase family in terms of domain organization [89]. The structure shows the conserved helicase core (HD1 and HD2), followed by a WH domain that displays a similar fold to the one observed in Hjm and RecQ helicases (Fig. 1.2). A novel structural module whose function is unknown is annotated as domain 4. Full length bacterial Lhr-like also possesses an additional C-terminal DEAD-associated helicase domain whose function is still unknown [87]. In Archaea, two Lhrs are commonly found, aLhr1 and aLhr2. A third distant phylogenetically group named Lhr-like is also commonly found in Sulfolobales. In terms of domain organization, archaeal Lhr resembles bacterial Lhr, with the aLhr2 proteins differing from other Lhr by the presence of a Cysteine-rich region which occurs in some metal binding proteins (Fig. 1.2) [27]. A deletion of the SiRe_1605 gene (corresponding to *lhr2 gene*) in *S. islandicus* render the strains more sensitive to MMS (methyl methanesulfonate) and leads to the deregulation of many nucleotide metabolism and DNA repair enzymes as well as the silencing of over 80 genes located at a specific region (Table 1.1) [90]

Sfth is widespread in Bacteria and in Archaea, and is limited to plants and fungi in eukaryotes [91]. Sfth is distinct from other helicases by the presence at its N-terminus of a DUF1998 domain containing a putative zinc-finger. Archaeal Sfth proteins display the same conserved helicase core and the characteristic DUF1998 domain (Fig. 1.2) [27]. The yeast homolog, named Hrq1, is a robust 3′−5′ helicase with remarkable biochemical properties as it forms a homo-heptameric ring [91−93]. A deletion of the *hrq1* gene is dispensable for vegetative growth in *Saccharomyces cerevisiae* [94]. In the fission yeast *Saccharomyces pombe*, evidence showed that Hrq1 plays a role in the repair of DNA Interstrand cross-links and in genome maintenance [92]. In addition, the human RecQ4 was proposed to be an orthologous member of Sfth helicases, however it lacks the characteristic DUF1998 domain [92]. The physiological function of Sfth in prokaryotes is yet to be demonstrated. However, a novel defense mechanism against phages was found associated with the DUF1998 domain characteristics of the Sfth helicases. This defense mechanism was named DISARM for defense island systems associated with restriction-modification (RM) [95]. The

DISARM system is widespread in Bacteria and Archaea and consists of a DNA methylase and four other genes annotated as a helicase domain, a DUF1998 domain, a phospholipase D domain, and a gene of unknown function. When this system was introduced into *Bacillus subtilis*, the engineered Bacteria became protected against all three major families of tailed dsDNA phages [95].

THE REVERSE GYRASE HELICASE FAMILY

Reverse gyrase enzymes are DNA topoisomerases that introduce positive supercoils into DNA and are specific to hyperthermophilic Archaea and Bacteria [96]. Reverse gyrase is one of the enzymes exclusively found in hyperthermophiles. Originally it was thought that this enzyme is essential for life at high temperature. However, *Thermococcus* cells knocked-out for the gene encoding the reverse gyrase are viable, meaning that this gene is not essential [97]. Nevertheless, the mutant strain shows a decrease in cellular growth rate and is thermosensitive, supporting its role at high temperature (Table 1.1). Reverse gyrases consist of a helicase core (HD1 and HD2) and a C-terminal type IA topoisomerase domain (Fig. 1.2). Crystal structures from *Thermotogata maritima* and *A. fulgidus* revealed domain arrangements including the functionally important zinc-finger motif [98,99]. A latch domain is inserted into HD2 and is supposed to play a role in the introduction of positive supercoiling by reverse gyrase. This domain connects the helicase and the topoisomerase domains and is required for their functional cooperation. An insert in HD1 adopts a helix-loop structure in the *T. maritima* enzyme and a β-hairpin structure in *A. fulgidus* reverse gyrase. Both inserts protrude from the same side of the helicase domain. The latch domain, the HD1 insert and the zinc-finger motif are required to bind and guide DNA during strand passage. The topoisomerase domain of the reverse gyrase is similar to the bacterial topoisomerase IA. The physiological function of the reverse gyrase was addressed in *Sulfolobus solfataricus*. A link to DNA repair was made by the identification of the translesion DNA polymerase SSoPolY/Dpo4 and of SSB as protein partner of the reverse gyrase [100,101].

THE DEAD-BOX LIKE HELICASE

DEAD-box like proteins are SF2 RNA helicases that participate in many processes involving RNA such as transcription, translation, editing, mRNA degradation, and ribosomal assembly [20,102,103]. These proteins exhibit RNA-dependent ATPase activity and helicase activity [104]. The crystal structure of a DEAD-box like helicase from *Methanococcus janaschii* was the first published prototype of helicase core organization into two RecA domains [57]. DEAD-box like helicases are the shortest helicase in length and in many cases the DEAD-box like proteins lack a characteristic C-terminal domain (Fig. 1.2). Some of these helicases have a C-terminal DbpA RNA-binding domain (RBD) known to mediate the recognition of hairpin 92 of the 23S ribosomal RNA [105–107]. DEAD-box helicases are found in most of the Euryarchaeota with the exception of Pyrococcales and Halobacteriales, and only in the Sulfolobales from the Crenarchaeota. The physiological function of DEAD-box like helicases in Archaea is not yet reported, but some studies performed in *Methanococcoides burtonii* and *T. kodakarensis* showed that these DEAD-box helicases are induced upon cold stress and showed reduced growth upon deletion (Table 1.1) [108–110].

THE Upf1-LIKE/Dna2 FAMILY

Upf1-like/Dna2 SF1 members have unwinding activity with a 5′→ 3′ polarity. Archaeal Upf1-like proteins possess similarities with the helicase/nuclease Dna2. In eukaryotes, Dna2 plays an important role in DNA replication by Okazaki fragment processing and in DNA repair pathways [105,111]. These proteins are sporadically widespread across archaeal genomes and their presence is restricted to Thermococcales, Archaeoglobales, and Halobacteriales. In addition, their N-terminal and C-terminal regions are extremely divergent (Fig. 1.2). The physiological function of archaeal Dna2 is uncertain. Dna2 from *Pyrococcus horikoshii* was characterized in vitro at the biochemical level and was shown to display 5′−3′ DNA helicase and nuclease activities [112].

THE Cas3, Snf2 AND T1R HELICASES FAMILIES

Cas3 helicase is an SF2 helicase that has been recognized as a novel SF2 clade, but it displays very weak conservation of the helicase core motifs [113]. T1R restriction helicase and cas3 helicases are part of prokaryotic defense mechanism [114]. Cas3 plays an essential role in prokaryotic adaptive immunity against foreign nucleic acids when invading viruses and plasmids [113]. The Type I restriction enzyme is a prokaryotic helicase and functions as part of a large multiprotein complex forming a restriction-modification system to protect Bacteria against foreign DNA [115,116].

SWI2/SNF2 like proteins have been documented in all domains of life [27]. However, in Bacteria and Archaea, the similarity is restricted to the helicase motifs and their function in archaeal physiology is not yet understood [117]. Whether they are involved in recombination—like Rad54 protein, or in transcription like Snf2 or Mot1, or in other chromatin remodeling processes—remains to be elucidated [117]. Crystal structures of SSoSwi2 in the absence or presence of its dsDNA substrate have been resolved [118]. The enzyme consists of two RecA-like domains (HD1 and HD2) and two specific subdomains inserted in each HD domain with a deep cleft separating the two domains (Fig. 1.2). The structure provided insights into how this enzyme translocates along the dsDNA where it was shown that SWI2/SNF2 ATPase moves along the minor groove of a dsDNA substrate without strand separation [119,120]. Studies on SWI2/SNF2 from *S. solfataricus* showed that the enzyme lacks helicase activity, but displays a dsDNA stimulated ATPase activity and translocation activity. It has been suggested that this translocation activity might be responsible for remodeling like activity of Sac7d/Alba:DNA complexes [119,120].

CONCLUSIONS AND FUTURE PERSPECTIVES

Enzymes belonging to the SF1 and SF2 helicase families are structurally similar, but functionally diverse due to the N- or C-terminal elements appended to the conserved helicase core. In this chapter, we have highlighted archaeal families and associated features that characterized the overall archaeal SF1 and SF2 enzymes in which, for the most, cellular functions are still to be discovered. Despite the subtle differences in sequence and structure, it is clear that these protein families participate in fundamental ways that involve substrate recognition and conformational rearrangements that could induce association or dissociation of nucleic acid, but also nucleic acid-binding protein displacement. The question remains to understand their impact on physiology of archaeal cells and

their basic mechanistic features in the context of the extreme conditions encountered by archaeal microorganisms. In addition, further structure—function studies will provide a framework for understanding more complex eukaryotic counterparts often associated in multiprotein complexes.

ACKNOWLEDGMENTS

This work is financed by the Lebanese University (UL) and National Center for research in Lebanon (CNRS-L). MH is a recipient of the AZM-UL excellency fellowship.

REFERENCES

[1] C.R. Woese, G.E. Fox, Phylogenetic structure of the prokaryotic domain: the primary kingdoms, Proc. Natl. Acad. Sci. U. S. A. 74 (11) (1977) 5088–5090.
[2] L. Eme, et al., Archaea and the origin of eukaryotes, Nat. Rev. Microbiol. 16 (2) (2018) 120.
[3] K. Koskinen, et al., First insights into the diverse human archaeome: specific detection of archaea in the gastrointestinal tract, lung, and nose and on skin, mBio 8 (6) (2017).
[4] C.R. Woese, O. Kandler, M.L. Wheelis, Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya, Proc. Natl. Acad. Sci. U. S. A. 87 (12) (1990) 4576–4579.
[5] M. Pester, C. Schleper, M. Wagner, The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology, Curr. Opin. Microbiol. 14 (3) (2011) 300–306.
[6] L. Guy, T.J. Ettema, The archaeal 'TACK' superphylum and the origin of eukaryotes, Trends Microbiol. 19 (12) (2011) 580–587.
[7] A. Spang, et al., Complex archaea that bridge the gap between prokaryotes and eukaryotes, Nature 521 (2015) 173.
[8] S.V. Albers, B.H. Meyer, The archaeal cell envelope, Nat. Rev. Microbiol. 9 (6) (2011) 414–426.
[9] K. Raymann, et al., Global phylogenomic analysis disentangles the complex evolutionary history of DNA replication in archaea, Genome Biol. Evol. 6 (1) (2014) 192–212.
[10] K.S. Makarova, E.V. Koonin, Archaeology of eukaryotic DNA replication, Cold Spring Harb. Perspect. Biol. 5 (11) (2013) a012963.
[11] M. Martinez-Pastor, et al., Transcriptional regulation in archaea: from individual genes to global regulatory networks, Annu. Rev. Genet. 51 (2017) 143–170.
[12] V. Anantharaman, E.V. Koonin, L. Aravind, Comparative genomics and evolution of proteins involved in RNA metabolism, Nucleic Acids Res. 30 (7) (2002) 1427–1464.
[13] E. Peeters, et al., The interplay between nucleoid organization and transcription in archaeal genomes, Nat. Rev. Microbiol. 13 (6) (2015) 333–341.
[14] A.M. Pyle, Translocation and unwinding mechanisms of RNA and DNA helicases, Annu. Rev. Biophys. 37 (1) (2008) 317–336.
[15] P. Linder, E. Jankowsky, From unwinding to clamping —the DEAD box RNA helicase family, Nat. Rev. Mol. Cell Biol. 12 (2011) 505.
[16] M.R. Singleton, M.S. Dillingham, D.B. Wigley, Structure and mechanism of helicases and nucleic acid translocases, Annu. Rev. Biochem. 76 (2007) 23–50.
[17] E. Jankowsky, M.E. Fairman, RNA helicases – one fold for many functions, Curr. Opin. Struct. Biol. 17 (3) (2007) 316–324.

[18] Mark S. Dillingham, Superfamily I helicases as modular components of DNA-processing machines, Biochem. Soc. Trans. 39 (2) (2011) 413–423.
[19] T.M. Lohman, E.J. Tomko, C.G. Wu, Non-hexameric DNA helicases and translocases: mechanisms and regulation, Nat. Rev. Mol. Cell Biol. 9 (2008) 391.
[20] S.S. Patel, K.M. Picha, Structure and function of hexameric helicases, Annu. Rev. Biochem. 69 (2000) 651–697.
[21] B. Medagli, S. Onesti, Structure and mechanism of hexameric helicases, Adv. Exp. Med. Biol. 767 (2013) 75–95.
[22] M. Abdel-Monem, H. Durwald, H. Hoffmann-Berling, Enzymic unwinding of DNA. 2. Chain separation by an ATP-dependent DNA unwinding enzyme, Eur. J. Biochem. 65 (2) (1976) 441–449.
[23] D.D. Leipe, et al., Classification and evolution of P-loop GTPases and related ATPases, J. Mol. Biol. 317 (1) (2002) 41–72.
[24] M.E. Fairman-Williams, U.-P. Guenther, E. Jankowsky, SF1 and SF2 helicases: family matters, Curr. Opin. Struct. Biol. 20 (3) (2010) 313–324.
[25] D.C. Beyer, M.K. Ghoneim, M. Spies, Structure and mechanisms of SF2 DNA helicases, in: M. Spies (Ed.), DNA Helicases and DNA Motor Proteins, Springer New York, New York, NY, 2013, pp. 47–73.
[26] K.D. Raney, A.K. Byrd, S. Aarattuthodiyil, Structure and mechanisms of SF1 DNA helicases, in: M. Spies (Ed.), DNA Helicases and DNA Motor Proteins, Springer New York, New York, NY, 2013, pp. 17–46.
[27] H. Chamieh, H. Ibrahim, J. Kozah, Genome-wide identification of SF1 and SF2 helicases from archaea, Gene 576 (1, Part 2) (2016) 214–228.
[28] A.K. Byrd, K.D. Raney, Structure and function of Pif1 helicase, Biochem. Soc. Trans. 45 (5) (2017) 1159–1171.
[29] C.J. Rudolph, et al., Is RecG a general guardian of the bacterial genome? DNA Repair (Amst.) 9 (3) (2010) 210–223.
[30] Y. He, G.R. Andersen, K.H. Nielsen, The function and architecture of DEAH/RHA helicases, Biomol. Concepts 2 (4) (2011) 315–326.
[31] K. Büttner, S. Nehring, K.-P. Hopfner, Structural basis for DNA duplex separation by a superfamily-2 helicase. Nat. Struct, Mol. Biol. 14 (2007) 647.
[32] H. Dürr, et al., X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA, Cell 121 (3) (2005) 363–373.
[33] L. Fan, et al., Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair, Mol. Cell. 22 (1) (2006) 27–37.
[34] L. Fan, et al., XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations, Cell 133 (5) (2008) 789–800.
[35] T. Nishino, et al., Structural and functional analyses of an archaeal XPF/Rad1/Mus81 nuclease: asymmetric DNA binding and cleavage mechanisms, Structure 13 (8) (2005) 1183–1192.
[36] T. Oyama, et al., Atomic structures and functional implications of the archaeal RecQ-like helicase Hjm, BMC Struct. Biol. 9 (1) (2009) 2.
[37] X. Zhang, et al., Crystal structure of an archaeal Ski2p-like protein from Pyrococcus horikoshii OT3, Protein Sci. 17 (1) (2008) 136–145.
[38] J.E. Cleaver, E.T. Lam, I. Revet, Disorders of nucleotide excision repair: the genetic and molecular basis of heterogeneity, Nat. Rev. Genet. 10 (11) (2009) 756–768.
[39] J. Kuper, C. Kisker, Damage recognition in nucleotide excision DNA repair, Curr. Opin. Struct. Biol. 22 (1) (2012) 88–93.
[40] C. Rouillon, M.F. White, The evolution and mechanisms of nucleotide excision repair proteins, Res. Microbiol. 162 (1) (2011) 19–26.
[41] A. Sancar, DNA excision repair, Annu. Rev. Biochem. 65 (1996) 43–81.

[42] W. Yang, Lessons learned from UvrD helicase: mechanism for directional movement, Annu. Rev. Biophys. 39 (2010) 367–385.
[43] H. Boubakri, et al., The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units in vivo, EMBO J. 29 (1) (2010) 145–157.
[44] C. Bruand, S.D. Ehrlich, UvrD-dependent replication of rolling-circle plasmids in Escherichia coli, Mol. Microbiol. 35 (1) (2000) 204–210.
[45] G.F. Moolenaar, C. Moorman, N. Goosen, Role of the Escherichia coli nucleotide excision repair proteins in DNA replication, J. Bacteriol. 182 (20) (2000) 5706–5714.
[46] A.B. Robertson, et al., MutL-catalyzed ATP hydrolysis is required at a post-UvrD loading step in methyl-directed mismatch repair, J. Biol. Chem. 281 (29) (2006) 19949–19959.
[47] R. Thakur, et al., Phospholipase D activity couples plasma membrane endocytosis with retromer dependent recycling, eLife 5 (2016).
[48] K. Theis, et al., The nucleotide excision repair protein UvrB, a helicase-like enzyme with a catch, Mutat. Res. 460 (3-4) (2000) 277–300.
[49] J. Eryilmaz, et al., Structural insights into the cryptic DNA-dependent ATPase activity of UvrB, J. Mol. Biol. 357 (1) (2006) 62–72.
[50] T.R. Waters, et al., Damage detection by the UvrABC pathway: crystal structure of UvrB bound to fluorescein-adducted DNA, FEBS Lett. 580 (27) (2006) 6423–6427.
[51] D.J. Crowley, et al., The uvrA, uvrB and uvrC genes are required for repair of ultraviolet light induced DNA photoproducts in Halobacterium sp. NRC-1, Saline Syst. 2 (2006) 11.
[52] D.L. Jones, B.K. Baxter, DNA repair and photoprotection: mechanisms of overcoming environmental ultraviolet radiation exposure in halophilic archaea, Front. Microbiol. 8 (2017) 1882.
[53] X. Giroux, S.A. MacNeill, A novel archaeal DNA repair factor that acts with the UvrABC system to repair mitomycin C-induced DNA damage in a PCNA-dependent manner, Mol. Microbiol. 99 (1) (2016) 1–14.
[54] L. Fan, et al., Conserved XPB core structure and motifs for DNA unwinding: implications for pathway selection of transcription or excision repair, Mol. Cell. 22 (1) (2006) 27–37.
[55] L. Fan, et al., XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations, Cell 133 (5) (2008) 789–800.
[56] T. Nishino, et al., Crystal structure and functional implications of Pyrococcus furiosus hef helicase domain involved in branched DNA processing, Structure 13 (1) (2005) 143–153.
[57] R.M. Story, H. Li, J.N. Abelson, Crystal structure of a DEAD box protein from the hyperthermophile Methanococcus jannaschii, Proc. Natl. Acad. Sci. U. S. A. 98 (4) (2001) 1465–1470.
[58] C. Rouillon, M.F. White, The XBP-Bax1 helicase-nuclease complex unwinds and cleaves DNA: implications for eukaryal and archaeal nucleotide excision repair, J. Biol. Chem. 285 (14) (2010) 11013–11022.
[59] P.C. Hanawalt, G. Spivak, Transcription-coupled DNA repair: two decades of progress and surprises, Nat. Rev. Mol. Cell Biol. 9 (12) (2008) 958–970.
[60] J.H. Min, N.P. Pavletich, Recognition of DNA damage by the Rad4 nucleotide excision repair protein, Nature 449 (7162) (2007) 570–575.
[61] F. Coin, V. Oksenych, J.M. Egly, Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair, Mol. Cell. 26 (2) (2007) 245–256.
[62] M. Newman, et al., Structure of an XPF endonuclease with and without DNA suggests a model for substrate recognition, EMBO J. 24 (5) (2005) 895–905.
[63] K. Komori, et al., Novel endonuclease in Archaea cleaving DNA with various branched structure, Genes Genet. Syst. 77 (4) (2002) 227–241.
[64] J.A. Roberts, S.D. Bell, M.F. White, An archaeal XPF repair endonuclease dependent on a heterotrimeric PCNA, Mol. Microbiol. 48 (2) (2003) 361–371.
[65] T. Nishino, et al., Crystal structure and functional implications of Pyrococcus furiosus hef helicase domain involved in branched DNA processing, Structure 13 (1) (2005) 143–153.

[66] M. Newman, et al., Structure of an XPF endonuclease with and without DNA suggests a model for substrate recognition, EMBO J. 24 (5) (2005) 895–905.
[67] K. Komori, et al., Cooperation of the N-terminal helicase and C-terminal endonuclease activities of Archaeal Hef protein in processing stalled replication forks, J. Biol. Chem. 279 (51) (2004) 53175–53185.
[68] R. Fujikane, et al., Genetic analysis of DNA repair in the hyperthermophilic archaeon, Thermococcus kodakaraensis, Genes Genet. Syst. 85 (4) (2010) 243–257.
[69] R. Lestini, Z. Duan, T. Allers, The archaeal Xpf/Mus81/FANCM homolog Hef and the Holliday junction resolvase Hjc define alternative pathways that are essential for cell viability in Haloferax volcanii, DNA Repair 9 (9) (2010) 994–1002.
[70] O.N. Voloshin, R.D. Camerini-Otero, The DinG protein from Escherichia coli is a structure-specific helicase, J. Biol. Chem. 282 (25) (2007) 18437–18447.
[71] J. Kuper, et al., Functional and structural studies of the nucleotide excision repair helicase XPD suggest a polarity for DNA translocation, EMBO J. 31 (2) (2012) 494–502.
[72] S.C. Wolski, et al., Crystal structure of the FeS cluster-containing nucleotide excision repair helicase XPD, PLoS Biol. 6 (6) (2008) e149.
[73] H. Liu, et al., Structure of the DNA repair helicase XPD, Cell 133 (5) (2008) 801–812.
[74] J. Rudolf, et al., The helicase XPD unwinds bubble structures and is not stalled by DNA lesions removed by the nucleotide excision repair pathway, Nucleic Acids Res. 38 (3) (2010) 931–941.
[75] C. Zhang, et al., Genetic manipulation in Sulfolobus islandicus and functional analysis of DNA repair genes, Biochem. Soc. Trans. 41 (1) (2013) 405–410.
[76] S.J. Johnson, R.N. Jackson, Ski2-like RNA helicase structures, RNA Biol. 10 (1) (2013) 33–43.
[77] C.P. Guy, E.L. Bolt, Archaeal Hel308 helicase targets replication forks in vivo and in vitro and unwinds lagging strands, Nucleic Acids Res. 33 (11) (2005) 3678–3690.
[78] R. Fujikane, et al., Identification of a novel helicase activity unwinding branched DNAs from the hyperthermophilic archaeon, Pyrococcus furiosus, J. Biol. Chem. 280 (13) (2005) 12351–12358.
[79] Z. Li, et al., Hjm/Hel308A DNA helicase from Sulfolobus tokodaii promotes replication fork regression and interacts with Hjc endonuclease in vitro, J. Bacteriol. 190 (8) (2008) 3006–3017.
[80] X. Song, J. Ni, Y. Shen, Structure-based genetic analysis of Hel308a in the hyperthermophilic archaeon Sulfolobus islandicus, J. Genet. Genomics 43 (6) (2016) 405–413.
[81] Y. Hong, et al., Dissection of the functional domains of an archaeal Holliday junction helicase, DNA Repair 11 (2) (2012) 102–111.
[82] K. Buttner, S. Nehring, K.P. Hopfner, Structural basis for DNA duplex separation by a superfamily-2 helicase, Nat. Struct. Mol. Biol. 14 (7) (2007) 647–652.
[83] S.J. Northall, et al., DNA binding and unwinding by Hel308 helicase requires dual functions of a winged helix domain, DNA Repair (Amst.) 57 (2017) 125–132.
[84] R. Fujikane, H. Shinagawa, Y. Ishino, The archaeal Hjm helicase has recQ-like functions, and may be involved in repair of stalled replication fork, Genes Cells 11 (2) (2006) 99–110.
[85] I.L. Woodman, K. Brammer, E.L. Bolt, Physical interaction between archaeal DNA repair helicase Hel308 and Replication Protein A (RPA), DNA Repair (Amst) 10 (3) (2011) 306–313.
[86] J.E. Walker, O. Luyties, T.J. Santangelo, Factor-dependent archaeal transcription termination, Proc. Natl. Acad. Sci. U. S. A. 114 (33) (2017) E6767–E6773.
[87] N.B. Reuven, et al., The gene for the longest known Escherichia coli protein is a member of helicase superfamily II, J. Bacteriol. 177 (19) (1995) 5393–5400.
[88] H. Ordonez, S. Shuman, Mycobacterium smegmatis Lhr Is a DNA-dependent ATPase and a 3′-to-5′ DNA translocase and helicase that prefers to unwind 3′-tailed RNA:DNA hybrids, J. Biol. Chem. 288 (20) (2013) 14125–14134.
[89] A. Ejaz, et al., Structure of mycobacterial 3'-to-5' RNA:DNA helicase Lhr bound to a ssDNA tracking strand highlights distinctive features of a novel family of bacterial helicases, Nucleic Acids Res. 46 (1) (2018) 442–455.

[90] X. Song, et al., Knockout and functional analysis of two DExD/H-box family helicase genes in Sulfolobus islandicus REY15A, Extremophiles 20 (4) (2016) 537–546.
[91] L. Yakovleva, S. Shuman, Mycobacterium smegmatis SftH exemplifies a distinctive clade of superfamily II DNA-dependent ATPases with 3′ to 5′ translocase and helicase activities, Nucleic Acids Res. 40 (15) (2012) 7465–7475.
[92] Matthew L. Bochman, et al., Hrq1, a homolog of the human RecQ4 helicase, acts catalytically and structurally to promote genome integrity, Cell. Rep. 6 (2) (2014) 346–356.
[93] S.-H. Kwon, et al., Saccharomyces cerevisiae Hrq1 requires a long 3′-tailed DNA substrate for helicase activity, Biochem. Biophys. Res. Commun. 427 (3) (2012) 623–628.
[94] A. Shiratori, et al., Systematic identification, classification, and characterization of the open reading frames which encode novel helicase-related proteins in Saccharomyces cerevisiae by gene disruption and Northern analysis, Yeast 15 (3) (1999) 219–253.
[95] G. Ofir, et al., DISARM is a widespread bacterial defence system with broad anti-phage activities, Nat. Microbiol. 3 (1) (2018) 90–98.
[96] P. Lulchev, D. Klostermeier, Reverse gyrase–recent advances and current mechanistic understanding of positive DNA supercoiling, Nucleic Acids Res. 42 (13) (2014) 8200–8213.
[97] H. Atomi, R. Matsumi, T. Imanaka, Reverse gyrase is not a prerequisite for hyperthermophilic life, J. Bacteriol. 186 (14) (2004) 4829–4833.
[98] A.C. Rodríguez, D. Stock, Crystal structure of reverse gyrase: insights into the positive supercoiling of DNA, EMBO J. 21 (3) (2002) 418–426.
[99] M.G. Rudolph, et al., Crystal structures of Thermotoga maritima reverse gyrase: inferences for the mechanism of positive DNA supercoiling, Nucleic Acids Res. 41 (2) (2013) 1058–1070.
[100] A. Valenti, et al., Inhibition of translesion DNA polymerase by archaeal reverse gyrase, Nucleic Acids Res. 37 (13) (2009) 4287–4295.
[101] A. Napoli, et al., Functional interaction of reverse gyrase with single-strand binding protein of the archaeon Sulfolobus, Nucleic Acids Res. 33 (2) (2005) 564–576.
[102] P. Linder, M.C. Daugeron, Are DEAD-box proteins becoming respectable helicases?, Nat. Struct. Biol. 7 (2) (2000) 97–99.
[103] S.S. Patel, I. Donmez, Mechanisms of helicases, J. Biol. Chem. 281 (27) (2006) 18265–18268.
[104] O. Cordin, et al., The DEAD-box protein family of RNA helicases, Gene 367 (2006) 17–37.
[105] F.V. Karginov, et al., YxiN is a modular protein combining a DExD/H core and a specific RNA-binding domain, J. Biol. Chem. 280 (42) (2005) 35499–35505.
[106] M.G. Rudolph, D. Klostermeier, The Thermus thermophilus DEAD box helicase Hera contains a modified RNA recognition motif domain loosely connected to the helicase core, RNA 15 (11) (2009) 1993–2001.
[107] K. Kossen, F.V. Karginov, O.C. Uhlenbeck, The carboxy-terminal domain of the DExDH protein YxiN is sufficient to confer specificity for 23S rRNA, J. Mol. Biol. 324 (4) (2002) 625–636.
[108] J. Lim, T. Thomas, R. Cavicchioli, Low temperature regulated DEAD-box RNA helicase from the antarctic archaeon, Methanococcoides burtonii11Edited by J. H. Miller, J. Mol. Biol. 297 (3) (2000) 553–567.
[109] Y. Shimada, et al., Property of cold inducible DEAD-box RNA helicase in hyperthermophilic archaea, Biochem. Biophys. Res. Commun. 389 (4) (2009) 622–627.
[110] E. Nagaoka, et al., Importance and determinants of induction of cold-induced DEAD RNA helicase in the hyperthermophilic archaeon Thermococcus kodakarensis, J. Bacteriol. 195 (15) (2013) 3442–3450.
[111] J.W. Gloor, et al., Biochemical analyses indicate that binding and cleavage specificities define the ordered processing of human Okazaki fragments by Dna2 and FEN1, Nucleic Acids Res. 40 (14) (2012) 6774–6786.
[112] H. Higashibata, et al., Helicase and nuclease activities of hyperthermophile Pyrococcus horikoshii Dna2 inhibited by substrates with RNA segments at 5′-end, J. Biol. Chem. 278 (18) (2003) 15983–15990.

[113] R.N. Jackson, et al., Fitting CRISPR-associated Cas3 into the helicase family tree, Curr. Opin. Struct. Biol. 24 (2014) 106–114.
[114] K.S. Makarova, Y.I. Wolf, E.V. Koonin, Comparative genomics of defense systems in archaea and bacteria, Nucleic Acids Res. 41 (8) (2013) 4360–4377.
[115] P.R. Bianco, C. Xu, M. Chi, Type I restriction endonucleases are true catalytic enzymes, Nucleic Acids Res. 37 (10) (2009) 3377–3390.
[116] S. Sistla, D.N. Rao, S-Adenosyl-L-methionine-dependent restriction enzymes, Crit. Rev. Biochem. Mol. Biol. 39 (1) (2004) 1–19.
[117] J. Yodh, ATP-dependent chromatin remodeling, Adv. Exp. Med. Biol. 767 (2013) 263–295.
[118] H. Durr, et al., X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA, Cell 121 (3) (2005) 363–373.
[119] L. Aravind, L.M. Iyer, V. Anantharaman, The two faces of Alba: the evolutionary connection between proteins participating in chromatin structure and RNA metabolism, Genome Biol. 4 (10) (2003) R64.
[120] H. Robinson, et al., The hyperthermophile chromosomal protein Sac7d sharply kinks DNA, Nature 392 (6672) (1998) 202–205.

THESIS OBJECTIVES

Helicases are proteins that use ATP energy to unwind nucleic acids and to remodel protein-nucleic acid complexes. They are involved in almost every aspect of the DNA and RNA metabolism and participate in numerous repair mechanisms maintaining cellular integrity. Helicases are classified into six superfamilies (SF1-6). The Lhr-type proteins belong to SF2 helicases that are poorly characterized to date, notably in archaea.

A phylogenomic study performed by the team of H. Chamieh in AZM center of the Lebanese University (Tripoli, Lebanon), classified SF1 and SF2 helicases from archaeal sequenced genomes and identified Lhr-type helicases as ubiquitous in Archaea. In this context, we wrote an overview of the functions of SF1 and SF2 helicases in Archaea (Introduction-PART I.4; Publication 1-Hajj et al, book chapter). The team of B. Clouet-d'Orval in Toulouse, France, is interested in RNA metabolism in archaea. In order to decipher the function of enzymes of the RNA metabolic pathways, protein-protein interaction networks were built (Phung et al. 2020). The presence of a protein annotated as Lhr2 in the interaction networks of the ASH-Ski2 helicase, and of the archaeal exosomal subunit of Rrp41 (Manon Batista thesis, 2020) raises the question about the significance of the helicase aLhr2 in these networks **(FIGURE 15)**. In addition to this work, it was previously shown that the archaeal Lhr2 was included in the protein interaction network of proteins involved in DNA replication and repair in *P. abyssi* (Pluchon et al. 2013) **(FIGURE 15)**. Moreover, previously published works proposed that bacterial and archaeal Lhr proteins are involved in DNA recombination and/or repair pathways and not in RNA metabolism (Buckley et al. 2020; Ejaz and Shuman 2018; Ordonez and Shuman 2013; Ejaz et al. 2018; De Felice et al. 2007).

In this context, this PhD project aims to understand the role of aLhr2 in cellular metabolism and the relevance of its presence in the RNA metabolism network in archaea, and to compare it to the function of *E. coli* Lhr protein. The PhD work presented in this manuscript is articulated around four objectives **(FIGURE 16)**. The first objective is to define the Lhr orthologous groups and their distribution in archaeal and bacterial domains by a phylogenomic approach (RESULTS-PARTI). Based on our phylogenetic classification and structural analysis of the different groups, we propose an evolution route and functions of Lhr enzymes in cellular metabolism. The second objective was to investigate the helicase activity of *Thermococcus barophilus* (*Tba*) aLhr2 by assessing its enzymatic properties as its capacity to hydrolyze ATP, to bind nucleic acid substrate, and to unwind or anneal nucleic acid duplex (Results-PART II;).

The third objective was to investigate the function(s) of *Tbar*-aLhr2 by performing functional studies. Pull-down assays identified the protein network of *Tbar*-aLhr2 and confirmed a link between aLhr2 and the RNA metabolism in *Tba*. Transcriptomes of wild-type and Δ*alhr2Tba* strains determined the effect of *alhr2* deletion (Δ*alhr2*) on gene expressions in *T. barophilus*. Interestingly, a significant number of genes affected by the deletion of aLhr2 are deregulated in similar proportions in a strain lacking the ribonuclease aRNase J (RESULTS-PART III). Finally, the fourth objective was to initiate a functional study of the bLhr helicase of *E.coli* in order to examine a putative interaction between *Eco*-bLhr and RNaseT, a ribonuclease known to be involved in DNA repair and tRNA/rRNA metabolism (RESULTS-PART IV).

Collectively, our data offers new perspectives on the role of Lhr helicases in Archaea. Different approaches of phylogenomic, proteomic, and transcriptomic and enzymatic assays will allow us to propose Lhr-type helicases as enzymes at the interface of RNA and DNA metabolism. The phylogenomics and biochemical characterization of aLhr2 will be part of a manuscript under preparation.

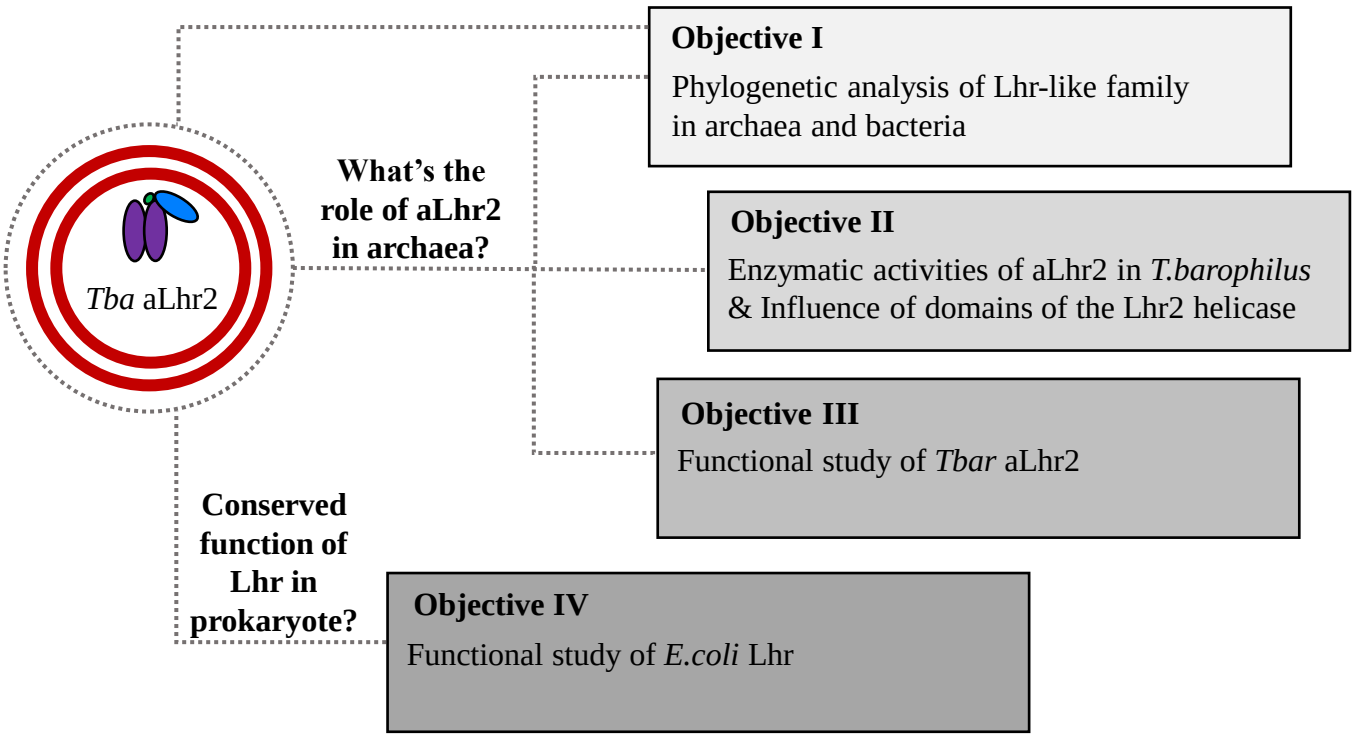


FIGURE 16. PhD Objectives.
This study has four mainly objectives. Objectives II, and III aim to decipher the role and function of *Tbar*-aLhr2. In objectives I and IV phylogenomic and functional studies were used to reveal common and differences functional and structural features of Lhr-type family in archaea and bacteria.

MATERIALS AND METHODS

I-Phylogenetic analysis

1-Building Lhr-type dataset

Completely sequenced and annotated genomes of 197 archaea and 2563 bacteria were downloaded from EBI (http://www.ebi.ac.uk/genomes/). The complete genomes of these 2760 strains, their proteomes and EMBL features were managed with an in-house MySQL database. Moreover, the protein sequences of these genomes have been annotated according to the conserved domain database downloaded from the NCBI (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) using the RPS-Blast program.
To retrieve Lhr-type proteins, we adopted three different approaches in order to collect the maximum of Lhr-type proteins. In the three approaches, very low threshold values were used again to avoid missing distantly related Lhr-type proteins even if false positives might be retrieved. These false-positives will then be discarded by analyzing the content of their functional domains. To avoid redundancy, the approaches were applied to one strain per species, resulting in a final set of 2760 analyzed genomes.

The family of Lhr helicases belongs to the SF2 helicase superfamily. Therefore, their members contain the two conserved RecA1 and RecA2 domains that constitute the helicase core and correspond to the pfam00270 and pfam00271 entries respectively. A third domain, located in the C-terminal region of the protein, the Domain 4 or the DEAD-associated domain, is specific of the Lhr helicases and corresponds to the pfam08494 entry.
In another functional domain collection, the COG database, the COG1201 profile covers the three previously described Pfam HMM profile entries.

In the first approach, we queried our MySQL database to retrieve the proteins that were annotated as containing the three Pfam domains. Only proteins showing an alignment with a score higher than 25 for pfam00270, 25 for pfam00271 and 40 for pfam08494 were retained. A first set of 615 sequences was obtained.

In the second approach, we collected from our MySQL database the proteins which possess the COG1201 domain with an alignment score greater than 40 and an alignment coverage of 30%. This low coverage value was used since in a preliminary analysis we observed that, in some proteins, insertion of another COG domain has occurred inside the protein region

corresponding to the COG1201 domain resulting in its split into two parts. A second set of 822 proteins was obtained.

Finally, in the third approach, we used results of previous studies performed with a fewer number of genomes that have allowed the construction of homemade HMM profiles for the characterization of the different orthologous groups of SF2 helicases. These HMM have been used to analyze and classify SF2 helicases from new genomes. Proteins of our MySQL database showing a score alignment greater than 40% with the HMM profile specific to the Lhr-type family have been retrieved. A very low score has been used to see if we can enrich our two previously obtained set of Lhr-type sequences. A third set of 894 proteins was obtained.

The three set of sequences were compared by building a Venn diagram. 823 proteins were retrieved either by the three approaches (614 proteins) or by two of them. The 71 proteins identified only by the third approach were further analyzed for their composition in functional domains and were discarded as false positive since they contain unrelated Lhr-type domains.

2-Identification of Lhr-type orthologous groups of proteins

To identify orthologous groups of Lhr-type proteins we first performed all-against-all BlastP comparisons of our initial set of 823 homologous sequences. Protein relationships were then converted into a graph in which the vertices represent protein sequences, and the edges represent their relationships that have been weighted according to their Log BlastP e-value. The graph was further processed by a graph-partitioning approach based on the Markov Clustering algorithm (MCL) using an Inflate Factor (IF) of 4. Different IF values were tested and an IF of 4 was required to identify the two already described Lhr1 and Lhr2 subfamilies of Lhr helicases among Archaea (Chamieh *et al.*, 2015). This identified five clusters of orthologous sequences, two clusters in archaea, and three clusters in bacteria.

3-Multiple Sequence Alignments and Phylogenetic Tree Constructions

The archaeal phylogenetic tree was inferred from a concatenated dataset of 81 protein families. If multiple copies occurred in a genome, all paralogs were removed. The alignments for each family were created using the MUSCLE program (Edgar, 2004) with the default parameters. We used the trimAl program (Capella-Gutiérrez et al, 2009) to remove spurious sequences and poorly aligned positions and to analyze the quality of the alignments according to gap numbers and residue conservation in the columns of the alignments. These parsed

alignments were concatenated to produce a single alignment of 17827 residues. When a species did not have a record for a family, the missing sequence was replaced by gaps in the alignment. The maximum-likelihood tree was computed with PhyML (Guindon & Gascuel, 2003) and the optimal combinations of parameters were selected using ProtTest3 program. The LG model of sequence evolution was used and the gamma-distributed substitution rate variation was approximated by eight discrete categories with shape parameter and proportion of invariant sites estimated from the data. The statistical branch support was inferred with the parametric bootstrap.

The Lhr-type helicase alignments were built by Mafft incorporating local pairwise refinement (L-INS-i) up to 2000 iteration (maxiterate 2000) (Katoh & Standley, 2013) and trimmed with trimAL as described above. The tree obtained on the full length aligned Lhr proteins was computed with PhyML using the combination of parameters determined with ProtTest3, *i.e.*, LG model of sequence evolution, the gamma-distributed substitution rate variation approximated by four discrete categories with shape parameter and proportion of invariant sites estimated from the data. The statistical branch support was inferred with the parametric bootstrap. The tree obtained on the helicase core-based tree and the PF08494 domain-based tree were computed with IQtree (Nguyen et al. 2015), using the combination of parameters determined with Ultrafast bootstrap (1000 replicates) approximation approach to computing the support of phylogenetic groups in maximum likelihood tree (Minh et al. 2013), LG+F+R10 and GTR20+F+R8 respectively.

The trees were arbitrarily rooted and were drawn with the online version of iTOL.

4-Genomic context analysis

The genomic contexts of Lhr encoding genes were visualized with Prokaryotic Syntax (Synteny & Taxonomy Explorer, https://archaea.i2bc.paris-saclay.fr) (Despalins et al. 2011). Genomic contexts were also examined manually since conserved clusters of genes encoding Lhr, MPE, and potential other proteins were not detected with an automated tool.

5-Structural model predictions

Protein secondary structures were predicted using the PSIPRED workbench as described in (McGuffin et al. 2000). Structure models and protein fold were obtained using Phyre2 (Kelley et al. 2015) under Intensive mode. Structural homology searches and predicted structure

models analyses were performed with DALI (Holm and Sander 1995). Model structure rendered in PyMOL (http://www.pymol.org) were superimposed against the *M. smegmatis* Lhr helicase structure (Ordonez and Shuman 2013) (PDB: 5V9X) using PyMOL. Sequence logos indicating sequence conservation were generated using WebLogo 3 (http://weblogo.threeplusone.com).

Table 7. Expression vectors constructed or already available (✓) and used (✓✓) in this study.

	aLhr2 WT	**alhr2 ΔDom4**	**alhr2 Dom4**	**aLhr2 K60A**	**aLhr2 T215A**	**aLhr2 I512A**	**aLhr2 W577A**	***Ecol*-Lhr-856**	***Ecol*-RNaseT**
pET11b	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	✓	
pET15b	✓✓	✓	✓						
pET21b	✓							✓✓	
pGEX-4T1									✓✓

Table 8. List of primers used in this study.

Primers	Sequences	Purpose	
B18-2	TTAGCAGCCGGATCCTCATTCAAGCTCCCCGATCA	WT aLhr2 In fusion cloning in pET vectors	
B18-3	CGCGGCAGCCATATGATGATAAGGTTTGCCCAGAGA		
B18-4	TTAGCAGCCGGATCCTCATTCAAGCTCCCCGATCA		
B18-5	GTGGTGCTCGAGTGCTCATTCAAGCTCCCCGATCA		
B18-6	CCTCCTATCAAGCGTGTAAAC	Sequencing *Tbar*-aLhr2	
B18-7	GTGTTGTTAGCTCAACATCGC		
B18-9	TGGAATTGTACCCGTGTTCATG	Inverse PCR aLhr2 ΔDom4	
B17-33	P- TAACAAAGCCCGAAAGGAAGCT		
B18-10	P- GATGAGGCTAAAATCGAAGTTTA	Inverse PCR Domain 4	
B15-1	CATATGTATATCTCCTTCTTAAAGTT		
B19-02	AGCGGAGCTACTTTATCAGCTTTTCTCGCTGCC	K60A	Site-directed mutagenesis *Tbar*-aLhr2
B19-03	AAAAGCTGATAAAGTAGCTCCGCTGCCCGTTGG		
B19-04	GATCCCAGCGGCGTTTTCTGAAATGTTGCCT	W577A	
B19-05	TTCAGAAAACGCCGCTGGGATCGTTGGCTT		
B19-06	AACACGGGTACAGCTCCAGATGAGGCTAAA	I512A	
B19-07	ATCTGGAGCTGTACCCGTGTTCATGTAGTA		
B19-22	AACCTCTTCGAGGGGATGAATAGCGGCACT	T215A	
B19-20	TTCGTTAGAATCGGTCTCAGTGCCGCTATTC		
B15-29	TAATACGACTCACTATAGGGGA	T7 promoter	
B15-30	GCTAGTTATTGCTCAGCGGT	T7 Terminator	
MH5	GGAGATATACATATGATGGCAGATAATCCAGACCCT	*Ecol*-Lhr-856 In fusion cloning in pET vectors	
MH7	GTGGTGCTCGAGTGCGGATATTTGAATTTCGCCACC		
MH8	TAATACGACTCACTATAGGGGA	Sequencing *Ecol*-Lhr-856	
MH9	GCTAGTTATTGCTCAGCGGT		
MH10	GGGGCGAAAAGCAGAAAAATC		
MH11	CCTCTAACCGTGTACAAAGTA		

The different expression vectors used in this study are listed in **Table 7**. Primers sequences with the purpose of each one are listed in **Table 8.** Buffers and media compositions are listed in **Table 9.** Nucleic acid substrates used for the enzymatic activities are listed in **Table 10.**

II-Construction of expression vectors

1-Construction of expression vectors carrying aLhr2 gene (TERMP_00533).

The pET vectors are amplified by reverse PCR with the Prime Star max enzyme (Ozyme) in the presence of 10ng of plasmid DNA and 0.3 µM of the appropriate pairs of primers (Thermocycler program: [98 ° C (15 '); 56 ° C (15 "); 72 ° C (1'30 ")] * 35cycles) (**Table 8)**. The PCR products were incubated at 37°C in the presence of 20 units of DpnI (ThermoScientific) for 1 h to remove the template. The gene encoding aLhr2 protein was amplified by PCR from genomic DNA of *Thermococcus barophilus* with 1 U of Phusion (High-Fidelity DNA polymerase from Finnzymes), 0.2 mM of dNTPs and 0.5 µM of the pairs of primers (Thermal cycler program: [98 ° C (30 "); 58 ° C (30 "); 72 ° C (1 ')] * 35cycles). To allow the assembly of the PCR products in the InFusion® reaction. The primers allowing the amplification of the coding sequences were synthesized to integrate, at their 5 'ends, 15 nucleotides complementary to that of the vectors. The InFusion® reactions were performed as recommended by the provider with equimolar amounts of PCR products. The entire reaction products were introduced by heat shock in 50 µL of chemocompetent Stellar® bacteria (Ozyme) and selected in the presence of ampicillin (100 µg/ml) after 1 hour of phenotypic expression at 37°C. The presence of the insert in the vectors was verified by PCR on colonies with the T7 Promoter and T7 Terminators primers (**Table 8**). The insert was verified by sequencing with the same pair of primers (MWG Eurofin).

2-Construction of expression vectors carrying aLhr2 Dom4 and aLhr2 ΔDom4.

This method was used to construct the expression vectors containing the genes aLhr2 ΔDom4 and aLhr2 Dom4 consists of excluding part of the WT gene by reverse PCR using phosphorylated primers in the 5' extremity on both sides of the domain to be eliminated. The pET11b and pET15b vectors were amplified by reverse PCR with the Prime Star max enzyme (Ozyme) in the presence of 10 ng of plasmid DNA and 0.3 µM of the appropriate phosphorylated pairs of primers (Thermocycler program: [98 ° C (15 '); 56 ° C (15 "); 72 ° C (1'30 ")] * 35cycles). The PCR reaction was incubated at 37°C in the presence of 20 U of DpnI (ThermoScientific) for 1 h to remove the template. The vectors were circularized using

30 U of T4 DNA ligase (BioLabs) for 1 h at 20 ° C. The entire reaction products were introduced by electric shock into 50 µL of electrocompetent DH5α bacteria and the transformants were selected in the presence of ampicillin (100 µg/ml) after 1 h of phenotypic expression at 37°C. The mutated inserts were verified by sequencing using T7 Promoter and T7 Terminator primers (MWG Eurofin).

3-Construction of expression vectors carrying aLhr2 with punctual mutations.

The pET11b vector expressing the recombinant proteins aLhr2 T215A, aLhr2 I512A, and aLhr2 W577A were constructed by directed mutagenesis using the QuikChangeXL® kit (Agilent) as recommended by the provider. 10 ng of the plasmid expressing aLhr2 T215A, aLhr2 I512A, and aLhr2 W577A were added to 2.5 U of Pfu Ultra HF and 125 ng of pairs of primers containing the point mutation to be introduced (Thermocycler program: [95 ° C (50 "); 60 ° C (50 "); 68 ° C (12 ')] * 18cycles). The PCR reaction was incubated at 37°C in the presence of 20 U of DpnI (ThermoScientific) for 1 h to remove the template. The entire reaction products were introduced by heat shock in 45 µL of chemocompetent XL10-Gold® bacteria (Agilent) and the transformants were selected in the presence of ampicillin (100 µg/ml). The integration of the mutations was confirmed by sequencing with the T7 Promoter and T7 Terminator primers (MWG Eurofin).

Table 9: List and composition of buffers and media used in this study

Medium/Buffer	Composition
LB medium	5g/L yeast extract; 10g/L NaCl; 10g/L Tryptone; 20mg/mL Thymine pH 7
SOC medium	SOB medium + 0.4% glucose; 0.01M $MgCl_2$; 0.1M $MgSO_4$
PBS 1X	137mM NaCl; 2,7mM KCl; 10mM Na2HPO4; 1,8mM KH2PO4
Lysis buffer	50mM Tris-HCl pH 7.5; 300mM NaCl; 10% glycerol
HisTrap buffer A	20mM NaP pH 7.4; 300mM NaCl; 10% glycerol
HisTrap buffer B	1mM Imidazole; 20mM NaP pH 7.4; 300mM NaCl; 10% glycerol
Heparin buffer A	50mM Tris-HCl pH 7.5; 10% glycerol
Heparin buffer B	50mM Tris-HCl pH 7.5; 300mM NaCl; 10% glycerol
Gel Filtration buffer	50mM HEPES pH 6.8; 500 mM NaCl; 10% glycerol
ATPase buffer 5X	50mM Hepes pH 7.5; 50mM KCl; 5mM MgCl2; 2mM DTT
SSC 1X buffer	150mM NaCl; 15mM NaCi
Helicase buffer 1X	25mM Tris-HCl pH 8, 50mM NaAc, 5mM MgCl2, 2.5mM β-Mercaptoethanol
STOP buffer 5X	0.1% SDS, 8mM EDTA, 0.1mg/ml Proteinase K, 4% glycerol, 0.02% bromophenol-blue
Cold buffer I	50mM Sodium phosphate pH 7; 200mM NaCl; 1mM PMSF
Cold buffer II	Cold buffer I + 20 mM Imidazole
Cold buffer III	50mM Tris-HCl pH6.8; 2% SDS; 0.1% bromophenol blue; 10% glycerol; 6M Urea; 100mM DTT; and 250 mM Imidazole
Lysis buffer (RNaseT)	20mM Tris pH 7.5; 0.1% Triton X-100;1 mg/ml lysozyme; 1µl/ml DNase
RNase T wash buffer	20mM Tris pH 7.5; 0.1% Triton X-100
RNase T elution buffer	10mM glutathione;50mM Tris pH 8

III-Production and purification of recombinant proteins

1-Expression and production of *Tbar*-aLhr2 recombinant proteins.

The pET vectors carrying aLhr2 WT or variants encoding genes were transformed into the *E.coli* BL21 (DE3) codon + expression strain by electroporation. The transformants were selected in the presence of ampicillin (100 µg/ml) and chloramphenicol (30 µg/ml). Overnight cultures, diluted to 1/100 in LB medium (Luria Berthani, 5g/L of yeast extract, 10g/L NaCl, 10g/L Tryptone, 20mg/mL Thymine pH 7), containing ampicillin and chloramphenicol, were incubated at 37°C with constant shaking (180 rpm) until the A600 reached 0.8. Protein production was induced by adding 0.2 mM of IPTG for 3h at 30°C. The cells were collected by centrifugation (15min, 6000rpm at 4°C. (rotor JA-14)). After a washing step with 1X PBS (137mM NaCl, 2.7mM KCl, 10mM Na2HPO4, 1.8mM KH2PO4), the cell pellets obtained after a second centrifugation (15min, 5000rpm at 4°C (JS rotor 5.3)) were stored at -20° C.

2-Purification of *Tbar*-aLhr2 recombinant proteins.

The cell pellets overexpressing the proteins of interest (2x400ml) were resuspended in 10mL of lysis buffer (50mM Tris HCl pH 7.5, 150mM NaCl, 10% glycerol) supplemented with lysosyme (1mg/ml) and 1/2 tablet of protease inhibitor (cOmplete ™ Roche) and lysed by three cycles of [5x10sec] of sonication (50% cycle) (Bioblock scientific, Vibra Cell). A treatment with a mix of RNase A (20ng/mL), RNase T1 (1000 units/µl) and DNase I (20ng/mL) containing 10mM of MgCl2 was performed at 37°C for 30 min. After a heating step at 70°C for 20 min, the extracts were furthered clarified by centrifugation (20 000 g, 4°C, 20 min (rotor JA-20)). The soluble fraction supernatant was then filtered (0.22 µm). The purification was carried out with FPLC (Fast Protein Liquid Chromatography) on an Akta system at room temperature.

2.1 Purification of aLhr2 proteins expressed in pET11b vectors

The cell extract diluted to a final concentration of 150 mM of NaCl was injected onto a 1ml cation exchange column (HiTrap ™ SP HP GE Healthcare) already equilibrated in the Heparin buffer (50mM Tris HCl pH 7.5, 150mM NaCl, 10% glycerol) on Akta FPLC (Fast Protein Liquid Chromatography) system at room temperature. After two washing step with 150 and 200 mM of NaCl, the elution was carried out with a linear gradient of 300 mM to 1M of NaCl. These steps are common to the purification of aLhr2 WT and aLhr2 Dom4, aLhr2 T215A, aLhr2 aLhr2 I512A, and aLhr2 W577A variants. An additional purification step was

necessary for of aLhr2 WT and aLhr2 Dom4, aLhr2 aLhr2 I512A, and aLhr2 W577A. The fractions of the peak corresponding to the proteins of interest were injected on HiLoad 16/60 Superdex 200 PG column (10 000 -600 000Da) already equilibrated in Gel filtration buffer (gel filtration (50mM HEPES pH 6.8, 300mM NaCl, 10% glycerol). The purity of the protein fractions was visualized after migration on a 4-15% SDS-PAGE gel (BioRad) and visualization with coomassie blue (Quick Coomassie, CliniSciences).

The exact purification steps (from lysis to injection onto columns on FPLC) were used to purify aLhr2 ΔDom4 using Heparin (Heparin FF GE Healthcare) and HiLoad 16/60 Superdex 200 PG columns.

2.2 Purification of aLhr2 proteins expressed in pET15b vectors

The cell extract was injected in a 1ml nickel column (HisTrap ™ GE Healthcare) already equilibrated in HisTrap buffer (20mM NaP pH 7.4, 300mM NaCl, 10% glycerol) on Akta FPLC system at room temperature. After two washing steps at 10 and 20 mM Imidazole, the elution was carried out with a linear gradient of Imidazole (50 to 500 mM). The fractions containing the proteins of interest were then injected in HiLoad 16/60 Superdex 200 PG column. The purity of the protein in the fractions was visualized after migration on a 4-15% SDS-PAGE gel (BioRad) and visualization with Coomassie blue (Quick Coomassie, CliniSciences).

2.3 Quantification and storage of proteins

The concentrations of aLhr2 WT and variants polypeptides were determined by SDS-PAGE analysis of aliquots of each preparation in parallel with a range of BSA standard (0.16 to 2.5 mg/ml) on 4-15% SDS-PAGE gel (Biorad) stained with Sypro Orange. The proteins were visualized on a Bioimager and quantified on the MultiGauge software.

The proteins are dialyzed (Slide-A-Lyser®, Thermo Scientific) in Gel filtration buffer (50mM HEPES pH 7.8, 300 or 500mM NaCl, 10% Glycerol) and stored at 16 ° C or aliquoted and frozen at -80 ° C.

Table 10: Nucleic acids substrates used in this study in ATPase, binding, unwinding and annealing assays.

Substrate type $_{lenght}$	Name	Sequence
DNA_{50}	S1	5’ ATCGATAGTCTCTAGACAGCATGTCCTAGCAAGCCAGAATTCGGCAGCGT 3’
RNA_{50}	US1	5’ AUCGAUAGUCUCUAGACAGCAUGUCCUAGCAAGCCAGAAUUCGGCAGCGU 3’
5’DNA_{26}	5RS1	5’ GGACATGCTGTCTAGAGACTATCGAT 3’
5’RNA_{26}	U5RS1	5’ GGACAUGCUGUCUAGAGACUAUCGAU 3’
3’RNA_{26}	U3RS1	5’ ACGCUGCCGAAUUCUGGCUUGCUAGG 3’
DNA_{59}	S2	5’ GACGCTGCCGAATTCTACCAGTGCCTTGCTAGGACATCTTTGCCCACCTGCAGGTTCAC 3’
5’DNA_{31}	5RS2	5’ TAGCAAGGCACTGGTAGAATTCGGCAGCGTC 3’
5’DNA_{59}/ DNA_{31}	S2/ 5RS2	5’GACGCTGCCGAATTCTACCAGTGCCTTGCTAGGACATCTTTGCCCACCTGCAGGTTCAC 3’ 3’ CTGCGACGGCTTAAGATGGTCACGGAACGAT 5’
3’DNA_{50}/ RNA_{26}	S1/ U5RS1	5’ATCGATAGTCTCTAGACAGCATGTCCTAGCAAGCCAGAATTCGGCAGCGT 3’ 3’UAGCUAUCAGAGAUCUGUCGUACAGG 5’
3’RNA_{50}/ RNA_{26}	US1/ U5RS1	5’AUCGAUAGUCUCUAGACAGCAUGUCCUAGCAAGCCAGAAUUCGGCAGCGU 3’ 3’UAGCUAUCAGAGAUCUGUCGUACAGG 5’
5’RNA_{50}/ RNA_{26}	US1/ U3RS1	5’AUCGAUAGUCUCUAGACAGCAUGUCCUAGCAAGCCAGAAUUCGGCAGCGU 3’ 3’GGAUCGUUCGGUCUUAAGCCGUCGCA 5’

IV-Enzymatic assays

1-Radiolabelled substrates

The substrates S1, 5RS1, U5RS1 and U3RS1 are phosphorylated at 5 ' extremity with 50µCi of ATP [γ32] using 5 units of T4 polynucleotide kinase (ThermoScientific) for 1h at 37°C. A The S1 substrate was amplified to obtain a PCR product supplemented with a T7 promoter sequence at the 5' end. The preparation of radiolabeled nucleic acid substrate US1 is obtained by *in vitro* transcription of the PCR product using the MEGAscript kit (Ambion) as recommended by the provider and in the presence of 50 µCi of UTP [α32]. The US1/U5RS1 and US1/U3RS1 duplexes are obtained by incubating the single strands at 90° C for 5 min in 1X SSC buffer (150 mM NaCl, 15 mM NaCi) followed by a step of cooling to room temperature of about 2h minimum. The duplexes are desalted on column (Micro-bio Spin, Biorad).

2-ATPase hydrolysis assay.

The purified protein (500nM) was incubated in presence or absence of 5nM of S1(ssDNA-50nt) or US1(ssRNA-50nt) in 5X buffer (50mM Hepes pH 7.5, 50mM KCl, 5mM MgCl2, 2mM DTT) at 65°C for 10min. ATP hydrolysis was stimulated by adding 4.5 µl of ATP mix containing 0.2mM of ATP and 0.085µl of [γ32P] ATP at 6000Ci/mmol in 5x buffer to obtain a final reaction of 17µl. The final reaction was incubated at 65°C. At 0, 2, 5, 10, 30, 60, and 90 min, 2µl of the reaction was spotted on TLC (Thin Layer Composite) membrane. ATP and released Pi were separated using 100ml of 250 mM KH_2PO_4. TLC membrane, already dried, was exposed in cassettes with a phosphor imaging plate (BAS-MS Imaging Plates-Fujifilm) overnight. The phosphorimager was then scanned using the Typhoon scanner (Typhoon Trio-Amersham Biosciences, 200 microns of resolution). The quantification of the hydrolysed ATP was done using MultiGauge software (Fujifilm). Curves and ATPase kinetic parameters were performed with the software GraphPad Prism 7.

The kinetic parameters for ATP hydrolysis, K_m and k_{cat}, were determined by nonlinear fitting to the Michaelis-Menten equation using GraphPad Prism for a range of ATP concentrations (0.1 M to 2 M) in the presence of excess DNA_{50} cofactor.

3-Nucleic acid binding assay.

A range of the protein concentration (0 to 150 or 350nM depending on the nucleic acid substrate) was used to test the binding activity of the WT aLhr2 or variant. Proteins were incubated for 10min at 60°C, a mix containing 0.5nM of labelled substrate and helicase buffer 1X (25mM Tris-HCl pH 8, 50mM NaAc, 5mM MgCl2, 2.5mM β-Mercaptoethanol) was added to obtain a final volume of 50µl. After incubation for 15min at 30°C, the reaction was stopped on ice. Nitrocellulose-filter binding technique was used to study protein-nucleic acid interactions. Briefly, reactions were spotted on nitrocellulose membrane (already incubated in 0.5M of KOH and in the helicase buffer 1X) and a nylon membrane (Hybond- already incubated in helicase buffer 1X) in the slot blot (Amersham Biosciences). Membranes were then dried and exposed in cassettes with a phosphor imaging plate (BAS-MS Imaging Plates-Fujifilm) overnight. The phosphorimager was then scanned using the Typhoon scanner (Typhoon Trio-Amersham Biosciences, 100 microns of resolution). The quantification was done using MultiGauge software (Fujifilm). Curves and KD analysis were performed with the software GraphPad Prism 7.

4-Annealing assay

A reaction mix containing 5nM of unlabelled single-strand RNA (US1) or DNA (S1 or S2), 5mM $MgCl_2$ and helicase buffer 1X (25mM Tris-HCl pH 8, 50mM NaAc, 2.5mM β-Mercaptoethanol) was pre-incubated separately from the protein (250nM) for 5min at 65°C. The radiolabelled single-strand RNA (U5RS1* or U3RS1*) or DNA (5RS1*) was then added to the reaction solution with the protein. A kinetic was performed at 65°C by collecting 5µl of the reaction at different times. The reaction was stopped by adding 1.25µl of STOP buffer 1X (0.1% SDS, 8mM EDTA, 0.1mg/ml proteinase K, 4% glycerol, and 0.02% Bromophenol blue). A reaction without the protein was also used as control. Duplex were separated from single-stranded substrates using electrophoresis in acrylamide gel 8% containing 0.1% SDS at 160V for 1h30. The dried gel was then exposed in cassettes with a phosphor imaging plate (BAS-MS Imaging Plates-Fujifilm) overnight. The phosphorimager was then scanned using the Typhoon scanner (Typhoon Trio-Amersham Biosciences, 200 microns of resolution). The quantification was done using MultiGauge software (Fujifilm). Curves were established using GraphPad Prism 7.

5-Unwinding assay

A starting solution containing 5mM ATP, 5mM $MgCl_2$ an excess of the unlabelled oligo trap (1µM of U5RS1, U3RS1 or 5RS2) and a reaction mix composed of 5nM of the radiolabelled

duplex (US1/U5RS1*, US1/U3RS1* or S2/5RS2*), the protein of interest (250nM) and the helicase buffer 1X (25mM Tris-HCl pH 8, 50mM NaAc, 2.5mM β-Mercaptoethanol) were incubated for 5min at 65°C. The unwinding reaction was triggered when the starting solution was added to the reaction mix. A kinetic was performed at 65°C by collecting 5µl of the reaction at 0, 2, 5, 10, 20, 40, 60, and 90min. The reaction was stopped by adding 1.25µl of STOP buffer 1X (0.1% SDS, 8mM EDTA, 0.1mg/ml proteinase K, 4% glycerol, and 0.02% Bromophenol blue). A reaction without the protein was also used as control. Duplex were separated from single-stranded substrates using electrophoresis in acrylamide gel 8% containing TBE 1X, 20mM EDTA and 0.1% SDS at 160V for 1h30min (Gel migration buffer: TBE 1X, 20mM EDTA and 0.1% SDS). The dried gel was then exposed in cassettes with a phosphor imaging plate (BAS-MS Imaging Plates-Fujifilm) overnight. The phosphorimager was then scanned using the Typhoon scanner (Typhoon Trio-Amersham Biosciences, 200 microns of resolution). The quantification was done using MultiGauge software (Fujifilm). Curves were established using GraphPad Prism 7.

V-Protein interaction networks and Transcriptomic analysis

1-Pull down assay

This method consists of incubating a $(His)_6$-tagged recombinant protein attached to cobalt beads with *T.barophilus* cell extract. Pull-down experiments were carried out in triplicate using *T.barophilus* cellular extracts cultivated in bioreactors at the exponential phase of growth under physiological conditions (85 °C, pH 7, anaerobic) at the Ifremer Laboratory (D. Flament; Brest). *T.barophilus* cell pellets are resuspended in (w/v 1/3) TK buffer (20mM TrisHCl pH7.4, 10mM MgCl2, 100mM KCl, 1mM DTT) supplemented with a mix of protease inhibitors (cOmplete ™ Roche). After sonication (VibraCell Biolock Scientific), the extract is clarified by centrifugation (10,000g, 60min, 4°C) and supplemented with 10mM MgCl2. The purified protein aLhr2 with a Histidine tag was used as bait. 20 µg of bait proteins are immobilized in 0.6 mg of magnetic cobalt beads (Dynabeads, Invitrogen). The bead-bait complex is incubated with 2mg of *T.barophilus* extract under rotation overnight at room temperature. To eliminate interactions mediated by DNA or RNA molecules, this step is carried out in the presence of a DNase/RNase mixture (10 µg /ml of RNase A and DNase I). The protein complexes formed *in vitro* are separated on a magnetic rack and washed with TK buffer (3x200 µl). A control reaction was also carried out under the same conditions using cobalt beads without bait protein. Proteins were then eluted in 25 µl of XT buffer (Biorad) containing 2 µl of 20X reducing agent for 10 min at 95 ° C.

2-Proteomic analysis

Mass spectrometry analyses were carried out at Pris Sud-Ouest PAPPSO (http://papso.inra.fr). After a short migration on SDS-PAGE gel (Criterion XT-Precast gel-BioRad), the protein bands are cut out and sent to the PAPPSO platform. Bands treatment method was performed as described previously (Branson and Freitas 2016). Data were processed to identify signals for specific interactions. The globality of the specific spectra was normalized between the series of replicates. A cut-off of 2 normalized spectra was used as the minimum MS signal for confirming network hits. The normalized spectra are then averaged between replicates and references versus control to calculate the number of “referenced spectra”. The calculation of the "specific index" score is the ratio of the normalized spectra versus controls. The "specificity index" varies from 0 to 1 (with a maximum threshold = 1), the closer the specificity index value to 0 the better is the specificity.

3-Transcriptomic RNA seq analysis

Total RNAs from wild type and *ΔTbar-alhr2* strains cells, at the exponential growth phase, were extracted using TRIZOL® reagent. The mutant strain constructions, cell cultures and RNA extractions were performed by Yann Moalic at the Ifremer in Brest (Birien et al. 2018; Thiel et al. 2014). For each strain, biological triplicates were produced. In Toulouse, DNA was removed from the samples using the TURBO DNA-free™ Kit (Invitrogen) as recommended by the provider. The quantity and the quality of the RNA samples were assessed using a Qubit fluorometer and a bioanalyser®, respectively, by the GeT-PlaGe platform (GenoToul). For each sample, 1µg of total RNA DNA-free were sent to the Genewiz company for strand-specific cDNA library construction and high-throughput sequencing (Illumina HiSeq® ; 2x150bp; 30 million reads per sample). The sequencing data were analyzed by Christine Gaspin (INRAE Toulouse - MIA-T lab) and Marta Kwapisz (Clouet d’orval team, IBCG Toulouse). Briefly, the data were normalized using the relative log-expression method. The differentially expressed genes between the wild-type and the mutant strains were identified using the DE-seq R package.

VI-*In vitro* interaction preliminary assays test of *Ecol*-Lhr-856 and RNAse T

1-Construction of expression vectors carrying *Ecol*-Lhr-856 gene (JW1645)

The *Ecol*-Lhr-856 gene was amplified from *E.coli* K-12 substrain MG1655 genomic DNA. Vectors were constructed with the kit InFusion® (Ozyme) using the same protocol as for

Tbar-aLhr2. Primers used for the pET11b and pET21b construction were listed in **Table 8**. The nucleotide sequences of the cloned insert were verified by sequencing using the same primer pair as for cloning (MWG Eurofins).

2-Expression and production of *Ecol*-Lhr-856 and *Ecol*-RNaseT recombinant proteins.

The PGEX4T1 vector carrying the synthetic gene encoding *Ecol*-RNAseT was provided by GenScript. The expression strain, *E. coli* BL21 (DE3) codon+, was transformed with the expression pET21b vector carrying *Ecol*-Lhr or the pGEX4T1 carrying *Ecol*-RNaseT gene by heat shock for 1'10s at 42°C. Transformed colonies were selected by the presence of ampicillin (100µg/ml) and chloramphenicol (0.03 mg/ml). An overnight culture using one single colony was diluted at 1/100 in fresh SOB medium (Himedia), containing ampicillin (100µg/ml) and chloramphenicol antibiotics (0.03 mg/ml), and incubated at 37°C with 180rpm of agitation. The cultures were induced by 0.1 mM of IPTG when the growth achieved an A_{600} of 0.6 by incubation for 3h (*Ecol*-Lhr) or overnight (*Ecol*-RNaseT) at 30°C. Cells were harvested by centrifugation (10min, 7000rpm à 4°C) and conserved at -20°C.

3-Purification of *Ecol*-Lhr-856 recombinant proteins.

Cell pellets containing the tagged protein of interest were resuspended in cold buffer I containing 50mM Sodium phosphate pH 7, 200mM NaCl, and PMSF. Crude cell extracts were obtained by sonication (5 cycles of 5s (level 3) and centrifugation (25 min, 7000rpm) at 4°C. The supernatant (soluble fraction) collected by centrifugation was loaded onto a 6ml Nickel (Ni^{2+})-column (His60 Ni Superflow^{TM} Resin, Takara) already equilibrated with buffer I and washed three times with 1ml of buffer II (same as buffer I with 20 mM Imidazole). Proteins were eluted with buffer III (50mM Tris-HCl pH6.8, 2% SDS, 0.1% bromophenol blue, 10% glycerol, 6M Urea, 100mM DTT, and 250 mM Imidazole). Eluted proteins were analyzed on 12% SDS-PAGE followed by coomassie brilliant blue (CBB-R250) staining.

4-Co-purification of recombinant protein on Nickel column

Cell pellets of 400 ml containing the bait (His_6-*Ecol*-Lhr) and prey (GST-*Ecol*-RNaseT) proteins were resuspended separately in cold buffer I containing 50mM Sodium phosphate pH 7, 200mM NaCl, and PMSF. Crude cell extracts were obtained by sonication (5 cycles of 5s (level 3) and centrifugation (25 min, 7000rpm) at 4°C. The supernatant (soluble fraction) of the bait protein collected by centrifugation was loaded onto a 400µl Nickel (Ni^{2+})-column (His60 Ni Superflow^{TM} Resin, Takara) already equilibrated with buffer I and washed three

times with 1ml of buffer II (same as buffer I with 20 mM Imidazole). After a wash step with buffer II, the supernatant of the prey proteins was added. Proteins were eluted with buffer III (50mM Tris-HCl pH6.8, 2% SDS, 0.1% bromophenol blue, 10% glycerol, 6M Urea, 100mM DTT, and 250 mM Imidazole). Eluted proteins were analyzed on 12% SDS-PAGE followed by coomassie brilliant blue (CBB-R250) staining.

RESULTS

I-Phylogenetic and structural analysis of Lhr-type helicase.

1-Archaeal and bacterial Lhr-type sequences are collected from three complementary search approaches

To search for all Lhr-type protein sequences in archaeal and bacterial genomes, we defined queries corresponding to Lhr specific domains and motifs. Lhr-type proteins are part of the SF2 helicase family. It has the characteristics helicase core formed by RecA1 and RecA2 conserved domains. These two domains correspond to the pfam00270 and pfam00271, respectively (**FIGURE 17A**).

Only a few Lhr-type helicases have been studied (Buckley et al. 2020; Ejaz and Shuman 2018; Ordonez and Shuman 2013; Ejaz et al. 2018; De Falco et al. 2018). The crystal structure of Lhr-*Msme* (Lhr-*Msme*; 1507aa) revealed the RecA1 (1-230) and RecA2 (231-435) and two other domains forming the Lhr core from the N-terminal region (Ejaz et al. 2018) (**FIGURE 17A**). The Domain 4 (530-856) corresponds to a novel conserved domain with unknown function that is specific to the Lhr-type proteins and that corresponds to the pfam08494. Moreover, using the Dali database and a tridimensional structure established for a truncated Lhr-*Msme* (1-856), a fourth domain has been uncovered and named WH (436-529) for Winged-Helix domain (**FIGURE 17A**). The Lhr WH domain has a similar fold as Hel308 Ski2-helicase WH but they only share six conserved amino acids (Ejaz et al. 2018). Msme-Lhr also possesses a C-terminal extension (856-1507). This region was not yet studied and no domains have been identified (**FIGURE 17A**).

By sequence homology, we defined the four structural domains of *Pyrococcus abyssi* aLhr2 (aLhr2-*Paby*; 866aa) as follows; the RecA1 (1-248)/RecA2 (248-421) helicase core, the WH domain (421-514), and the Domain 4 (514-) (**FIGURE 17A**). We used the pfam00270, pfam00271, and pfam08494 corresponding to the RecA1 and RecA2 domains, and Domain 4, respectively, that share conserved amino acid sequence in the Lhr-type family to identify orthologous groups.

In order to collect Lhr-type protein sequences from the bacterial and archaeal genomes, we used an in-house Complete Genome DataBase (CGDB) that contain 2563 genomes of bacteria and 197 genomes of archaea provided by the fichant team. These genome entries were automatically retrieved from EMBL and processed by a set of Perl program into a MySQL database. Protein sequences from the CGDB have been annotated according to the conserved domain database available at NCBI using the RPS-Blast program. Orthologous groups have been determined using BlastP local alignment. With the objective to collect the maximum

possible of Lhr-type protein sequences, we implemented three approaches using different queries (In collaboration with Fichant team-LMGM) (**FIGURE 17B**). In order to limit the redundancy, our search was restricted to one strain per species.

As all the SF2 helicases have in their N-terminal region the conserved helicase core domain, to limit the risk of collecting false-positive Lhr sequences, in the first query, we used the cluster of orthologous group COG1201 that covers the Lhr core formed by the Pfam00270, Pfam00271, and Pfam08494 (**FIGURE 17A**). We collected 822 protein sequences (**FIGURE 17B-**Approach 1). The second challenge we faced, was the difference in sequence lengths of the Lhr-type proteins that varies from 505 to 1583 amino acids. Thus, we also used the three Pfam00270, Pfam00271, and Pfam08494 of Lhr core as independent queries for searching for Lhr-type protein sequences in the CGDB. This second approach allowed us to collect 615 Lhr-type sequences (**FIGURE 17B**-Approach 2). Finally, in a third approach, to avoid missing putative Lhr-type protein sequences from the CGDB, the helicase core combining the RecA1 and RecA2 domains has been used as a query for retrieving all the SF1 and SF2 helicase proteins from the CGDB (**FIGURE 17B-**Approach 3). The collected sequences were clustered using the Markov clustering method. 894 protein sequences corresponding to the family annotated as Lhr-type were then retrieved using the Hidden Markov Modeling (HMM).

To compare the sequences collected using the three approaches, we built a Venn diagram (**FIGURE 17B**, lower panel). The intersection between the three approaches shows 614 Lhr-type sequences in common. All the sequences recovered by the first and second approaches were retrieved by at least two of our three approaches. The third approach was the most prone in collecting false-positive sequences as 71 identified sequences were not recovered by the first and/or the second approach. These sequences were checked for the presence of the characteristic Lhr core domains and determined as false-positives as they either belong to other SF2 subfamily or correspond to protein fragments.

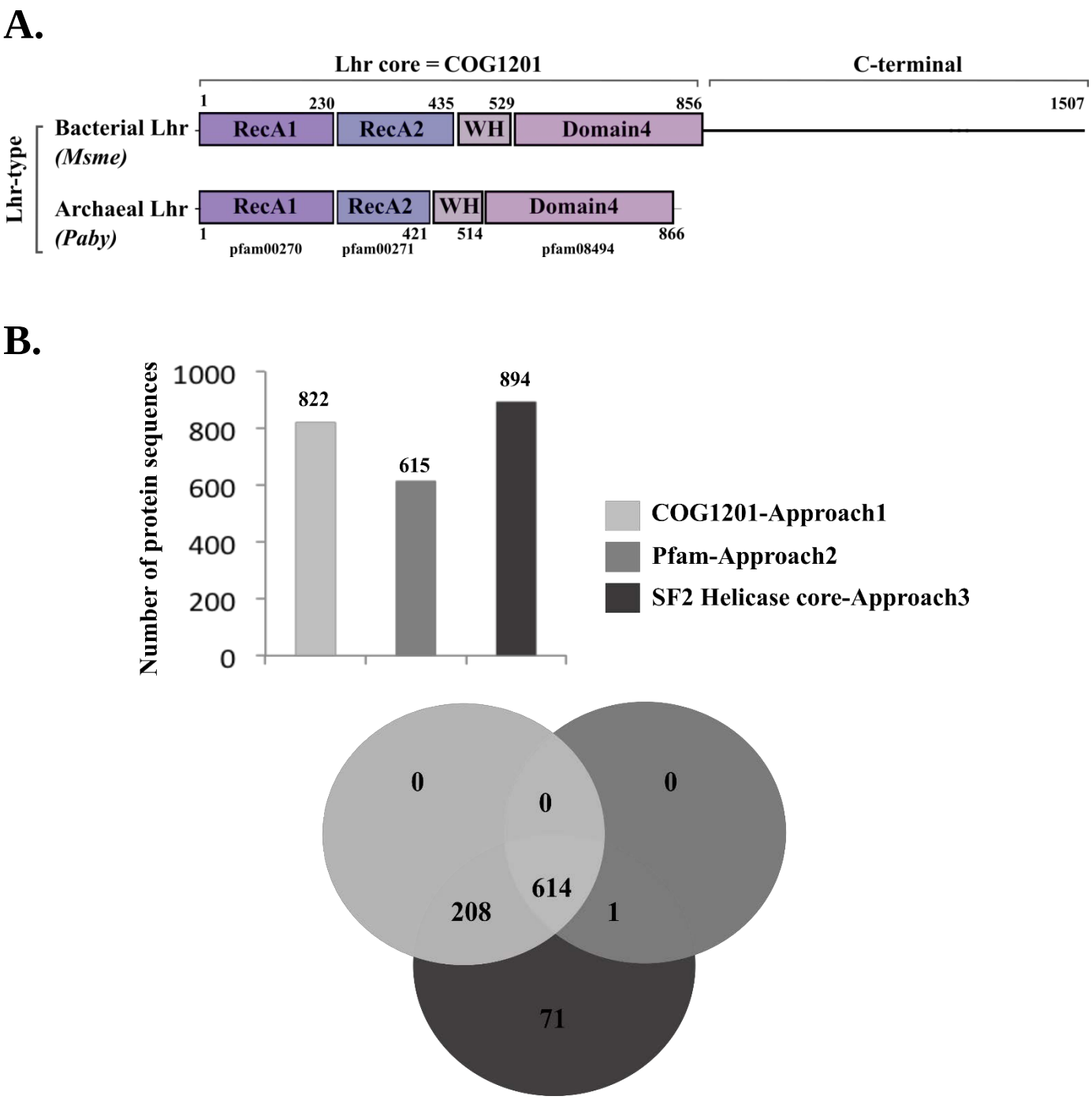


FIGURE 17. The Lhr-type domains and approaches used in collecting Lhr–type protein sequences.

A. The typical SF2 helicase domain representation is shown using the representatives bLhr-*Msme* of *Mycobacterium smegmatis* and aLhr2-*Paby* of *Pyrococcus barophilus* of the Lhr-type protein family. The Lhr core is formed by the RecA1 (pfam00270; purple) and RecA2 (pfam00271; blue) domains. the Winged Helix (WH) domain and the Domain 4 (pfam08494). The COG1201 covers the pfam00270, pfam00271 and pfam08494.

B. The COG1201 (Approach 1), the three Lhr pfam (Approach 2), and the SF2 helicase core (Approach 3) were used as distinct queries to collect the Lhr-type protein sequences in a complete genome database. The histogram indicates the number of Lhr-type sequences retrieved by each approach. The Venn diagram illustrates the number of common sequences obtained using the three approaches.

Altogether, and after thorough verification, 823 Lhr-type protein sequences have been collected from the CGDB, with 274 archaeal sequences, and 549 bacterial sequences.

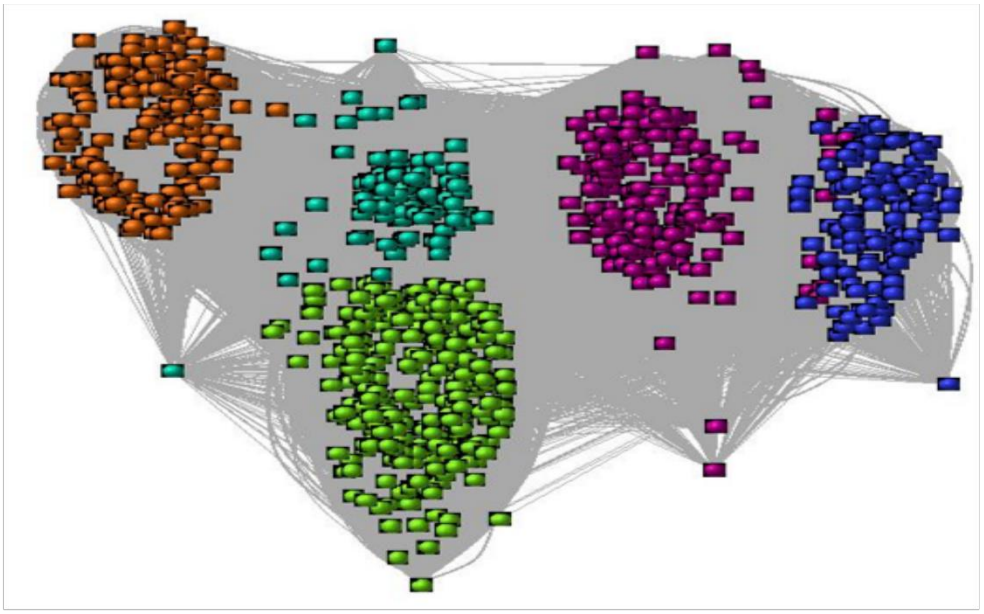
In addition, no Lhr-type protein sequence has been found in eukaryotic genomes using any of these three approaches (database kingdom 15, eukaryotes). Therefore, we systematically searched for putatively unannotated genes (missed by annotation or pseudogenes) with the tblastN program. As a result, no corresponding genes have been found in eukaryotes.

2-Five Lhr–type orthologous groups are identified by MCL clustering.

The 823 selected Lhr-type protein sequences were clustered by applying the Markov clustering (MCL) algorithm after all-against-all protein comparison using BlastP. The protein sequences of the similarity network were clustered into five well-defined groups of Lhr-type proteins (**FIGURE 18A**). In the interaction graph, vertices represent individual proteins, while edges weight according to their BlastP E-Value. We identified the two archaeal groups named aLhr1 and aLhr2 that were previously reported (Chamieh et al. 2016). They respectively have 169 and 105 protein sequences (**FIGURE 18B**). For the first time, three bacterial groups were also defined. We assigned them the following nomenclature: bLhr-HTH, bLhr, and Lhr-like (**FIGURE 18A**). They respectively contain 251, 182, and 107 protein sequences (**FIGURE 18B**). Interestingly, we show that the aLhr1 cluster is connected to the Lhr-like cluster while the aLhr2 cluster is closely linked to both bLhr-HTH and bLhr clusters (**FIGURE 18A**). We should note here that bacterial bLhr-HTH family is named after the conserved HTH domain of the C-terminal region **(FIGURE 19A).**

To determine the orthologous groups (OGs) and to analyze the evolutionary relationships of the Lhr protein families, the 823 archaeal and bacterial Lhr-type protein sequences were used to construct phylogenetic trees via the maximum-likelihood method. The computed PhyML tree of **FIGURE 19A**, constructed using the full-length protein sequences, shows that Lhr-type proteins cluster into five OGs: aLhr1 and aLhr2 in archaea, bLhr, bLhr-HTH and Lhr-like in bacteria. These five OGs are coherent with the groups retrieved by our previous MCL clustering (**FIGURE 18**), with the exception of 12 aLhr1 proteins that cluster with the Lhr-like group proteins.

A.



B.

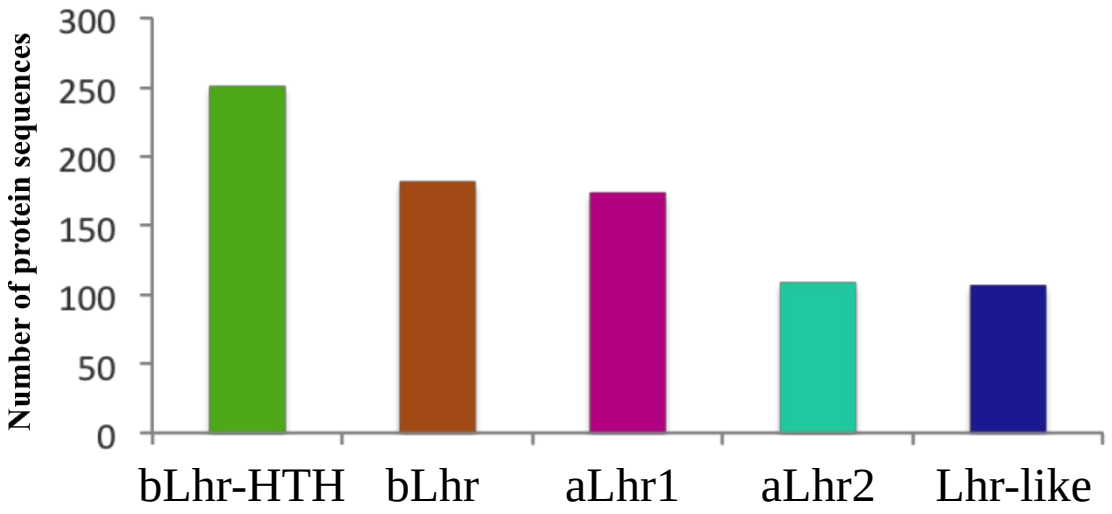


FIGURE 18. MCL clustering for defining Lhr-type protein clusters.
A. A set of 823 Lhr-type protein sequences were selected and clustered using the Markov clustering algorithm (MCL). Two archaeal clusters aLhr1 (pink) and aLhr2 (blue-green), and three bacterial clusters Lhr-like (blue), bLhr (orange), and bLhr-HTH (green)- were identified. In the graph, vertices represent individual proteins and edges weight to their BlastP E-Value. **B.** The histogram represents the number of protein sequences found in each Lhr-type group.

To evaluate the reliability of the Lhr-type clustering, two phylogenetic trees based either on the helicase core (**FIGURE 19B**) or Domain 4 (**FIGURE 19C**) were also constructed. Despite some differences in their topology, the trees exhibit the same archaeal and bacterial OGs as the tree with the complete protein sequence (**FIGURE 19A**). The tree based on the Domain 4 lacks the Lhr-like group as their members do not contain a Domain 4 (**FIGURE 19B**). Altogether, these results show the robustness of the clustering in orthologous groups.
To explore the Lhr helicases in the Asgard archaeal group, we computed another PhyML tree with the same protein dataset but with Asgard genomes included in our complete genome database. We found that the Asgard encode in their genomes Lhr-type helicases from both, the aLhr1 and aLhr2 OGs (**FIGURE ANNEX 1**).

Furthermore, to ensure the belonging of the Lhr-like group with an incomplete Lhr core lacking the Domain 4 to the Lhr-type helicase family, we included the HRQ-like helicase family as an outgroup in our protein dataset to construct a new phylogenetic tree (**FIGURE ANNEX 2**). This tree confirms that proteins of Lhr-like cluster belong to the Lhr-type helicase family. Indeed, the Lhr-like group is separated from the HRQ-like branch and segregate from a branch common to all the Lhr-type helicase clusters.
Interestingly, Lhr-like cluster is the first to diverge from the others **FIGURE 19A**. Its long branch suggests a lateral gene transfer. The aLhr1 group that shows a large variation of branch length most likely emerged from an ancient duplication event **(FIGURE 19A)**. The aLhr2 group that exhibits less variability with shorter branch length probably emerged from a recent duplication event within the bLhr cluster **(FIGURE 19A).**
Hence, we propose that an early gene duplication gave rise to two main Lhr clusters: one archaeal cluster aLhr1, and one sister bacterial cluster, bLhr-HTH. A later gene duplication within the sister cluster gave rise to two more archaeal and bacterial clusters, aLhr2 and bLhr respectively **(FIGURE 19A)**.

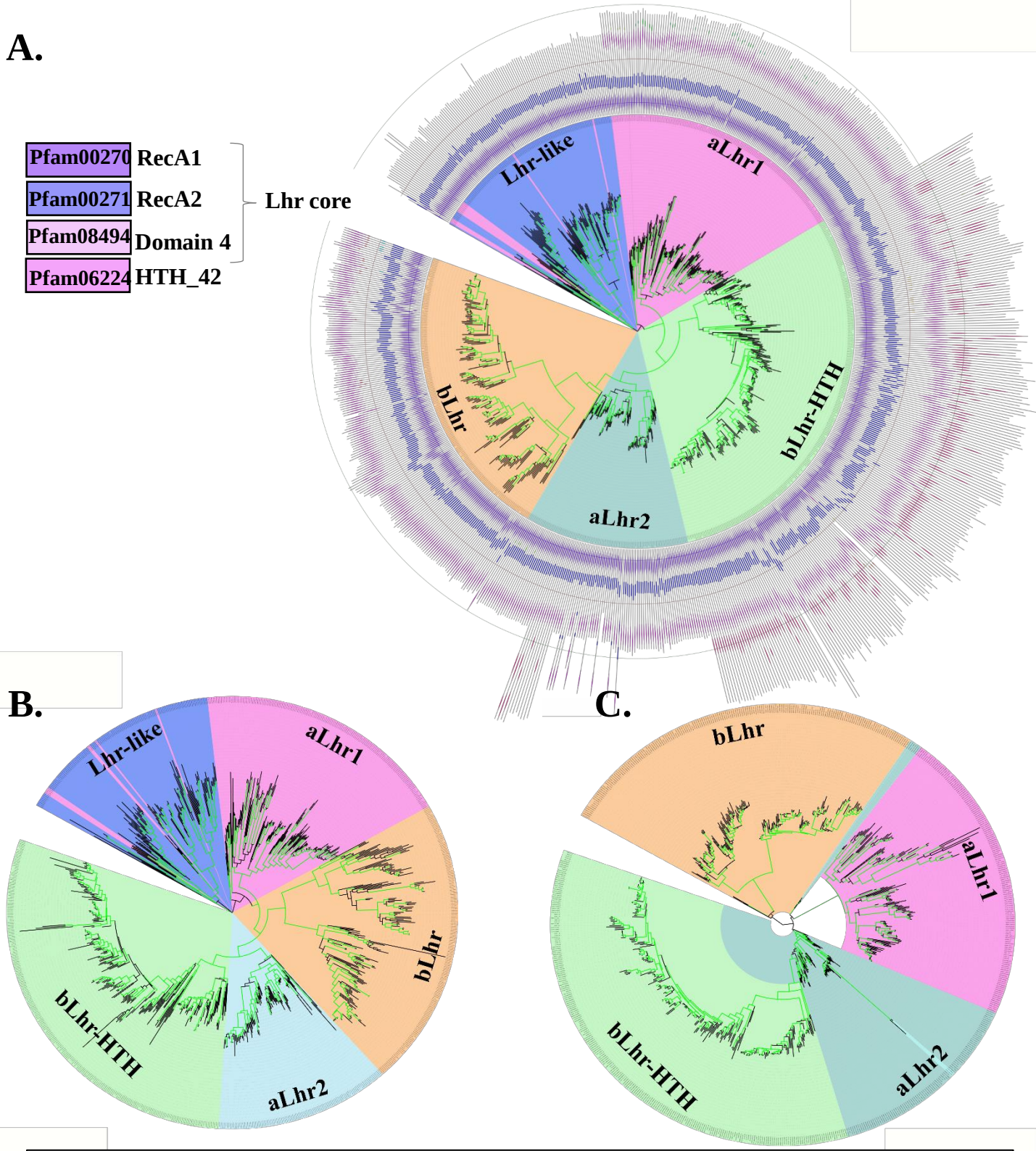


FIGURE 19. Phylogenetic trees of the five Lhr-type orthologous groups.
A. The maximum likelihood tree computed with PhyML using a dataset of 823 Lhr-type protein of full-length sequences. The tree branches are supported by high bootstrap values. Bootstrap above 0.7 are represented in green. The Lhr protein family is segregated in five orthologous groups. Lhr-like OG harbors mostly bacterial members with some archaeal members initially identified as aLhr1 proteins by MCL clustering. The Pfam domains of each Lhr-type OG are color-coded and represented at the leaves of the tree. **B.** Phylogenetic tree was constructed using the dataset of 823 Lhr-type members but restricted to the SF2 helicase core **C.** Phylogenetic tree restricted to the 691 Lhr-type members containing a Domain 4 present in most Lhr-type members except for the Lhr-like group.

3-Taxonomic occurrence identifies aLhr1 & aLhr2 proteins as ubiquitous in archaeal genomes

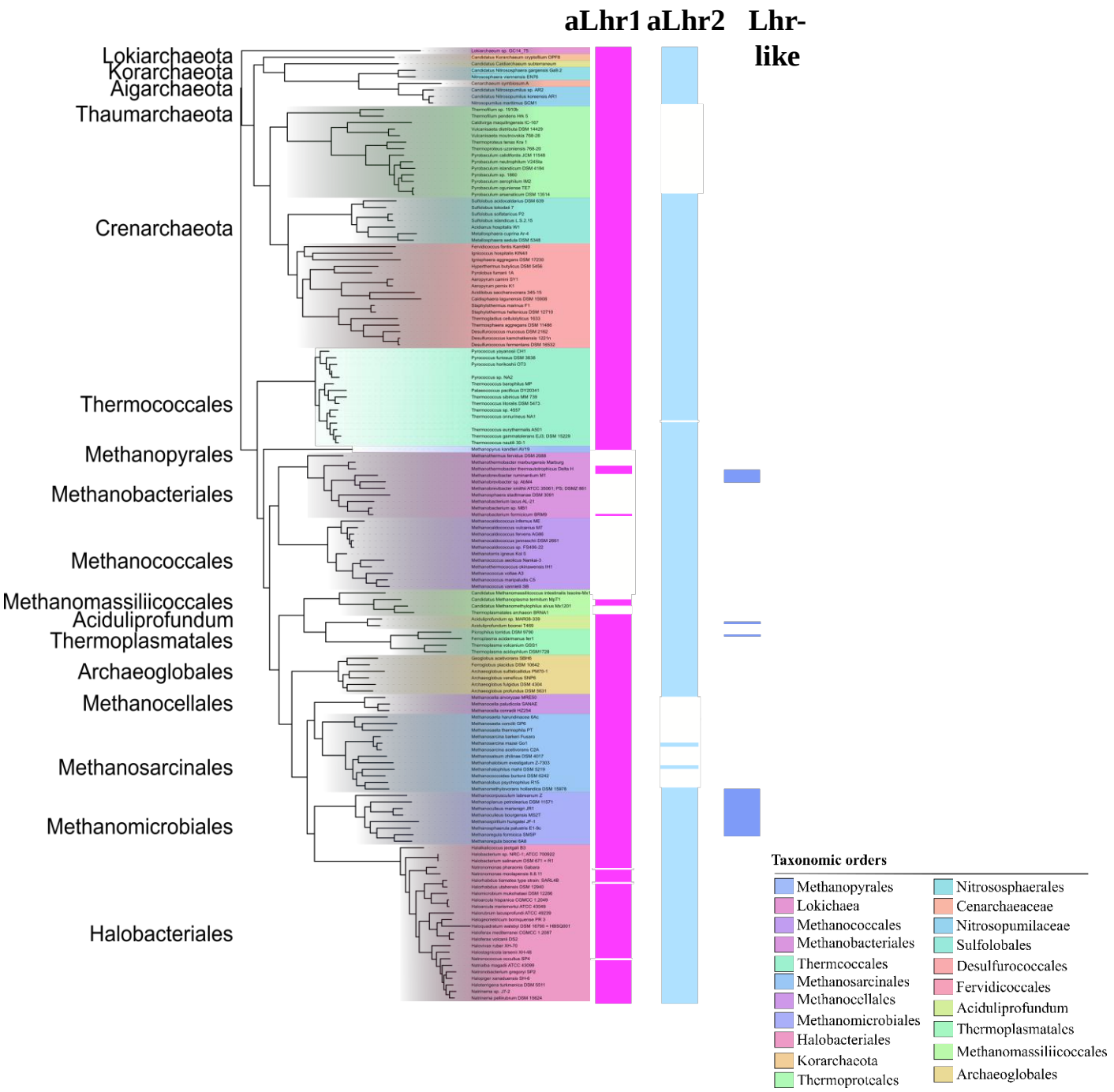
After determining the Lhr-type protein OGs, we looked for their taxonomic occurrence in the different archaeal and bacterial taxa. Interestingly, all the archaeal genomes encode at least one Lhr-type protein. aLhr1 members were found in all the TACK, and in most of the Euryarchaeota with the exception of the Methanococcales and most of the Methanobacteriales (**FIGURE 20A**). Moreover, the Sulfolobales and Acidilobales, and few Desulfurococcales, encode two aLhr1 members. Members of the aLhr2 group were also found in most of the TACK and Euryarcheota with the exception of the Thaumarchaeota, Methanosarcinales and Methanocellales. Our results are consistent with the previously obtained distribution of the archaeal Lhr-type proteins (Chamieh et al. 2016). 12 Lhr-like archaeal proteins, initially identified as belonging to aLhr1 group by MCL clustering, are distributed as following: 7 in methanomicrobia, 3 in methanobacteria, 1 in thermoplasmata, and 1 in aciduliprofundum. In Asgard, Lhr-type proteins belong either to the aLhr1 or aLhr2 group. Odinarchaeota, Bathyarchaeota, and Thorarchaeota encode an aLhr1 member while Heimdallarchaeota encodes an aLhr2 member.

As we identified an aLhr2 member in the Thermococcales RNA metabolism network, we focused on the evolution route of the aLhr2 group. The phylogenetic tree in **FIGURE 20A** shows the emerging of aLhr2 into two sub-groups, one that occurs only in Halobacteria, Methanococci, and Methabacteria, and the second in Thermoplasmata, Aciduliprofundum, Methanobacteria, Archaeoglobi, Methanococci, Thermoprotei and Korarchaeota. In order to examine the possibility of a vertical transfer of aLhr2 genes, we also evaluated the phylogenetic congruence between the phylogeny of the aLhr2 proteins (based on 109 alhr2 genes) and the phylogeny of the corresponding archaeal phyla constructed from a concatenated sequence of 53 ribosomal proteins (**FIGURE 20B**). Despite the presence of some inconsistencies between both trees, notably in Methanocci and Archaeoglobi, the phylogenies are largely congruent which suggests an ancient origin of *alhr2* gene predating the emergence of archaea.

The present phylogenetic analysis reports for the first time three well-defined clusters of Lhr-type in bacteria (**FIGURE 21A**). Unlike the archaeal Lhr, the distribution of the bacterial Lhr members is more erratic (**FIGURE 21**). bLhr-HTH proteins are found in almost all Actinobacteria and Acidobacteria, and in some Firmicutes, Proteobacteria, and Spirochetes. Whereas bLhr proteins are mainly distributed in Bacteroidetes and Proteobacteria with also

some members in Cyanobacteria, Acidobacteria, and Chloroflexi. Finally, bLhr-like members are scattered across the bacterial taxa, with more representatives in Proteobacteria, Cyanobacteria, Firmicutes, and Actinobacteria. Indeed, the phylogenetic placement of Lhr-like proteins was rather scattered among different phyla also suggesting HGT **(FIGURE 21)**. The occurrence of Lhr-like in cyanobacteria suggests that this family is acquired by Horizontal Gene Transfer from a cyanobacterial symbiont.

A.


aLhr1
aLhr2
Lhr-like
Lokiarchaeota
Korarchaeota
Aigarchaeota
Thaumarchaeota
Crenarchaeota
Thermococcales
Methanopyrales
Methanobacteriales
Methanococcales
Methanomassiliicoccales
Aciduliprofundum
Thermoplasmatales
Archaeoglobales
Methanocellales
Methanosarcinales
Methanomicrobiales
Halobacteriales
Taxonomic orders
Methanopyrales
Lokichaea
Methanococcales
Methanobacteriales
Thermococcales
Methanosarcinales
Methanocellales
Methanomicrobiales
Halobacteriales
Korarchaeota
Thermoprotcales
Nitrososphaerales
Cenarchaeaceae
Nitrosopumilaceae
Sulfolobales
Desulfurococcales
Fervidicoccales
Aciduliprofundum
Thermoplasmatales
Methanomassiliicoccales
Archaeoglobales

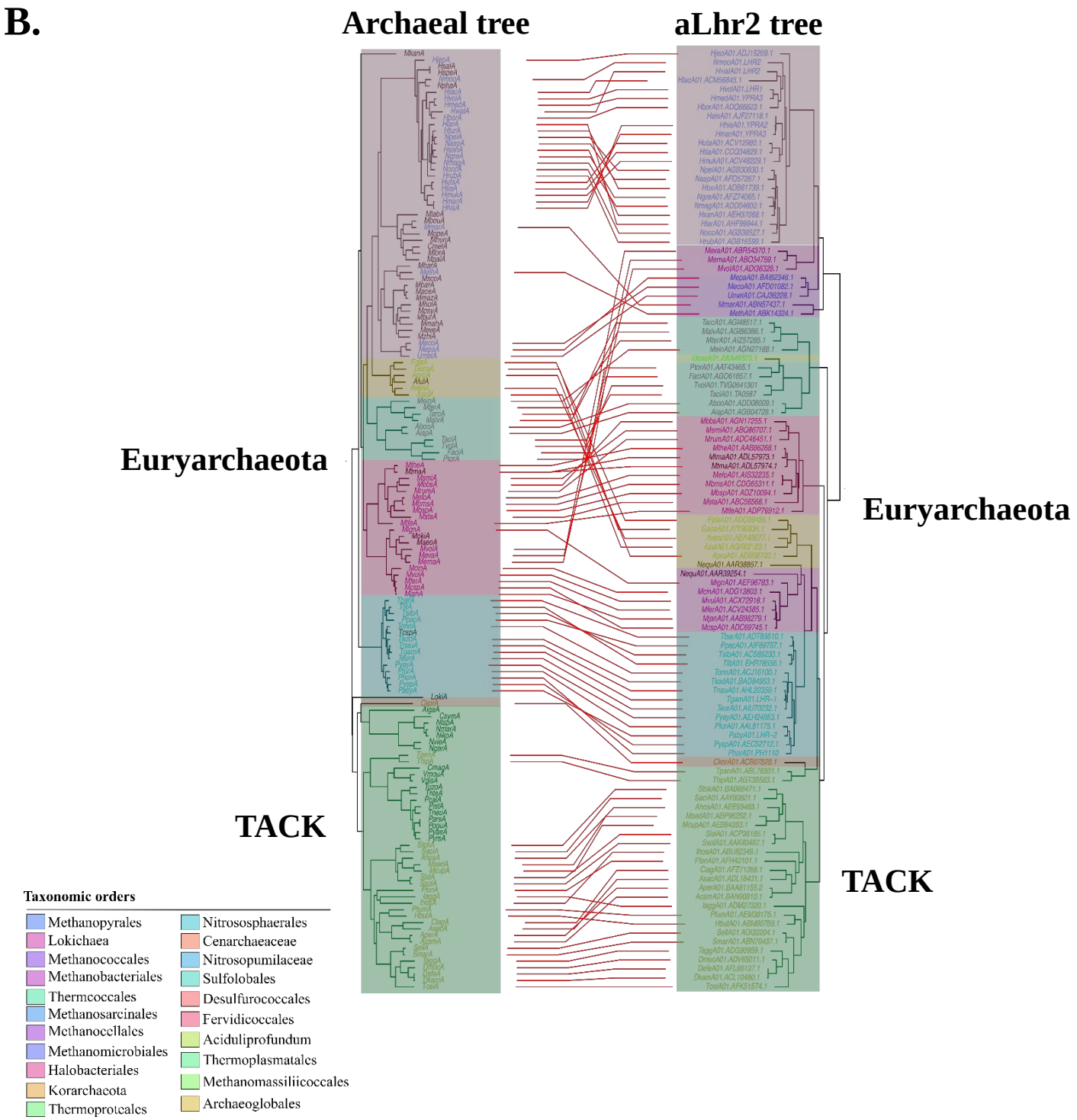


FIGURE 20. Distribution of the archaeal Lhr-type groups throughout the phylogeny of Archaea. **A.** The occurrence of archaeal Lhr–type proteins belonging to aLhr1, aLhr2 or Lhr-like among archaeal phylogenetic groups is shown. **B.** Congruence of the archaeal and aLhr2 phylogenetic trees. The maximum likelihood tree has been computed with PhyML using the 109-archaeal-aLhr2 sequences dataset collected from the CGDB in this analysis. The archaeal tree was constructed from a concatenated sequence of 53 ribosomal proteins.
Taxonomic orders are represented in different colors.

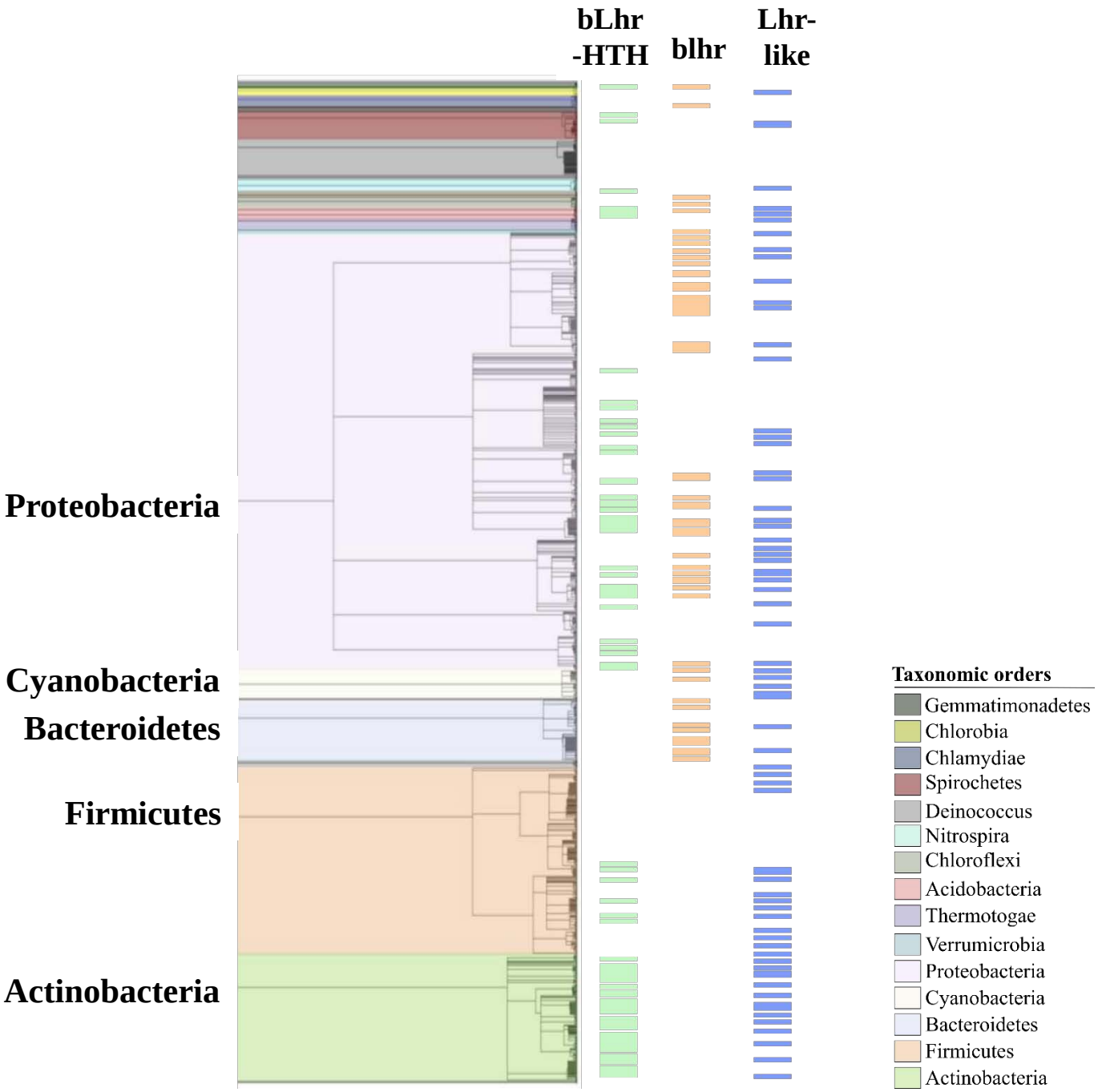


FIGURE 21. Distribution of the bacterial Lhr-type groups throughout the phylogeny of Bacteria. The occurrence of bacterial Lhr–type proteins belonging to bLhr-HTH, bLhr or Lhr-like among bacterial phylogenetic groups is shown.

4-Specificities at the structural level are observed for each Lhr-type orthologous group

To gain further insight into the structural diversity of the five Lhr-type OGs, we analyzed the conservation of their domains at the sequence and structural levels. The Lhr core: RecA1 (Pfam00270) and RecA2 (Pfam00271) domains, and the Domain 4 (Pfam08494), are highly conserved, except for the bacterial Lhr-like group that does not contain a Domain 4 but instead an additional C-terminal region **(FIGURE 19 &23)**. In addition, members of the bLhr-HTH and aLhr1 groups, and few of the aLhr2 group, show an extended C-terminal region after the Lhr core. Interestingly, sequence alignments of these extensions show highly conserved sequences within each OGs **(FIGURE 22 and ANNEX 3)**. We can note the conservation of SF2 motifs in all Lhr-type families. The Trp597 (Domain 4) of *Msme*-bLhr-HTH identified as important for coupling ATP hydrolysis to DNA translocation and therefore duplex unwinding is also conserved in all Lhr-type families including Lhr-like. The Ile528 (WH domain) of *Msme*-bLhr-HTH, essential for duplex nucleic acid unwinding, in addition to a motif with the consensus sequence 'DVL' are conserved in all Lhr-type excepting Lhr-like family **(FIGURE 22)**.

To get an overview of the structure-function of the members of each group, we chose to compare the predicted or resolved structures of one member per group. Structure models of aLhr2 and aLhr1 of *Pyrococcus abyssi* (*Paby*), bLhr of *Pseudomonas putida* (*Pput*), Lhr-like of *Streptomyces coelicolor* (*Scoe*), and the C-terminal region of bLhr-HTH of *Mycobacterium smegmatis (Ecol),* have been predicted using Phyre2. Structural homology searches were performed with the DALI server (**FIGURE 23A**). These models were deduced using the crystal structure of the truncated bLhr-HTH of *Msme* (1-856; PDB5V9X) with 28-, 29-, 31- and 34-% of identity for Lhr-like-*Scoe*, aLhr1-*Paby*, bLhr-*Pput*, and aLhr2-*Paby*, respectively. The Lhr protein folds of the Lhr core (helicase core, WH domain and Domain 4) seem to be conserved both in archaea and bacteria. Interestingly, despite the low amino acid sequence identity, the C-terminal region of the Lhr-like-*Scoe* that is deprived of an obvious Domain 4 adopts a structure that is similar to the predicted structure of the Domain 4 (25% of identity). Note that the high similarity of structure between the different Lhr-type groups does not necessarily mean that these proteins exhibit the same cellular function *in vivo*.

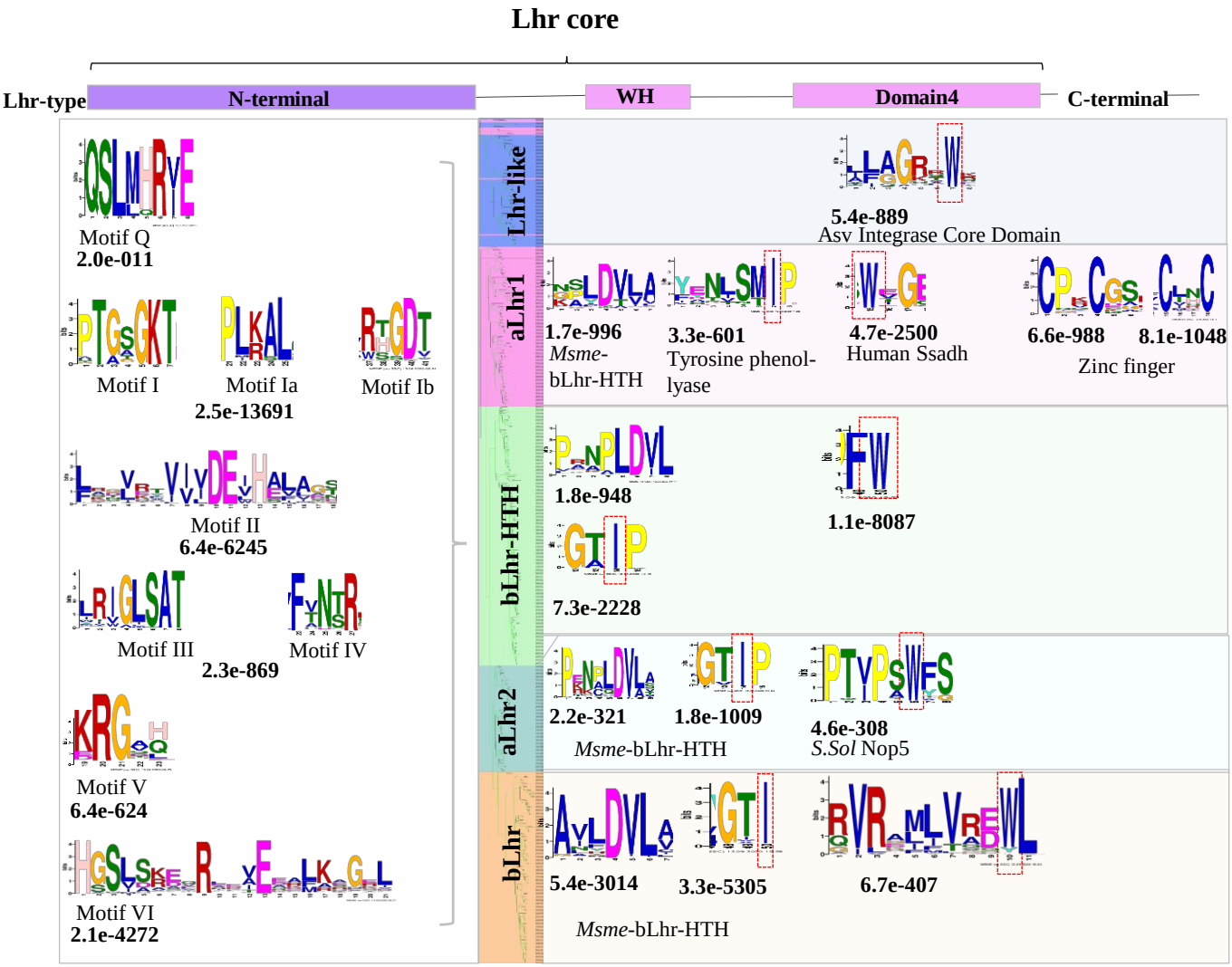


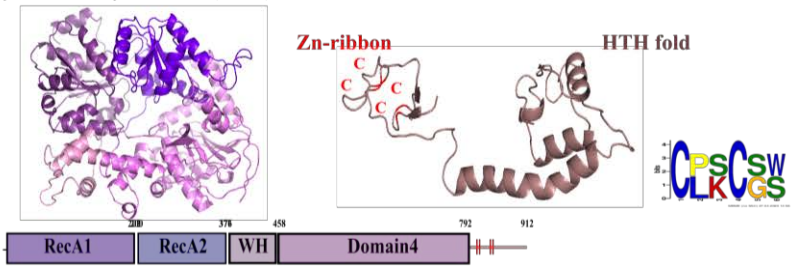
FIGURE 22. Weblogo sequences of the Lhr-type orthologous groups.
The Lhr core domains and the C-terminal regions of each Lhr-type group were aligned using Mafft and used for motifs revealing using MEME http://meme-suite.org/tools/meme. Sequences that match the discovered motifs were found using MAST http://meme-suite.org/tools/mast. The conserved Trp (Domain 4) and Ile (WH domain) residues are highlighted.
The relative height of each letter is proportional to the relative entropy of the corresponding base at the given position, and the p-values were indicated below of each conserved motifs.

A. Structural features

B. Genomic context

aLhr1

Pyrococcus abyssi aLhr1 (1-912) *



15%

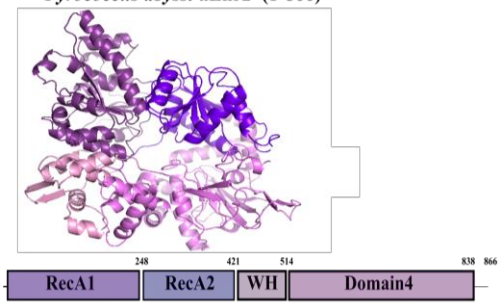
Methanomicrobia — MPE — Lhr

Methanoinales — Lhr — MPE

Halobacteria — Lhr — MPE; Lhr — MPE

aLhr2

Pyrococcus abyssi aLhr2 (1-866) *



16%

Sulfolobales — GTS — Lhr

Desulfurococcales — GTS — Lhr

Thermoproteales — Lhr — GTS

Thermococcales — GTS — Lhr

16%

Thermococcales — MPE — Lhr; Lhr — MPE

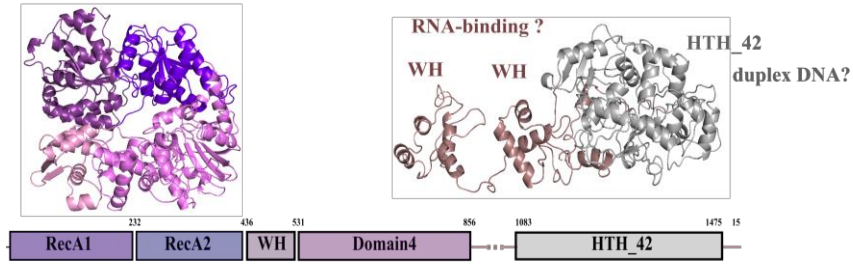
Methanobacteriales — MPE — Lhr

Methanococcales — Lhr — MPE; Lhr — MPE

Archaeoglobales — Lhr — MPE — Ligase

bLhr-HTH

Escherichia coli bLhr-HTH (1-856)*



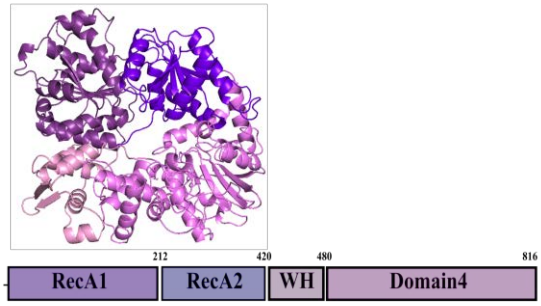
3%

Proteobacteria 82%

RNT — Lhr

bLhr

Pseudomonas putida bLhr (1-816) *



70%

Lhr — MPE

Exo — Ligase — Lhr — MPE

Exo — Ligase — Lhr — MPE

Exo — Ligase — Lhr — MPE

Exo-ligase fusion — Lhr — MPE

Ligase — Exo — Lhr — MPE

Ligase — Exo — Lhr — MPE

Lhr-like

Streptomyces coelicolor Lhr-like (1-731) *

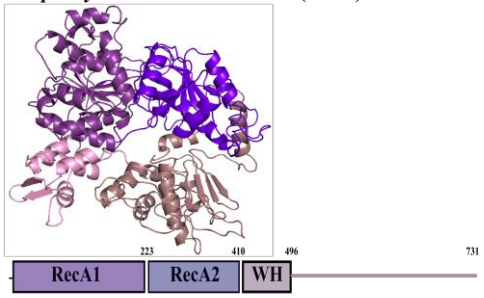


FIGURE 23. Features of Lhr-type orthologous groups.
A. Structural specificities of the five Lhr-type group members are shown. The Lhr core formed by the RecA1 (violet-blue), RecA2 (purple-blue) and the WH domain (light pink) are conserved in all Lhr-type families. The Domain 4 of the Lhr core (violet) is conserved in the bacterial bLhr, bLhr-HTH, and in the archaeal aLhr1 and aLhr2 members. The C-terminal region with no identified conserved domain is represented in dirty pink. All 3D structural models were obtained using Phyre2, stereo view of the cartoon models was analyzed using PyMOL (*). **B.** Common genomic context of genes encoding Lhr-type members. Genes encoding the Lhr-type proteins are signified by pink arrows. At the vicinity of the *lhr* genes, genes encoding a metallophospho-esterase (MPE; Red arrows), an ATP-dependent DNA ligase (Ligase; Blue arrows), a glycosyltransferase (GTS; Yellow arrows), an exonuclease of the metallo-β-lactamase enzyme family (Exo; Light green arrows) and/or the Ribonuclease T (RNT; Dark green arrows) are found.

The bLhr-HTH group is characterized by an extended C-terminal region. We scanned this region for Pfam domains and identified a helix-turn-helix structural motif named HTH_42 belonging to the Pfam06224. Then, we used the C-terminal region (888-1507) of *Msme* bLhr-HTH to search for homologous structures. The most similar structure found in the PDB database with 19% of identity corresponds to the DNA glycosylase AlkZ of *Streptomyces sahachiroi* (*Ssah*; PDB5UUJ) (Mullins et al. 2017). The 3D structure of *Ssah*-AlkZ exhibits a channel formed by a cluster of three WH motifs in tandem and by a short C-terminal β-hairpin that is highly accommodated for binding dsDNA. Conversely, in the published structure of truncated *Msme*-bLhr (1-856), the pocket composed of the four conserved domains of Lhr core (RecA1, RecA2, WH, and Domain4) was found to be too narrow to permit passage of a dsDNA (Ejaz et al. 2018). It is possible that the C-terminal HTH_42 assume dsDNA binding. Finally, the region between the Lhr core and the HTH_42 domains resemble with 21% of identity to the structure of the SelB elongation factor of *Moorella thermoacetica* (PDB1LVA), which contains two winged-helix domains involved in RNA binding.

Proteins of the aLhr1 group possess, in addition to Lhr core, a short C-terminal extension. We searched for conserved motifs and identified a Zn^{2+} binding cysteine-rich motif **C**xx**C**x_{14}**C**xx**C** (**FIGURE 22, FIGURE 23A**). By structure prediction, we identified a protein fold similar to the DNA binding motif of the uncharacterized PF0610 protein of *Pyrococcus furiosus* (PDB 2GMG, 38% of identity). This protein consists of a winged helix-turn-helix wHTH, and a putative rubredoxin-like Zn ribbon (**C**XX**C**) motifs (Wang et al. 2007). While HTH motifs are often involved in DNA binding, Zn ribbon is a metal-binding motif important for protein/protein, protein/nucleic acids (DNA or RNA), or protein/lipids interactions (Miłoch and Krężel 2014).

5-Genomic context analysis of the genes coding for Lhr–type proteins helps in predicting the function(s) of some Lhr helicases

If conserved, the genomic context of a gene can inform about its functions and can predict functional interactions with surrounding genes. Here, we have systematically analyzed the genomic environment of all Lhr-type encoding genes in Archaea and Bacteria (**FIGURE 23B**). In the neighborhood of 70% of bLhr encoding genes, three genes, encoding an ATP-dependent DNA ligase, a putative metallo-β-lactamase exonuclease (Exo), and a metallo-phosphoesterase (MPE) are found with various organization. Some *blhr* genes are only surrounded by an MPE encoding gene. While not conserved in other phylum, for 82% of

Proteobacteria, the bLhr-HTH encoding genes are followed by the *rnt* gene encoding the ribonuclease RNase T.

Regarding the archaeal Lhr-type, in 15% of the genomes, the open reading frame (ORF) encoding an aLhr1 member is typically located within a gene cluster that includes a gene encoding a MPE protein. For aLhr2, the encoding gene can be found within a *mpe* gene (16% of genomes) or within a gene encoding a putative glycosyltransferase enzyme (GTS) (16% of genomes). In Archaeoglobales, in addition to the *gts* gene, an ORF encoding an ATP-dependent DNA ligase is also found in the neighborhood of *alhr2* gene. It is important to consider that for many archaea, a substantial fraction of the genes are annotated as 'hypothetical' proteins in the databases. This could prevent the identification of genes with conserved function within an *lhr*-type gene in archaeal genomes. Nevertheless, in bacteria and in archaea, the genomic context analysis shows a wide prevalence of genes encoding ligase and/or nuclease in the neighborhood of the lhr-type ORFs (**FIGURE 23B**)

II-Biochemical studies of *Tbar*-aLhr2 protein.

1-*Thermococcus barophilus* as model of study

Almost all Thermococcales genomes encode two aLhr, aLhr1 and aLhr2 (RESULT- PART I). One aim of this thesis work is to decipher the physiological role of aLhr2 in the cellular metabolism of archaeal microorganisms. The group in Toulouse is focusing on Thermococcales archaea to capture main actors of RNA processing and maturation pathways. Previous work from the group in Toulouse (described in the introduction, see FIGURE 10) have used *Pyrococccus abyssi* strain as a study model and were able to show that aLhr2 is retrieved in the aRNase J ribonuclease and ASH-Ski2 RNA helicase networks using *P. abyssi* cellular extract. These partners are proposed to be main actors of the RNA metabolism in Thermococcales (Phung et al. 2020).

At first, a former student in the group in Toulouse tried without any success to clone *P. abyssi Paby*-alhr2 gene (PAB0744) using the pET system in *E.coli*. This is certainly due to a cellular toxicity that we have not been able to explain. We then logically turned to the *T. barophilus* strain, a model recently used for transcriptomic and genetic studies by the group in Toulouse. As mentioned in the introduction section (INTRODUCTION-Archaeal study model), *T. barophilus* is an hyperthermophilic archaea of the order of Thermococcales, closely related to *P. abyssi*, which serves now as a standard genetic model which was developed by M. Jebbar's team (Thiel et al. 2014). In addition, *Paby*-aLhr2 and *Tbar*-aLhr2 are orthologous (see RESULT- Part I) and have 78.86% of amino-acids sequence conservation and superimposable structural models built with Phyre2 **(FIGURE 24A & B)**. In this context, we proceeded to perform biochemical, transcriptomics and proteomics studies on aLhr2 from *T. barophilus* strain.

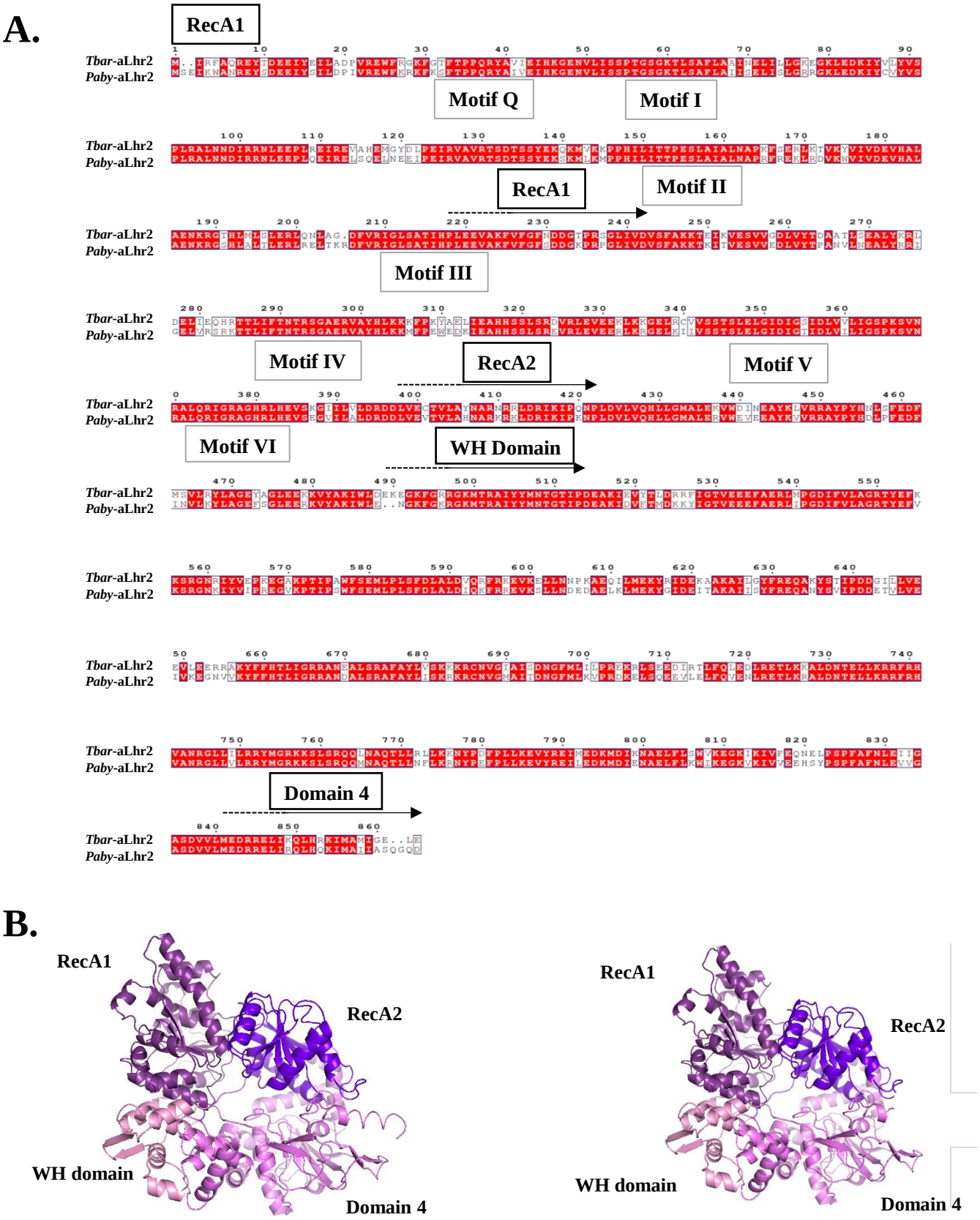


FIGURE 24. Homology of aLhr2 of *Paby* and *Tbar*.
A. Multiple sequence alignment of *Tbar*-aLhr2 and *Paby*-aLhr2 protein sequences showing the conservation of the characteristic motifs (grey boxes) and domains (black boxes). Arrows are placed at the end of each domain. The MSA was obtained using mafft and visualized with Psipred. **B.** The 3D structural models of *Tbar*-aLhr2 and *Paby*-aLhr2 were generated using Phyre2 intensive model showing the conservation of the Lhr core domain.

2-Production and Purification of *Tbar*-aLhr2 recombinant proteins

In order to characterize the role of aLhr2 in archaea, we performed *in vitro* enzymatic assays using recombinant proteins purified to homogeneity. We settled up an adapted methodology to characterize the ability of the hyperthermophilic *Tbar*-aLhr2 enzyme to hydrolyze ATP, bind nucleic acids and unwind or anneal DNA or RNA duplexes (**FIGURE 25A**). To determine the importance of Lhr core domains and conserved motifs in these activities, we compared the enzymatic activities of wild-type *Tbar*-aLhr2 to variants.

2-1 Constructions of pET11b-*Tbar*-aLhr2 expression vectors

As SF2-like proteins, *Tbar*-aLhr2 protein harbors a helicase core formed by the two RecA-like domains which include conserved motifs known to be crucial for the activity of helicases (**see FIGURE 11,** Introduction). In addition to this catalytic core, *Tbar*-aLhr2, as member of the aLhr2-type subfamily, possesses a typical Lhr core composed of a winged-helix (WH) domain and a long domain with unknown function named Domain 4 **(FIGURE 25B).** Therefore, we set out to construct expression vectors to obtain recombinant proteins corresponding to the wild version of *Tbar*-aLhr2 and altered versions for protein domains or motifs using the pET expression system for protein over-expression **(FIGURE 25B).** In view of the degree of conservation of each amino acid determined by our phylogenomics analysis **(FIGURE 22)**, the mutations at the following positions have been chosen: Lhr2-K60A (motif I), Lhr2-T215A (motif III), Lhr2-W577A (domain 4), and Lhr2-I512A (WH domain). We also constructed vectors to express a version of *Tbar*-aLhr2 that are lacking domain 4 and a version containing domain 4 alone. Briefly, the TERMP_00533 gene was amplified by PCR and inserted into the pET11b vectors to obtain pET11b-*Tbar*-aLhr2. The pET11b-*Tbar*-aLhr2ΔDom4 and pET11b-*Tbar*-aLhr2Dom 4 vectors were obtained by reverse PCR and the pET11b-*Tbar*-aLhr2K60A, T215A, I512A and W577A containing punctual mutations within the coding sequence by site-direct mutagenesis (**Table 8**). For details, refer to the section "Materials and Methods".

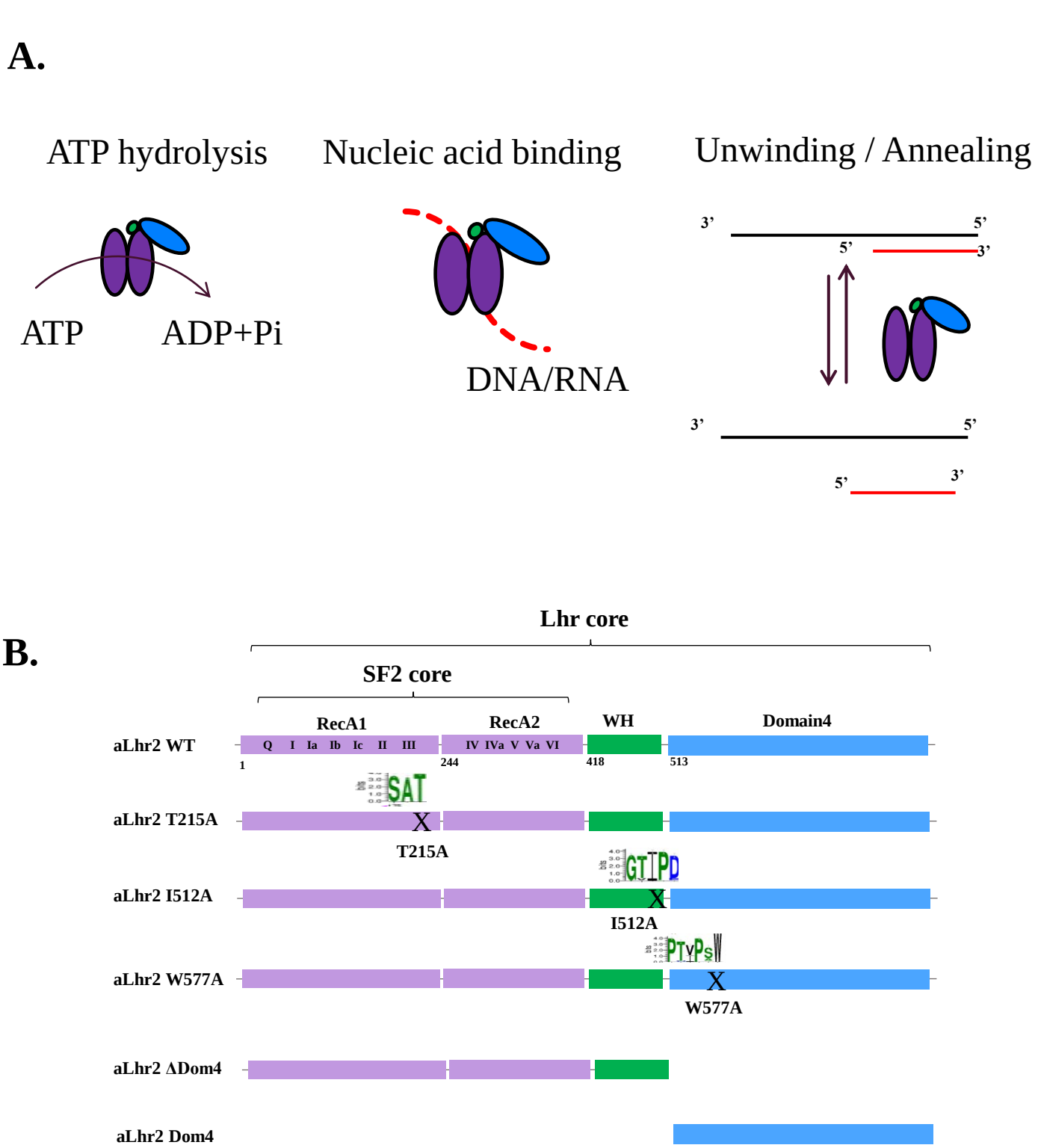


FIGURE 25. Helicase enzymatic activity and protein variants used in this study.
A. A model of helicase activities to hydrolyse ATP, bind nucleic acid, unwind or rewind double-stranded nucleic acid. Note that helicases possess one or both translocation polarities. **B.** Domain organization of Lhr core of *Tbar*-aLhr2 WT and variants. The helicase core is shown in purple, WH domain in green and Domain 4 in blue. Punctual mutations are represented by a cross with the corresponding weblogo of the conserved motifs in the aLhr2 family.

2-2 Optimization of expression and purification experimental conditions

To obtain purified recombinant protein at homogeneity, we settled up the heterologous overexpression and purification steps of *Tbar*-aLhr2 recombinant protein through the pET expression system in *E. coli* BL21 DE3 codon+ strain (**FIGURE 26**).

Initially, the routinely used protocol of the laboratory was assayed. Briefly, after electro-transformation of BL21-CodonPlus (DE3) *E.coli* strains with pET vectors, the overexpression of recombinant proteins were induced in exponential growth (OD_{600nm} of 0.6) by addition of IPTG (0.1 mM), for 3 hours at 30°C (Phung et al. 2020, 2013). We noticed that, in this condition, the overexpression of the protein of interest was not optimal and appeared to be stressful for the bacterial cells. We noticed that even before the induction of the expression of the pET-embedded gene the growth rate of the transformed bacteria (OD of cells density measured at 600 nm plotted over the time) was reduced in presence of the pET11b-*Tbar*-aLhr2 compared to pET11b- *Tbar*-aLhr2Domain 4 and pET11b- *Tbar*-aLhr2T215A in which conserved motifs are altered **(FIGURE 27A)**. There is probably an intrinsic leakage of this expression system which somehow affects bacterial growth indicating a toxicity of *Tbar*-aLhr2 protein for *E. coli* in its wild type version.

In this context and in order to optimize the overexpression of *Tbar*-aLhr2 recombinant protein, we optimized the parameters for the over-expression step. We tested the effect of cell density, IPTG concentration, time and temperature of incubation after induction, and addition of ethanol 10% to the culture during the growth time (Chhetri et al. 2015) (**FIGURE 27B**). From these assays, we identified that the finest overexpression conditions are high cell density (A_{600} of 0.9) at the time of induction, induction with 0.2mM of IPTG, and over-expression time of 3 hours at 30°C (**FIGURE 27B).** A similar procedure was used to express recombinant *Tbar*-aLhr2 variants **(FIGURE 28),** except for *Tbar*-aLhr2 K60A which could not be produced in any of these conditions.

Subsequently, the aLhr2 recombinant protein was sequentially purified by a heating step at 70°C, purification steps on cation exchange (HiTrap SP) followed by a Superdex 200 size-exclusion chromatographies using an FPLC (Fast Protein Liquid Chromatography) facility (**FIGURE 26,** for details, refer to the section "Materials and Methods"). Fractions of elution with highly pure recombinant protein were pulled together, dialyzed and quantified against a BSA protein concentration range (**FIGURE 26**). The predicted characteristics of *Tbar*-aLhr2

wild-type and variant proteins and the purification steps performed to obtain recombinant proteins to homogeneity are summarized in **Table 11 (FIGURE ANNEX 4&5)**.

Table 11. Biochemical characteristics of *Tbar*-aLhr2 WT and variants

	WT	**ΔDom4**	**Dom4**	**T215A**	**I512A**	**W577A**
Length (aa)	863	513	350	863	863	863
PI	9.06	8.64	9.31	9,06	9,06	9,06
MW (Da)	99 553.53	58 388.45	41 183.09	99 523.50	99 511.45	99 438.39
ε($M^{-1}cm^{-1}$)	75 180	47 790	27 390	75 180	75 180	69 680
Purif. steps	Cation S200	Heparine S200	Cation S200	Cation	Cation S200	Cation S200
Conc. (mg/ml)	0.47	0.59	0.12	0.325	0.42	0.28

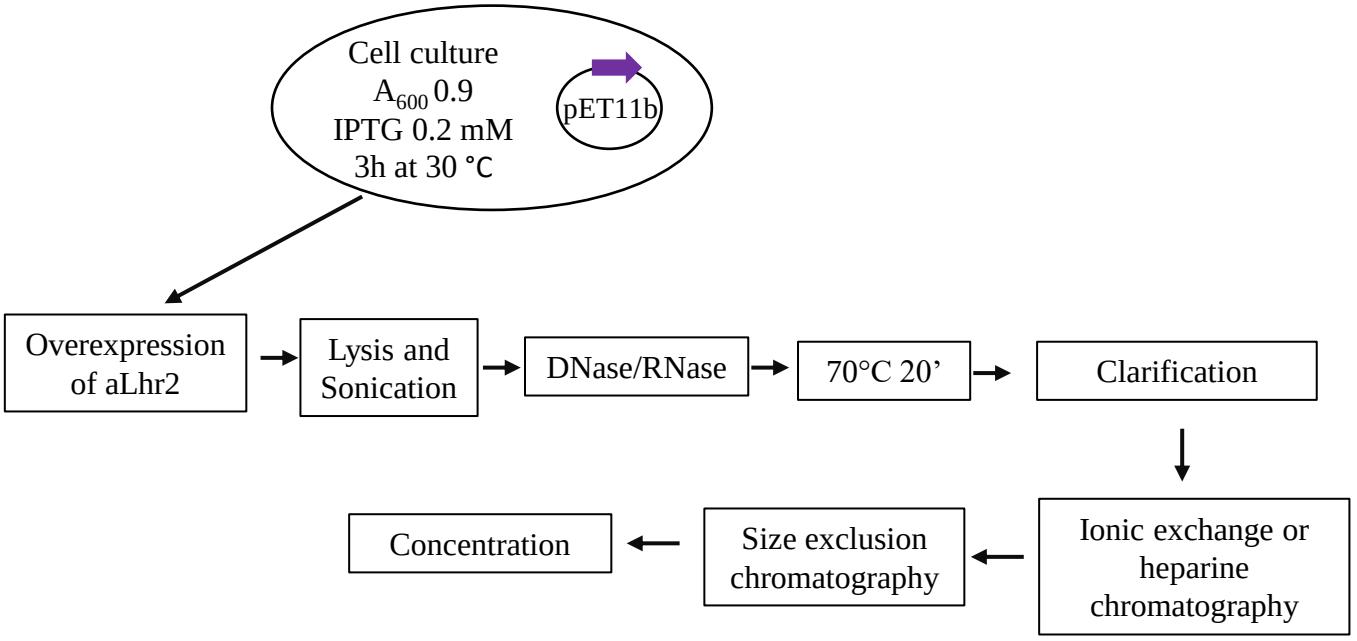
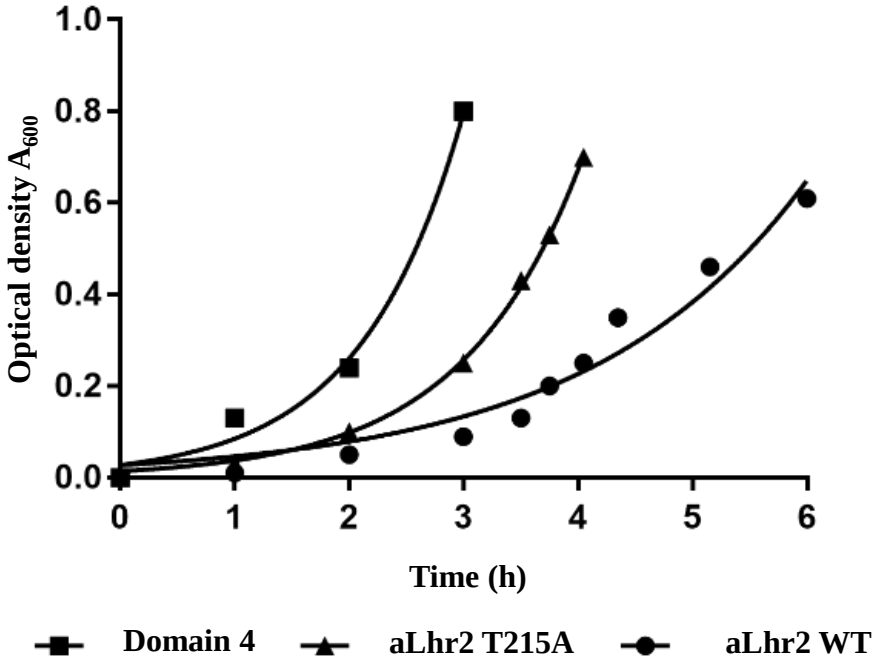


FIGURE 26. Experimental workflow for production and purification of recombinant proteins. Schematic overview of the established workflow used in this study to produce and purify recombinant proteins from *T.barophilus* in *E.coli*. The purification was carried out using liquid affinity chromatography on AKTA system.

A.



B.

-	0,1	0,1	0,1	0,1	0,2	**IPTG (mM)**
0,6	0,6	0,6	0,6	0,6	0,9	**OD A_{600}**
-	3	3	3	24	3	**Time (hour)**
-	30	30	16	16	30	**Temp (°C)**
-	-	+	-	-	-	**Ethanol (mM) 10%**

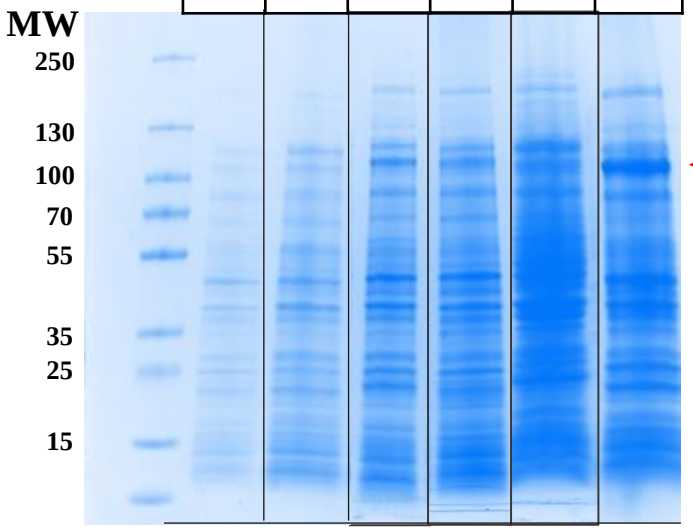


FIGURE 27. Set up heterologous overexpresssion of *Tbar* aLhr2 WT in *E.coli.*
A. Growth curves of *E. coli* cells expressing aLhr2 WT and variants. BL21 DE3 codon+ cells carrying the pET11b-aLhr2 WT (circle) pET11b-Dom4 (square) and pET11b-aLhr2-T215A (triangle) were grown at 37°C in LB medium. The doubling time of each cell culture calculated using GraphPrism are as follows: 1.3h for the WT aLhr2, 0.6 for aLhr2 T215A, and 0.7h for the Domain 4. **B.** The overexpression of the WT aLhr2 recombinant protein was optimized after testing different set of parameters and conditions. Samples (10µl/lane) before and after adding IPTG were loaded on SDS-PAGE gel and visualized after coloration with coomassie blue.

Briefly, the homogenous recombinant proteins *Tbar*-aLhr2 WT (**FIGURE 29**), were purified in two steps by using a cation exchange column in accordance with the pI of 9.06 of *Tbar*-Lhr2 and a S-200 size exclusion chromatography. The size exclusion chromatogram shows a single peak at 77.43 min (**FIGURE 29A**) which could correspond to the monomer form of *Tbar*-aLhr2 by comparison with another recombinant protein purified in the same condition in the laboratory. However, the oligomerization state of the *Tbar*-aLhr2 should be confirmed using analytical size exclusion chromatography. To verify that the observed protein with the predicted molecular weight was indeed *Tbar*-aLhr2, the main (100 KDa) band and two minor bands (70 and 30 KDa) were sequenced (**FIGURE 29A**). The mass spectrometry results confirmed that the 100KDa-band corresponds to the full-length *Tbar*-aLhr2 and the minor bands to proteolysis-product of *Tbar*-aLhr2.

A similar protocol was used for *Tbar*-aLhr2 Dom 4, *Tbar*-aLhr2 W577A, and *Tbar*-aLhr2 I512A recombinant proteins (**FIGURE 29B, ANNEX 4&5**) In the case of *Tbar*-aLhr2-T215A, the first purification step on a cation exchange chromatography was sufficient to obtain recombinant protein purified to homogeneity. In the case of *Tbar*-aLhr2 ΔDom4, the first step was a Heparin chromatography (**ANNEX 5**).

We obtained final concentrations of recombinant proteins that allowed us to perform the enzymatic tests described in the section below and the pull-down assays described in section 4 of the manuscript. Moreover, we obtain an amount of *Tbar*-aLhr2WT that allowed us in obtaining polyclonal anti-*Tbar*-Lhr2 antibody (Eurogentec).

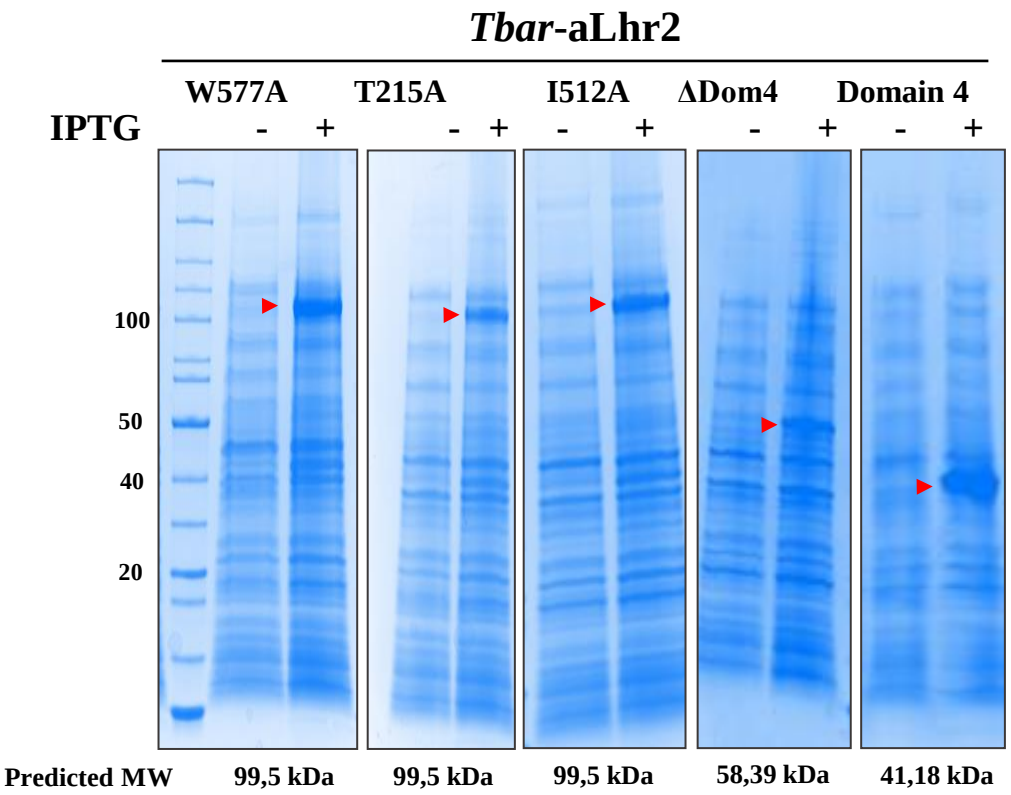


FIGURE 28. Over Expression of recombinant *Tbar*-aLhr2 variants.
Confirmation of recombinant proteins expression in pET11b/BL21 DE3 codon+ system on SDS-PAGE gel (4-15%) colored with Coomassie blue. Overexpression of recombinant aLhr2 variants were induced with 0.2 mM IPTG in *E. coli* cells grown to an OD600 of 0.7–0.9 in LB broth, at 30°C, with agitation at 180 rpm, for 3 h (10 µL of each fraction was loaded per lane).
Lane (-) before adding IPTG, Lane (+) after adding IPTG.

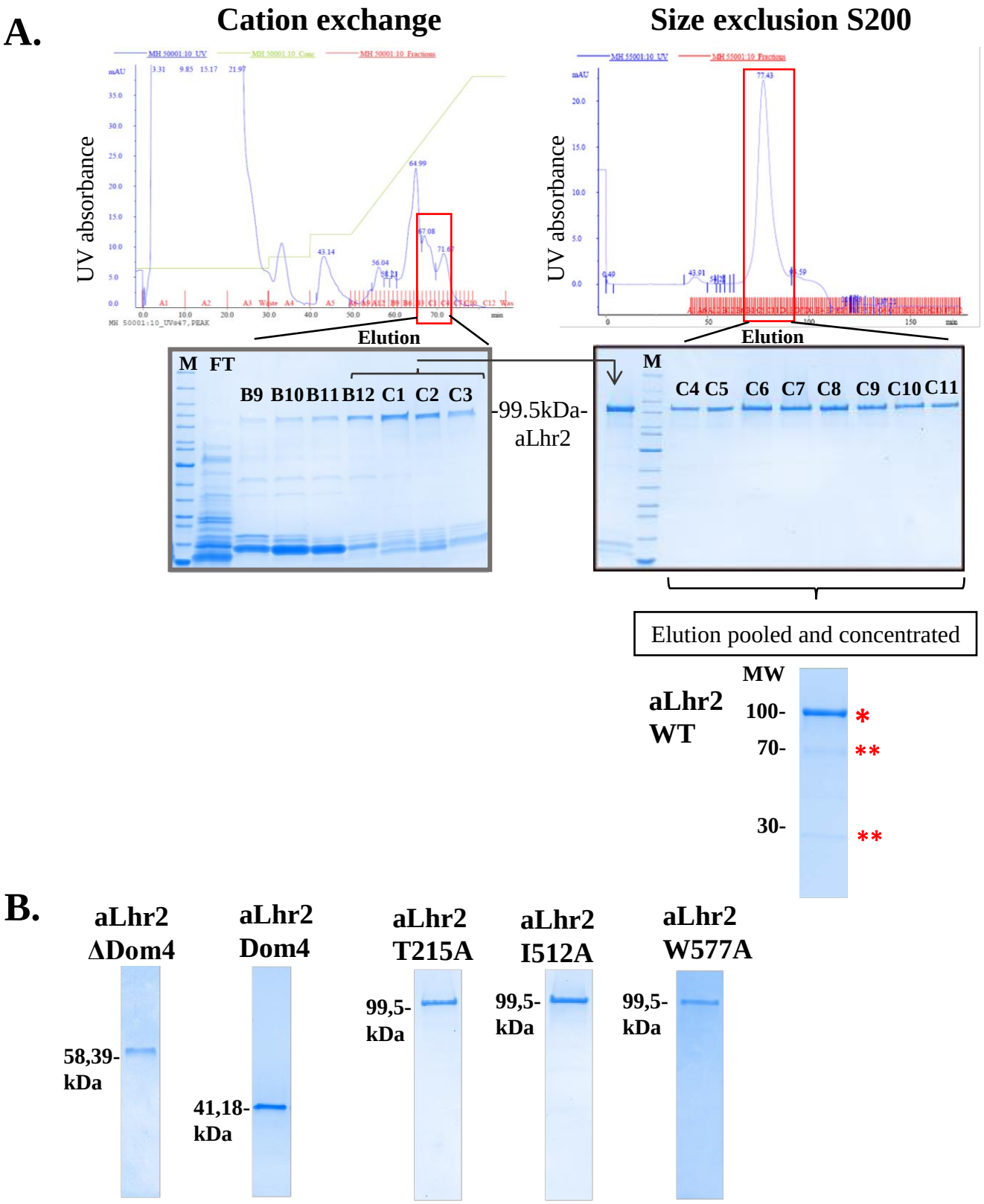


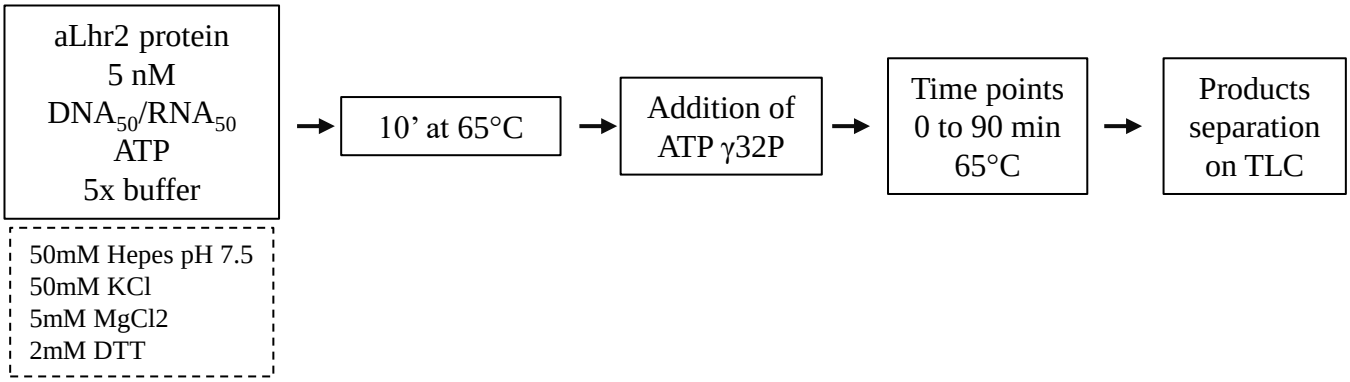
FIGURE 29. Purification of recombinant aLhr2 WT and variants.
A. Purification steps of *Tbar*-aLhr2 WT by cation exchange & size-exclusion chromatographies. Chromatograms and the corresponding SDS PAGE showing the FT (Flow through) and the elution fractions visualized by staining with Coomassie blue are shown. Elution fractions after Size-exclusion S200 step were pooled and concentrated. Protein bands indicated by an * (corresponding to aLhr2 MW) and ** were analysed by sequencing followed by mass spectrometry. **B.** The recombinant proteins purified to homogeneity are visualized on SDS-PAGE gel (4-15%) stained with Coomassie blue. The predicted molecular weight (kDa) of each protein is indicated.

3-Biochemical activities of *Tbar*-aLhr2 wild-type and variants

To decipher the biochemical properties of *Tbar*-aLhr2, the ability to hydrolyze ATP, bind DNA or RNA substrates, unwind nucleic acid duplex and anneal complementary nucleic acid molecules were experienced. The experimental assays performed with *Tbar*-aLhr2 were settled at 65°C as aLhr2 is a protein of hyperthermophilic archaea.

A panel of substrates listed in **Table 10** have been used and consists of different types of single-stranded DNA or RNA molecules of 50 or 26 nucleotides (nt) long, named DNA_{50}, RNA_{50}, DNA_{26} and RNA_{26} respectively. These substrates refer to those used in the study of the enzymatic properties of *Mthe*-Hel308 (Guy and Bolt 2005). This strategy allowed comparing the activities of the three helicases with identical substrates: the *Paby*-ASH-Ski2 RNA helicase and *Paby*-Hel 308 DNA helicase reported recently in the M. Batista PhD thesis (Batista et al, manuscript in preparation) and *Tbar*-aLhr2 helicase reported in here. The DNA_{50} (S1-50nt), 5'DNA_{26} (5RS1-26nt) and 5'RNA_{26} (U5RS-26nt) substrates are synthetic oligonucleotides and the RNA_{50} (US1) substrate is a molecule obtained by *in vitro* transcription (for details, refer to the section "Materials and Methods").

A.



B.

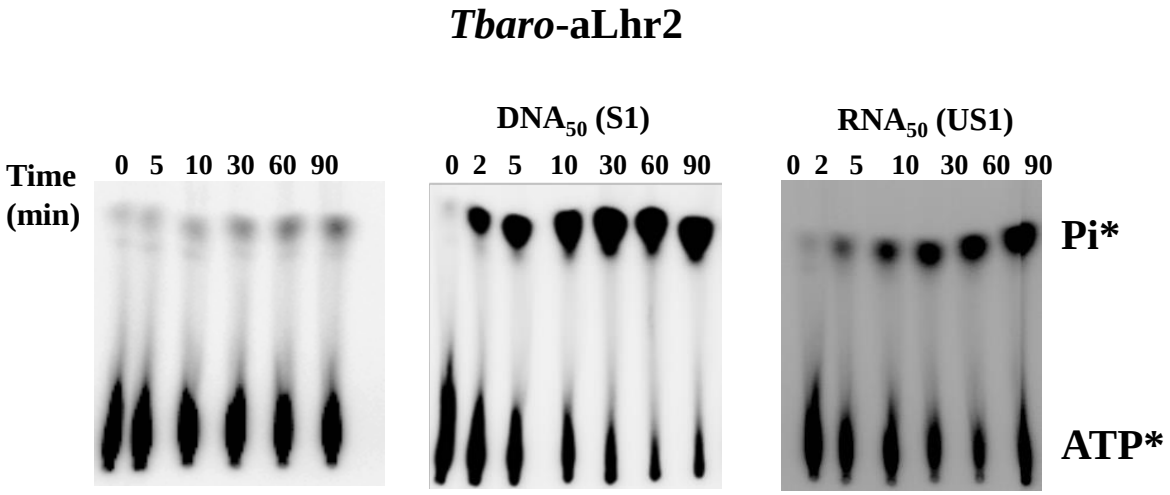


FIGURE 30. ATP hydrolysis assays.
A. Workflow of ATP hydrolysis experimental strategy. **B.** Autoradiography profile corresponding to ATPase reactions stopped at different time, taken from a TLC plate. The inorganic Pi released from ATP hydrolysis was followed over time and quantified after separation on TLC in KH_2PO_4 buffer.

3-1 ATPase activity

The methodology adopted to measure the ATPase hydrolysis of *Tbar*-aLhr2 was developed in the group of Toulouse. For this study, we repeated these experiments in triplicate using wild type and variant versions of recombinant proteins following the protocol schematized in **FIGURE 30A**.

A kinetic of the ATP hydrolysis overtime was stetted up at 65°C using 500nM of purified recombinant protein in the presence (5nM) or absence of nucleic acid substrates DNA_{50} or RNA_{50} **(Table 10)**, and a mix solution containing radiolabelled [ATP γP^{32}]. The release of inorganic phosphate (Pi) (product of the reaction) was followed over time and quantified after separation on thin-layer chromatography (TLC) **(FIGURE 30B)**. We also performed experiments with 250 nM of protein concentration but only 60 % of the radiolabelled ATP was hydrolysed in 90 min. Therefore, we choose a protein concentration of 500 nM in presence of an excess of ATP substrate.

In our experimental conditions, we observed that *Tbar*-Lhr2 WT has the ability to hydrolyse ATP only in presence of nucleic acids **(FIGURE 31A)**. The ATP hydrolysis of the WT protein reaches a plateau at 90 min with the formation of ~90% of Pi in the presence of DNA_{50} and 60% in the presence of RNA_{50} **(FIGURE 31A).** As expected, *Tbar*-Lhr2-Domain 4 does not show any ATPase activity whereas *Tbar*-Lhr2- ΔDom4 shows a residual ATPase activity compare to the WT version of the protein. As well, a reduced activity is observed in the case of *Tbar*-Lhr2-I512A and *Tbar*-Lhr2-W577A which are invalidated for amino acid residues in the Winged-Helix and Domain 4, respectively, with only ~30% of ATP hydrolysed after 10 minutes of incubation. **(FIGURE 31B).** In all the cases, the induction of the ATPase activity in the presence of DNA_{50} molecules is faster than in the presence of RNA_{50} molecules. However, this preference is reduced in the context of the *Tbar*-Lhr2 variants (**FIGURE 31B**).

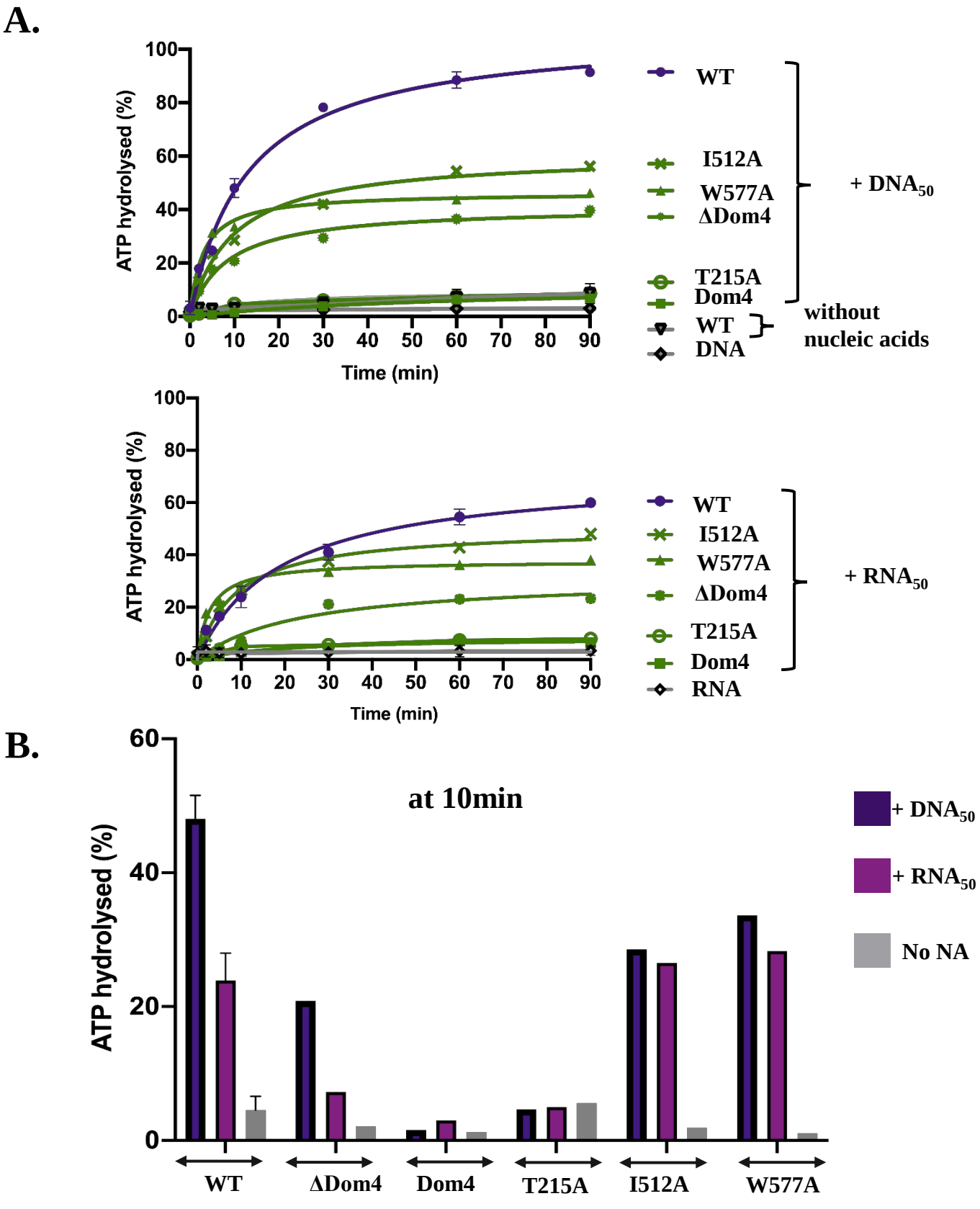
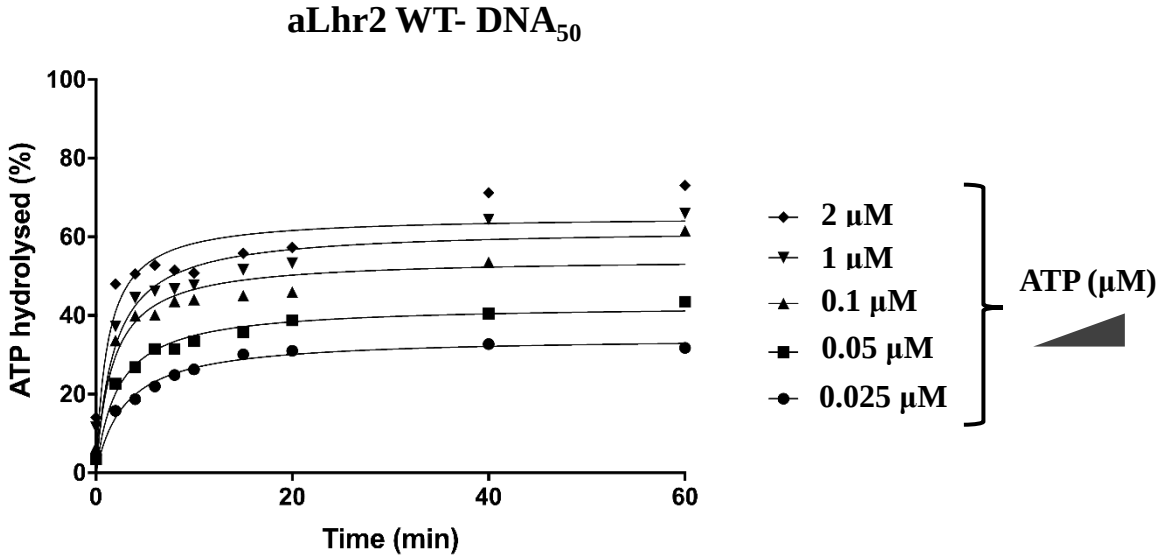


FIGURE 31. ATP hydrolysis assay.

A. Time-dependent ATPase activity of aLhr2 WT and variants. The protein was incubated with [γ-32 P] ATP for various time intervals. The amount of ATP hydrolyzed at each time point is shown as a percentage of the original [γ-32 P] ATP before incubation at 65°C without (grey curves) and with DNA_{50} (top panel) or RNA_{50} (lower panel). For the WT aLhr2 each point represents the average of 3 independent experiments. **B.** A histogram showing the % of ATP hydrolysis of aLhr2 WT and variants in the absence (grey) and presence of DNA_{50} (purple) or RNA_{50} (pink) after 10 min of reaction incubation.

A.



B.

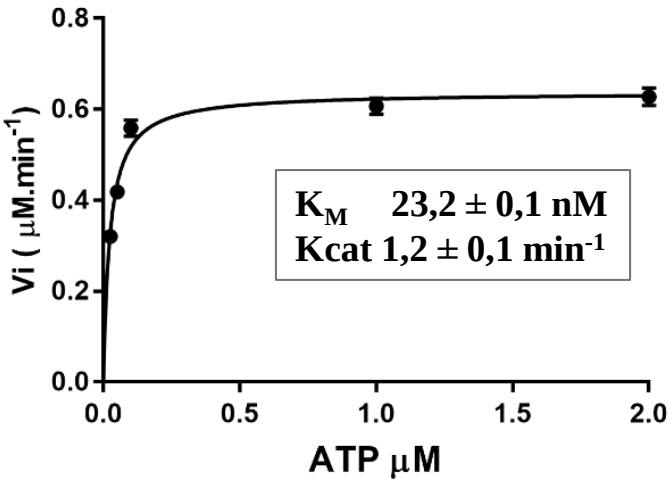


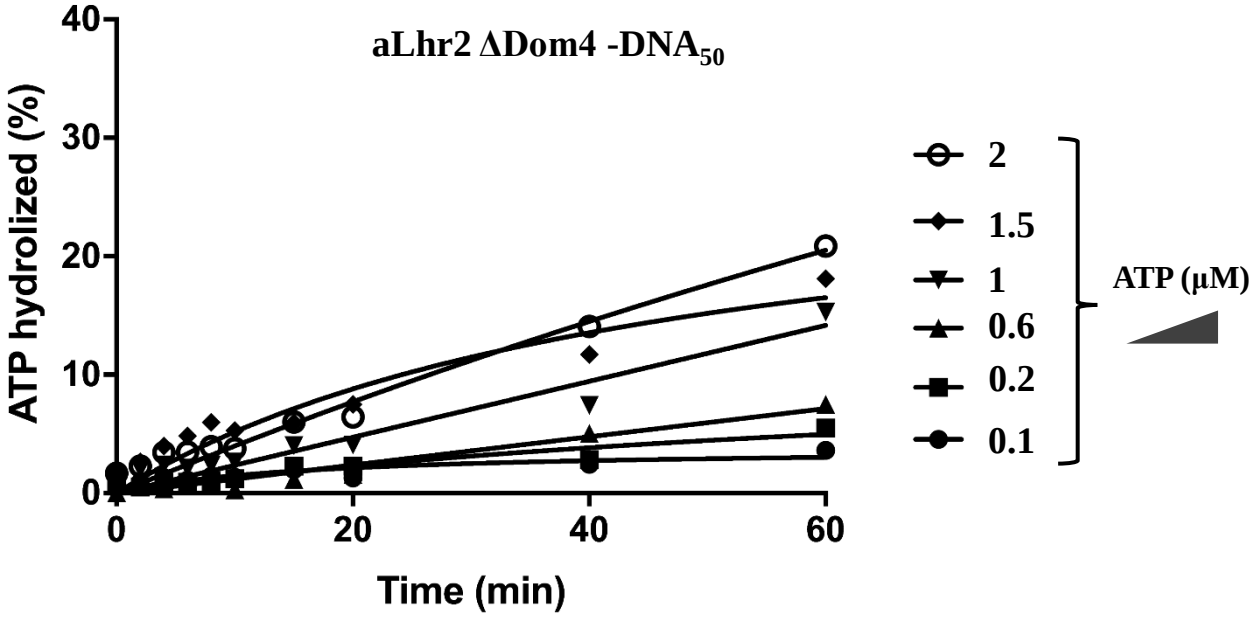
FIGURE 32. Kinetic of ATP hydrolysis of aLhr2 WT.
A. Michaelis–Menten kinetic analyses of aLhr2 WT ATPase activity. The rate of ATP hydrolysis in the presence of different substrate concentrations was measured in the presence of 500 nM of protein and the appropriate buffer (see Materials and methods). **B.** Initial velocity of ATP hydrolysis of aLhr2 WT in the presence of different substrate (ATP) concentrations. Curves and kinetic parameters (Vi, KM, and Kcat) were obtained using Graph Prism.

In conclusion, *Tbar*-aLhr2 is a nucleic acid-dependent ATPase which is induced, *in vitro*, preferentially in the presence of DNA substrate rather than RNA substrate. The helicase core of the protein that contains the active site formed by the two RecA1 and RecA2 domains is critical for the ATPase activity. Indeed, invalidation of the conserved motif III T215A is detrimental for the activity. Finally, no activity was observed for the Domain 4 alone and a very low activity for the catalytic mutant in motif III (**FIGURE 31A &B**) which is reported to be coordination hydrolyse ATP (Banroques et al. 2010).

To go further, we determined the steady-state kinetic parameters of *Tbar*-Lhr2 WT by measuring the velocity of ATP hydrolysis as a function of cold ATP concentration (0.025, 0.05, 0.1, 1, and 2µM) in the presence of 5nM of DNA_{50} (**FIGURE 32A**). From a nonlinear regression curve fit of Michaelis-Menten equation, we calculated the K_m (Michaelis constant) and the k_{cat} (catalytic constant or turnover) using Graph-Prism (see Material & Methods) (**FIGURE 32B**). The derived K_m/k_{cat} value issignificantly lower than that of the other reported bacterial Lhr (Ejaz and Shuman 2018; Ordonez and Shuman 2013). We should note that bacterial Lhr showed a preference for calcium (*Msme*-Lhr) and manganese (*Pput*-Lhr) for hydrolysing ATP which could explain the slower catalysis obtained in our conditions with only magnesium as divalent ions. It will be interesting to test if other divalent ions affect *Tbar*-Lhr2 ATPase activity.

We performed the same analysis with *Tbar*-aLhr2 ΔDom4 which is deprived of the domain 4 (**FIGURE 33A**) and with *Tbar*-Lhr2-T215A which is invalidated in the Motif III of the RecA2 domain (**FIGURE 33B**). The absence of the C-terminal and the Motif III mutation reduced considerably the capacity of the enzyme to bind ATP which compromises the accurate calculation of Michaelis-Menten constants.

A.



B.

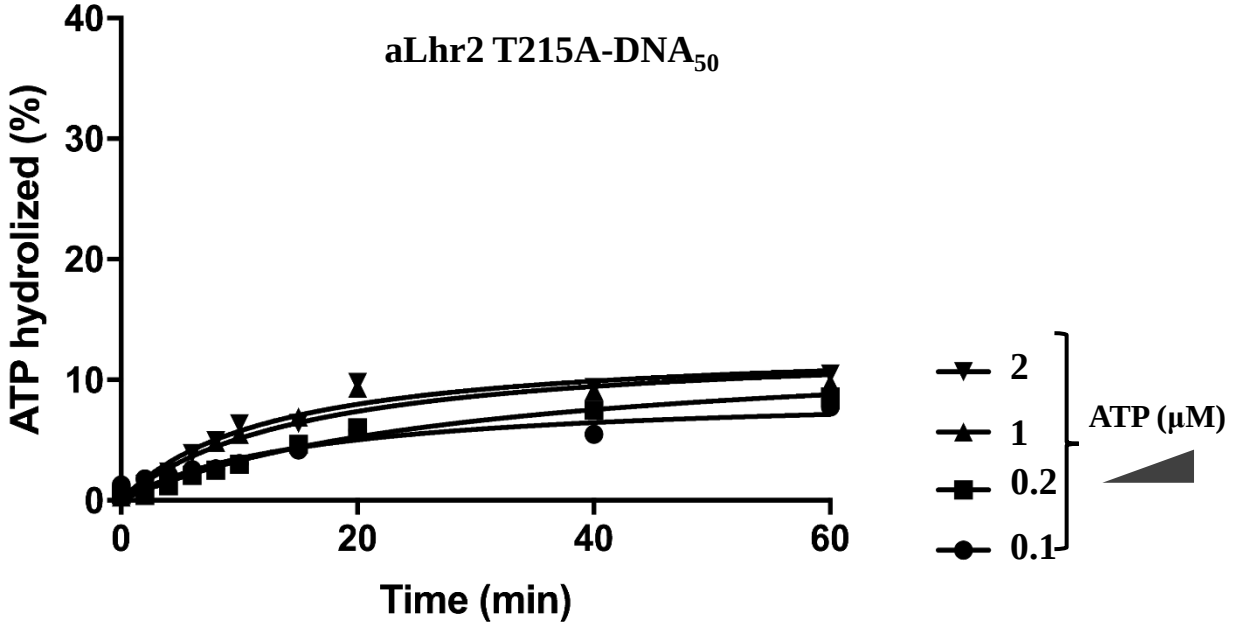


FIGURE 33. Kinetic of ATP hydrolysis of aLhr2 ΔDom4 and T215A.
Michaelis–Menten kinetic analyses of the ATPase activity of aLhr2 ΔDom4 **A.** and T215A **B.** The rate of ATP hydrolysis in the presence of different substrate concentrations was measured in the presence of 500 nM of protein and the appropriate buffer (see Materials and methods).

3-2 Nucleic acid affinity

The binding capacity of *Tbar*-aLhr2 WT and variants for nucleic acid was tested using a set of single-stranded substrates: a 50-mer DNA (DNA_{50}) or RNA (RNA_{50}), and a 26-mer DNA (5'DNA_{26}) or RNA (5'DNA_{26}) (**Table 10**). The binding assay was carried out as shown in **FIGURE 34A**. A range of protein concentration from 0 to 350 nM was incubated in the presence of 5nM of each labelled substrate for 15min at 30°C. The nucleoprotein complexes were separated from free nucleic acid by a double filtration system **(FIGURE 34B)**. Nucleic acid-protein complexes were retained on the nitrocellulose membrane (bound fractions), while the free DNA and RNA molecules (unbound fractions) were retained on the nylon membrane which is positively charged. The percentage of bound fraction was plotted against the protein concentration **(FIGURE 35)**. In all conditions, we observed that the binding curves are sigmoidal with Hill coefficients superior to 1 (S-shape curves) indicating a positive binding cooperativity. Most likely, more than one molecule of protein binds to multiple sites on a single molecule of nucleic acids. This was observed in the case of WT (**FIGURE 35A**) and variants of *Tbar*-Lhr2 (**ANNEX 6-7&8**).

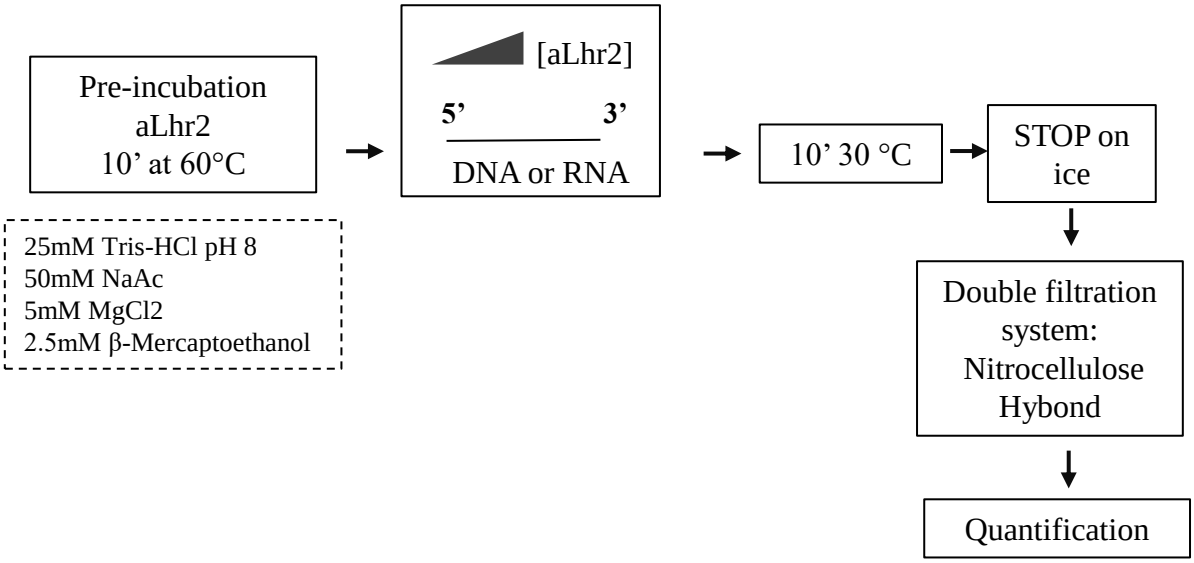
The apparent dissociation constant K_D was determined for each protein/substrate complex **(Table 12).** First, *Tbar*-aLhr2WT harbours comparable affinity constants for both DNA and RNA substrates and the presence of ATP in the binding reaction mixture does not affect the binding affinity. The length of substrates (50nt versus 26nt) does not alter the binding constant affinity **(FIGURE 35B).** In all the case, similar values of K_D have been obtained for DNA or RNA substrates and the *Tbar*-aLhr2 variants recombinant proteins appeared to bind with slightly higher affinity longer substrates (**Table 12**). Interestingly, the domain4 of *Tbar*-Lhr2 (*Tbar*-Lhr2 Dom4) has an intrinsic affinity of nucleic acid independently of the presence of the helicase core. Therefore, we can propose that *Tbar*-Lhr2 has two binding sites for nucleic acids, one located in the helicase core and one in the Domain4.

Table 12. Apparent K_D values.

Substrate	apparent dissociation constant K_D (nM)						
	WT	**ΔDom4**	**Dom4**	**T215A**	**I512A**	**W577A**	**WT+ATP**
DNA_{50}	39 ± 1	54 ± 1	162 ± 5	68 ± 2	22 ± 2	69 ± 2	38 ± 1
5'DNA_{26}	58 ± 2	211 ± 9	277 ± 16	141 ± 2	31 ± 1	118 ± 3	*n.d*
RNA_{50}	41 ± 1	69 ± 1	161 ± 5	77 ± 2	19 ± 1	65 ± 2	36 ± 1
5'RNA_{26}	80 ± 2	150 ± 1	241 ± 17	164 ± 3	33 ± 1	125 ± 6	*n.d*

K_D values of nucleic acid binding affinity are obtained using Graph Prism.

A.



B.

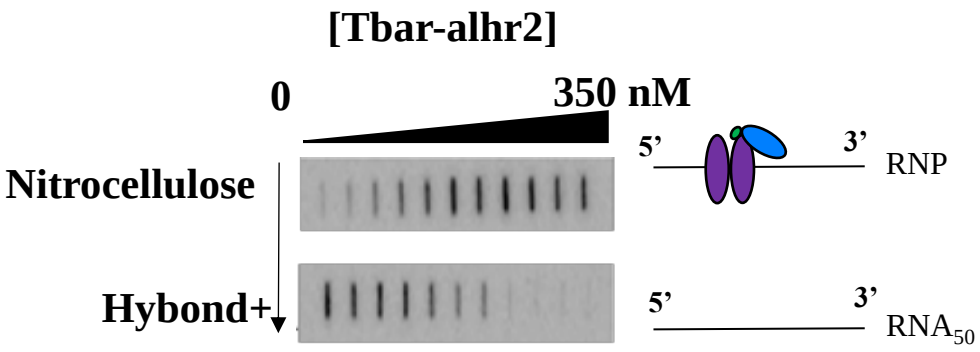


FIGURE 34. *Tbar*-aLhr2 affinity for nucleic acids.
A. Workflow of the experimental strategy which uses a range of protein concentration in the presence of 5nM of radiolabelled nucleic acid substrate **B.** Affinity tests by double filtration on a nitrocellulose DEAE and HydondN membrane with 5nM of radiolabelled RNA (RNA_{50} and 5'RNA_{26}) or DNA (DNA_{50} and 5'DNA_{26}) substrates incubated with a range of aLhr2 WT and variants protein concentrations at 30 ° C for 15min in presence of the appropriate buffer (see Materials and methods).

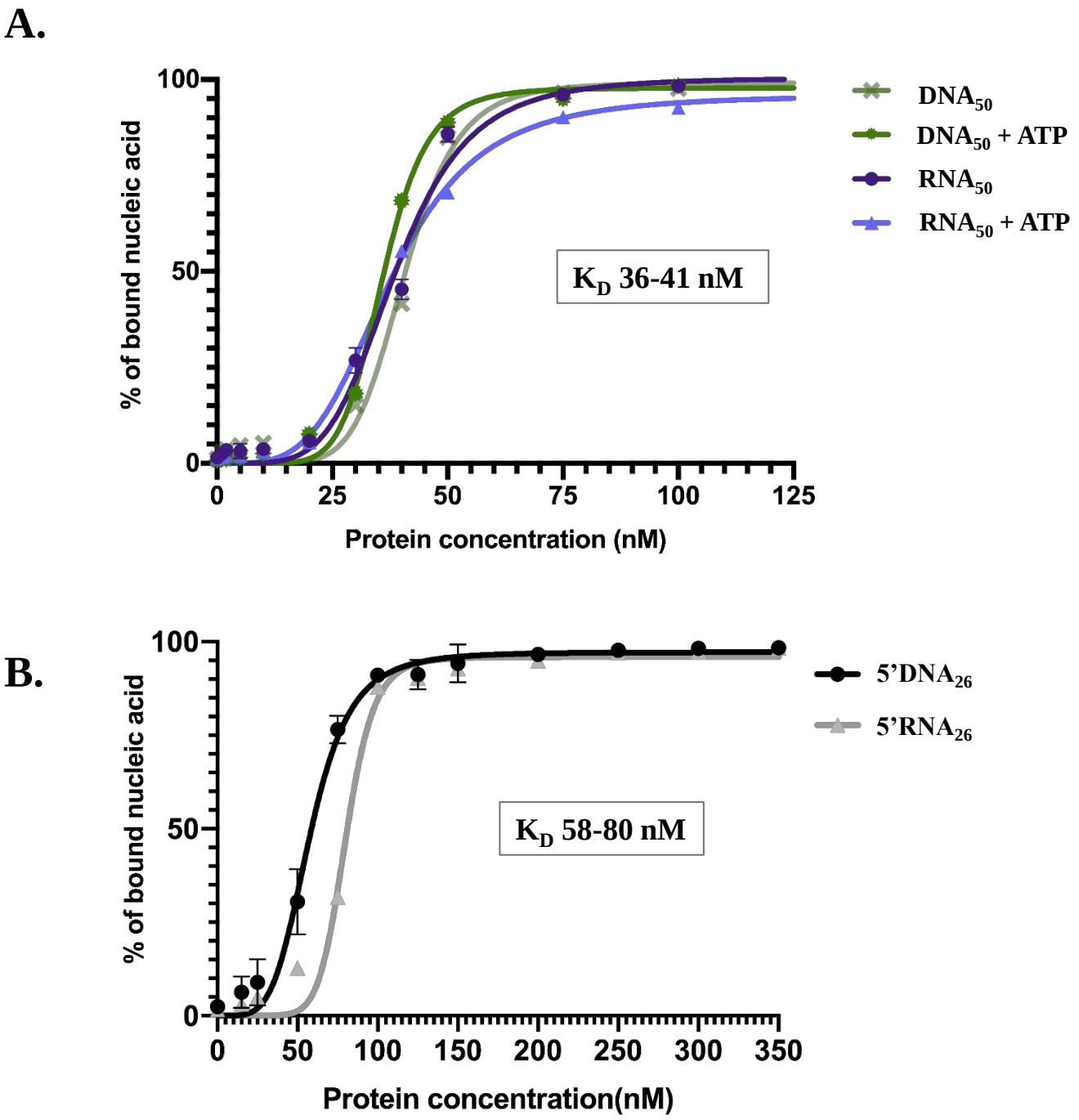


FIGURE 35. Binding affinity curves of *Tbar*-aLhr2 WT for single strand-nucleic acids.
The percentage of RNP complex formation is plotted versus protein concentration to calculate the apparent affinity constant (KD). **A.** *Tbar*-aLhr2 WT affinity for DNA_{50} (S1) and RNA_{50} (US1) in the presence (+) or absence (-) of ATP **B.** *Tbar*-aLhr2 WT affinity for 5'DNA_{26} (5RS1) and 5'RNA_{26} (U5RS1).
Values are the average of three independent experiments.

3-3 Unwinding activity

To assess the helicase activity of *Tbar*-aLhr2, we tested the capacity of recombinant protein to unwind nucleic acid duplexes (**FIGURE 36A**). To perform these assays, a panel of DNA and RNA oligonucleotides or transcribed molecules were used (**Table 10 Materials and methods**). Briefly, semi-homo (DNA/DNA or RNA/RNA) or heteroduplexes (DNA/RNA) were *in vitro* pre-formed using 50nt-single-stranded DNA (S1) or RNA (US1) substrate complementary to a 26nt-single-stranded RNA (U5RS1, U3RS1) or DNA (5RS1). In the context of hetero-(DNA/RNA) or homo-duplexes (RNA/RNA), U5RS1 and U3RS1 were used to form 3'- or 5'-overhang duplexes, respectively. In addition, to obtain stable DNA/DNA homo-duplexes, longer DNA molecules of 59nt-single-stranded DNA (S2) and a 31nt-radiolabeled single-stranded DNA (5RS2) were employed. In each case, one of the molecules of the duplex was beforehand 5'end radiolabelled by addition of a αP^{32} (See Material & methods)

Using this panel of RNA/RNA or DNA/RNA duplexes, the dissociation of duplexes was followed over time at 65°C in presence of 250nM of purified recombinant proteins, 5 mM of ATP, and an excess of unlabelled oligo or Trap oligo **(FIGURE 36B)**. The Trap oligo prevents the re-hybridization of single-stranded molecules formed during the course of the reaction by capturing the unlabelled complementary molecule. The radiolabelled-single stranded product was separated from radiolabelled-duplexes by electrophoresis in native conditions **(FIGURE 36C, left panel)**. At the same time, we carried out a protein-free control experiment **(FIGURE 36C, right panel)**.

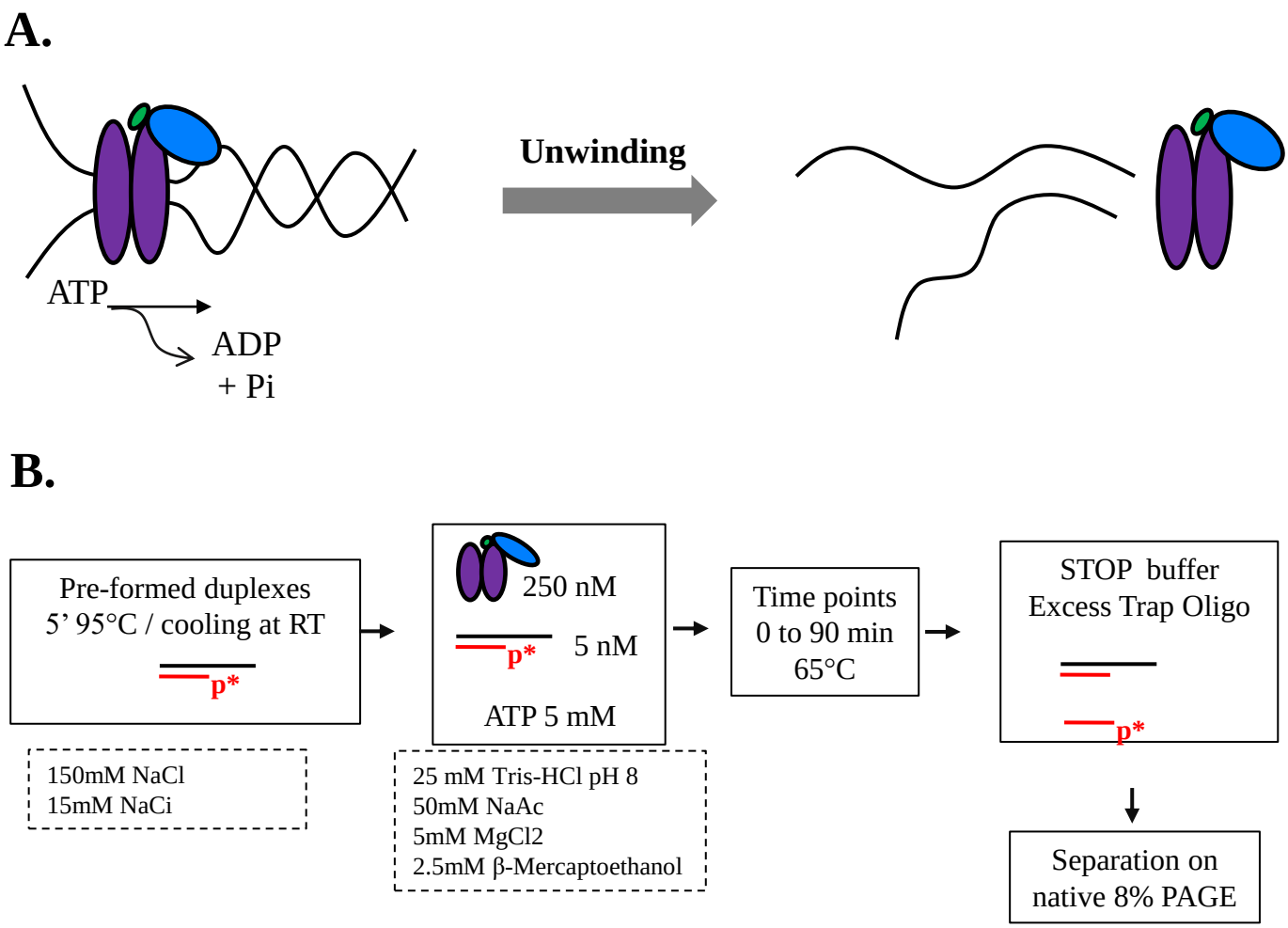


FIGURE 36. Unwinding activity assay.
A. Schematic representation of the unwinding reaction **B.** Experimental workflow. The substrate duplex formed by incubation at 95°C for 5 min followed by a step of cooling to RT of single-stranded radiolabelled and unlabelled substrates. **C.** Radiolabelled duplex are incubated for 90' at 65 ° C with 25 mM of ATP and in the presence (left panel) /absence (Control-right panel) of 250nM of protein and the appropriate buffer (see Materials and Methods section). Reaction was quenched at different times and analyzed after migration in a native 8% acrylamide gel.

After quantification of the signal, the percentages of unwound duplex were plotted over time (**FIGURE 37**). In the case of *Tbar*-aLhr2, more than 60% of 3' DNA_{50}/RNA_{26} or RNA_{50}/RNA_{26} duplexes was unwound after 90 min of reaction, in our experimental conditions **(FIGURE 37A)**. Only 40 % was observed in the case of 5' RNA_{50}/RNA_{26} duplex (US1/U3RS1) **(FIGURE 37A).** This indicates *Tbar*-aLhr2 has a preference for 3'-overhang duplex in the cases of RNA/RNA duplex. We observed that *Tbar*-aLhr2 was not able to unwound 3' DNA_{59}/DNA_{31} duplex in our conditions *in vitro* **(FIGURE 37A)**. In this case, the percentage of unwound duplexes is comparable to that observed for the control sample without protein. In contrast, *Paby*-Hel308 which was reported to be a DNA helicase involved in DNA repair (Li et al. 2008) was able to rapidly unwind 80% of DNA_{59}/DNA_{31} duplexes after only 2 minutes of incubation but not RNA_{50}/RNA_{26} duplexes **(FIGURE 37B)**. Finally, in the case of *Tbar*-aLhr2 ΔDom4 recombinant protein, no unwinding activity was observed **(FIGURE 37C).**

In addition, a series of experiments in presence of non-hydrolysable ATP analogues were performed **(FIGURE 38A)**. We observed that in presence of AMP-PNP or ATPγS, *Tbar*-aLhr2 WT was no longer able to unwind 3' RNA_{50}/RNA_{26} duplex **(FIGURE 38A).** It is possible that the binding of ATP analogue freezes *Tbar*-aLhr2 WT within a conformation unable to unwind duplexes.

From each experimental condition performed in triplicate, we could derive an initial velocity of unwinding (**FIGURE 38B**). Comparison of values for *Tbar*-aLhr2 WT and *Paby*-Hel308 illustrate a divergent behaviour that may reflect that *Tbar*-aLhr2 may not be involved in DNA metabolism such as reported for Hel308 DNA helicase (Li et al. 2008). The absence of activity for *Tbar*-aLhr2 ΔDom4 suggests that the Domain4 is required for the formation of active unwind conformation of the protein.

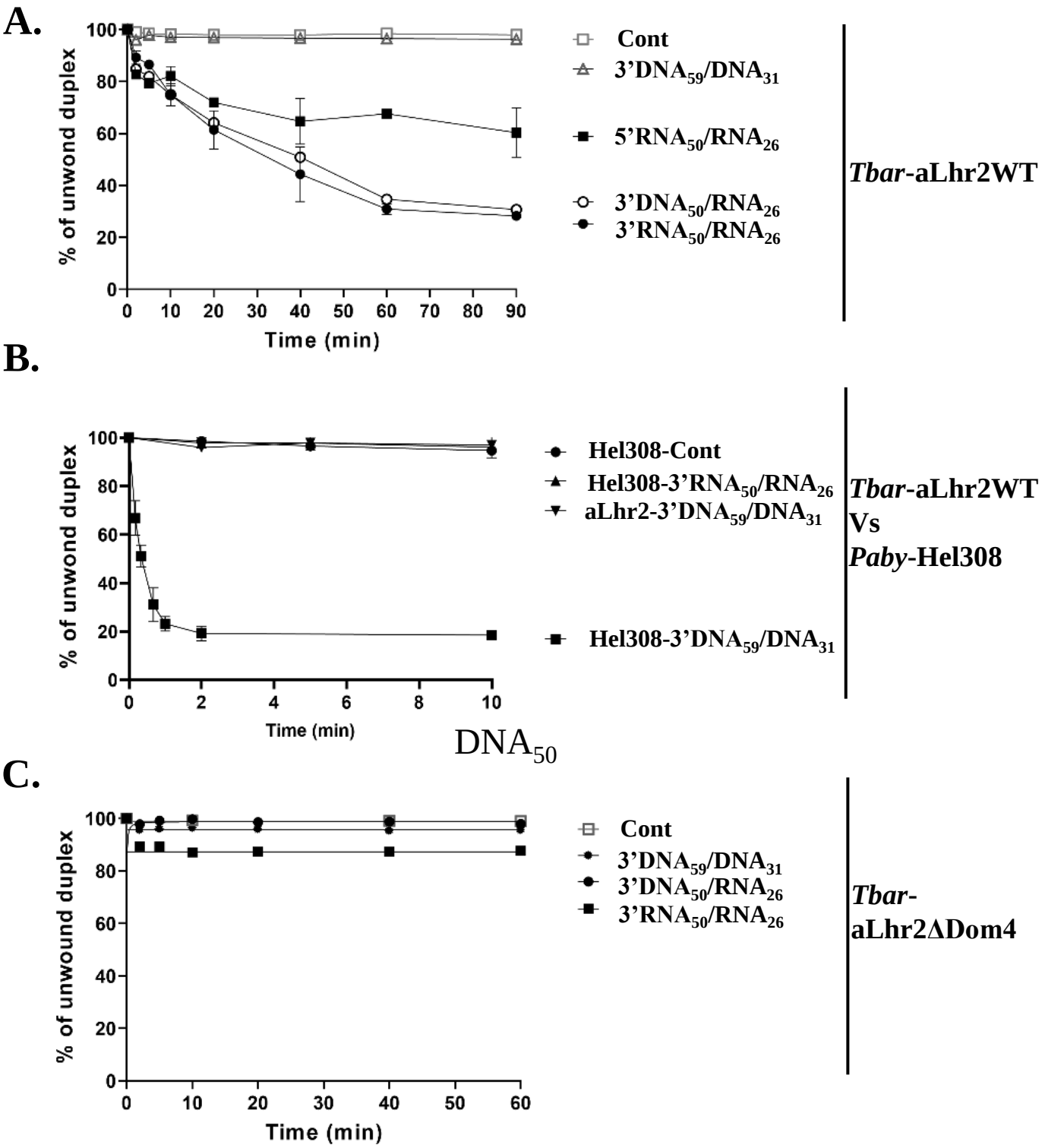
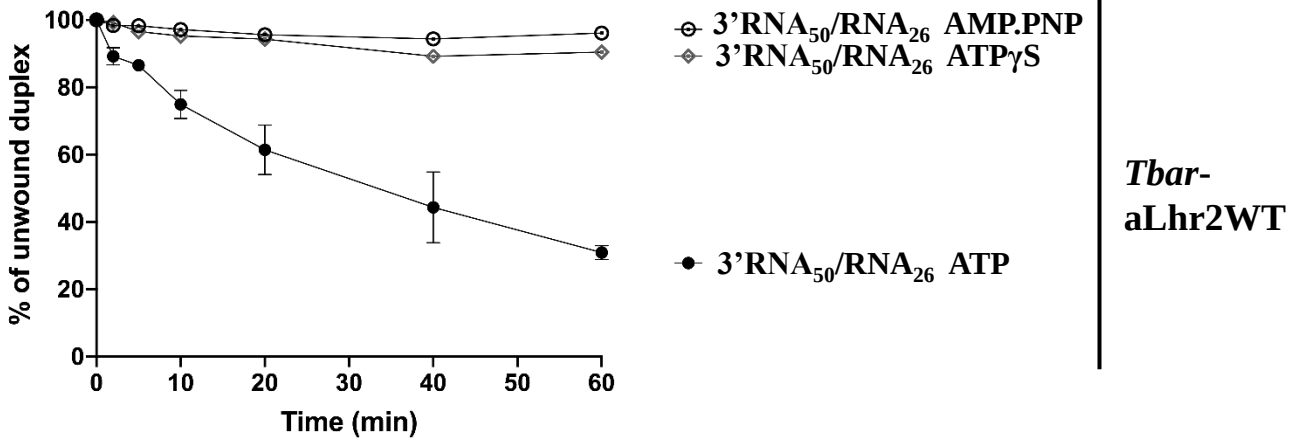


FIGURE 37. Unwinding activity of *Tbar*-aLhr2WT and *Tbar*-aLhr2ΔDom4 in presence of ATP. **A.** Kinetics of strand dissociation activity in the presence of ATP in absence (Cont) or presence of 250 nM of *Tbar*-aLhr2 protein **B.** Kinetics of unwinding of 3'RNA_{50}/RNA_{26} and 3'DNA_{59}/DNA_{31} duplex of *Tbar*-aLhr2 WT vs *Paby*-Hel308 **C.** Kinetics of unwinding of *Tbar*-aLhr2 ΔDom4.

A.



B.

Vi (nM.min^{-1})	*Tbar*-aLhr2WT	*Tbar*-aLhr2ΔDom4	*Paby*-Hel308
3'DNA$_{59}$/DNA$_{31}$	-	-	4.3
3'RNA$_{50}$/RNA$_{26}$	0.82	-	0.17
5'RNA$_{50}$/RNA$_{26}$	0.34	-	*n.d.*
3'DNA$_{50}$/RNA$_{26}$	0.72	-	*n.d.*

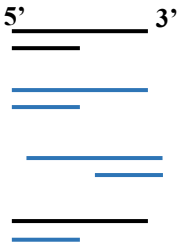


FIGURE 38. Unwinding in presence of analogs of ATP & Initial velocity of unwinding reaction. **A.** Kinetics of 3'RNA$_{50}$/RNA$_{26}$ Duplex unwinding activity of *Tbar*-aLhr2 WT in presence of ATP, AMP.PNP and ATPγS. **B.** Michaelis–Menten kinetic (Vi) analyses of the unwinding activity of *Tbar*-aLhr2 WT and *Paby*-Hel308. Curves and kinetic parameters were obtained using Graph Prism. **(-) :** no activity, **n.d:** activity not determined.

3-4 Annealing activity

With the same panel of RNA and DNA substrates (**Table 10, Materials and Methods**), we tested the capacity of the enzyme to anneal single-stranded of nucleic acids with some complementary sequences (**FIGURE 39A**). In this context, we settled up an experimental workflow that allowed to follow the formation of nucleic acid duplex over a time course at 65°C in presence of 250nM of purified protein (**FIGURE 39B**). Quantification of each species allowed plotting the percentage of duplex versus time (**FIGURE 40A**). We performed each assay in triplicate in order to derive an initial velocity of annealing. The radiolabelled-duplex was separated from the radiolabelled-single stranded product by electrophoresis in native conditions **(FIGURE 39C, left panel)**. At the same time, we carried out a protein-free control experiment **(FIGURE 39C, right panel)**.

In absence of ATP, *Tbar*-aLhr2WT is able to rapidly anneal 3'overhang RNA_{50}/RNA_{26} and DNA_{50}/RNA_{26} duplexes with 80% of formed duplexes after 3 minutes of reaction (**FIGURE 40A, top panel)**. More precisely, we observed that *Tbar*-aLhr2WT was also able to rapidly and efficiently assemble DNA_{50}/RNA_{26} hetero-duplex, and DNA_{59}/DNA_{26} and RNA_{50}/RNA_{26} homo-duplex with up to 60, 75, and 70% of duplex product, respectively (**FIGURE 39A, top panel).** The annealing velocity of *Tbar*-aLhr2 WT is drastically reduced in the case of 5'overhang RNA_{50}/RNA_{26} duplex (**FIGURE 40B).** In presence of ATP, *Tbar*-aLhr2WT lost its annealing capacity with almost no duplexes formed (**FIGURE 40A, top panel)**.

Finally, similar experiments were performed with the *Tbar*-aLhr2 ΔDom4 version (**FIGURE 40A**, lower panel). In this case, the annealing activity is significantly reduced independently of the nature (DNA or RNA) of the substrate.

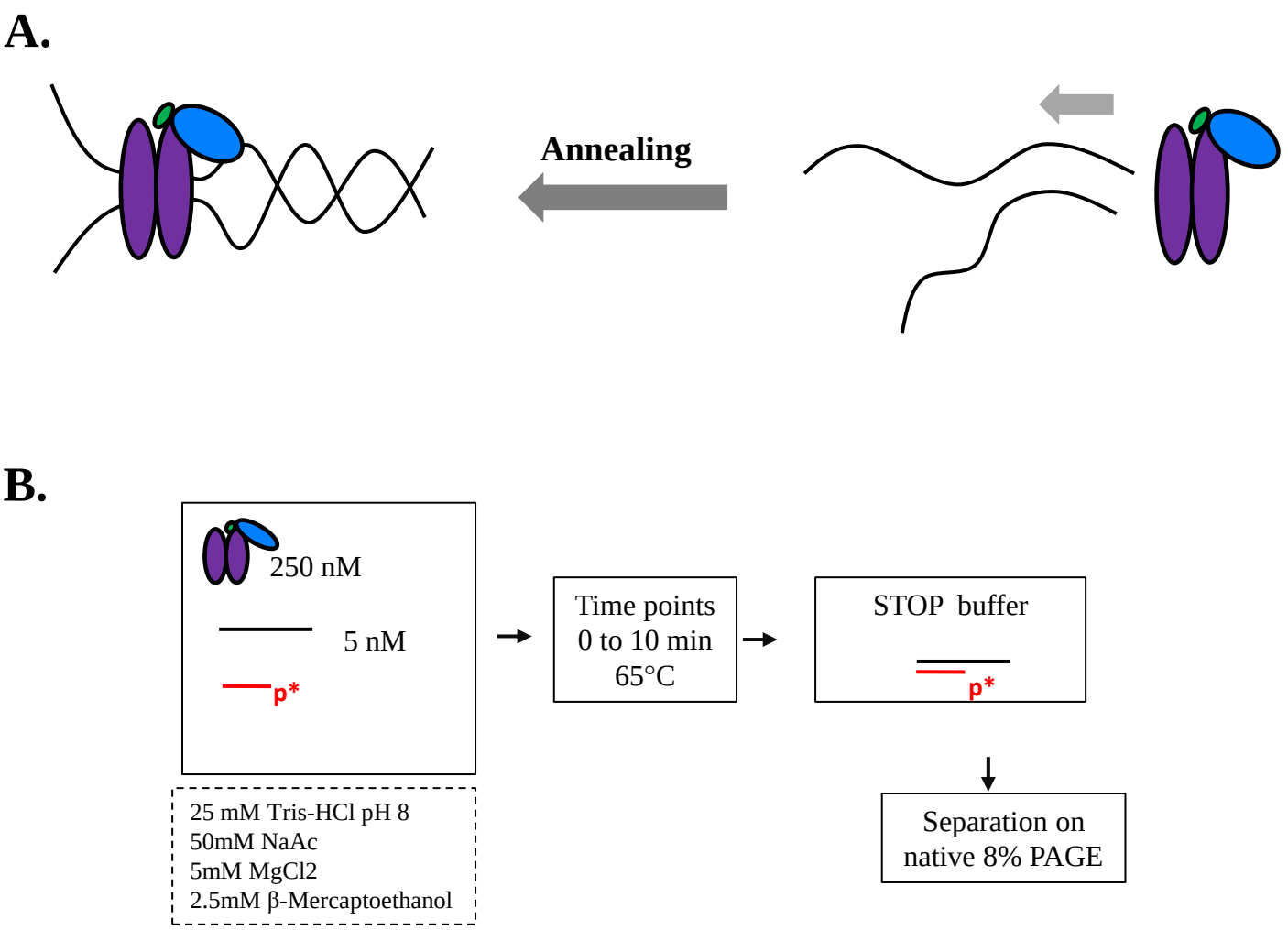
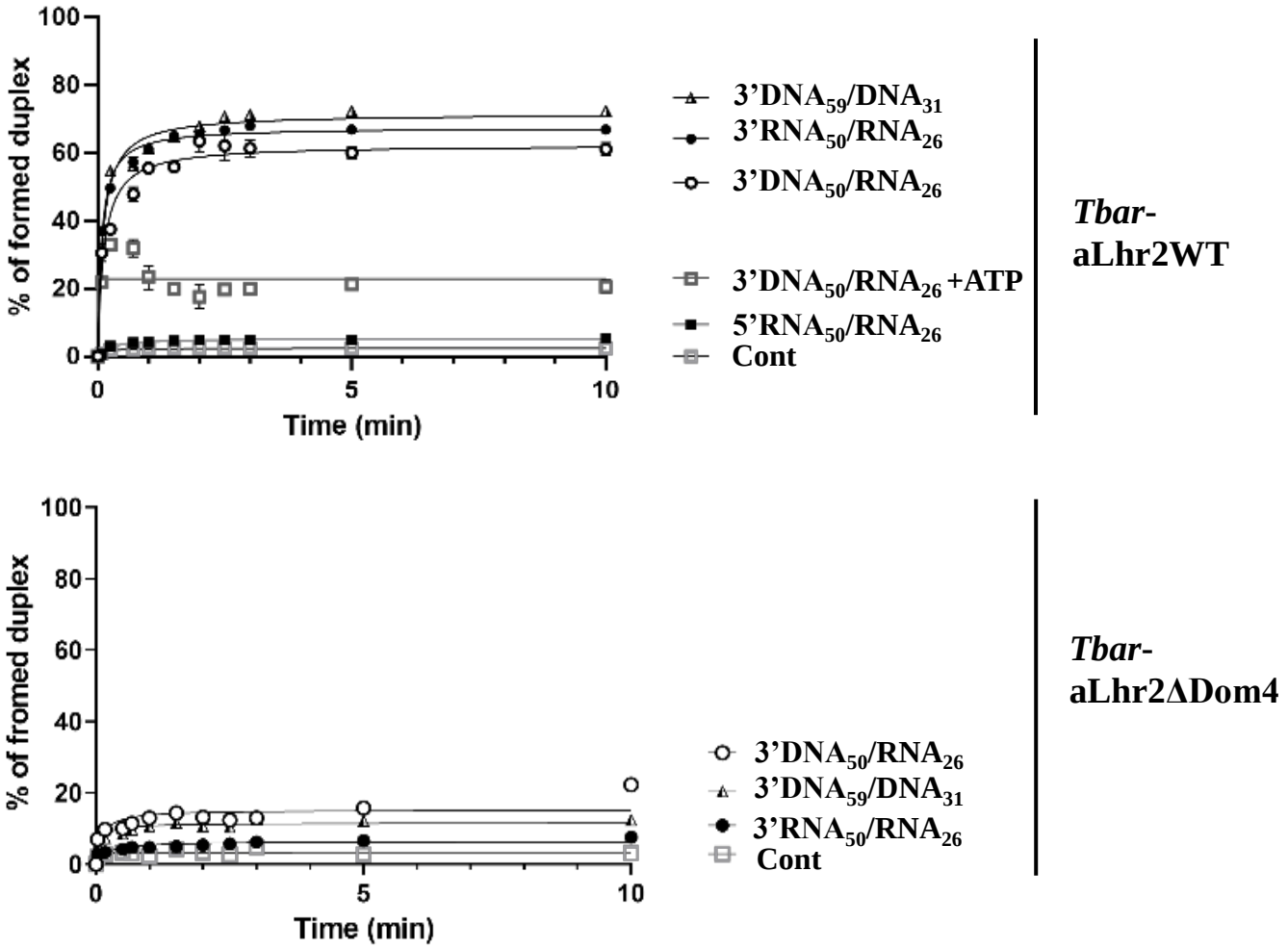


FIGURE 39. Duplex annealing assay.
A. Schematic representation of the annealing reaction **B.** Experimental workflow. **C.** Single-stranded radiolabelled and unlabelled substrates are incubated for 10‘ at 65°C in the presence(left panel) /absence (Control-right panel) of 250nM of protein and the appropriate buffer (see Materials and Methods section). Reaction was quenched at different times and analyzed after migration in a native 8% acrylamide gel.

A.



B.

Vi (nM.min^{-1})	***Tbar*-aLhr2WT**		***Tbar*-aLhr2ΔDom4**
	-ATP	+ATP	
3'DNA$_{59}$/DNA$_{31}$	4	*n.d.*	-
3'RNA$_{50}$/RNA$_{26}$	3	-	-
5'RNA$_{50}$/RNA$_{26}$	0.3	*n.d.*	-
3'DNA$_{50}$/RNA$_{26}$	3	*n.d.*	-



FIGURE 40. Annealing activity assay of *Tbar*-aLhr2WT and ΔDom4.
A. Kinetics of duplex re-association activity in absence (Cont) or in presence of 250 nM of *Tbar*-aLhr2 WT **(top panel)** or 250 nM of *Tbar*-aLhr2 ΔDom4 **(lower panel). B.** Michaelis–Menten kinetic (Vi) analyses of the annealing activity of *Tbar*-aLhr2 WT in presence or absence of ATP . Curves and kinetic parameters were obtained using Graph Prism. (**-**) : no activity, **n.d:** activity not determined.

4-Conclusions

A summary of the results obtained in this series of assays is shown in **FIGURE 41**. We propose that *Tbar*-aLhr2 is a nucleic-acid dependent ATPase with a slight preference for DNA template and has an unwinding activity with a 3'-5' polarity in presence ATP that is inhibited by ATP analogues. More specifically, *Tbar*-aLhr2 is only able to displace RNA strand but not DNA strand. Finally, *Tbar*-aLhr2 is able to anneal DNA and RNA homoduplex, as well as RNA/DNA heteroduplex with the same efficiency.

If we compare the properties of *Tbar*-aLhr2 with the Lhr-type reported for *Sulfolobus solfataricus* (*Ssol*-Hel112), we can observe some divergences. *Ssol*-Hel112 was shown to be DNA-independent ATPase with a very low affinity for ssDNA. Nonetheless *Ssol*-Hel112 and *Tbar*-aLhr2 harbour similar annealing activities (De Felice et al. 2007).

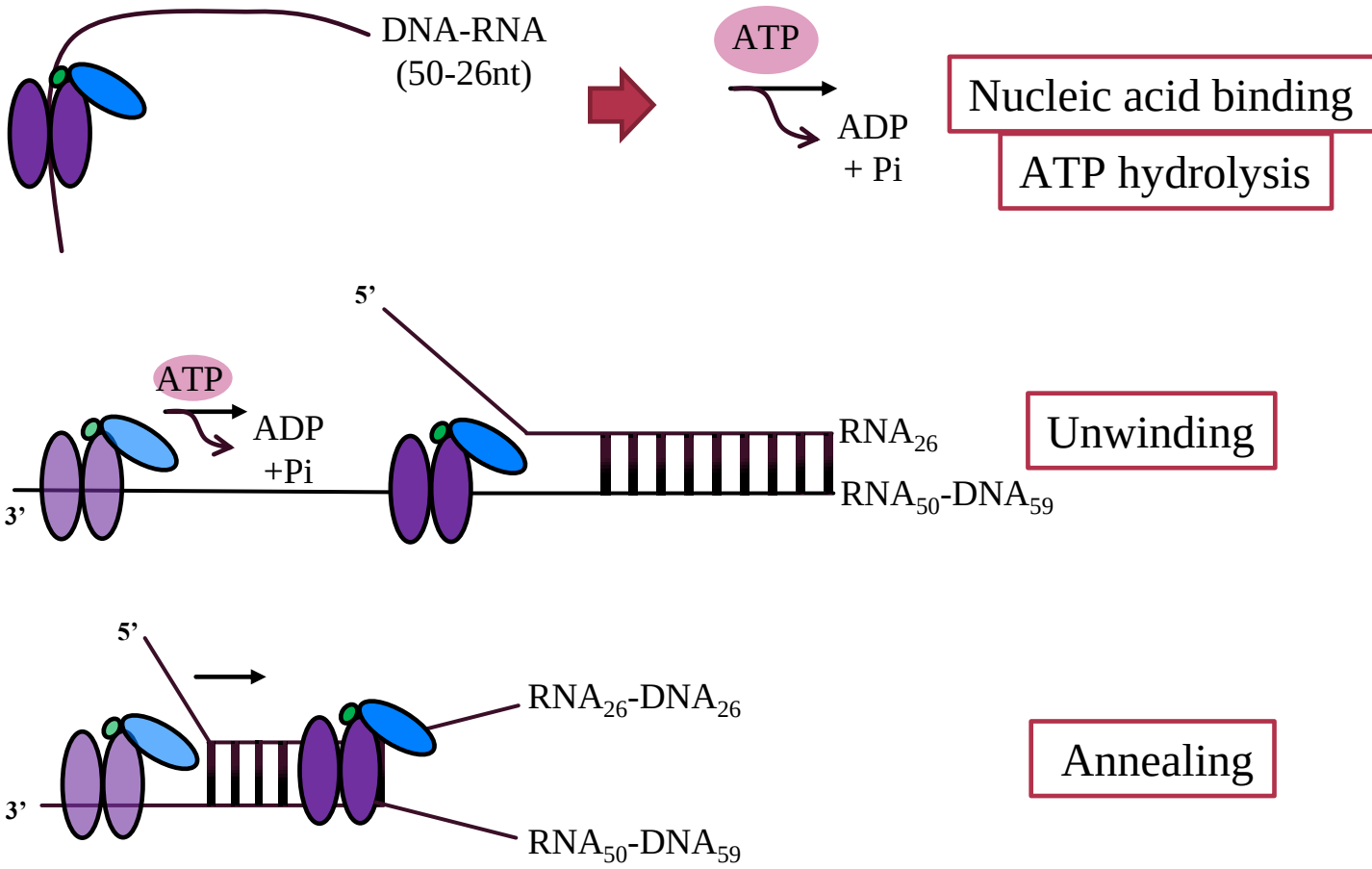


FIGURE 41. Mechanistic activities of *Tbar*-aLhr2.
Schematic representation summarizing the enzymatic activities of *Tbar*-aLhr2 WT validated in this study.

III-Towards the significance of aLhr2 of *T.barophilus*

Functions of aLhr2 protein in *Thermococcus barophilus* was tackled through proteomic and transcriptional exploratory studies. The purpose of these analyses was the identification of protein partners and cellular targets to tackle the role of *Tbar*-aLhr2. Consequently, we developed in parallel a proteomic approach using *T .barophilus* cellular extract with purified recombinant *Tbar*-aLhr2 protein and a transcriptomic approach by comparing the transcriptomes of *ΔTbar-alhr2* strains.

1-Identification of protein interaction network of *Tbar*-$(His)_6$-aLhr2

1-1 The Pull-down methodology

In this part of the manuscript, we have determined the interaction network of the *Tbar*-$(His)_6$-aLhr2 using the methodology initially developed by the team of D. Flament (Ifremer, Brest) and is currently used in the laboratory in Toulouse (Phung et al. 2020). This technique has made it possible to establish interaction networks of proteins involved in DNA maintenance and RNA metabolism in *P. abyssi*, respectively (Phung et al. 2020; Pluchon et al. 2013). Briefly, the method consists in fixing recombinant protein bearing a Histidine-Tag on nickel beads, previously purified to homogeneity (**FIGURE 42**). Subsequently, the purified recombinant protein is incubated with a cellular extract from a culture in the late exponential phase. The protein species of the extract interacting with the bait protein are eluted and identified by mass spectrometry (**FIGURE 43**). In our case, the recombinant protein *Tbar*-$(His)_6$-aLhr2 protein was expressed with an N-terminal Histidine-Tag from a pET15b expression vector in *E. coli*. After homogeneity purification on HisTrap and Size exclusion columns (**FIGURE 42**), 20µg of *Tbar*-$(His)_6$-aLhr2 protein was fixed on nickel beads to be used as a bait. The *T. barophilus* cellular extract is then added to the beads and incubated overnight at 4°C followed or not by a treatment at room temperature with a mix of DNase I and RNase A (**FIGURE 43**). The nuclease treatment allows eliminating a certain number of protein species of which the interaction with the bait protein is mediated by DNA or RNA**.** After washing in TK buffer, the eluate is deposited on an SDS-PAGE gel (short migration), then cut up and sent to the Paris Sud Ouest Proteomic Analysis Platform (PAPPSO) for identification of the associated protein species by mass spectrometry. In order to eliminate non-specific interactions, control experiments are carried out in parallel by incubating the cell

extract with the nickel beads in the absence of the bait protein. All of these experiments were carried out in triplicate. The MS/MS spectrometry data were analyzed according to an algorithm (established by S. Laurent, Ifremer) used to determine the partners specifically interacting with the bait protein. The "Spec, Index" value takes into account specificity (normalization with respect to control samples), the closer this value is to zero the better the specificity. The "Ref, Spectra" value reflects the amount of protein in each sample, i.e. the abundance of spectra of a given peptide, the higher the number, the greater the amount of peptide. The two values "Spec, Index" and "Ref, Spectra" make it possible to establish a ranking from the most specific to the least specific interactions (see the method in (Phung et al. 2020). For the mass spectrometry experiments with *Tbar*-$(His)_6$-aLrh2 as bait, nuclease and nuclease-free assays were run in triplicate (**FIGURE 43**). Because we could not detect any significant difference between conditions, we decided to treat the AP-MS nuclease and nuclease-free data as a unique set of data. Therefore, our data set includes 4 replicates including 2 controls (one of the 3 control experiments was excluded from the analysis since the number of spectra was not coherent to the overall set of controls).

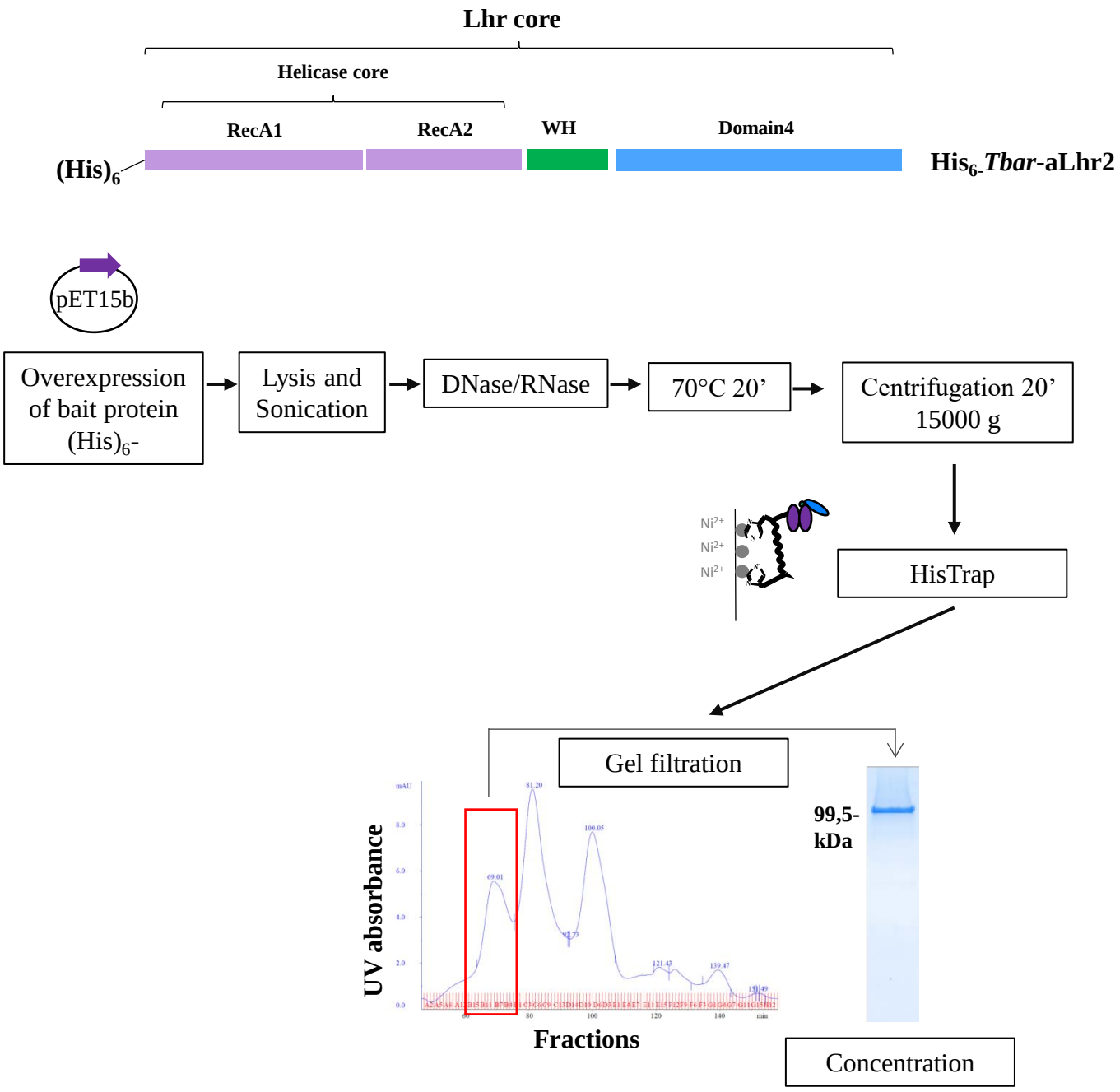


Figure 42. Expression and Purification of *Tbar*-(His)$_6$-aLhr2.
Schematic representation of the over-expression in pET15b and the two purification step of *Tbar*-(His)$_6$-aLhr2 using FPLC. The size-exclusion chromatogram and SDS-PAGE colored with Coomassie blue are shown.

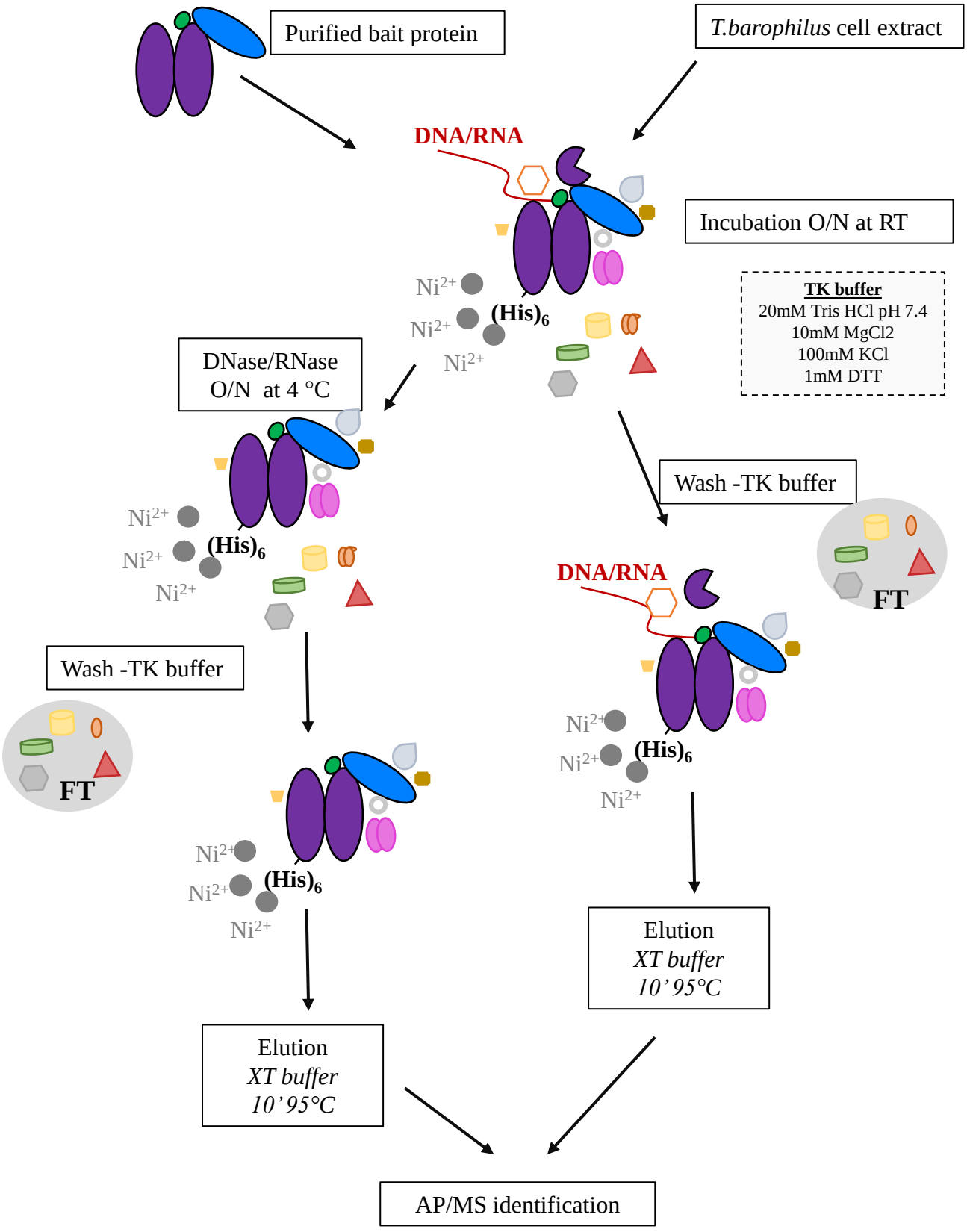


Figure 43. Pull-down assay.
Schematic representation of the "Pull-Down" technique using *Tbar*-His$_6$-aLhr2 as bait protein adsorbed onto cobalt beads, and *T.barophilus* cell extract (colored proteins) using TK buffer. This method provides a list of candidates identified by Mass spectrometry.

1-2 Overview of the interaction network of *Tbar*-$(His)_6$-aLhr2

Altogether, the pull-down assays performed with *Tbar*-$(His)_6$-aLhr2 identified 72 partner proteins with a significant score. The list of partners of this exploratory analysis from the most to the least specific is displayed in **Table 13.**

Co-purified partners with aLhr2 are shown in **FIGURE 44** with each protein category grouped according to central functions in physiological processes. Unexpectedly, 31 % of the proteins are annotated as ribosomal proteins from the small (30S) and large (50S) subunits (**Table 13**). Another 14% are annotated as implicated in RNA modification or reported as ribonuclease (RNase). Altogether partners related to RNA metabolism and translation compose 45 % of the aLhr2 proteome retrieved by this approach. Finally, more than 14 % of the partners are reported with functions in DNA replication and repair. From this distribution, we can propose that *T. barophilus* aLhr2 helicase is highly connected to the ribosome and also at the interface of the DNA and RNA metabolism Nevertheless, the percentage in each category should be taken with care since a significant number of *T. barophilus* proteins are of unknown function considering the current annotation of *T. barophilus* genome.

Table 13 List of potential proteins partners identified by pulldown-MS/MS with *Tbar*-aLhr2 as bait protein and *T. barophilus* cellular extract.

ID	Name	Description	ORF	Ref. Spectra	Spec, Index
F0LK05	TRAM	PRC domain-containing protein	TERMP_00565	0.95	0
F0LLF3	S6e	30S ribosomal protein S6e	TERMP_02008	0.41	0.10
F0LLL5	S17e	30S ribosomal protein S17e	TERMP_00815	0.51	0.11
F0LHC0	S4e	30S ribosomal protein S4e	TERMP_00104	0.37	0.13
F0LJ21	Elp3	tRNA acetyltransferase of elongator complex	TERMP_00390	13.06	0.14
F0LIY0	Peroxiredoxin	Alkyl hydroperoxide reductase sub. C-like	TERMP_00349	0.10	0.16
F0LLE1	RPA41	Replication A factor	TERMP_01996	3.11	0.19
F0LM78	RIO	Ribosomal biogenesis kinase	TERMP_00949	9.34	0.21
F0LGY9	ASCH	ASCH domain-containing isomerase	TERMP_01221	1.38	0.23
F0LHB2	L22	50S ribosomal protein L22	TERMP_00096	0.69	0.24
F0LMW5	L7Ae	50S ribosomal protein L7Ae (C/D & HACA guide RNP component)	TERMP_01118	2.42	0.27
F0LKY6		Unknown DUF460	TERMP_01918	7.10	0.28
F0LJ76	DNAPolB	DNA polymerase B	TERMP_01623	4.96	0.29
F0LJT8	TopoVI	DNA topo IV Toposimerase subunit A	TERMP_00498	10.93	0.32
F0LHR0	nrdF1	Brix-domain-containing-ribosomal biogenesis protein	TERMP_00167	0.58	0.33
F0LMR0	PINA	ATPase/ Holliday junction	TERMP_01063	1.75	0.33
F0LKF6		Unknown DUF504	TERMP_00637	1.37	0.36
F0LMS7	Peptidase	Peptidase/caspase catalytic subunit	TERMP_01080	2.75	0.39
F0LH70	L10e	50S ribosomal protein L10e	TERMP_00054	1.63	0.41
F0LHB7	S17	30S ribosomal protein S17	TERMP_00101	4.19	0.42
F0LKE3		Unknown	TERMP_00624	1.86	0.44
F0LKL3	DNA polD	DNA polymerase D	TERMP_01872	10.56	0.44
F0LI15	Spt5	Transcription initiation factor Spt5	TERMP_00192	0.50	0.44
F0LKH1	Topo I	DNA Topoisomerase I	TERMP_00652	0.25	0.44
F0LLE3	RPA32	Replication A factor	TERMP_01998	1.81	0.45
F0LHL4	L14e	50S ribosomal protein L14e	TERMP_00120	1.46	0.45
F0LJI2	Cas6	CRISPR endoribonuclease	TERMP_00471	1.19	0.50
F0LI18	L10	50S ribosomal protein L10	TERMP_00195	2.21	0.51
F0LJ19	S-layer protein	S-layer structural protein	TERMP_00388	0.26	0.52
F0LK43	Methyl-transferase	SAM-dependent methyltransferase	TERMP_01780	0.37	0.53
F0LHB0	L2	50S ribosomal protein L2	TERMP_00094	0.76	0.53
F0LHK7	L30	50S ribosomal protein L30	TERMP_00113	0.53	0.55
F0LN34	Metallophospho-esterase	Metallophosphoesterase-calcineurin superfamily	TERMP_01187	0.42	0.55
F0LMV8	PRC	PRC domain-containing protein	TERMP_01111	0.73	0.55
F0LI01	Peptidase	M48 family metallopeptidase	TERMP_01426	1.00	0.57
F0LI17	L1	50S ribosomal protein L1	TERMP_00194	0.97	0.58
F0LM50	EndA	tRNA splicing endoribonuclease	TERMP_00921	2.53	0.59
F0LLV5	rfbB	dTDP-glucose 4,6-dehydratase	TERMP_02080	9.72	0.60
F0LHC5	L32	50S ribosomal protein L32e	TERMP_00109	0.25	0.61
F0LHQ6	Rrp41	RNA exosome catalytic subunit	TERMP_00163	1.83	0.61
F0LI70	S7	30S ribosomal protein S7	TERMP_00247	1.04	0.61
F0LHK5	L18	50S ribosomal protein L18	TERMP_00111	1.14	0.62
F0LKC2	S10	30S ribosomal protein S10	TERMP_00603	0.99	0.65
F0LKG4	NUDIX	NUDIX-like hydrolase	TERMP_00645	1.03	0.69
F0LLG4	Taw1	Wyosine tRNA modification	TERMP_02019	4.69	0.70

F0LLE2	RPA14	Replication factor A subunit	TERMP_01997	1.15	0.71
F0LHF9		Uncharacterized protein	TERMP_01311	0.48	0.71
F0LKV2	S8e	30S ribosomal protein S8e	TERMP_00704	0.14	0.72
F0LL87	ArsR	HTH arsR-type domain-containing protein	TERMP_00761	0.68	0.73
F0LJX9	HydB	Sulfo hydrogenase II subunit b	TERMP_00539	3.38	0.74
F0LMB9	Proteasome	Proteasome subunit β2	TERMP_02168	0.66	0.74
F0LM28	S3Ae	30S ribosomal protein S3Ae	TERMP_00899	0.06	0.75
F0LM11	thpR	RNA 2',3'-cyclic phosphodiesterase	TERMP_02136	0.50	0.77
F0LKR2	ParB	Partition B protein	TERMP_00664	4.11	0.78
F0LHF4	Hydroxyl.-CoA synthase	hydroxymethylglutaryl-CoA synthase	TERMP_01306	3.49	0.79
F0LI69	S12	30S ribosomal protein S12	TERMP_00246	0.51	0.80
F0LLG0	S24e	30S ribosomal protein S24e	TERMP_02015	0.81	0.81
F0LJ08	Nop10	Box H/ACA pseudouridylation guide RNP	TERMP_00377	0.70	0.81
F0LJL0		Unknown UPF0216	TERMP_01677	0.29	0.82
F0LN06	XerA	DNA recombination repair	TERMP_01159	1.02	0.84
F0LKL8	NucS	DNA repair nuclease	TERMP_01877	0.19	0.85
F0LHY7	Lrp/AsnC	Transcription regulator factor	TERMP_01412	0.61	0.85
F0LMQ4	mce	Methylmalonyl-CoA epimerase	TERMP_01057	0.38	0.86
F0LHK4	L19	50S ribosomal protein L19e	TERMP_00110	0.20	0.86
F0LIK3	L37e	50S ribosomal protein L37e	TERMP_00300	0.47	0.87
F0LLT9	Nucleotidyl-transferase	NTP_transf_2_domain-containing protein	TERMP_02064	6.45	0.88
F0LJ25	L21e	50S ribosomal protein L21e	TERMP_00394	2.65	0.88
F0LHT9	HypA	Hydrogenase maturation factor HypA	TERMP_01364	0.07	0.91
F0LHM3	S11	30S ribosomal protein S11	TERMP_00130	1.31	0.92
F0LHM2	S4	30S ribosomal protein S4	TERMP_00129	0.43	0.95
F0LJC4	KaiC	RecA superfamily ATPase implicated in signal transduction	TERMP_01671	0.42	0.95
F0LKD3	oxidoreductase α	oxidoreductase subunit alpha	TERMP_00614	0.58	0.95

The identity (ID) and the description are from the databank Uniprot (UP000007478). The reference spectra number " Re.Spectra" reflects the peptide quantity and the index "Spec. Index" the specificity of the interaction.

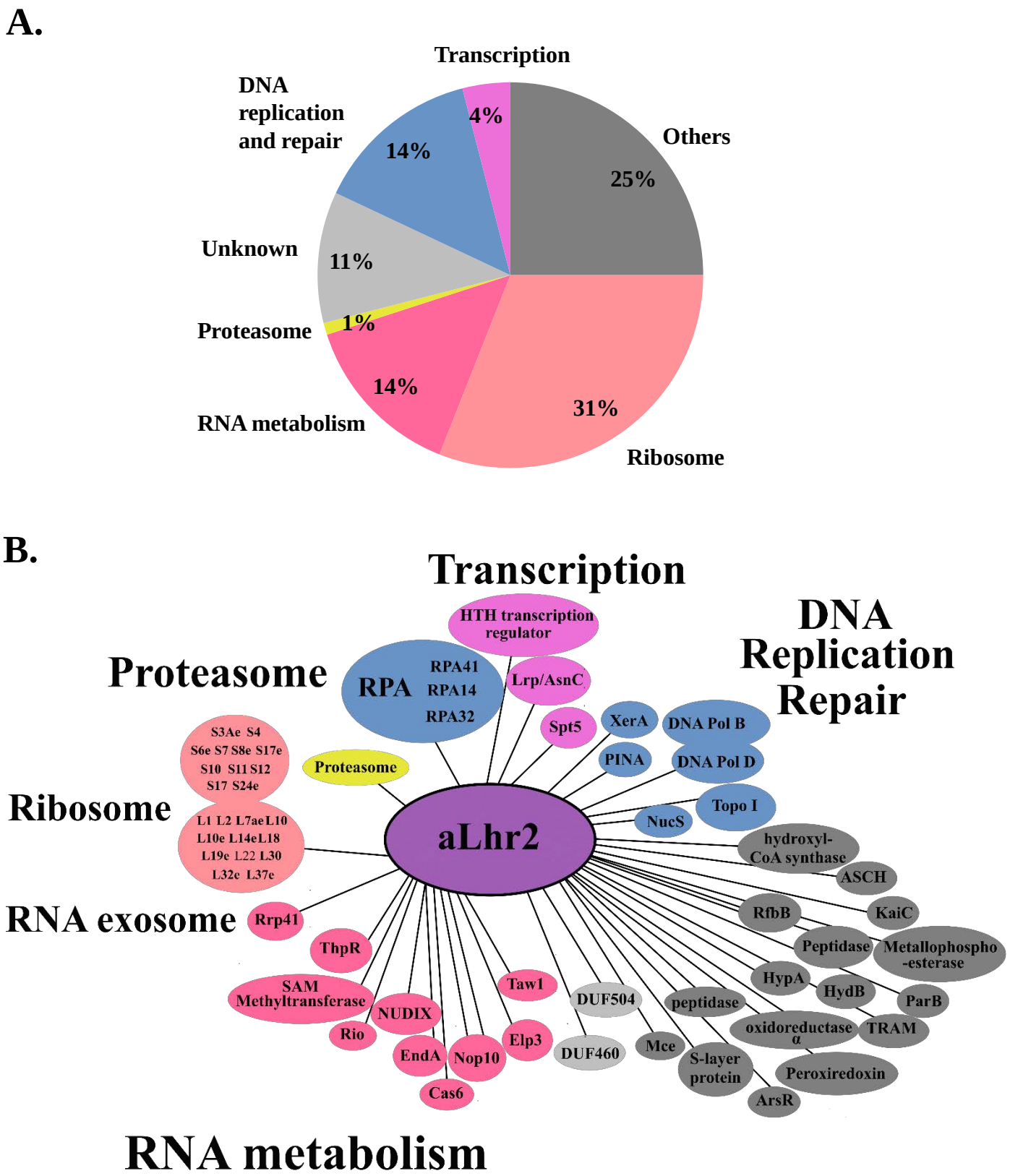


Figure 44. Interaction network obtained using *Tbar*-(His)$_6$-aLhr2 as bait protein.
A. The distribution according to the function of the candidates according to the current annotation of *T.barophilus* genome (Uniprot and NCBI): RNA metabolism in dark pink, DNA replication and repair in blue, ribosomal proteins in light pink, transcription in purple, and the proteasome in yellow, proteins of unknown function in light grey or involved in other processes in dark gray. **B.** The diagram of the interaction networks indicating the link with each partner with colours according to the protein function.

- **Factors involved in translation**

In total, 12 ribosomal proteins out of the 25 that constitute the 30S subunit and 13 out of the 39 that constitute the 50S subunit are significantly detected **(Figure 44B)**. We should note that the L7Ae protein, in addition to its function as a ribosomal protein, is also a component of the box H/ACA and C/D guide RNP complexes involved in pseudouridylation and 2'O-ribose methylation, respectively, of rRNAs and tRNAs (Clouet-d'Orval et al. 2018). In light of these data, we propose a cross-talk between aLhr2 helicase and the translating-ribosome which must be viewed in the context of the coupling between translation and transcription that exists in archaeal cells.

Table 14 List of 30S and 50S ribosomal proteins of *Tbar*-aLhr2 network.

ID	Name	Description	ORF
F0LLF3	S6e	30S ribosomal protein S6e	TERMP_02008
F0LLL5	S17e	30S ribosomal protein S17e	TERMP_00815
F0LHC0	S4e	30S ribosomal protein S4e	TERMP_00104
F0LHB7	S17	30S ribosomal protein S17	TERMP_00101
F0LI70	S7	30S ribosomal protein S7	TERMP_00247
F0LKC2	S10	30S ribosomal protein S10	TERMP_00603
F0LKV2	S8e	30S ribosomal protein S8e	TERMP_00704
F0LM28	S3Ae	30S ribosomal protein S3Ae	TERMP_00899
F0LI69	S12	30S ribosomal protein S12	TERMP_00246
F0LLG0	S24e	30S ribosomal protein S24e	TERMP_02015
F0LHM3	S11	30S ribosomal protein S11	TERMP_00130
F0LHM2	S4	30S ribosomal protein S4	TERMP_00129
F0LHB2	L22	50S ribosomal protein L22	TERMP_00096
F0LMW5	L7Ae	50S ribosomal protein L7Ae	TERMP_01118
F0LH70	L10e	50S ribosomal protein L10e	TERMP_00054
F0LHL4	L14e	50S ribosomal protein L14e	TERMP_00120
F0LI18	L10	50S ribosomal protein L10	TERMP_00195
F0LHB0	L2	50S ribosomal protein L2	TERMP_00094
F0LHK7	L30	50S ribosomal protein L30	TERMP_00113
F0LI17	L1	50S ribosomal protein L1	TERMP_00194
F0LHC5	L32	50S ribosomal protein L32e	TERMP_00109
F0LHK5	L18	50S ribosomal protein L18	TERMP_00111
F0LHK4	L19	50S ribosomal protein L19e	TERMP_00110
F0LIK3	L37e	50S ribosomal protein L37e	TERMP_00300
F0LJ25	L21e	50S ribosomal protein L21e	TERMP_00394

- **Factors involved in RNA metabolism**

Amongst protein partners, we detected the component of the barrel catalytic Rrp41 of the RNA exosome which is a 3'-5' RNA degrading machinery conserved in most Archaea (Clouet-d'Orval et al. 2018) is significantly detected. This is consistent with unpublished results of the team in Toulouse (Batista et al unpublished data) which shows that *Paby*-aLhr2 was detected in the interaction network of *Paby*-Rrp41. In addition, the RNA exosome machinery which is proposed to be of major importance in all aspects of RNA metabolism (Clouet-d'Orval et al. 2018) was shown to be in proximity of the ribosome by cellular fractionation assay performed with *P. abyssi* cell (Phung et al. 2020).

In this data set, we also observed many factors annotated as critical enzymes for tRNA maturation such as EndA, the tRNA endonuclease involved in the splicing of archaeal tRNA intron (Clouet-d'Orval et al. 2018) and several tRNA modification enzymes (**Table 13**). Importantly, components (Nop10 and L7Ae) of the box H/ACA pseudouridylation guide machinery are detected in this network. The box H/ACA guide machinery is critical for pseudouridylation of ribosomal RNAs but also of tRNAs in Archaea (Charpentier et al. 2007). Finally, we notice the presence of tRNA acetyltransferase and Wyosine (imG) tRNA modification enzyme which is responsible of the tRNA(Phe) at position 37 adjacent to the anticodon in Archaea and Eucaryotes (de Crécy-Lagard et al. 2010).

Finally, an endoribonuclease important for cellular immunity involving CRISPR RNAs is detected in aLhr2 network. The Cas6 enzyme is reported to process the CRISPR RNA precursors into small crRNAs that will be integrated into CRISPR effector RNP complex to protect the cell against infection of virulent genetic elements (Koonin and Makarova 2017).

- **Factors involved in transcription**

In aLhr2 network, factors described as Lrp/AsnC-like transcription regulation factors are observed (Lemmens et al. 2019)**.** The Lrp family is one of the best studied Transcription factor (TF) families in archaea The Archaeal Lrp-like regulators act locally or globally and regulate genes involved in amino acid metabolism, central metabolism and/or transport processes (Peeters and Charlier 2010). In addition the Stp5 factor which is a factor that regulates RNA polymerase elongation is identified in aLhr2 network. Stpt5/Spt4 complex is

reported as involved in the coupling of chromatin modification and RNA processing to transcription elongation (Werner 2012).

- **Factors involved in DNA repair and replication**

Several factors described as critical for DNA replication are part of aLhr2 network. Both DNA polymerases, PolD and PolB, which are responsible of the replication of DNA on the leading and lagging strand (Zatopek et al. 2018), respectively, are detected as well as and the DNA Topoisomerase I and IV which control the topology of DNA during the replication process. Finally, PINA which has been recently described as a novel ATPase highly conserved in Archaea, working together with Hel308 helicase to function in replication fork regression is also detected (Zhai et al. 2018).

Importantly the three subunits (RPA14/RPA32/RPA41) of the Replication A (RPA) complex are significantly retrieved in this network. RPA is a coated single-stranded DNA binding complex that contributes in maintaining open the replication bubble that has also been proposed to enhance transcription rate *in vitro* (Pluchon et al. 2013).

Some factors involved in DNA recombination and repair mechanisms are also part of aLhr2 network. NucS is an endonuclease with high specificity for DNA mismatches than for branched or single-stranded DNA and was shown to establish a specific physical interaction with the sliding clamp PCNA (proliferating cell nuclear antigen) (White and Allers 2018) and XerA is tyrosine recombinase propose to play a role in chromosome resolution (Cortez et al. 2010).

2-Transcriptomic analysis of aLhr2 helicase-lacking strain

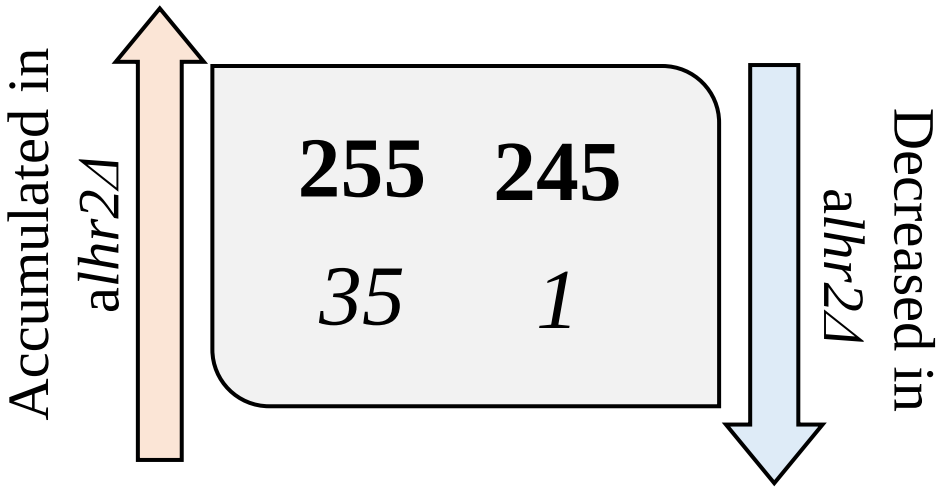
Note that this part of the work are preliminary analyses of *alhr2*Δ transcriptome which were performed in close collaboration with Marta Kwapisz (Team in Toulouse) and the bigA platform at the CBI (Toulouse) which performed statistical analyses.

Briefly, a strain deleted for *aLHR2* gene was constructed in collaboration with the team of M. Jebbar by using the PopIn/PopOut methodology (Birien et al. 2018; Thiel et al. 2014). Total RNAs were extracted from three independent exponentially grown wild type (WT) and *alhr2*-deleted (*alhr2*Δ) cultures and analysed by RNA-sequencing (RNA-seq). RNA-seq is a high-throughput sequencing assay, which permits the identification of transcripts and the quantification of gene expression. Pair-ended, sense-specific reads were aligned on *T. barophilus* genome and differential expression analysis was performed using DESeq. Amounts of transcripts were normalized to the wild type level and expressed as log2-FC (log2-fold change). We have arbitral chosen log2-FC higher than the absolute value of 0.8 as significantly changed.

Deletion of *aLHR2* gene has impacted the amounts of 500 coding transcripts out of 2268 identified in *T. barophilus* (25%). 255 transcripts accumulated in *alhr2*Δ strain and the amounts of 245 were significantly decreased (FIGURE 5A). This indicated that aLhr2 was implicated in the regulation of the cellular amounts of these transcripts. It is important to keep in mind that RNA-seq measures steady-state RNA levels, which is the output of RNA transcription and degradation rates. Thus, aLhr2 helicase could regulate transcription or degradation of identified transcripts directly as a transcriptional regulator or as a decay factor, alone or as a partner of other proteins and complexes. These two possibilities could not be discriminated on the basis of RNA-seq data as well as indirect effects of a*LHR2* deletion could not be excluded.

For 255 mRNA targets accumulating in aLHR2-deleted strain, aLhr2 acts likely as a transcriptional repressor or degrading factor. For 245 underrepresented RNAs, aLhr2 helicase could activate their transcription or act as a protecting factor *i.e.* target RNA is stabilized by aLhr2-binding. Deletion of *aLHR2* gene has also impacted non-coding RNAs; 35 out of 59 accumulated in *alhr2*Δ strain but only 1 was significantly decreased **(FIGURE 45A)**. This result suggested some specificity; aLhr2 would not bind non-coding substrates to protect them but would rather participate in their degradation.

A.



B.

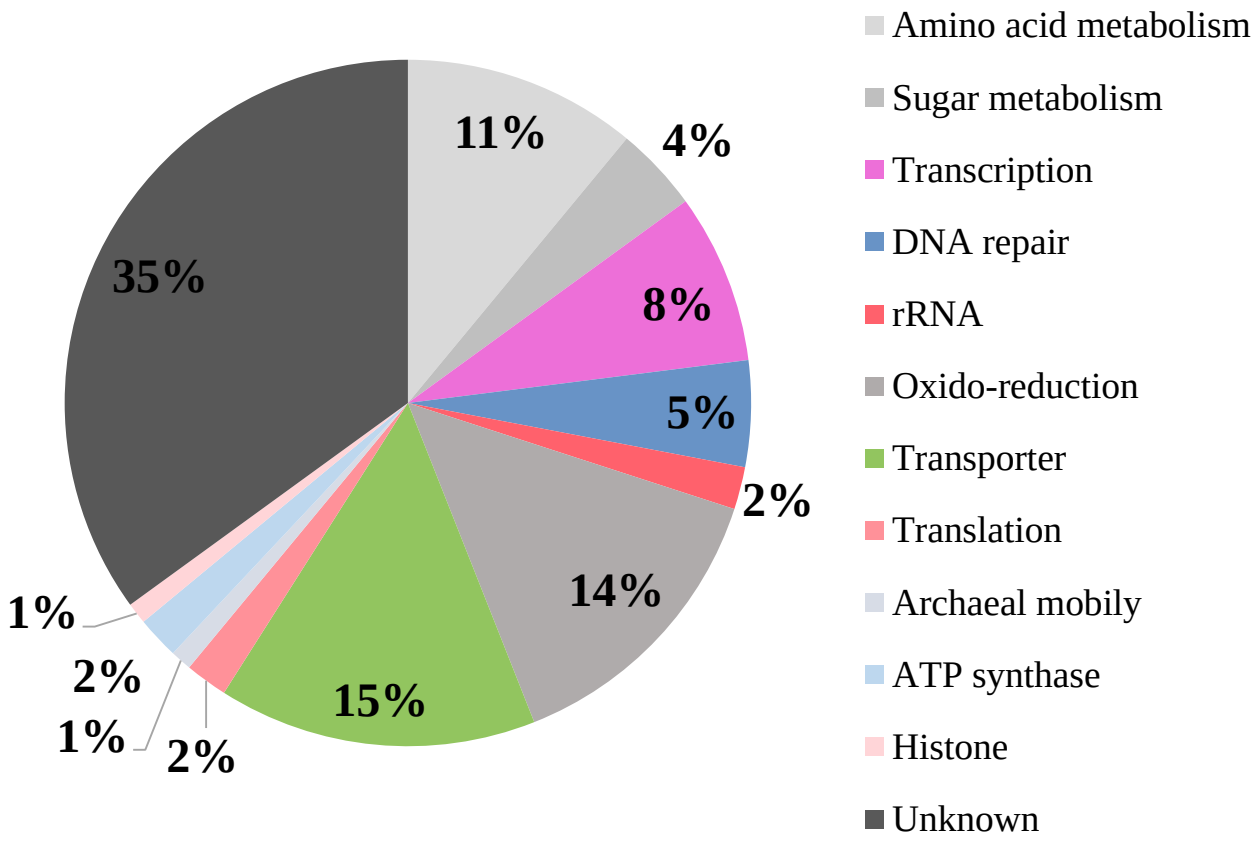


FIGURE 45. Changes in transcriptional profile of *aLhr2* helicase-lacking strain.
A. Quantity of **coding** (in bold) and *non-coding* (in italic) transcripts accumulated (pink arrow) or decreased (blue arrow) in *T.barophilus* strain deleted for *alhr2* gene. **B.** Pie chart showing the relative distribution of mRNAs deregulated in aLhr2 mutant. Numbers indicate the percentage of the implicated function.

Detailed analyses of mRNAs deregulated in *aLhr2* mutant are shown in **FIGURE 45B**. The most abundant fraction of deregulated transcripts (35%) belongs to the category of unknown proteins. Gene Ontology (GO) analysis has shown significant changes for 5 GO classes: IMP metabolism (GO:0006189), structural components of the ribosome (GO:0003735), ribosome (GO:0005840), translation (GO:0006412), and rRNA binding (GO:0019843). More than 25% of transcripts deregulated are related to RNA metabolism (translation & translation). This indicates that in the absence of aLhr2 helicase, the translational apparatus, including ribosomal proteins and translation factors, is potentially affected.

3-Conclusions

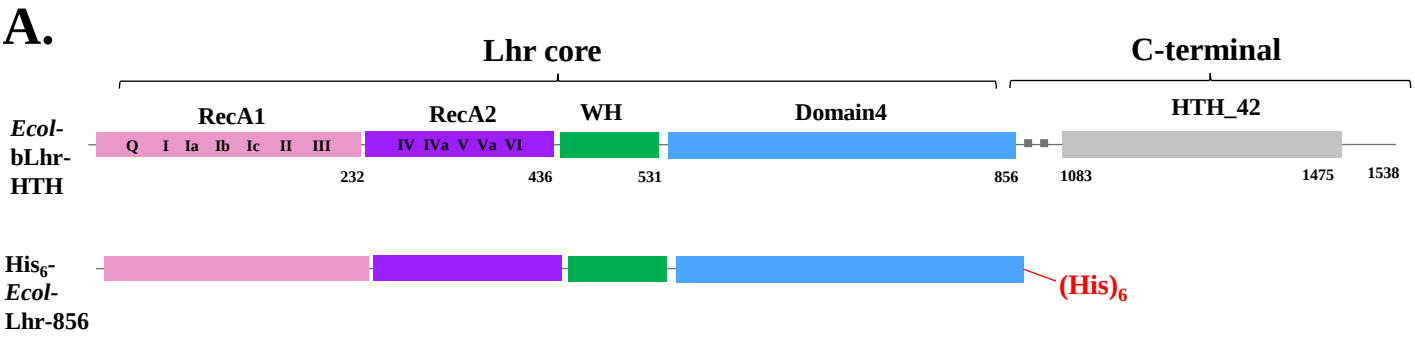
The interaction network of *Tbar*-aLhr2 indicated that aLhr2 cross-talk with factors involved in fundamental pathways as maintaining the integrity of the genome (DNA repair/replication) and regulating gene expression (ribosome & RNA processing/decay) in the cell. Interestingly, the aLhr2 interactome showed also that ribosomal proteins are among the major aLhr2-interacting proteins **(Table 14)**. In coherence, transcriptomic analyses also highlight a relationship between aLhr2 and the translation (ribosome & translation factors). Taken together, these results strongly suggest that aLhr2 helicase cross-talk with pathways connected to translation or/and ribosome biogenesis.

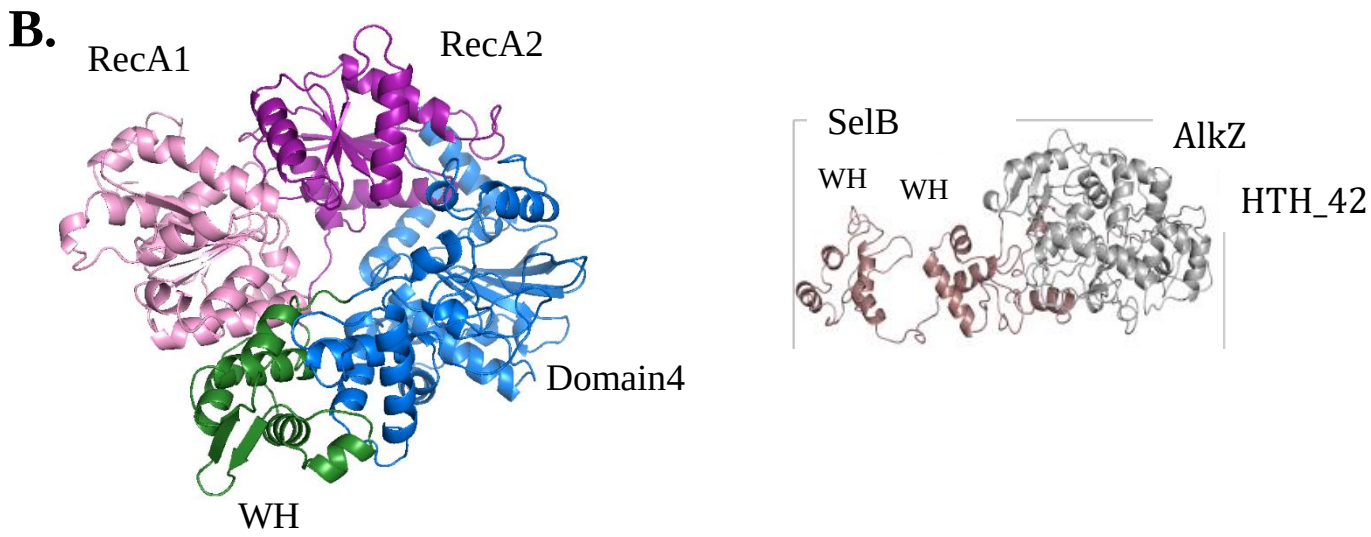
IV-Functional study of *Ecol*-Lhr

1-Constructions of pET21b-*Ecol*-bLhr-856 expression vectors

E.coli possesses one Lhr-type encoding gene that belongs to the bLhr-HTH family. This family harbours the Lhr core (from 1-856) composed of a winged-helix (WH) domain (from 437-531) and Domain 4 (DEAD_associated_domain 532-856), and a typical C-terminal domain (from 857-1538) found only in bLhr-HTH and some members of aLhr2 families **(FIGURE 46A)** (see RESULT-PART I). The Pfam analysis of this region followed by MAST research revealed the presence of a conserved HTH_42 domain. Homologous structures using DALI search resulted in a structure in the PDB database with 19% of identity that corresponds to the DNA glycosylase AlkZ of *Streptomyces sahachiroi* (*Ssah*; PDB ID: 5UUJ) and that exhibits a channel formed by a cluster of three WH motifs in tandem and by a short C-terminal β-hairpin highly accommodated for binding dsDNA (Mullins et al. 2017) **(FIGURE 46B)**. We also performed modelling of the structure of the region between Domain 4 and the HTH_42 domains using Phyre2. Using DALI, we found that this region resembles with 21% of identity to the structure of the SelB elongation factor of *Moorella thermoacetica* (PDB1LVA), which contains two winged-helix domains involved in RNA binding **(FIGURE 46B)**. In *E.coli* genome and 82% of Proteobacteria, the bLhr-HTH encoding genes are followed by the *rnt* gene encoding the ribonuclease RNase T (Reuven et al. 1995) **(FIGURE 46C, Figure 23)**. An interaction between *Ecol*-Lhr and *Ecol*-RNase T has been demonstrated at the transcriptional level, as the Lhr and RNase T are co-transcribed, but at the protein level this interaction was not established neither *in vitro* nor *in vivo.*

.





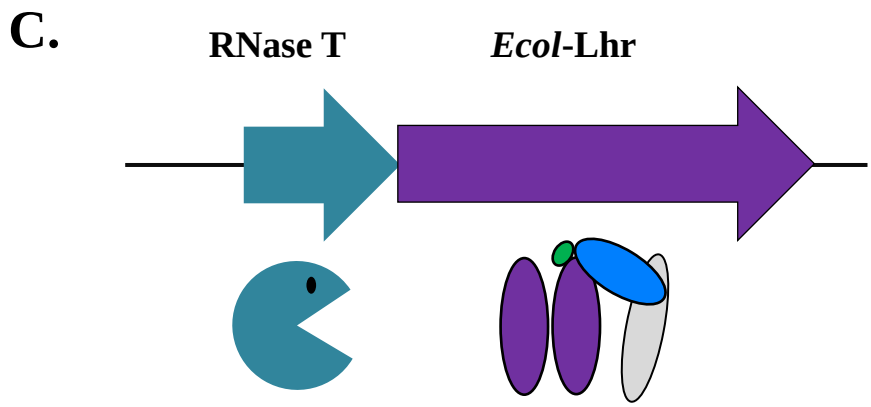


Figure 46. bLhr-HTH of *E.coli.*
A. Domain organization of bLhr-HTH and bLhr-HTH-856-His_6 (Lhr core) of *Ecoli*. **B.** Structure of *Ecol*- bLhr-HTH predicted using Phyre2 in cartoon representation. The Lhr core comprises the RecA1 (pink), RecA2 (purple), WH (green), and Domain 4 (blue). The C-terminal region is formed by the HTH domain (grey), and SelB-like domain (dirty pink). **C.** Genetic cluster of *Ecol*- bLhr-HTH and *Ecol*-RNaseT.

RNase T is a member of the large DEDD family of exonucleases important for the final step of the maturation of many stable RNAs **(FIGURE 47A)**. RNAse T carries out the 3′ processing of most bacterial precursor tRNAs by exonucleolytic trimming to yield a mature CCA end (Zuo et al. 2007) **(FIGURE 47B)**. Moreover, RNase T specifically trims the 3′ end of structured DNA, including bulge, bubble, and Y-structured DNA, and is critical for *E.coli* resistance to various DNA-damaging agents and UV radiation (Hsiao et al. 2014) **(FIGURE 47C)**. The presence of protein interaction between RNaseT and *Ecol*-bLhr-HTH place this helicase at the interface of DNA repair and RNA metabolism which is coherent with what we suggest for *Tbar*-aLhr2 (see RESULT-PART III).

In order to define a protein interaction between *Ecol*-Lhr and *Ecol*-RNase T proteins in *E.coli*, we settled up the heterologous overexpression of *Ecol*-Lhr-856 lacking the C-terminal extension and *Ecol*-RNase T recombinant proteins through the pET and pGX expression system respectively in *E. coli* BL21 DE3 codon+ strain with the aim to purify each of the proteins separately and study their interaction and complex assembly in vitro

Therefore, we set out to construct expression vectors to obtain recombinant proteins corresponding to *Ecol*-His_6-Lhr-856 that comprise the helicase core, the WH, and the Domain 4**.** Briefly, the encoding sequence of the first 856 amino-acids of JW1645 protein (Lhr core) were amplified by PCR and inserted into the pET21b vector to obtain pET21b-*Ecol*-Lhr-856 (**Table 8**). For details, refer to the section "Materials and Methods".

For proteins overexpression, we used the same conditions of expression setup for *Tbar*-aLhr2 (except for the optical density of the cultures at the time of IPTG induction). Briefly, after heat shock transformation of BL21-CodonPlus (DE3) *E.coli* strains with pET/pGX vectors, the overexpression of recombinant proteins was induced in exponential growth (OD_{600nm} of 0.4-0.6) by addition of IPTG (0.1 mM), for 3 hours at 30°C which resulted in relatively high protein expression (**FIGURE 48A**).

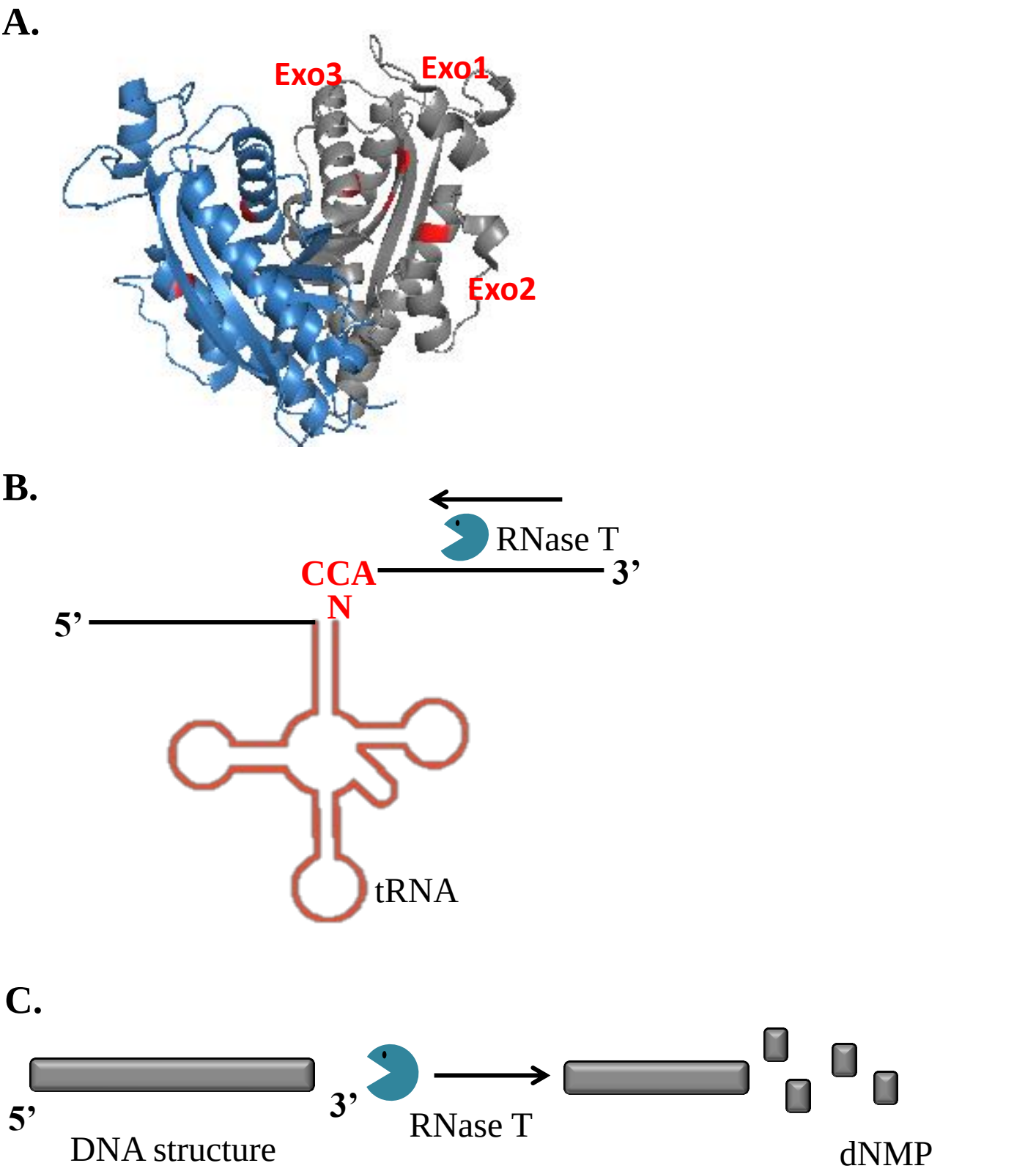


Figure 47. RNAse T of *E.coli.*
A. Crystal structure of *E. coli* RNase T (PDB2IS3) in a cartoon representation, with the two subunits in each dimer coloured differently. Red colour indicates exonuclease sites. **B.** 3’-5’ Exonucleolytic activity of RNase T in tRNA processing pathways in *Escherichia coli*. Exonucleolytic digestion is represented by black arrows parallel to the transcript. **C.** Rnase T catalyzes the removal of nucleotides from linear ssDNA or RNA or DNA structures in the 3' to 5' direction.

2-Optimization of Co-purification experimental conditions

Initially, the aim of our work was to purify each of the proteins separately and then study their interaction in vitro after assembly. However, we initiated some co-purification experiments to test for a possible interaction between the proteins in cellular extracts **(FIGURE 46C)**, A protocol for co-purification on affinity chromatography on Ni-NTA beads using a bait/prey system of recombinant proteins from *E.coli* was therefore tested **(FIGURE 48B)**. The *Ecol*-His_6-Lhr-856 bait protein is expressed from pET21 expression vectors to obtain a protein with an N-terminal $(His)_6$-tag. The *Ecol*-GST-RNAse T prey protein is expressed from the pGEX-41T expression vector. The prey/bait proteins are expressed independently in *E. coli* cell culture extracts **(FIGURE 48B)**. Cellular extracts of *E. coli* culture which expressed either the bait or prey protein are lysed and sonicated. After a centrifugation step, the soluble fraction of *Ecol*-His_6-Lhr-856 is then incubated with His60 Ni Resin beads which retain the bait protein with the $(His)_6$ tag **(FIGURE 48B)**. After a wash step, the soluble fraction containing *Ecol*-GST-RNAseT was added to the His60 Ni Resin beads. The low concentration Imidazole wash steps remove non-specific binding contaminants. The *Ecol*-His_6-Lhr-856 protein was then eluted using two imidazole concentration (250 and 500mM) potentially with the *Ecol*-GST-RNAseT prey protein if the two proteins interact under these experimental conditions **(FIGURE 48B)**. The protein content of each fraction collected is analyzed on an SDS-PAGE gel and stained with coomassie blue.

Our preliminary results show the presence of very weak bands that could correspond to the molecular weight of *Ecol*-Lhr-856 (94KDa) and *Ecol*-RNAseT (22.5 KDa) in addition to many other contaminants in the elution fractions **(FIGURE 49)**. The identity of these bands and whether they correspond to our proteins should be determined by band excision from the gel followed by mass spectrometry

Many optimizations steps should be performed to confirm or rule out this interaction. It might be more suitable indeed if the two proteins are purified first in Ni and GST columns, and the interaction could then be assessed with or without the cleavable tags. The presence of the tags could prevent indeed this interaction. A full length version and different variants of Lhr should be also tested for the interaction with the RNase as the Lhr-856 protein used is a truncated version without the C-terminal. It is possible that the C-terminal domain mediates in the interaction with RNaseT.

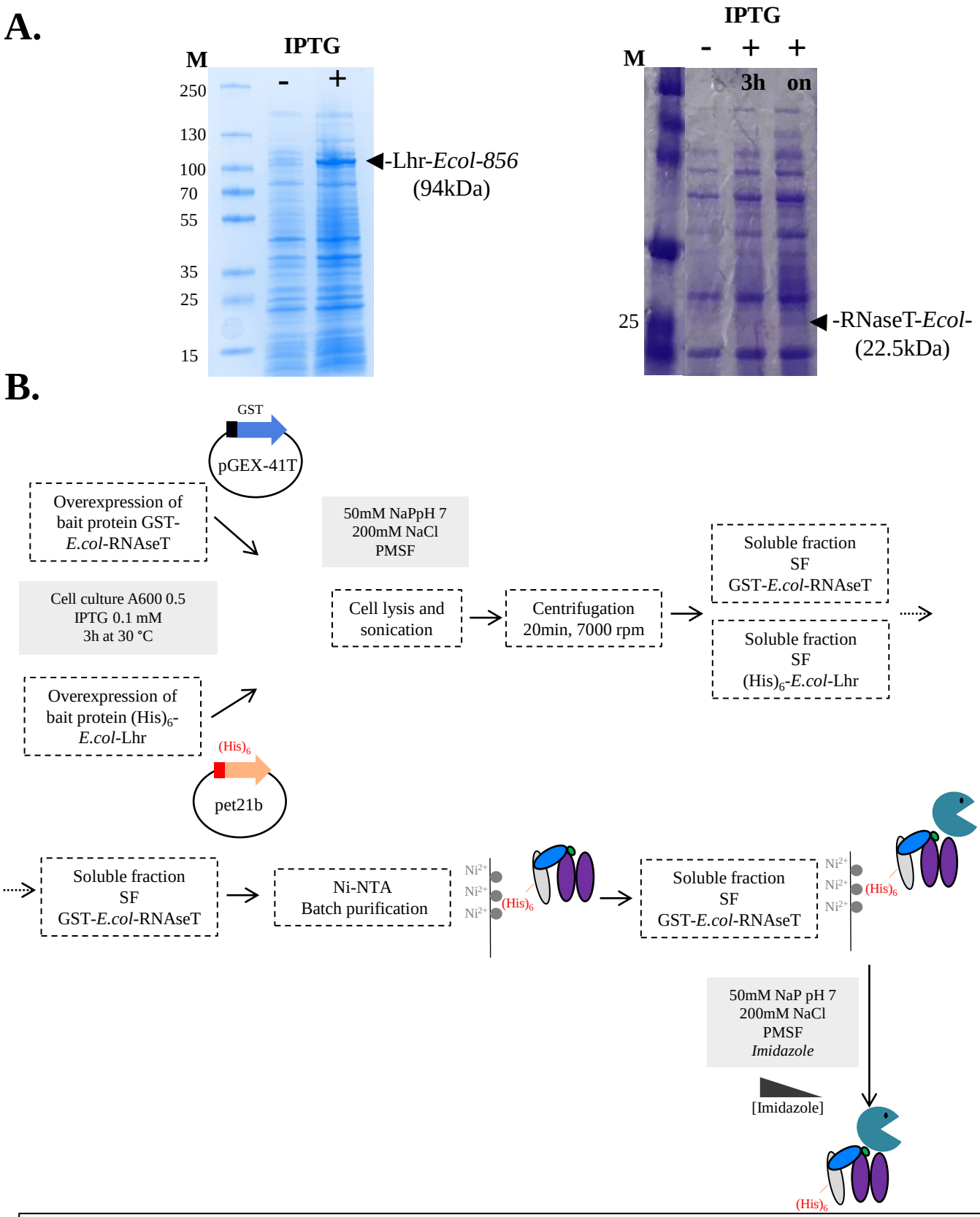


Figure 48. Co-purification assay.
A. Confirmation of recombinant protein expression in pET21b/pGX-BL21 (DE3) system on SDS-PAGE gel (4-15%) colored with Coomassie blue. Overexpression of recombinant *Ecol*-Lhr-856 and *Ecol*-RNAseT were induced with 0.1 mM IPTG in *E. coli* cells in mid-phase growth (OD600 of 0.4-0.6) in LB broth, at 30°C, with agitation at 180 rpm, for 3 h or overnight (on) (10 µL of each fraction was loaded per lane). **B.** Co-purification experiment workflow on His60 Ni Resin beads using a bait/prey system of recombinant proteins of *E.coli*. The different buffers are indicated and described in the Materials and Methods section.

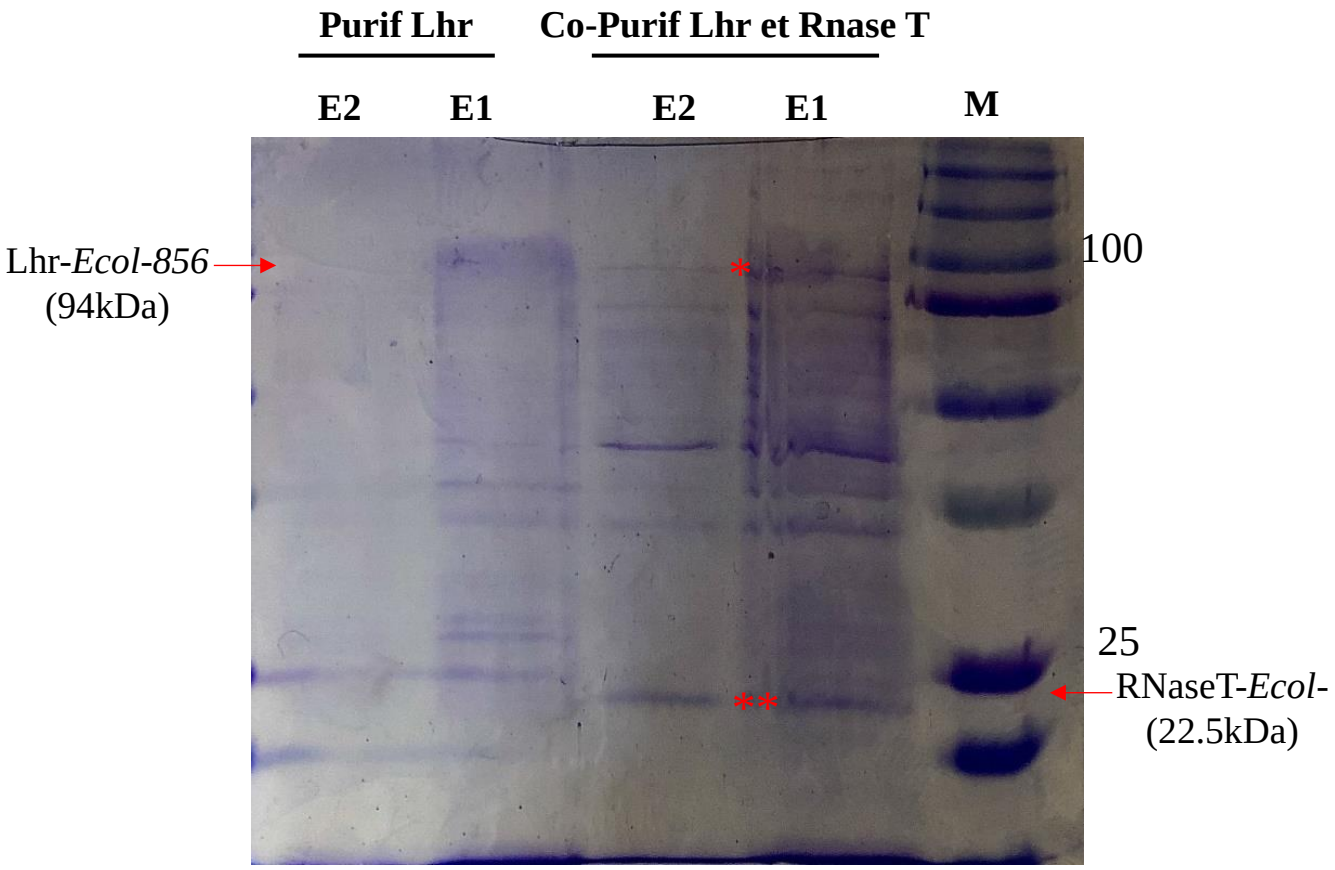


Figure 49. Co-purification of *Ecol* bLhr-HTH-856 and RNaseT.
Ecol bLhr-HTH-856 (tagged $(His)_6$), was used as bait protein. Fractions were analysed on SDS PAGE colored with Coomassie blue. E1 and E2 are the elution with 250 mM and 500 mM of imidazole respectively; Putative bands of RNAse T are indicated with ** and of Lhr with *.

Discussion

Lhr-type family are SF2 helicases exclusively found in prokaryotes and detected in almost every archaeal genome sequenced to date. Our work provided new insights into the classification and molecular functions of the Lhr-type proteins.

We have first provided a review on the state of the art of SF1 and SF2 helicases in archaea (Introduction PART II-6 published book chapter; Hajj et al, 2019).The published book chapter shed light on the structure-function of these helicases in archaea, presented their biochemical and mechanistic properties and highlighted their contribution at the physiological level in different cellular processes.

We have explored the role of Lhr-type proteins in cellular metabolism using different complementary approaches. Starting with phylogenomic approach, we defined five distinct orthologous groups of the Lhr-type helicases, two clusters were found in archaea named aLhr1 and aLhr2 and three clusters found in Bacteria, named bLhr, bLhr-HTH, and Lhr-like (RESULTS-Part I). In the context of this thesis, we focused our work to elucidate the contribution of Lhr-type helicases in cellular mechanisms by studying aLhr2 of the hyperthermophilic archaea *T. barophilus* (*Tbar*-aLhr2) and bLhr-HTH of *E.coli*. .

Through enzymatic assays, aLhr2 from *T. barophilus* was identified as a helicase able to unwind RNA/DNA hybrids (RESULTS-Part II). In addition, the interaction network of *Tbar*-aLhr2 suggests a cross-talk between aLhr2 helicase and critical players of both RNA and DNA metabolism (RESULTS-Part III). We can propose that *Tbar*-aLhr2 helicase is at the nexus of critical DNA and RNA metabolic processes and could participate in the coordination of these pathways during the cell cycle. In line, we examine by co-purification assay the functional link between bLhr-HTH and RNase T of which encoding genes form a gene cluster in *E.coli*. (RESULTS-Part IV). RNAse T has been also shown to be either involved in RNA and DNA metabolism (Zuo et al. 2007; Hsiao et al. 2014).

1-Landscape of Lhr-type proteins the three domains of life

Lhr-type proteins are SF2 helicases found in archaea and bacteria but not in eukaryotes **(FIGURE 50)**. Even if was reported a PDB structure of the helicase core of the eukaryotic DDX52 helicase overlapping the structure of *Mthe*-aLhr2 helicase core (Buckley et al. 2020) our phylogenomic search did not identify any counterpart of Lhr-type proteins in yeast and human genomes.

From the topology of the phylogenetic trees constructed with our collection of Lhr-type protein (RESULTS-PART I; **FIGURE 50**), it is difficult to picture a simple evolutionary scenario that will help in understanding the emergence of Lhr-type helicases in archaea and in bacteria, respectively. Probably early gene duplication and lateral gene transfer events from archaea to bacteria or vice versa may have taken place since archaeal aLhr2 and bacterial bLhr-HTH and bLhr groups appear to emerge from a unique ancestor (**FIGURE 50**).

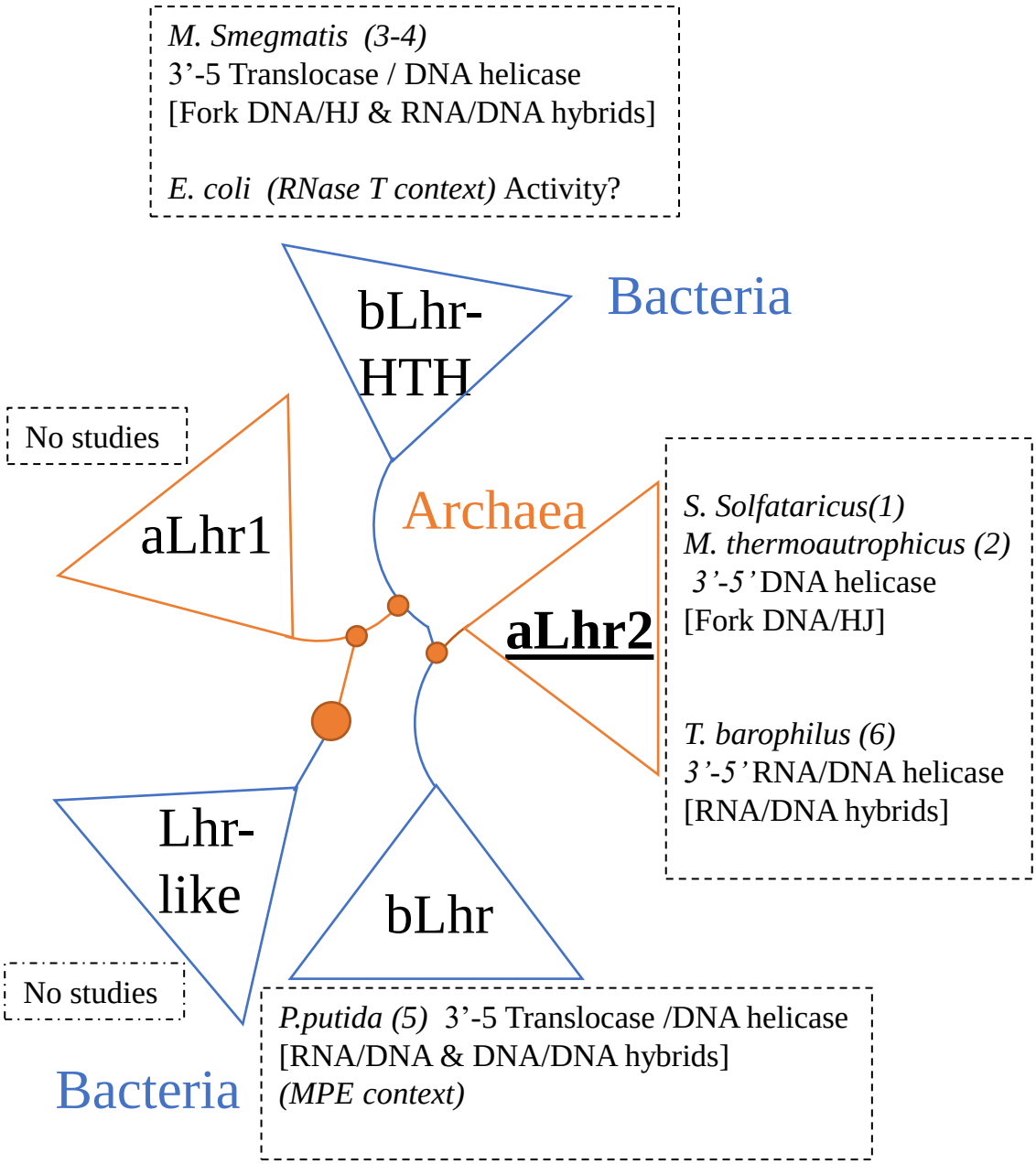


FIGURE 50. enzymatic activity and gene cluster of Lhr-type families.
Schematic representation of the phylogenetic tree obtained in this study. For each Lhr-type family was indicated the determined enzymatic activities *in vitro* and the putative function of studied Lhr-type member in archaea (orange) and bacteria (blue).
(1) De Felice et al 2007; (2) Buckley et al 2020; (3) Ordonez and Shuman 2013; (4) Ejaz and Shuman 2018; (5) Ejaz et al 2018 (6) Hajj et al, in preparation.

Our study shows that Lhr-type proteins are ubiquitous in archaea with two groups, aLhr1 and aLhr2, with at least one representative of Lhr-type in each archaeal genome suggesting a fundamental role in cellular processes (RESULTS-PART I; Manuscript -Hajj et al, in preparation) . Our results define Hel112 of *S.solfataricus* (De Felice et al. 2007), Lhr of *M. thermautotrophicus (Buckley et al. 2020)* as members of the aLhr2 group. *S.solfataricus* genome encodes one aLhr2 and two aLhr1 type-helicases whereas *M. thermautotrophicus* and *T.barophilus* genomes harbour only one gene of each group. To date, no study has yet reported the cellular function or enzymatic activity of archaeal aLhr1 helicases (**FIGURE 50**). However, in the case of aLhr2, several studies focusing on DNA repair in *S. solfataricus* and *M. thermautotrophicus* proposed a role in the resolution of holiday junctions or/and at DNA forks (Buckley et al. 2020; De Falco et al. 2018) (**Table 15**). Interestingly, previous work of the group in Toulouse identified also aLhr2 as part of the network of critical players of the RNA metabolism in Thermococcales (Phung et al. 2020). Altogether from this set of data, it is still not clear what is the specific role of aLhr2-type helicases in archaeal cells.

Three distinct families of Lhr-type are identified in bacteria of which one harbours a longer C-terminal extension carrying at least two conserved domains **(**RESULTS-Part I **FIGURE 23**). *P.putida* genome encodes both bLhr and bLhr-HTH proteins while *M.smegmatis* and *E.coli* genome encode only one Lhr-type (**FIGURE 50**). It is proposed from reported in vitro activities of these enzymes that they are part of DNA repair /recombination pathways. The occurrence of Lhr-type was found to be more erratic in bacterial genomes (RESULTS-Part I). In addition, Lhr-type enzymes from *P.putida* and *M. smegmatis* were identified as ATP-dependent ssDNA translocase with the 3’ to 5’ directionality (Ejaz and Shuman 2018; Ordonez and Shuman 2013; Ejaz et al. 2018) (**Table 15**) (**FIGURE 50**). Bacterial Lhr-type also differs in terms of gene cluster (RESULTS-Part 1-**FIGURE 23**). *Pput*-bLhr as for most of bLhr encoding genes is part of a cluster which group a gene encoding a metallophosphoesterase endonuclease (MPE) and a ligase. However, protein interaction between *Pput*-bLhr and *Pput*-MPE was not yet being shown. In addition, bLhr-HTH from *E.coli* and almost all from proteobacteria are in a gene cluster with RNase T encoding gene (RESULTS-Part I). RNase T has been shown in tRNA and rRNA maturation in bacteria, as well as in DNA repair pathways. In this context, we tested if *E.coli* bLhr-HTH and RNase T could directly interact *in vitro* (Zuo et al. 2007; Hsiao et al. 2014). (RESULTS-Part IV).

Table 15 Biochemical activities of studied Lhr-type proteins.

		aLhr2				bLhr-HTH	bLhr
		T.bar	***S.sol***		***M.the***	***M.sme***	***P.put***
Enzymatic activities		Monomer ?	Monomer	Dimer	ND	Monomer	Monomer
	ATPase	DNA/RNA-dep	DNA-Independent		ND	-DNA/RNA-dep -ATP/dATPase -calcium	-DNA-dep -ATP/dATPase -Manganese
		-K_M 23,26 nM -K_{cat} 1,27 min^{-1}	ND		ND	-K_M 10^5 nM -K_{cat} 366 min^{-1}	-K_M 37.10^4 nM -K_{cat} 198 min^{-1}
	Polarity	3'-5'	3'-5'		3'-5'	3'-5'	3'-5'
	Tranlocase	n.d	n.d		n.d	Yes	Yes
	Binding	-ssDNA (K_D 58 nm) -ssRNA (K_D 80 nm)	-Fork -3′-dsDNA	-Fork -3′dsDNA	Fork DNA	-SsDNA: (K_D 11 nm)	-SsDNA (K_D 44 nm)
	Unwinding	-3’ dsRNA -3’DNA:RNA	-fork -3’dsDNA -HJ		-3’dsDNA -3'RNA:DNA -HJ/Forked	-3’dsDNA, -Fork -3’DNA/RNA	-3'RNA:DNA -3’dsDNA
	Annealing	-3’tail RNA -3’tail DNA -3’DNA:RNA -Reduced by ATP	-Fork -HJ	-Fork	ND	ND	ND
	REF	This Work	(1)		(2)	(3-4)	(5)

(1) De Felice et al 2007; (2) Buckley et al 2020; (3) Ordonez and Shuman 2013; (4) Ejaz and Shuman 2018; (5) Ejaz et al 2018.

While the SF2 core of Lhr-type members shows highly conserved motifs across the five Lhr groups identified in this study, we observed many divergences in Domain 4 and WH motifs of the Lhr core except for a highly conserved Trp residue (RESULTS-Part I- **FIGURE 22**). The specific conserved residues in each group may confer definite properties as interaction with specific substrates or protein partners. In particular, the C-terminal domain of aLhr1 harbors four highly conserved cysteine residues that could form a zinc finger motif. Therefore it is possible that Lhr-type proteins have specific functions according to the group they belong to.

It is important to mention that for both archaea and bacteria, the genes encoding Lhr-type enzymes are not essential in optimal growth conditions. Indeed, null mutation in *T. barophilus* (this work), *S.islandicus* (Song et al. 2016) identify aLhr2 as non-essential genes. This is also the case in bacteria, *E.coli* and *M.tuberculosis*, for bLhr-HTH encoding genes (Song et al. 2016; Rand et al. 2003; Boshoff et al. 2003; Reuven et al. 1995). Therefore the absence of Lhr-type activity can certainly be compensated in the cell. In this context, it is possible that for example, aLhr1 can compensate the absence of Lhr2 in some archaea.

2-Is archaeal aLhr2 enzymes acting on DNA or RNA hybrids?

To further gain insight into the function of Lhr-type enzymes, we focused on *Tbar*-aLhr2 *in vitro* activity. *Tbar*-aLhr2 has an ATPase activity carried by the SF2 catalytic core composed of the RecA1 and RecA2 domains. In absence of ATP, *Tbar*-aLhr2 has the capacity to anneal single-stranded nucleic acid substrates. In presence of ATP a conformational change in the protein is proposed to operate with the ATP molecule acting as a molecular switch from a strand-annealing to an unwinding mode **(FIGURE 51A)**.

Our results indicate that the conserved Domain 4 of *Tbar*-aLhr2 is important for a full ATPase activity (RESULTS-Part II-**FIGURE 31)**. The Trp577 (Domain 4) and the I512 (WH) amino-acid residues are highly conserved across Lhr-type families (RESULTS-Part II-**FIGURE 22)**. These residues in *Msme*-bLhr-HTH are important for ATPase activity and the coordination between ATP hydrolysis and the mechanical activity (Ejaz et al. 2018). Our study shows that the alteration of these residues reduce the ATPase activity but do not affect nucleic-acid binding capacity (RESULTS-Part-II-**FIGURE 31, FIGURE ANNEX 7)**. The Domain 4 may help the enzyme core in adopting an optimal conformation for ATP hydrolysis. However, we still need to test the strand annealing/unwinding activities of corresponding protein variants to conclude of the importance of these residues on unwinding/ strand annealing reactions. In

addition, more advanced enzymatic activities of *Tbar*-aLhr2 helicase would be necessary to dissect in-depth mechanistic elements important of unwinding and strand annealing reactions, respectively. For example, picturing the protein- nucleic acid complexes with single-stranded or hybrid substrates by gel electrophoresis mobility shift assay (EMSA) will help to determine the DNA binding properties**.** It is also important to examine if *Tbar*-aLhr2 can displace proteins from nucleic acids while translocating along with the DNA/RNA substrates, for example, using the biotin-streptavidin complex assay as in the case of *Msme*-bLhr-HTH and *Pput*-bLhr (Ejaz and Shuman 2018; Ordonez and Shuman 2013) **(Table 15)**

In addition, *Tbar*-aLhr2 unwinds more efficiently *in vitro* hybrids with a 3' overhang strand indicative of a 3' to 5' polarity (RESULTS-Part II-**FIGURE 37)**. This is consistent with the polarity previously observed for other archaeal and bacterial counterparts (Buckley et al. 2020; Ejaz and Shuman 2018; Ordonez and Shuman 2013; Ejaz et al. 2018; De Felice et al. 2007) **(Table 15)**. Nevertheless, the main substrates used in the mechanistic study of *Ssol*-aLhr2 and *Mthe*-aLhr2 are DNA fork and Holliday junction type structures **(FIGURE 51B) (Table 15)**. As *Tbar*-aLhr2, *Mthe*-aLhr2 unwind RNA/DNA duplex but with less efficiency (**FIGURE 51B**). We should note that the structure of the RNA/DNA duplex used for testing the helicase activity of *Mthe*-aLhr2 is different from that used in this study. The RNA/DNA unwinding activity of *Ssol*-aLhr2 was not yet tested. *Ssol*-aLhr2 shows a nucleic-acid independent ATPase activity while *Tbar*-aLhr2 is unable to hydrolyse ATP in absence of nucleic acids **(Table 15)**. *Ssol*-aLhr2 has no affinity for ssDNA substrate while *Tbar*-aLhr2 shows a high affinity for such substrate. Both *Ssol*-aLhr2 and *Tbar*-aLhr2 were tested for their *in vitro* strand-annealing activity but using a different set of substrates, 3'tailed hybrids and DNA fork and Holliday Junction structure, respectively. Although our preliminary results suggest a monomeric form of *Tbar*-aLhr2, *Ssol*-aLhr2 have both monomeric and dimeric form that were shown to have specific biochemical activities (Buckley et al. 2020; De Felice et al. 2007) **(Table 15)**. Because the experimental conditions and the substrates used are specific to each study, it is difficult to compare and conclude if aLhr2 proteins have the exact same biochemical properties in all archaeal organisms. It is possible that enzymes from Euryarchaea and Crenarchaea do not have the same biochemical characteristics and activities and may be part of different metabolic pathways.

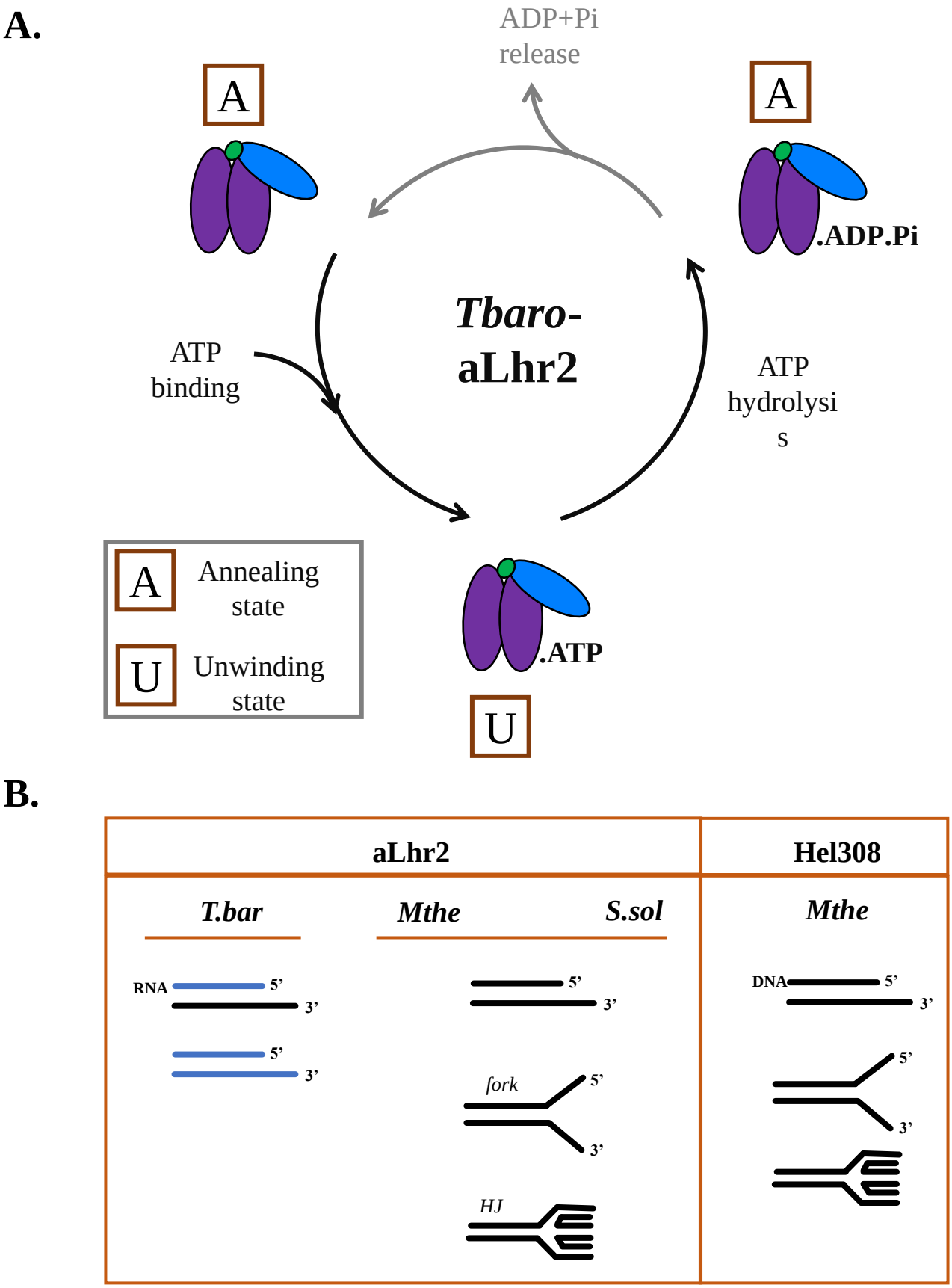


FIGURE 51. *Tbar*-aLhr2 duplex unwinding and strand annealing.
A. ATP provoke changes in *Tbar*-aLhr2 conformation to switch from annealing (A) to unwinding mode (U). **B.** Unwinding substrates of *Tbar*-aLhr2 (This work), *Mthe*-aLhr2 (Buckley et al 2020), *Ssol*-aLhr2 (De Felice et al 2007), and *Pfur*-Hel308 (Guy and Bolt 2005).

In our study, we choose substrates that allowed comparing *in vitro* unwinding and strand-annealing activities of *Tbar*-aLhr2 with the Hel308 DNA helicase of *P. abyssi* studied in parallel in the group in Toulouse (Guy and Bolt 2005) **(FIGURE 51B)**. Hel308 has been shown to be a bonafide DNA helicase recognizing DNA-forks in DNA replication and repair (Guy and Bolt 2005). As mentioned, *Tbar*-aLhr2 could efficiently unwind and anneal DNA/RNA and RNA/RNA hybrids *in vitro*. This is not the case of *Paby*-Hel308 which, in the same experimental condition, cannot unwind RNA hybrids. As expected, only DNA hybrids were very efficiently unwound by *Paby*-Hel308 **(FIGURE 37B)**. Therefore *Tbar*-aLhr2 may not be a bonafide DNA helicase as proposed for *Ssol*-aLhr2 and *Mmth*-aLhr2 (De Felice et al. 2007; Buckley et al. 2020). However, the specific substrates of *Tbar*-aLhr2 remain to be determined.

3-*Tbar*-aLhr2 network at the crossroad of translation, RNA processing/decay and DNA repair pathways

To tackle the physiological role of *Tbar*-aLhr2 in nucleic acid metabolism, we choose, in this exploratory work, to determine its interaction network using *T. barophilus* cellular extract and recombinant protein. In coherence, with studies proposing that aLhr2 helicase in *S.solfataricus* and *M. thermautotrophicus* are involved in DNA recombination and repair (De Falco et al. 2018; Buckley et al. 2020), we observed that some *Tbar*-aLhr2 protein partners are annotated as factors of DNA metabolism pathways (RESULTS-Part III). Indeed the topoisomerase TopoI, the replication A protein RPA (RPA31, RPA42, RPA14) and the PINA ATPase are significantly detected in *Tbar*-aLhr2 interaction network (RESULTS- Part III) **(FIGURE 44B)**. Some of the other partners of *Tbar*-aLhr2 are factors reported to be important in RNA processing and decay. Indeed, the catalytic subunit (Rrp41) of the RNA exosome degrading machinery is part of this network. The role of RNA exosome with a 3'-5' exoribonucleolytic activity which is present in most archaea remains to be determined (RESULTS-Part III) (Clouet-d'Orval et al. 2018; Evguenieva-Hackenberg et al. 2017). Equivalent machinery in eukaryotes has been shown to be involved in a multitude of RNA pathways but also important in DNA repair pathway (Lingaraju et al. 2020).

A.

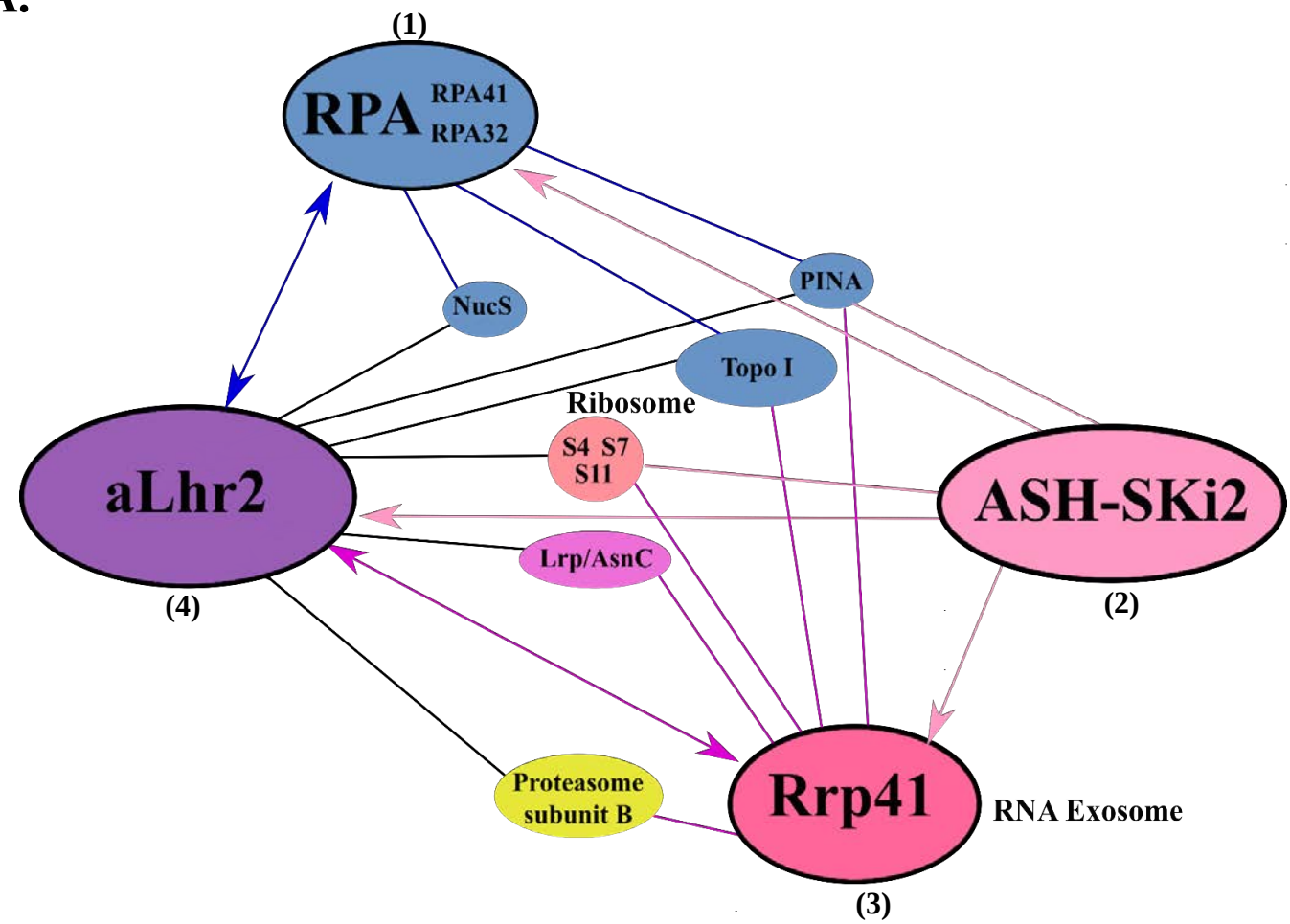


FIGURE 52. Common protein candidates of aLhr2 and putative protein partners.
The diagram of the common candidates of $(His)_6$-aLhr2 (black link), $(His)_6$-Rrp41 (dark pink link), $(His)_6$-RPA (blue link), and $(His)_6$-ASH-Ski2 (light pink link) interaction networks with colours according to the function. Double arrows indicate a reciprocal link between both proteins.
(1) Pluchon et al 2013 (2) Phung et al 2020 (3) Batista et al unpublished (4) This work.

In previous work, the protein networks of *P. abyssi* RPA (replication A protein), ASH-Ski2 (RNA helicase) and Rrp41 (catalytic subunit of the RNA exosome) were identified, respectively (Phung et al. 2020; Pluchon et al. 2013) (Batista et al unpublished data). Interestingly, *Paby*-aLhr2 was significantly detected as a partner protein of these networks. Although these networks have been established with cellular extracts from *P. abysssi* whereas in our study it was done with *T. barophilus* cellular extract. We derived a common network shown in **FIGURE 52,** whose partners are listed in **Table 16**. This network highlight a cross-talk of partners involved in DNA repair, translation and RNA metabolic pathways (**FIGURE 52**). The aLhr2 helicase may play an important role in the coordination of these pathways by acting on DNA/RNA duplexes or RNA-loops which are encountered in these processes.

Table 16 Common partners of protein networks.

Name	ORF *Tbar*	ORF *Paby*	*Tbar*-Lhr2 *This work*	Paby-Rrp41 *Batista et al*	Paby-RPA *Pluchon et al*
Lhr2	**TERMP_00533**	**PAB0744**		+	+
RPA41	**TERMP_01996**	**PAB2163**	+	+	
RPA32	**TERMP_01998**	**PAB2165**	+	+	
Rrp41	TERMP_00163	PAB0420	+		
DNA Topo I	TERMP_00652	PAB1430	+	+	+
PINA	**TERMP_01063**	**PAB1633**	+	+	+
30S ribosomal prot. S7	**TERMP_00247**	**PAB0428**	+	+	
30S ribosomal prot. S11	**TERMP_00130**	**PAB0362**	+	+	
30S ribosomal prot. S4	**TERMP_00129**	**PAB0361**	+	+	
Lrp/AsnC	TERMP_00902	PAB2299	+	+	
NucS	TERMP_01877	PAB2263	+		+
Proteasome subunit β2	TERMP_02168	PAB1867	+	+	

Common partners of $(His)_6$-aLhr2 network using *T. barophilus* cellular extract and $(His)_6$-Rrp41 (Batista et al, unpublished data) and $(His)_6$-RPA (Pluchon et al, 2013) networks using *P. abyssi* cellular extract.

Interestingly, three ribosomal proteins of the 30S ribosomal subunits are common partners between the RNA exosome Rrp41 subunit and aLhr2 networks. A close relationship between the RNA exosome and the ribosome which includes the ASH-Ski2 RNA helicase was recently established in Thermococcales (Phung et al. 2020). These factors were shown to be at the vicinity of the ribosome together with the 5'-3' exoribonuclease aRNaseJ (Phung et al. 2020) (**FIGURE 7**). This relationship appears to be also emphasized in aLhr2 network of this study. We also noticed that ribosomal proteins form a major part of aLhr2-interacting proteins and transcriptomic analysis indicates that in the absence of aLhr2 helicase, transcripts corresponding to ribosomal protein-encoding genes are potentially affected. We keep in mind that mRNA steady-state level is a poor indicator of protein amount in the cell and that aLhr2 protein level needs to be quantified by Western blot.

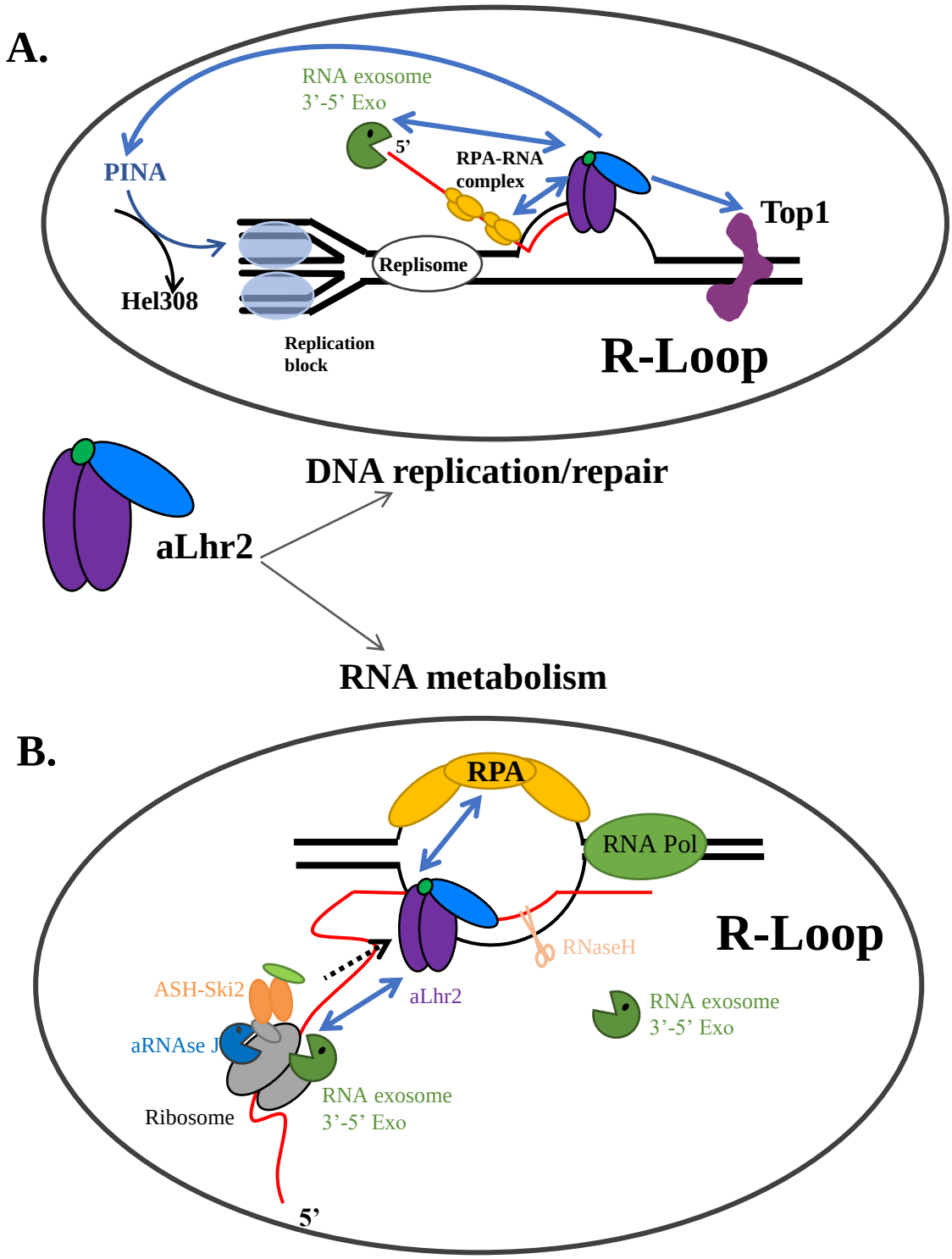


FIGURE 53. Proposition of the implication of aLhr2 in R-loops processing based on the protein interaction profile.

Protein candidates in aLhr2 interaction network are indicated with blue arrows and reciprocal links are indicated with blue double arrows.

A. aLhr2 in replication machinery. Hel308 form HJ structure at stalled replication fork. PINA interacts with Hel308 and mediate HJ structure. Top I relax DNA-negative supercoiling. RPA was proposed to promote R-loops formation in RNA-RPA complex. **B.** aLhr2 in transcription and translation machinery. aLhr2 is detected in ASH-Ski2 network (dotted arrow); RPA was proposed to stimulate RNaseH1 to remove R-loops structure. The complex formed by ASH-Ski2-aRNAseJ-RNA exosome is proposed to be implicated in RNA decay process.

4-Archaeal aLhr2 helicases acting on R-loop structures?

We have identified that aLhr2 has the capacity to unwind RNA strand from RNA/DNA hybrids. In the cell, multiple processes require removing the RNA strand from RNA/DNA duplex. Via its DNA/RNA unwinding activity *Tbar*-aLhr2 be critical in DNA replication and repair by removing Okazaki fragments, in transcription, and in restricting pervasive transcription (Zheng et al. 2019; Han et al. 2017; Han and Porrua 2018). As well, through its aptitude to unwind RNA hybrids, *Tbar*-aLhr2 could participate at nearly every level of RNA metabolism including RNA surveillance, RNA maturation and RNA interference related pathways (Jankowsky 2011; Song and Ji 2019; Zatopek et al. 2018; Ambrus and Frolov 2009). Recently, it has emerged an intimate link between RNA metabolism and genome stability. It has been proposed that dysfunctional RNA metabolism precipitates endogenous transcription-associated DNA damage that drives genome instability in human disease (Zong et al. 2020). Transcription potentially exposes the genome to damage by permitting the formation of unusual RNA-DNA hybrid structures, such as R-loops. R-loops are three-stranded nucleic acid structure consisting of an RNA-DNA hybrid. The accumulation of R-loops structure provokes genome instability. So far, several proteins are known to reduce R-loop accumulation in the living cell. Therefore, regulatory mechanisms are in place to minimize transcription-associated DNA damage.

It is now accepted that, in eukaryotes, RNA/DNA hybrids or R-loops should form at the right place and time to achieve regulatory processes. Important enzymes in R-loop metabolism are RNA/DNA helicases. We can cite the yeast Sen1 or homologous human Senataxin (SETX) (Skourti-Stathaki et al. 2011), the DEAD-box helicase 5 (DDX5) (Mersaoui et al. 2019), or the human DHX9 helicase, which also acts on G4 structures (Chakraborty and Grosse 2011). It is possible that *Tbar*-aLhr2 could play such a role in archaeal cells (**FIGURE 53A**). In addition to DNA/RNA helicases, other factors as Topoisomerases and RPA are involved in R-loop processing (Skourti-Stathaki and Proudfoot 2014; Villarreal et al. 2020). Indeed, deficiency in these enzymes causes the accumulation of R-loops (Skourti-Stathaki and Proudfoot 2014; Crossley et al. 2019). Topoisomerases relax DNA-negative supercoiling that otherwise causes persistent R- loop formation (El Hage et al. 2010; Yang et al. 2014). The Replication A protein (RPA) is also a key sensor of ssDNA and participates in the coordination of DNA damage signaling and DNA repair/recombination at stalled replication forks and sites of DNA damage (**FIGURE 53A**). RPA stimulates RNaseH1 activity at R-loop

site and enhances the association of RNaseH1 with RNA/DNA hybrids (Zou et al. 2006; Wold 1997; Chen and Wold 2014; Borgstahl et al. 2014) (**FIGURE 53B**)**.** Indeed, it was proposed that R-loops may play an important role during DNA repair by serving as a primer to restart DNA replication at stalled replication forks at damaged DNA lesions (Zaitsev and Kowalczykowski 2000; Kogoma 1997). Although, the eukaryotic RPA protein was recently shown to promotes R-loops formation by invading RNA into covalently closed duplex DNA (Mazina et al. 2020). More recently, it was proposed that the RNA exosome machinery is recruited at R-loops to promote degradation of the unwounded RNA released by SETX, and therefore to prevents possible rehybridization and the resultant DNA damage (Skourti-Stathaki et al. 2011; Richard et al. 2013). Moreover, the EJC complex and the RNA exosome also reduce R-loop accumulation by sterically prohibiting interactions between the nascent mRNA and exposed ssDNA (Pefanis et al. 2015; Li et al. 2007). The interaction network of *Tbar*-aLhr2 which includes among others TopoI, RPA and RNA exosome subunit could reflect a cross-talk between these partners at R-loop sites in thermococcales cells (**FIGURE 53A &B**).

5- Is there a Cross-talk between bLhr-HTH and RNase T in *E.coli*

Bacterial bLhr-HTH of *M.smegmatis* was proposed to be implicated in DNA repair pathway. Here we initiate a study to examine the function of bLhr-HTH of *E.coli* and hence to inspect a conserved function of Lhr-type in bacteria. The first Lhr-type encoding protein was detected in *E.coli* in a gene cluster with the ribonuclease T encoding gene. Indeed we showed conservation of this gene cluster in proteobacteria (RESULTS-Part I **FIGURE 23**). We aimed to test a putative protein interaction between bLhr-HTH and RNase T *in vitro* (RESULTS-Part IV). RNAse T is a member of the DEDD 3-5′ exonuclease superfamily (Deutscher and Marlor 1985; Deutscher et al. 1984). This exonuclease is involved in the processing of the 5S and 23S rRNAs by removal of the 3′-end precursor residues (Li and Deutscher 1995; Li et al. 1999). RNase T has both DNase and RNase activity (Viswanathan et al. 1998) and was shown to trims the 3′ end of structured DNA in a DNA repair pathway (Hsiao et al. 2014). Although the protein interaction suggested by our preliminary assays still needs to be confirmed by other additional experiments particularly optimizing the protein level expression and purification steps. The occurrence of such interaction raises the question about the role of bLhr-HTH in RNA metabolism and reput the protein at the nexus of DNA and RNA metabolic pathways

6-General conclusion

The exploratory work, described in this thesis, focused on the physiological function of the helicase of the Lhr-type family. This work brings new elements about aLhr2 in *T. barophilus* and Lhr-HTH in *E. coli*. *In silico* approaches suggest an essential role of Lhr-type helicases in archaea and reveal three distinct families of Lhr-type in bacteria. Our results open new perspectives on the role of thermococcales aLhr2 in the interface of the DNA and RNA metabolism. We discuss these data in a manuscript that is under preparation (Hajj et al, in preparation, to be submitted to *Biomolecules*).

Additional experiments should be conducted to assess specific functions of *Tbar*-aLhr2. Co-purification and co-immunoprecipitation assays need to be performed to indeed identify if specific protein established direct protein interaction with *Tbar*-aLhr2. Cellular fractionation assays on sucrose gradient of *Thermococcus barophilus* of wild type and aLhr2Δ cellular extract will allow identifying if *Tbar*-aLhr2 is part of a multiprotein complex. Proteins and RNAs from the fractions containing *Tbar*-aLhr2 protein will be further analyzed.

Acknowledgment

Undertaking this PhD has been a truly life-changing experience for me.
All the work presented in this thesis would not have been possible without the support and guidance that I received from many people.

Firstly, I would like to thank my thesis jury Ghislaine Henneke, Jacques Oberto, Herve le Hir Matthieu Arlat, and Jamila Borjac-Natour for generously offering their time, support, guidance, and motivation throughout the preparation and review of this document.

I would like to express my sincere gratitude to my advisors Béatrice, Marie, Hala, and Ziad for the continuous support of my PhD study.

Thanks also to my thesis committee, Tamara Basta Le Berre and Remy Dulermo, for their scientific advice, suggestions, and discussions.

I would like to thank the entire team of Béatrice & AJ Carpousis, Manon, Clarisse, Julien, Marta, Léo, Isa and Lydia for their feedbacks, cooperation, their good humor, and to all the internship students that I met or supervised especially Zoé.

Thanks to Manon Batista for her careful reading of the manuscript and helpful comments**.**

Many thanks to Marta and Marie for their work on the transcriptomic part of this project.

I also would like to thank Gwennaële Fichant, Petra Langendijk-Genevaux, and all the Fichant' team for their collaboration for the phylogenomic study and for reviewing this part of this thesis.

I would like to thank all IBCG members for this warm and expressive environment for work.

Special thanks to Sebastien Laurent for his help in analysing the Mass spectrometry results of Pull down and for his advice.

I would like to extend my sincere thanks to Isabelle Saves for helping me enormously to resolve administrative issues. She was always ready to give her timely help and support whenever required. My thanks also go to Catherine Stasiulis for helping me thoroughly with the administrative paperwork both for the University and Prefecture.

I express my deep gratitude to my very close friend, Rawya for her warm friendship, kindness, support and love during all these years.

I owe my deepest gratitude towards my better half Tarek who has always supported me and helped me overcoming the difficulties without complaining. Day by day, you take care of me, you filled up my moments with all these small things. Thank you.

I must express my very profound gratitude to my mom Fida and my dad Hicham for all of the sacrifices that they've made on my behalf, my siblings, Linda, Sarah, Mohammad, and Omar, my brother in law Firas, and my little nephews for providing me with unfailing support and continuous encouragement throughout my years of study, and through the process of researching and writing this thesis. This accomplishment would not have been possible without them.

Finally, my sincere and foremost thanks go to Béatrice Clouet D'Orval who provided me the opportunity to join her team and gave access to the laboratory and research facilities. Her guidance helped me in all the time of research and writing of this thesis. Without her precious support, patience, motivation, and immense knowledge it would not be possible to conduct this research. I could not have imagined having a better advisor and mentor for my PhD.

Bibliography

Aagaard C, Dalgaard JZ, Garrett RA. 1995. Intercellular mobility and homing of an archaeal rDNA intron confers a selective advantage over intron- cells of Sulfolobus acidocaldarius. *Proc Natl Acad Sci USA* **92**: 12285–12289.

Aagaard C, Leviev I, Aravalli RN, Forterre P, Prieur D, Garrett RA. 1996. General vectors for archaeal hyperthermophiles: strategies based on a mobile intron and a plasmid. *FEMS Microbiol Rev* **18**: 93–104.

Abdel-Monem M, Dürwald H, Hoffmann-Berling H. 1976. Enzymic unwinding of DNA. 2. Chain separation by an ATP-dependent DNA unwinding enzyme. *Eur J Biochem* **65**: 441–449.

Abdel-Monem M, Hoffmann-Berling H. 1976. Enzymic unwinding of DNA. 1. Purification and characterization of a DNA-dependent ATPase from Escherichia coli. *Eur J Biochem* **65**: 431–440.

Abdel-Monem M, Taucher-Scholz G, Klinkert MQ. 1983. Identification of Escherichia coli DNA helicase I as the traI gene product of the F sex factor. *Proc Natl Acad Sci USA* **80**: 4659–4663.

Adam PS, Borrel G, Brochier-Armanet C, Gribaldo S. 2017. The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *ISME J* **11**: 2407–2425.

Alabert C, Groth A. 2012. Chromatin replication and epigenome maintenance. *Nat Rev Mol Cell Biol* **13**: 153–167.

Albers S-V, Driessen AJM. 2008. Conditions for gene disruption by homologous recombination of exogenous DNA into the Sulfolobus solfataricus genome. *Archaea* **2**: 145–149.

Albers SV, Meyer BH. 2011. The archaeal cell envelope. *Nat Rev Microbiol* **9**: 414–426.

Allers T, Mevarech M. 2005. Archaeal genetics - the third way. *Nat Rev Genet* **6**: 58–73.

Ambrus AM, Frolov MV. 2009. The diverse roles of RNA helicases in RNAi. *Cell Cycle* **8**: 3500–3505.

Aminov RI. 2013. Role of archaea in human disease. *Front Cell Infect Microbiol* **3**: 42.

Aoyama A, Hayashi M. 1986. Synthesis of bacteriophage ϕX174 in vitro: Mechanism of switch from DNA replication to DNA packaging. *Cell* **47**: 99–106.

Aravind L, Koonin EV. 1999. DNA-binding proteins and evolution of transcription regulation in the archaea. *Nucleic Acids Res* **27**: 4658–4670.

Armache J-P, Anger AM, Márquez V, Franckenberg S, Fröhlich T, Villa E, Berninghausen O, Thomm M, Arnold GJ, Beckmann R, et al. 2013. Promiscuous behaviour of archaeal

ribosomal proteins: implications for eukaryotic ribosome evolution. *Nucleic Acids Res* **41**: 1284–1293.

Atomi H, Imanaka T, Fukui T. 2012. Overview of the genetic tools in the Archaea. *Front Microbiol* **3**: 337.

Azeroglu B, Leach DRF. 2017. RecG controls DNA amplification at double-strand breaks and arrested replication forks. *FEBS Lett* **591**: 1101–1113.

Azzalin CM, Lingner J. 2006. The double life of UPF1 in RNA and DNA stability pathways. *Cell Cycle* **5**: 1496–1498.

Babski J, Haas KA, Näther-Schindler D, Pfeiffer F, Förstner KU, Hammelmann M, Hilker R, Becker A, Sharma CM, Marchfelder A, et al. 2016. Genome-wide identification of transcriptional start sites in the haloarchaeon Haloferax volcanii based on differential RNA-Seq (dRNA-Seq). *BMC Genomics* **17**: 629.

Baker BJ, De Anda V, Seitz KW, Dombrowski N, Santoro AE, Lloyd KG. 2020. Diversity, ecology and evolution of Archaea. *Nat Microbiol*.

Balakrishnan L, Bambara RA. 2013. Flap endonuclease 1. *Annu Rev Biochem* **82**: 119–138.

Ballut L, Marchadier B, Baguet A, Tomasetto C, Séraphin B, Le Hir H. 2005. The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nat Struct Mol Biol* **12**: 861–869.

Ban N, Beckmann R, Cate JHD, Dinman JD, Dragon F, Ellis SR, Lafontaine DLJ, Lindahl L, Liljas A, Lipton JM, et al. 2014. A new system for naming ribosomal proteins. *Curr Opin Struct Biol* **24**: 165–169.

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* **289**: 905–920.

Bang C, Schmitz RA. 2015. Archaea associated with human surfaces: not to be underestimated. *FEMS Microbiol Rev* **39**: 631–648.

Banroques J, Cordin O, Doère M, Linder P, Tanner NK. 2008. A conserved phenylalanine of motif IV in superfamily 2 helicases is required for cooperative, ATP-dependent binding of RNA substrates in DEAD-box proteins. *Mol Cell Biol* **28**: 3359–3371.

Banroques J, Doère M, Dreyfus M, Linder P, Tanner NK. 2010. Motif III in superfamily 2 “helicases” helps convert the binding energy of ATP into a high-affinity RNA binding site in the yeast DEAD-box protein Ded1. *J Mol Biol* **396**: 949–966.

Bansbach CE, Bétous R, Lovejoy CA, Glick GG, Cortez D. 2009. The annealing helicase SMARCAL1 maintains genome integrity at stalled replication forks. *Genes Dev* **23**: 2405–2414.

Barranco-Medina S, Galletto R. 2010. DNA binding induces dimerization of Saccharomyces cerevisiae Pif1. *Biochemistry* **49**: 8445–8454.

Barthelme D, Dinkelaker S, Albers S-V, Londei P, Ermler U, Tampé R. 2011. Ribosome recycling depends on a mechanistic link between the FeS cluster domain and a conformational switch of the twin-ATPase ABCE1. *Proc Natl Acad Sci USA* **108**: 3228–3233.

Bartos JD, Wang W, Pike JE, Bambara RA. 2006. Mechanisms by which Bloom protein can disrupt recombination intermediates of Okazaki fragment maturation. *J Biol Chem* **281**: 32227–32239.

Bedinger P, Hochstrasser M, Jongeneel CV, Alberts BM. 1983. Properties of the T4 bacteriophage DNA replication apparatus: the T4 dda DNA helicase is required to pass a bound RNA polymerase molecule. *Cell* **34**: 115–123.

Bell SD, Jaxel C, Nadal M, Kosa PF, Jackson SP. 1998. Temperature, template topology, and factor requirements of archaeal transcription. *Proc Natl Acad Sci USA* **95**: 15218–15222.

Bell SD, Kosa PL, Sigler PB, Jackson SP. 1999. Orientation of the transcription preinitiation complex in archaea. *Proc Natl Acad Sci USA* **96**: 13662–13667.

Benelli D, Londei P. 2009. Begin at the beginning: evolution of translational initiation. *Res Microbiol* **160**: 493–501.

Benelli D, Londei P. 2011. Translation initiation in Archaea: conserved and domain-specific features. *Biochem Soc Trans* **39**: 89–93.

Bergerat A, de Massy B, Gadelle D, Varoutas PC, Nicolas A, Forterre P. 1997. An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature* **386**: 414–417.

Bernstein KA, Gangloff S, Rothstein R. 2010. The RecQ DNA helicases in DNA repair. *Annu Rev Genet* **44**: 393–417.

Besseling MA, Hopmans EC, Koenen M, van der Meer MTJ, Vreugdenhil S, Schouten S, Sinninghe Damsté JS, Villanueva L. 2019. Depth-related differences in archaeal populations impact the isoprenoid tetraether lipid composition of the Mediterranean Sea water column. *Org Geochem* **135**: 16–31.

Bird LE, Brannigan JA, Subramanya HS, Wigley DB. 1998. Characterisation of Bacillus stearothermophilus PcrA helicase: evidence against an active rolling mechanism. *Nucleic Acids Res* **26**: 2686–2693.

Birien T, Thiel A, Henneke G, Flament D, Moalic Y, Jebbar M. 2018. Development of an Effective 6-Methylpurine Counterselection Marker for Genetic Manipulation in Thermococcus barophilus. *Genes (Basel)* **9**.

Bitan-Banin G, Ortenberg R, Mevarech M. 2003. Development of a gene knockout system for the halophilic archaeon Haloferax volcanii by use of the pyrE gene. *J Bacteriol* **185**: 772–778.

Bjornson KP, Amaratunga M, Moore KJ, Lohman TM. 1994. Single-turnover kinetics of helicase-catalyzed DNA unwinding monitored continuously by fluorescence energy transfer. *Biochemistry* **33**: 14306–14316.

Bocquier AA, Liu L, Cann IK, Komori K, Kohda D, Ishino Y. 2001. Archaeal primase: bridging the gap between RNA and DNA polymerases. *Curr Biol* **11**: 452–456.

Borgstahl GEO, Brader K, Mosel A, Liu S, Kremmer E, Goettsch KA, Kolar C, Nasheuer H-P, Oakley GG. 2014. Interplay of DNA damage and cell cycle signaling at the level of human replication protein A. *DNA Repair (Amst)* **21**: 12–23.

Borrel G, Brugère J-F, Gribaldo S, Schmitz RA, Moissl-Eichinger C. 2020. The host-associated archaeome. *Nat Rev Microbiol* **18**: 622–636.

Borrel G, McCann A, Deane J, Neto MC, Lynch DB, Brugère J-F, O'Toole PW. 2017. Genomics and metagenomics of trimethylamine-utilizing Archaea in the human gut microbiome. *ISME J* **11**: 2059–2074.

Boshoff HIM, Reed MB, Barry CE, Mizrahi V. 2003. DnaE2 polymerase contributes to in vivo survival and the emergence of drug resistance in Mycobacterium tuberculosis. *Cell* **113**: 183–193.

Boubakri H, de Septenville AL, Viguera E, Michel B. 2010. The helicases DinG, Rep and UvrD cooperate to promote replication across transcription units in vivo. *EMBO J* **29**: 145–157.

Branson OE, Freitas MA. 2016. Tag-Count Analysis of Large-Scale Proteomic Data. *J Proteome Res* **15**: 4742–4746.

Bräsen C, Esser D, Rauch B, Siebers B. 2014. Carbohydrate metabolism in Archaea: current insights into unusual enzymes and pathways and their regulation. *Microbiol Mol Biol Rev* **78**: 89–175.

Briegel A, Ortega DR, Huang AN, Oikonomou CM, Gunsalus RP, Jensen GJ. 2015. Structural conservation of chemotaxis machinery across Archaea and Bacteria. *Environ Microbiol Rep* **7**: 414–419.

Briggs GS, Mahdi AA, Wen Q, Lloyd RG. 2005. DNA binding by the substrate specificity (wedge) domain of RecG helicase suggests a role in processivity. *J Biol Chem* **280**: 13921–13927.

Brinkman AB, Ettema TJG, de Vos WM, van der Oost J. 2003. The Lrp family of transcriptional regulators. *Mol Microbiol* **48**: 287–294.

Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P. 2008. Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Microbiol* **6**: 245–252.

Brochier-Armanet C, Forterre P, Gribaldo S. 2011. Phylogeny and evolution of the Archaea: one hundred genomes later. *Curr Opin Microbiol* **14**: 274–281.

Brosh RM, Matson SW. 2020. History of DNA helicases. *Genes (Basel)* **11**.

Brugère J-F, Borrel G, Gaci N, Tottey W, O'Toole PW, Malpuech-Brugère C. 2014. Archaebiotics: proposed therapeutic use of archaea to prevent trimethylaminuria and cardiovascular disease. *Gut Microbes* **5**: 5–10.

Brügger K, Redder P, She Q, Confalonieri F, Zivanovic Y, Garrett RA. 2002. Mobile elements in archaeal genomes. *FEMS Microbiol Lett* **206**: 131–141.

Brüning L, Hackert P, Martin R, Davila Gallesio J, Aquino GRR, Urlaub H, Sloan KE, Bohnsack MT. 2018. RNA helicases mediate structural transitions and compositional changes in pre-ribosomal complexes. *Nat Commun* **9**: 5383.

Buckley RJ, Kramm K, Cooper CDO, Grohmann D, Bolt EL. 2020. Mechanistic insights into Lhr helicase function in DNA repair. *Biochem J* **477**: 2935–2947.

Bujalowski W, Klonowska MM, Jezewska MJ. 1994. Oligomeric structure of Escherichia coli primary replicative helicase DnaB protein. *J Biol Chem* **269**: 31350–31358.

Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, et al. 1996. Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. *Science* **273**: 1058–1073.

Burns DG, Camakaris HM, Janssen PH, Dyall-Smith ML. 2004. Cultivation of Walsby's square haloarchaeon. *FEMS Microbiol Lett* **238**: 469–473.

Büttner K, Nehring S, Hopfner K-P. 2007. Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nat Struct Mol Biol* **14**: 647–652.

Byrd AK, Raney KD. 2017. Structure and function of Pif1 helicase. *Biochem Soc Trans* **45**: 1159–1171.

Cantor SB, Guillemette S. 2011. Hereditary breast cancer and the BRCA1-associated FANCJ/BACH1/BRIP1. *Future Oncol* **7**: 253–261.

Carbone V, Schofield LR, Zhang Y, Sang C, Dey D, Hannus IM, Martin WF, Sutherland-Smith AJ, Ronimus RS. 2015. Structure and Evolution of the Archaeal Lipid Synthesis Enzyme sn-Glycerol-1-phosphate Dehydrogenase. *J Biol Chem* **290**: 21690–21704.

Carpousis AJ. 2002. The Escherichia coli RNA degradosome: structure, function and relationship to other ribonucleolytic multienyzme complexes. *Biochem Soc Trans* **30**: 150–155.

Castelle CJ, Banfield JF. 2018. Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell* **172**: 1181–1197.

Cavicchioli R. 2011. Archaea--timeline of the third domain. *Nat Rev Microbiol* **9**: 51–61.

Chakrabarti S, Jarzynski C, Thirumalai D. 2019. Processivity, velocity, and universal characteristics of nucleic acid unwinding by helicases. *Biophys J* **117**: 867–879.

Chakrabarti S, Jayachandran U, Bonneau F, Fiorini F, Basquin C, Domcke S, Le Hir H, Conti E. 2011. Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. *Mol Cell* **41**: 693–703.

Chakraborty P, Grosse F. 2011. Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair (Amst)* **10**: 654–665.

Chamieh H, Ibrahim H, Kozah J. 2016. Genome-wide identification of SF1 and SF2 helicases from archaea. *Gene* **576**: 214–228.

Chamot D, Colvin KR, Kujat-Choy SL, Owttrim GW. 2005. RNA structural rearrangement via unwinding and annealing by the cyanobacterial RNA helicase, CrhR. *J Biol Chem* **280**: 2036–2044.

Chandran V, Poljak L, Vanzo NF, Leroy A, Miguel RN, Fernandez-Recio J, Parkinson J, Burns C, Carpousis AJ, Luisi BF. 2007. Recognition and cooperation between the ATP-dependent RNA helicase RhlB and ribonuclease RNase E. *J Mol Biol* **367**: 113–132.

Charoensawan V, Wilson D, Teichmann SA. 2010. Genomic repertoires of DNA-binding transcription factors across the tree of life. *Nucleic Acids Res* **38**: 7364–7377.

Charpentier B, Fourmann J, Branlant C. 2007. Reconstitution of Archaeal H/ACA sRNPs and Test of their Activity. In *RNA Modification*, Vol. 425 of *Methods in Enzymology*, pp. 389–405, Elsevier.

Chemnitz Galal W, Pan M, Kelman Z, Hurwitz J. 2012. Characterization of DNA primase complex isolated from the archaeon, Thermococcus kodakaraensis. *J Biol Chem* **287**: 16209–16219.

Chen R, Wold MS. 2014. Replication protein A: single-stranded DNA's first responder: dynamic DNA-interactions allow replication protein A to direct single-strand DNA intermediates into different pathways for synthesis or repair. *Bioessays* **36**: 1156–1161.

Chhetri G, Kalita P, Tripathi T. 2015. An efficient protocol to enhance recombinant protein expression using ethanol in Escherichia coli. *MethodsX* **2**: 385–391.

Chu WK, Hickson ID. 2009. RecQ helicases: multifunctional genome caretakers. *Nat Rev Cancer* **9**: 644–654.

Clouet-d'Orval B, Batista M, Bouvier M, Quentin Y, Fichant G, Marchfelder A, Maier L-K. 2018. Insights into RNA-processing pathways and associated RNA-degrading enzymes in Archaea. *FEMS Microbiol Rev* **42**: 579–613.

Clouet-d'Orval B, Rinaldi D, Quentin Y, Carpousis AJ. 2010. Euryarchaeal beta-CASP proteins with homology to bacterial RNase J Have 5'- to 3'-exoribonuclease activity. *J Biol Chem* **285**: 17574–17583.

Cohen GN, Barbe V, Flament D, Galperin M, Heilig R, Lecompte O, Poch O, Prieur D, Quérellou J, Ripp R, et al. 2003. An integrated analysis of the genome of the hyperthermophilic archaeon Pyrococcus abyssi. *Mol Microbiol* **47**: 1495–1512.

Coin F, Oksenych V, Egly J-M. 2007. Distinct roles for the XPB/p52 and XPD/p44 subcomplexes of TFIIH in damaged DNA opening during nucleotide excision repair. *Mol Cell* **26**: 245–256.

Comstock MJ, Whitley KD, Jia H, Sokoloski J, Lohman TM, Ha T, Chemla YR. 2015. Protein structure. Direct observation of structure-function relationship in a nucleic acid-processing enzyme. *Science* **348**: 352–354.

Constantinesco F, Forterre P, Koonin EV, Aravind L, Elie C. 2004. A bipolar DNA helicase gene, herA, clusters with rad50, mre11 and nurA genes in thermophilic archaea. *Nucleic Acids Res* **32**: 1439–1447.

Constantinou A, Tarsounas M, Karow JK, Brosh RM, Bohr VA, Hickson ID, West SC. 2000. Werner's syndrome protein (WRN) migrates Holliday junctions and co-localizes with RPA upon replication arrest. *EMBO Rep* **1**: 80–84.

Cooper DL, Boyle DC, Lovett ST. 2015. Genetic analysis of Escherichia coli RadA: functional motifs and genetic interactions. *Mol Microbiol* **95**: 769–779.

Corcelli A, Chong PL-G, Koga Y. 2012. Lipid biology of archaea. *Archaea* **2012**: 710836.

Cordin O, Beggs JD. 2013. RNA helicases in splicing. *RNA Biol* **10**: 83–95.

Cordin O, Tanner NK, Doère M, Linder P, Banroques J. 2004. The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *EMBO J* **23**: 2478–2487.

Cortez D, Quevillon-Cheruel S, Gribaldo S, Desnoues N, Sezonov G, Forterre P, Serre M-C. 2010. Evidence for a Xer/dif system for chromosome resolution in archaea. *PLoS Genet* **6**: e1001166.

Costa A, Onesti S. 2009. Structural biology of MCM helicases. *Crit Rev Biochem Mol Biol* **44**: 326–342.

Crossley MP, Bocek M, Cimprich KA. 2019. R-Loops as Cellular Regulators and Genomic Threats. *Mol Cell* **73**: 398–411.

Cui S, Eisenächer K, Kirchhofer A, Brzózka K, Lammens A, Lammens K, Fujita T, Conzelmann K-K, Krug A, Hopfner K-P. 2008. The C-terminal regulatory domain is the RNA 5'-triphosphate sensor of RIG-I. *Mol Cell* **29**: 169–179.

Curti E, Smerdon SJ, Davis EO. 2007. Characterization of the helicase activity and substrate specificity of Mycobacterium tuberculosis UvrD. *J Bacteriol* **189**: 1542–1555.

Da Cunha V, Gaia M, Nasir A, Forterre P. 2018. Asgard archaea do not close the debate about the universal tree of life topology. *PLoS Genet* **14**: e1007215.

Dacks JB, Field MC, Buick R, Eme L, Gribaldo S, Roger AJ, Brochier-Armanet C, Devos DP. 2016. The changing view of eukaryogenesis - fossils, cells, lineages and how they all come together. *J Cell Sci* **129**: 3695–3703.

Dahlke I, Thomm M. 2002. A Pyrococcus homolog of the leucine-responsive regulatory protein, LrpA, inhibits transcription by abrogating RNA polymerase recruitment. *Nucleic Acids Res* **30**: 701–710.

De Falco M, Massa F, Rossi M, De Felice M. 2018. The Sulfolobus solfataricus RecQ-like DNA helicase Hel112 inhibits the NurA/HerA complex exonuclease activity. *Extremophiles* **22**: 581–589.

De Felice M, Aria V, Esposito L, De Falco M, Pucci B, Rossi M, Pisani FM. 2007. A novel DNA helicase with strand-annealing activity from the crenarchaeon Sulfolobus solfataricus. *Biochem J* **408**: 87–95.

de Crécy-Lagard V, Brochier-Armanet C, Urbonavicius J, Fernandez B, Phillips G, Lyons B, Noma A, Alvarez S, Droogmans L, Armengaud J, et al. 2010. Biosynthesis of wyosine derivatives in tRNA: an ancient and highly diverse pathway in Archaea. *Mol Biol Evol* **27**: 2062–2077.

Dehghani-Tafti S, Levdikov V, Antson AA, Bax B, Sanders CM. 2019. Structural and functional analysis of the nucleotide and DNA binding activities of the human PIF1 helicase. *Nucleic Acids Res* **47**: 3208–3222.

DeLong EF. 1992. Archaea in coastal marine environments. *Proc Natl Acad Sci USA* **89**: 5685–5689.

DeLong EF, Pace NR. 2001. Environmental diversity of bacteria and archaea. *Syst Biol* **50**: 470–478.

DeMaere MZ, Williams TJ, Allen MA, Brown MV, Gibson JAE, Rich J, Lauro FM, Dyall-Smith M, Davenport KW, Woyke T, et al. 2013. High level of intergenera gene exchange shapes the evolution of haloarchaea in an isolated Antarctic lake. *Proc Natl Acad Sci USA* **110**: 16939–16944.

Despalins A, Marsit S, Oberto J. 2011. Absynte: a web tool to analyze the evolution of orthologous archaeal and bacterial gene clusters. *Bioinformatics* **27**: 2905–2906.

Deutscher MP, Marlor CW. 1985. Purification and characterization of Escherichia coli RNase T. *J Biol Chem* **260**: 7067–7071.

Deutscher MP, Marlor CW, Zaniewski R. 1984. Ribonuclease T: new exoribonuclease possibly involved in end-turnover of tRNA. *Proc Natl Acad Sci USA* **81**: 4290–4293.

Dey G, Thattai M, Baum B. 2016. On the Archaeal Origins of Eukaryotes and the Challenges of Inferring Phenotype from Genotype. *Trends Cell Biol* **26**: 476–485.

Dhote V, Sweeney TR, Kim N, Hellen CUT, Pestova TV. 2012. Roles of individual domains in the function of DHX29, an essential factor required for translation of structured mammalian mRNAs. *Proc Natl Acad Sci USA* **109**: E3150–9.

Dickey TH, Altschuler SE, Wuttke DS. 2013. Single-stranded DNA-binding proteins: multiple domains for multiple functions. *Structure* **21**: 1074–1084.

Dillingham MS. 2011. Superfamily I helicases as modular components of DNA-processing machines. *Biochem Soc Trans* **39**: 413–423.

Ding H, Guo M, Vidhyasagar V, Talwar T, Wu Y. 2015. The Q motif is involved in DNA binding but not ATP binding in chlr1 helicase. *PLoS One* **10**: e0140755.

Dombrowski N, Williams TA, Sun J, Woodcroft BJ, Lee J-H, Minh BQ, Rinke C, Spang A. 2020. Undinarchaeota illuminate DPANN phylogeny and the impact of gene transfer on archaeal evolution. *Nat Commun* **11**: 3939.

Dominski Z, Carpousis AJ, Clouet-d'Orval B. 2013. Emergence of the β-CASP ribonucleases: highly conserved and ubiquitous metallo-enzymes involved in messenger RNA maturation and degradation. *Biochim Biophys Acta* **1829**: 532–551.

Doolittle WF. 2020. Evolution: two domains of life or three? *Curr Biol* **30**: R177–R179.

Drancourt M, Nkamga VD, Lakhe NA, Régis J-M, Dufour H, Fournier P-E, Bechah Y, Scheld WM, Raoult D. 2017. Evidence of archaeal methanogens in brain abscess. *Clin Infect Dis* **65**: 1–5.

Duggin IG, Dubarry N, Bell SD. 2011. Replication termination and chromosome dimer resolution in the archaeon Sulfolobus solfataricus. *EMBO J* **30**: 145–153.

Durand S, Condon C. 2018. RNases and Helicases in Gram-Positive Bacteria. *Microbiol Spectr* **6**.

Dürr H, Körner C, Müller M, Hickmann V, Hopfner K-P. 2005. X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA. *Cell* **121**: 363–373.

Dutta A, Hromas R, Sung P. 2020. Senataxin: A putative RNA: DNA helicase mutated in als4—emerging mechanisms of genome stability in motor neurons. In *Amyotrophic Lateral Sclerosis - Recent Advances and Therapeutic Challenges [Working Title]*, IntechOpen.

Egelman EH, Yu X, Wild R, Hingorani MM, Patel SS. 1995. Bacteriophage T7 helicase/primase proteins form rings around single-stranded DNA that suggest a general structure for hexameric helicases. *Proc Natl Acad Sci USA* **92**: 3869–3873.

Eggleston AK, O'Neill TE, Bradbury EM, Kowalczykowski SC. 1995. Unwinding of nucleosomal DNA by a DNA helicase. *J Biol Chem* **270**: 2024–2031.

Ejaz A, Ordonez H, Jacewicz A, Ferrao R, Shuman S. 2018. Structure of mycobacterial 3'-to-5' RNA:DNA helicase Lhr bound to a ssDNA tracking strand highlights distinctive features of a novel family of bacterial helicases. *Nucleic Acids Res* **46**: 442–455.

Ejaz A, Shuman S. 2018. Characterization of Lhr-Core DNA helicase and manganese-dependent DNA nuclease components of a bacterial gene cluster encoding nucleic acid repair enzymes. *J Biol Chem* **293**: 17491–17504.

El Hage A, French SL, Beyer AL, Tollervey D. 2010. Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev* **24**: 1546–1558.

Eme L, Ettema TJG. 2018. The eukaryotic ancestor shapes up. *Nature* **562**: 352–353.

Eme L, Spang A, Lombard J, Stairs CW, Ettema TJG. 2018. Archaea and the origin of eukaryotes. *Nat Rev Microbiol* **16**: 120.

Enemark EJ, Joshua-Tor L. 2006. Mechanism of DNA translocation in a replicative hexameric helicase. *Nature* **442**: 270–275.

Eoff RL, Raney KD. 2006. Intermediates revealed in the kinetic mechanism for DNA unwinding by a monomeric helicase. *Nat Struct Mol Biol* **13**: 242–249.

Eoff RL, Raney KD. 2010. Kinetic mechanism for DNA unwinding by multiple molecules of Dda helicase aligned on DNA. *Biochemistry* **49**: 4543–4553.

Epshtein V, Dutta D, Wade J, Nudler E. 2010. An allosteric mechanism of Rho-dependent transcription termination. *Nature* **463**: 245–249.

Erauso G, Reysenbach A-L, Godfroy A, Meunier J-R, Crump B, Partensky F, Baross J, Marteinsson V, Barbier G, Pace N, et al. 1993. Pyrococcus abyssi sp. nov., a new hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. *Arch Microbiol* **160**.

Eun Y-J, Ho P-Y, Kim M, LaRussa S, Robert L, Renner LD, Schmid A, Garner E, Amir A. 2018. Archaeal cells share common size control with bacteria despite noisier growth and division. *Nat Microbiol* **3**: 148–154.

Evguenieva-Hackenberg E, Gauernack S, Klug G. 2017. The Archaeal Exosome: Degradation and Tailing at the 3′ -End of RNA. In *RNA metabolism and gene expression in archaea* (ed. B. Clouet-d'Orval), Vol. 32 of *Nucleic acids and molecular biology*, pp. 115–128, Springer International Publishing, Cham.

Fairman ME, Maroney PA, Wang W, Bowers HA, Gollnick P, Nilsen TW, Jankowsky E. 2004. Protein displacement by DExH/D "RNA helicases" without duplex unwinding. *Science* **304**: 730–734.

Fairman-Williams ME, Guenther U-P, Jankowsky E. 2010. SF1 and SF2 helicases: family matters. *Curr Opin Struct Biol* **20**: 313–324.

Farkas JA, Picking JW, Santangelo TJ. 2013. Genetic techniques for the archaea. *Annu Rev Genet* **47**: 539–561.

Fernández A, Guo HS, Sáenz P, Simón-Buela L, Gómez de Cedrón M, García JA. 1997. The motif V of plum pox potyvirus CI RNA helicase is involved in NTP hydrolysis and is essential for virus RNA replication. *Nucleic Acids Res* **25**: 4474–4480.

Fernández-Pevida A, Kressler D, de la Cruz J. 2015. Processing of preribosomal RNA in Saccharomyces cerevisiae. *Wiley Interdiscip Rev RNA* **6**: 191–209.

Fiala G, Stetter KO. 1986. Pyrococcus furiosus sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100�C. *Arch Microbiol* **145**: 56–61.

Forterre P. 2013. The common ancestor of archaea and eukarya was not an archaeon. *Archaea* **2013**: 372396.

Fournier G. 2009. Horizontal gene transfer and the evolution of methanogenic pathways. *Methods Mol Biol* **532**: 163–179.

Frick DN, Banik S, Rypma RS. 2007. Role of divalent metal cations in ATP hydrolysis catalyzed by the hepatitis C virus NS3 helicase: magnesium provides a bridge for ATP to fuel unwinding. *J Mol Biol* **365**: 1017–1032.

Frick DN, Richardson CC. 2001. DNA primases. *Annu Rev Biochem* **70**: 39–80.

Fuhrman JA, McCallum K, Davis AA. 1992. Novel major archaebacterial group from marine plankton. *Nature* **356**: 148–149.

Fuller-Pace FV, Lane DP. 1992. RNA Helicases. In *Nucleic acids and molecular biology* (eds. F. Eckstein and D.M.J. Lilley), Vol. 6 of *Nucleic acids and molecular biology*, pp. 159–173, Springer Berlin Heidelberg, Berlin, Heidelberg.

Fuller-Pace FV, Nicol SM, Reid AD, Lane DP. 1993. DbpA: a DEAD box protein specifically activated by 23s rRNA. *EMBO J* **12**: 3619–3626.

Gaci N, Borrel G, Tottey W, O'Toole PW, Brugère J-F. 2014. Archaea and the human gut: new beginning of an old story. *World J Gastroenterol* **20**: 16062–16078.

Gambelli L, Meyer BH, McLaren M, Sanders K, Quax TEF, Gold VAM, Albers S-V, Daum B. 2019. Architecture and modular assembly of Sulfolobus S-layers revealed by electron cryotomography. *Proc Natl Acad Sci USA* **116**: 25278–25286.

Gangloff S, McDonald JP, Bendixen C, Arthur L, Rothstein R. 1994. The yeast type I topoisomerase Top3 interacts with Sgs1, a DNA helicase homolog: a potential eukaryotic reverse gyrase. *Mol Cell Biol* **14**: 8391–8398.

Gao J, Wang J. 2012. Re-annotation of two hyperthermophilic archaea Pyrococcus abyssi GE5 and Pyrococcus furiosus DSM 3638. *Curr Microbiol* **64**: 118–129.

Gao Y, Yang W. 2020. Different mechanisms for translocation by monomeric and hexameric helicases. *Curr Opin Struct Biol* **61**: 25–32.

Garcia JL, Patel BK, Ollivier B. 2000. Taxonomic, phylogenetic, and ecological diversity of methanogenic Archaea. *Anaerobe* **6**: 205–226.

Garcia PL, Liu Y, Jiricny J, West SC, Janscak P. 2004. Human RECQ5beta, a protein with DNA helicase and strand-annealing activities in a single polypeptide. *EMBO J* **23**: 2882–2891.

Gehring AM, Walker JE, Santangelo TJ. 2016. Transcription regulation in archaea. *J Bacteriol* **198**: 1906–1917.

Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI, Relman DA, Fraser-Liggett CM, Nelson KE. 2006. Metagenomic analysis of the human distal gut microbiome. *Science* **312**: 1355–1359.

Gilman B, Tijerina P, Russell R. 2017. Distinct RNA-unwinding mechanisms of DEAD-box and DEAH-box RNA helicase proteins in remodeling structured RNAs and RNPs. *Biochem Soc Trans* **45**: 1313–1321.

Giraldo R. 2003. Common domains in the initiators of DNA replication in Bacteria, Archaea and Eukarya: combined structural, functional and phylogenetic perspectives. *FEMS Microbiol Rev* **26**: 533–554.

Goetz GS, Dean FB, Hurwitz J, Matson SW. 1988. The unwinding of duplex regions in DNA by the simian virus 40 large tumor antigen-associated DNA helicase activity. *J Biol Chem* **263**: 383–392.

Gong Z, Dong C-H, Lee H, Zhu J, Xiong L, Gong D, Stevenson B, Zhu J-K. 2005. A DEAD box RNA helicase is essential for mRNA export and important for development and stress responses in Arabidopsis. *Plant Cell* **17**: 256–267.

Gorbalenya AE, Koonin EV. 1993. Helicases: amino acid sequence comparisons and structure-function relationships. *Curr Opin Struct Biol* **3**: 419–429.

Gorbalenya AE, Koonin EV, Donchenko AP, Blinov VM. 1988a. A conserved NTP-motif in putative helicases. *Nature* **333**: 22–22.

Gorbalenya AE, Koonin EV, Donchenko AP, Blinov VM. 1988b. A novel superfamily of nucleoside triphosphate-binding motif containing proteins which are probably

involved in duplex unwinding in DNA and RNA replication and recombination. *FEBS Lett* **235**: 16–24.

Gorbalenya AE, Koonin EV, Donchenko AP, Blinov VM. 1989. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res* **17**: 4713–4730.

Graves-Woodward KL, Gottlieb J, Challberg MD, Weller SK. 1997. Biochemical analyses of mutations in the HSV-1 helicase-primase that alter ATP hydrolysis, DNA unwinding, and coupling between hydrolysis and unwinding. *J Biol Chem* **272**: 4623–4630.

Graves-Woodward KL, Weller SK. 1996. Replacement of gly815 in helicase motif V alters the single-stranded DNA-dependent ATPase activity of the herpes simplex virus type 1 helicase-primase. *J Biol Chem* **271**: 13629–13635.

Greenough L, Kelman Z, Gardner AF. 2015. The roles of family B and D DNA polymerases in Thermococcus species 9°N Okazaki fragment maturation. *J Biol Chem* **290**: 12514–12522.

Gribaldo S, Brochier-Armanet C. 2020. Evolutionary relationships between archaea and eukaryotes. *Nat Ecol Evol* **4**: 20–21.

Grogan DW. 2003. Cytosine methylation by the SuaI restriction-modification system: implications for genetic fidelity in a hyperthermophilic archaeon. *J Bacteriol* **185**: 4657–4661.

Grogan DW. 1996. Exchange of genetic markers at extremely high temperatures in the archaeon Sulfolobus acidocaldarius. *J Bacteriol* **178**: 3207–3211.

Grogan DW, Carver GT, Drake JW. 2001. Genetic fidelity under harsh conditions: analysis of spontaneous mutation in the thermoacidophilic archaeon Sulfolobus acidocaldarius. *Proc Natl Acad Sci USA* **98**: 7928–7933.

Grohmann D, Nagy J, Chakraborty A, Klose D, Fielden D, Ebright RH, Michaelis J, Werner F. 2011. The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. *Mol Cell* **43**: 263–274.

Guéron M, Leroy J-L. 1995. [16] Studies of base pair kinetics by NMR measurement of proton exchange. In *Nuclear magnetic resonance and nucleic acids*, Vol. 261 of *Methods in Enzymology*, pp. 383–413, Elsevier.

Gugliandolo C, Maugeri TL. 2019. Phylogenetic diversity of archaea in shallow hydrothermal vents of eolian islands, italy. *Diversity (Basel)* **11**: 156.

Guy CP, Atkinson J, Gupta MK, Mahdi AA, Gwynn EJ, Rudolph CJ, Moon PB, van Knippenberg IC, Cadman CJ, Dillingham MS, et al. 2009. Rep provides a second motor at the replisome to promote duplication of protein-bound DNA. *Mol Cell* **36**: 654–666.

Guy CP, Bolt EL. 2005. Archaeal Hel308 helicase targets replication forks in vivo and in vitro and unwinds lagging strands. *Nucleic Acids Res* **33**: 3678–3690.

Guy L, Ettema TJG. 2011. The archaeal "TACK" superphylum and the origin of eukaryotes. *Trends Microbiol* **19**: 580–587.

Hajj M, El-Hamaoui S, Batista M, Bouvier M, Abdel-Razzak Z, Clouet-d'Orval B, Chanieh H. 2019. Archaea helicases: unwinding in the extreme. In *Helicases from All Domains of Life* (ed. R. Tutejac), pp. 1–13, Elsevier.

Hall MC, Matson SW. 1999. Helicase motifs: the engine that powers DNA unwinding. *Mol Microbiol* **34**: 867–877.

Han Z, Libri D, Porrua O. 2017. Biochemical characterization of the helicase Sen1 provides new insights into the mechanisms of non-coding transcription termination. *Nucleic Acids Res* **45**: 1355–1370.

Han Z, Porrua O. 2018. Helicases as transcription termination factors: Different solutions for a common problem. *Transcription* **9**: 152–158.

Hanson PI, Whiteheart SW. 2005. AAA+ proteins: have engine, will work. *Nat Rev Mol Cell Biol* **6**: 519–529.

Haque RU, Paradisi F, Allers T. 2020. Haloferax volcanii for biotechnology applications: challenges, current state and perspectives. *Appl Microbiol Biotechnol* **104**: 1371–1382.

Hardwick SW, Luisi BF. 2013. Rarely at rest: RNA helicases and their busy contributions to RNA degradation, regulation and quality control. *RNA Biol* **10**: 56–70.

Hartman AL, Norais C, Badger JH, Delmas S, Haldenby S, Madupu R, Robinson J, Khouri H, Ren Q, Lowe TM, et al. 2010. The complete genome sequence of Haloferax volcanii DS2, a model archaeon. *PLoS One* **5**: e9605.

Hausner W, Thomm M. 2001. Events during initiation of archaeal transcription: open complex formation and DNA-protein interactions. *J Bacteriol* **183**: 3025–3031.

He Y, Andersen GR, Nielsen KH. 2010. Structural basis for the function of DEAH helicases. *EMBO Rep* **11**: 180–186.

He Y, Andersen GR, Nielsen KH. 2011. The function and architecture of DEAH/RHA helicases. *Biomol Concepts* **2**: 315–326.

Hehman GL, Hauswirth WW. 1992. DNA helicase from mammalian mitochondria. *Proc Natl Acad Sci USA* **89**: 8562–8566.

Held NL, Whitaker RJ. 2009. Viral biogeography revealed by signatures in Sulfolobus islandicus genomes. *Environ Microbiol* **11**: 457–466.

Henneke G. 2012. In vitro reconstitution of RNA primer removal in Archaea reveals the existence of two pathways. *Biochem J* **447**: 271–280.

Herrmann M, Saunders AM, Schramm A. 2008. Archaea dominate the ammonia-oxidizing community in the rhizosphere of the freshwater macrophyte Littorella uniflora. *Appl Environ Microbiol* **74**: 3279–3283.

Hilbert BJ, Hayes JA, Stone NP, Duffy CM, Sankaran B, Kelch BA. 2015. Structure and mechanism of the ATPase that powers viral genome packaging. *Proc Natl Acad Sci USA* **112**: E3792–9.

Hilliker A, Gao Z, Jankowsky E, Parker R. 2011. The DEAD-box protein Ded1 modulates translation by the formation and resolution of an eIF4F-mRNA complex. *Mol Cell* **43**: 962–972.

Hirata A, Klein BJ, Murakami KS. 2008. The X-ray crystal structure of RNA polymerase from Archaea. *Nature* **451**: 851–854.

Hishida T, Han Y-W, Shibata T, Kubota Y, Ishino Y, Iwasaki H, Shinagawa H. 2004. Role of the Escherichia coli RecQ DNA helicase in SOS signaling and genome stabilization at stalled replication forks. *Genes Dev* **18**: 1886–1897.

Holm L, Sander C. 1995. Dali: a network tool for protein structure comparison. *Trends Biochem Sci* **20**: 478–480.

Hotta Y, Stern H. 1978. DNA unwinding protein from meiotic cells of Lilium. *Biochemistry* **17**: 1872–1880.

Hsiao Y-Y, Fang W-H, Lee C-C, Chen Y-P, Yuan HS. 2014. Structural insights into DNA repair by RNase T--an exonuclease processing 3’ end of structured DNA in repair pathways. *PLoS Biol* **12**: e1001803.

Hua Z-S, Qu Y-N, Zhu Q, Zhou E-M, Qi Y-L, Yin Y-R, Rao Y-Z, Tian Y, Li Y-X, Liu L, et al. 2018. Genomic inference of the metabolism and evolution of the archaeal phylum Aigarchaeota. *Nat Commun* **9**: 2832.

Huber MD, Duquette ML, Shiels JC, Maizels N. 2006. A conserved G4 DNA binding domain in RecQ family helicases. *J Mol Biol* **358**: 1071–1080.

Huber R, Kristjansson JK, Stetter KO. 1987. Pyrobaculum gen. nov., a new genus of neutrophilic, rod-shaped archaebacteria from continental solfataras growing optimally at 100�C. *Arch Microbiol* **149**: 95–101.

Hübscher U, Stalder HP. 1985. Mammalian DNA helicase. *Nucleic Acids Res* **13**: 5471–5483.

Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, et al. 2016. A new view of the tree of life. *Nat Microbiol* **1**: 16048.

Imachi H, Nobu MK, Nakahara N, Morono Y, Ogawara M, Takaki Y, Takano Y, Uematsu K, Ikuta T, Ito M, et al. 2020. Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* **577**: 519–525.

Inagaki F, Suzuki M, Takai K, Oida H, Sakamoto T, Aoki K, Nealson KH, Horikoshi K. 2003. Microbial communities associated with geological horizons in coastal subseafloor sediments from the sea of okhotsk. *Appl Environ Microbiol* **69**: 7224–7235.

Itou H, Yao M, Watanabe N, Tanaka I. 2008. Crystal structure of the PH1932 protein, a unique archaeal ArsR type winged-HTH transcription factor from Pyrococcus horikoshii OT3. *Proteins* **70**: 1631–1634.

Iyer LM, Leipe DD, Koonin EV, Aravind L. 2004a. Evolutionary history and higher order classification of AAA+ ATPases. *J Struct Biol* **146**: 11–31.

Iyer LM, Makarova KS, Koonin EV, Aravind L. 2004b. Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. *Nucleic Acids Res* **32**: 5260–5279.

Jäger D, Förstner KU, Sharma CM, Santangelo TJ, Reeve JN. 2014. Primary transcriptome map of the hyperthermophilic archaeon Thermococcus kodakarensis. *BMC Genomics* **15**: 684.

Jain S, Caforio A, Driessen AJM. 2014. Biosynthesis of archaeal membrane ether lipids. *Front Microbiol* **5**: 641.

Jankowsky A, Guenther U-P, Jankowsky E. 2011. The RNA helicase database. *Nucleic Acids Res* **39**: D338–41.

Jankowsky E. 2011. RNA helicases at work: binding and rearranging. *Trends Biochem Sci* **36**: 19–29.

Jankowsky E, Fairman ME. 2007. RNA helicases--one fold for many functions. *Curr Opin Struct Biol* **17**: 316–324.

Jankowsky E, Jankowsky A. 2000. The DExH/D protein family database. *Nucleic Acids Res* **28**: 333–334.

Janssen PH, Kirs M. 2008. Structure of the archaeal community of the rumen. *Appl Environ Microbiol* **74**: 3619–3625.

Jarmoskaite I, Russell R. 2014. RNA helicase proteins as chaperones and remodelers. *Annu Rev Biochem* **83**: 697–725.

Jarrell KF, Albers S-V. 2012. The archaellum: an old motility structure with a new name. *Trends Microbiol* **20**: 307–312.

Jarrell KF, Ding Y, Meyer BH, Albers S-V, Kaminski L, Eichler J. 2014. N-linked glycosylation in Archaea: a structural, functional, and genetic analysis. *Microbiol Mol Biol Rev* **78**: 304–341.

Jezewska MJ, Lucius AL, Bujalowski W. 2005. Binding of six nucleotide cofactors to the hexameric helicase RepA protein of plasmid RSF1010. 2. Base specificity, nucleotide structure, magnesium, and salt effect on the cooperative binding of the cofactors. *Biochemistry* **44**: 3877–3890.

Johnson SJ, Jackson RN. 2013. Ski2-like RNA helicase structures: common themes and complex assemblies. *RNA Biol* **10**: 33–43.

Joo S, Chung BH, Lee M, Ha TH. 2019. Ring-shaped replicative helicase encircles double-stranded DNA during unwinding. *Nucleic Acids Res* **47**: 11344–11354.

Kalde M, Elliott L, Ravikumar R, Rybak K, Altmann M, Klaeger S, Wiese C, Abele M, Al B, Kalbfuß N, et al. 2019. Interactions between Transport Protein Particle (TRAPP) complexes and Rab GTPases in Arabidopsis. *Plant J* **100**: 279–297.

Kanaan J, Raj S, Decourty L, Saveanu C, Croquette V, Le Hir H. 2018. UPF1-like helicase grip on nucleic acids dictates processivity. *Nat Commun* **9**: 3752.

Karow JK, Constantinou A, Li JL, West SC, Hickson ID. 2000. The Bloom's syndrome gene product promotes branch migration of holliday junctions. *Proc Natl Acad Sci USA* **97**: 6504–6508.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**: 845–858.

Kellner S, Spang A, Offre P, Szöllősi GJ, Petitjean C, Williams TA. 2018. Genome size evolution in the Archaea. *Emerg Top Life Sci* **2**: ETLS20180021.

Kelly TJ, Simancek P, Brush GS. 1998. Identification and characterization of a single-stranded DNA-binding protein from the archaeon Methanococcus jannaschii. *Proc Natl Acad Sci USA* **95**: 14634–14639.

Kelman LM, Kelman Z. 2014. Archaeal DNA replication. *Annu Rev Genet* **48**: 71–97.

Kelman LM, Kelman Z. 2018. Do archaea need an origin of replication? *Trends Microbiol* **26**: 172–174.

Kelman LM, O'Dell WB, Kelman Z. 2020. Unwinding 20 years of the archaeal minichromosome maintenance helicase. *J Bacteriol* **202**.

Kelman Z, Lee JK, Hurwitz J. 1999. The single minichromosome maintenance protein of Methanobacterium thermoautotrophicum DeltaH contains DNA helicase activity. *Proc Natl Acad Sci USA* **96**: 14783–14788.

Khelaifia S, Raoult D. 2016. Haloferax massiliensis sp. nov., the first human-associated halophilic archaea. *New Microbes New Infect* **12**: 96–98.

Khemici V, Linder P. 2018. RNA helicases in RNA decay. *Biochem Soc Trans* **46**: 163–172.

Kim SH, Smith J, Claude A, Lin RJ. 1992. The purified yeast pre-mRNA splicing factor PRP2 is an RNA-dependent NTPase. *EMBO J* **11**: 2319–2326.

Kimble JC, Winter AS, Spilde MN, Sinsabaugh RL, Northup DE. 2018. A potential central role of Thaumarchaeota in N-Cycling in a semi-arid environment, Fort Stanton Cave, Snowy River passage, New Mexico, USA. *FEMS Microbiol Ecol* **94**.

Klinger CM, Spang A, Dacks JB, Ettema TJG. 2016. Tracing the Archaeal Origins of Eukaryotic Membrane-Trafficking System Building Blocks. *Mol Biol Evol* **33**: 1528–1541.

Klingl A. 2014. S-layer and cytoplasmic membrane - exceptions from the typical archaeal cell wall with a focus on double membranes. *Front Microbiol* **5**: 624.

Knittel K, Lösekann T, Boetius A, Kort R, Amann R. 2005. Diversity and distribution of methanotrophic archaea at cold seeps. *Appl Environ Microbiol* **71**: 467–479.

Knoll A, Puchta H. 2011. The role of DNA helicases and their interaction partners in genome stability and meiotic recombination in plants. *J Exp Bot* **62**: 1565–1579.

Kobayashi K, Saito K, Ishitani R, Ito K, Nureki O. 2012. Structural basis for translation termination by archaeal RF1 and GTP-bound EF1α complex. *Nucleic Acids Res* **40**: 9319–9328.

Kocsis ZS, Sarlós K, Harami GM, Martina M, Kovács M. 2014. A nucleotide-dependent and HRDC domain-dependent structural transition in DNA-bound RecQ helicase. *J Biol Chem* **289**: 5938–5949.

Kogoma T. 1997. Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiol Mol Biol Rev* **61**: 212–238.

Kohler PRA, Metcalf WW. 2012. Genetic manipulation of Methanosarcina spp. *Front Microbiol* **3**: 259.

Komori K, Fujikane R, Shinagawa H, Ishino Y. 2002. Novel endonuclease in Archaea cleaving DNA with various branched structure. *Genes Genet Syst* **77**: 227–241.

Komori K, Ishino Y. 2001. Replication protein A in Pyrococcus furiosus is involved in homologous DNA recombination. *J Biol Chem* **276**: 25654–25660.

Koonin EV, Makarova KS. 2017. Mobile Genetic Elements and Evolution of CRISPR-Cas Systems: All the Way There and Back. *Genome Biol Evol* **9**: 2812–2825.

Koonin EV, Wolf YI. 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res* **36**: 6688–6719.

Korkhin Y, Unligil UM, Littlefield O, Nelson PJ, Stuart DI, Sigler PB, Bell SD, Abrescia NGA. 2009. Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure. *PLoS Biol* **7**: e1000102.

Koskinen K, Pausan MR, Perras AK, Beck M, Bang C, Mora M, Schilhabel A, Schmitz R, Moissl-Eichinger C. 2017. First Insights into the Diverse Human Archaeome: Specific Detection of Archaea in the Gastrointestinal Tract, Lung, and Nose and on Skin. *MBio* **8**.

Kozubal MA, Romine M, Jennings R deM, Jay ZJ, Tringe SG, Rusch DB, Beam JP, McCue LA, Inskeep WP. 2013. Geoarchaeota: a new candidate phylum in the Archaea from high-temperature acidic iron mats in Yellowstone National Park. *ISME J* **7**: 622–634.

Kuhn B, Abdel-Monem M, Krell H, Hoffmann-Berling H. 1979. Evidence for two mechanisms for DNA unwinding catalyzed by DNA helicases. *J Biol Chem* **254**: 11343–11350.

Kumarevel T, Tanaka T, Umehara T, Yokoyama S. 2009. ST1710-DNA complex crystal structure reveals the DNA binding mechanism of the MarR family of regulators. *Nucleic Acids Res* **37**: 4723–4735.

Kuper J, Braun C, Elias A, Michels G, Sauer F, Schmitt DR, Poterszman A, Egly J-M, Kisker C. 2014. In TFIIH, XPD helicase is exclusively devoted to DNA repair. *PLoS Biol* **12**: e1001954.

Kyrpides NC, Woese CR. 1998. Universally conserved translation initiation factors. *Proc Natl Acad Sci USA* **95**: 224–228.

Lagrange T, Kapanidis AN, Tang H, Reinberg D, Ebright RH. 1998. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev* **12**: 34–44.

Langer D, Hain J, Thuriaux P, Zillig W. 1995. Transcription in archaea: similarity to that in eucarya. *Proc Natl Acad Sci USA* **92**: 5768–5772.

Langworthy TA, Smith PF, Mayberry WR. 1972. Lipids of Thermoplasma acidophilum. *J Bacteriol* **112**: 1193–1200.

Le Hir H, Izaurralde E, Maquat LE, Moore MJ. 2000. The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J* **19**: 6860–6869.

Lecompte O, Ripp R, Thierry J-C, Moras D, Poch O. 2002. Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. *Nucleic Acids Res* **30**: 5382–5390.

Lee JY, Yang W. 2006. UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. *Cell* **127**: 1349–1360.

Lee MS, Marians KJ. 1989. The Escherichia coli primosome can translocate actively in either direction along a DNA strand. *J Biol Chem* **264**: 14531–14542.

Lee S-J, Surma M, Hausner W, Thomm M, Boos W. 2008. The role of TrmB and TrmB-like transcriptional regulators for sugar transport and metabolism in the hyperthermophilic archaeon Pyrococcus furiosus. *Arch Microbiol* **190**: 247–256.

Lee TI, Young RA. 2013. Transcriptional regulation and its misregulation in disease. *Cell* **152**: 1237–1251.

Lehnik-Habrink M, Newman J, Rothe FM, Solovyova AS, Rodrigues C, Herzberg C, Commichau FM, Lewis RJ, Stülke J. 2011. RNase Y in Bacillus subtilis: a Natively disordered protein that is the functional equivalent of RNase E from Escherichia coli. *J Bacteriol* **193**: 5431–5441.

Leigh JA, Albers S-V, Atomi H, Allers T. 2011. Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfolobales. *FEMS Microbiol Rev* **35**: 577–608.

Leigh JA, Whitman WB. 2013. Archaeal Genetics. In *Brenner's encyclopedia of genetics*, pp. 188–191, Elsevier.

Leipe DD, Koonin EV, Aravind L. 2003. Evolution and classification of P-loop kinases and related proteins. *J Mol Biol* **333**: 781–815.

Leipe DD, Wolf YI, Koonin EV, Aravind L. 2002. Classification and evolution of P-loop GTPases and related ATPases. *J Mol Biol* **317**: 41–72.

Lemmens L, Maklad HR, Bervoets I, Peeters E. 2019. Transcription Regulators in Archaea: Homologies and Differences with Bacterial Regulators. *J Mol Biol* **431**: 4132–4146.

Lestini R, Duan Z, Allers T. 2010. The archaeal Xpf/Mus81/FANCM homolog Hef and the Holliday junction resolvase Hjc define alternative pathways that are essential for cell viability in Haloferax volcanii. *DNA Repair (Amst)* **9**: 994–1002.

Levy S, Portnoy V, Admon J, Schuster G. 2011. Distinct activities of several RNase J proteins in methanogenic archaea. *RNA Biol* **8**: 1073–1083.

Li F, Herrera J, Zhou S, Maslov DA, Simpson L. 2011. Trypanosome REH1 is an RNA helicase involved with the 3'-5' polarity of multiple gRNA-guided uridine insertion/deletion RNA editing. *Proc Natl Acad Sci USA* **108**: 3542–3547.

Li H. 2015. Structural principles of CRISPR RNA processing. *Structure* **23**: 13–20.

Li X, Niu T, Manley JL. 2007. The RNA binding protein RNPS1 alleviates ASF/SF2 depletion-induced genomic instability. *RNA* **13**: 2108–2115.

Li Y, Pan S, Zhang Y, Ren M, Feng M, Peng N, Chen L, Liang YX, She Q. 2016. Harnessing Type I and Type III CRISPR-Cas systems for genome editing. *Nucleic Acids Res* **44**: e34.

Li Z, Deutscher MP. 1995. The tRNA processing enzyme RNase T is essential for maturation of 5S RNA. *Proc Natl Acad Sci USA* **92**: 6883–6886.

Li Z, Lu S, Hou G, Ma X, Sheng D, Ni J, Shen Y. 2008. Hjm/Hel308A DNA helicase from Sulfolobus tokodaii promotes replication fork regression and interacts with Hjc endonuclease in vitro. *J Bacteriol* **190**: 3006–3017.

Li Z, Pandit S, Deutscher MP. 1999. Maturation of 23S ribosomal RNA requires the exoribonuclease RNase T. *RNA* **5**: 139–146.

Linder P, Jankowsky E. 2011. From unwinding to clamping - the DEAD box RNA helicase family. *Nat Rev Mol Cell Biol* **12**: 505–516.

Linder P, Lasko PF, Ashburner M, Leroy P, Nielsen PJ, Nishi K, Schnier J, Slonimski PP. 1989. Birth of the D-E-A-D box. *Nature* **337**: 121–122.

Linder P, Lemeille S, Redder P. 2014. Transcriptome-wide analyses of 5'-ends in RNase J mutants of a gram-positive pathogen reveal a role in RNA maturation, regulation and degradation. *PLoS Genet* **10**: e1004207.

Lingaraju M, Schuller JM, Falk S, Gerlach P, Bonneau F, Basquin J, Benda C, Conti E. 2020. To process or to decay: A mechanistic view of the nuclear RNA exosome. *Cold Spring Harb Symp Quant Biol*.

Lipscomb GL, Keese AM, Cowart DM, Schut GJ, Thomm M, Adams MWW, Scott RA. 2009. SurR: a transcriptional activator and repressor controlling hydrogen and elemental sulphur metabolism in Pyrococcus furiosus. *Mol Microbiol* **71**: 332–349.

Liu H, Rudolf J, Johnson KA, McMahon SA, Oke M, Carter L, McRobbie A-M, Brown SE, Naismith JH, White MF. 2008. Structure of the DNA repair helicase XPD. *Cell* **133**: 801–812.

Liu Y, Whitman WB. 2008. Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. *Ann N Y Acad Sci* **1125**: 171–189.

Lohman TM. 1992. Escherichia coli DNA helicases: mechanisms of DNA unwinding. *Mol Microbiol* **6**: 5–14.

Lohman TM. 1993. Helicase-catalyzed DNA unwinding. *J Biol Chem* **268**: 2269–2272.

Lohman TM, Bjornson KP. 1996. Mechanisms of helicase-catalyzed DNA unwinding. *Annu Rev Biochem* **65**: 169–214.

Lohman TM, Fazio NT. 2018. How Does a Helicase Unwind DNA? Insights from RecBCD Helicase. *Bioessays* **40**: e1800009.

Lohman TM, Tomko EJ, Wu CG. 2008. Non-hexameric DNA helicases and translocases: mechanisms and regulation. *Nat Rev Mol Cell Biol* **9**: 391–401.

Londei P. 2005. Evolution of translational initiation: new insights from the archaea. *FEMS Microbiol Rev* **29**: 185–200.

Lucius AL, Maluf NK, Fischer CJ, Lohman TM. 2003. General methods for analysis of sequential "n-step" kinetic mechanisms: application to single turnover kinetics of helicase-catalyzed DNA unwinding. *Biophys J* **85**: 2224–2239.

Lundberg KS, Shoemaker DD, Adams MW, Short JM, Sorge JA, Mathur EJ. 1991. High-fidelity amplification using a thermostable DNA polymerase isolated from Pyrococcus furiosus. *Gene* **108**: 1–6.

Lurie-Weinberger MN, Gophna U. 2015. Archaea in and on the human body: health implications and future directions. *PLoS Pathog* **11**: e1004833.

Lyu Z, Whitman WB. 2017. Evolution of the archaeal and mammalian information processing systems: towards an archaeal model for human disease. *Cell Mol Life Sci* **74**: 183–212.

MacGregor BJ, Moser DP, Alm EW, Nealson KH, Stahl DA. 1997. Crenarchaeota in Lake Michigan sediment. *Appl Environ Microbiol* **63**: 1178–1181.

Machwe A, Lozada EM, Xiao L, Orren DK. 2006. Competition between the DNA unwinding and strand pairing activities of the Werner and Bloom syndrome proteins. *BMC Mol Biol* **7**: 1.

MacNeill SA. 2010. Structure and function of the GINS complex, a key component of the eukaryotic replisome. *Biochem J* **425**: 489–500.

Maier L-K, Stachler A-E, Brendel J, Stoll B, Fischer S, Haas KA, Schwarz TS, Alkhnbashi OS, Sharma K, Urlaub H, et al. 2019. The nuts and bolts of the Haloferax CRISPR-Cas system I-B. *RNA Biol* **16**: 469–480.

Makarova KS, Koonin EV, Kelman Z. 2012. The CMG (CDC45/RecJ, MCM, GINS) complex is a conserved component of the DNA replication system in all archaea and eukaryotes. *Biol Direct* **7**: 7.

Maluf NK, Fischer CJ, Lohman TM. 2003. A Dimer of Escherichia coli UvrD is the active form of the helicase in vitro. *J Mol Biol* **325**: 913–935.

Manosas M, Xi XG, Bensimon D, Croquette V. 2010. Active and passive mechanisms of helicases. *Nucleic Acids Res* **38**: 5518–5526.

Marie L, Rapisarda C, Morales V, Bergé M, Perry T, Soulet A-L, Gruget C, Remaut H, Fronzes R, Polard P. 2017. Bacterial RadA is a DnaB-type helicase interacting with RecA to promote bidirectional D-loop extension. *Nat Commun* **8**: 15638.

Marinsek N, Barry ER, Makarova KS, Dionne I, Koonin EV, Bell SD. 2006. GINS, a central nexus in the archaeal DNA replication fork. *EMBO Rep* **7**: 539–545.

Marintcheva B, Weller SK. 2003. Helicase motif Ia is involved in single-strand DNA-binding and helicase activities of the herpes simplex virus type 1 origin-binding protein, UL9. *J Virol* **77**: 2477–2488.

Marteinsson VT, Birrien JL, Reysenbach AL, Vernet M, Marie D, Gambacorta A, Messner P, Sleytr UB, Prieur D. 1999. Thermococcus barophilus sp. nov., a new barophilic and hyperthermophilic archaeon isolated under high hydrostatic pressure from a deep-sea hydrothermal vent. *Int J Syst Bacteriol* **49 Pt 2**: 351–359.

Martin R, Straub AU, Doebele C, Bohnsack MT. 2013. DExD/H-box RNA helicases in ribosome biogenesis. *RNA Biol* **10**: 4–18.

Mathy N, Bénard L, Pellegrini O, Daou R, Wen T, Condon C. 2007. 5'-to-3' exoribonuclease activity in bacteria: role of RNase J1 in rRNA maturation and 5' stability of mRNA. *Cell* **129**: 681–692.

Matlock DL, Yeruva L, Byrd AK, Mackintosh SG, Langston C, Brown C, Cameron CE, Fischer CJ, Raney KD. 2010. Investigation of translocation, DNA unwinding, and protein displacement by NS3h, the helicase domain from the hepatitis C virus helicase. *Biochemistry* **49**: 2097–2109.

Matson SW. 1989. Escherichia coli DNA helicase II (uvrD gene product) catalyzes the unwinding of DNA.RNA hybrids in vitro. *Proc Natl Acad Sci USA* **86**: 4430–4434.

Matson SW, Tabor S, Richardson CC. 1983. The gene 4 protein of bacteriophage T7. Characterization of helicase activity. *J Biol Chem* **258**: 14017–14024.

Mattiroli F, Bhattacharyya S, Dyer PN, White AE, Sandman K, Burkhart BW, Byrne KR, Lee T, Ahn NG, Santangelo TJ, et al. 2017. Structure of histone-based chromatin in Archaea. *Science* **357**: 609–612.

Mazina OM, Somarowthu S, Kadyrova LY, Baranovskiy AG, Tahirov TH, Kadyrov FA, Mazin AV. 2020. Replication protein A binds RNA and promotes R-loop formation. *J Biol Chem*.

McGlynn P, Lloyd RG. 2001. Rescue of stalled replication forks by RecG: simultaneous translocation on the leading and lagging strand templates supports an active DNA unwinding model of fork reversal and Holliday junction formation. *Proc Natl Acad Sci USA* **98**: 8227–8234.

McGuffin LJ, Bryson K, Jones DT. 2000. The PSIPRED protein structure prediction server. *Bioinformatics* **16**: 404–405.

McInerney JO, Wilkinson M, Patching JW, Embley TM, Powell R. 1995. Recovery and phylogenetic analysis of novel archaeal rRNA sequences from a deep-sea deposit feeder. *Appl Environ Microbiol* **61**: 1646–1648.

Meagher M, Epling LB, Enemark EJ. 2019. DNA translocation mechanism of the MCM complex and implications for replication initiation. *Nat Commun* **10**: 3117.

Medagli B, Onesti S. 2013. Structure and mechanism of hexameric helicases. *Adv Exp Med Biol* **767**: 75–95.

Merrikh H, Machón C, Grainger WH, Grossman AD, Soultanas P. 2011. Co-directional replication-transcription conflicts lead to replication restart. *Nature* **470**: 554–557.

Mersaoui SY, Yu Z, Coulombe Y, Karam M, Busatto FF, Masson J-Y, Richard S. 2019. Arginine methylation of the DDX5 helicase RGG/RG motif by PRMT5 regulates resolution of RNA:DNA hybrids. *EMBO J* **38**: e100986.

Mihajlovski A, Alric M, Brugère J-F. 2008. A putative new order of methanogenic Archaea inhabiting the human gut, as revealed by molecular analyses of the mcrA gene. *Res Microbiol* **159**: 516–521.

Mihajlovski A, Doré J, Levenez F, Alric M, Brugère J-F. 2010. Molecular evaluation of the human gut methanogenic archaeal microbiota reveals an age-associated increase of the diversity. *Environ Microbiol Rep* **2**: 272–280.

Miłoch A, Krężel A. 2014. Metal binding properties of the zinc finger metallome--insights into variations in stability. *Metallomics* **6**: 2015–2024.

Minh BQ, Nguyen MAT, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* **30**: 1188–1195.

Miyata T, Yamada K, Iwasaki H, Shinagawa H, Morikawa K, Mayanagi K. 2000. Two different oligomeric states of the RuvB branch migration motor protein as revealed by electron microscopy. *J Struct Biol* **131**: 83–89.

Mohaghegh P, Karow JK, Brosh RM, Bohr VA, Hickson ID. 2001. The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic Acids Res* **29**: 2843–2849.

Moissl-Eichinger C, Pausan M, Taffner J, Berg G, Bang C, Schmitz RA. 2018. Archaea are interactive components of complex microbiomes. *Trends Microbiol* **26**: 70–85.

Moissl-Eichinger C, Probst AJ, Birarda G, Auerbach A, Koskinen K, Wolf P, Holman H-YN. 2017. Human age and skin physiology shape diversity and abundance of Archaea on skin. *Sci Rep* **7**: 4039.

Mullins EA, Warren GM, Bradley NP, Eichman BF. 2017. Structure of a DNA glycosylase that unhooks interstrand cross-links. *Proc Natl Acad Sci USA* **114**: 4400–4405.

Myers TW, Romano LJ. 1988. Mechanism of stimulation of T7 DNA polymerase by Escherichia coli single-stranded DNA binding protein (SSB). *J Biol Chem* **263**: 17006–17015.

Nagata M, Ishino S, Yamagami T, Simons J-R, Kanai T, Atomi H, Ishino Y. 2017. Possible function of the second RecJ-like protein in stalled replication fork repair by interacting with Hef. *Sci Rep* **7**: 16949.

Narasingarao P, Podell S, Ugalde JA, Brochier-Armanet C, Emerson JB, Brocks JJ, Heidelberg KB, Banfield JF, Allen EE. 2012. De novo metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. *ISME J* **6**: 81–93.

Neuwald AF, Aravind L, Spouge JL, Koonin EV. 1999. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res* **9**: 27–43.

Newman M, Murray-Rust J, Lally J, Rudolf J, Fadden A, Knowles PP, White MF, McDonald NQ. 2005. Structure of an XPF endonuclease with and without DNA suggests a model for substrate recognition. *EMBO J* **24**: 895–905.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**: 268–274.

Nishikiori M, Sugiyama S, Xiang H, Niiyama M, Ishibashi K, Inoue T, Ishikawa M, Matsumura H, Katoh E. 2012. Crystal structure of the superfamily 1 helicase from Tomato mosaic virus. *J Virol* **86**: 7565–7576.

Nishino T, Komori K, Tsuchiya D, Ishino Y, Morikawa K. 2005. Crystal structure and functional implications of Pyrococcus furiosus hef helicase domain involved in branched DNA processing. *Structure* **13**: 143–153.

Noble CG, Song H. 2007. MLN51 stimulates the RNA-helicase activity of eIF4AIII. *PLoS One* **2**: e303.

Nunoura T, Takaki Y, Kakuta J, Nishi S, Sugahara J, Kazama H, Chee G-J, Hattori M, Kanai A, Atomi H, et al. 2011. Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res* **39**: 3204–3223.

Ochs SM, Thumann S, Richau R, Weirauch MT, Lowe TM, Thomm M, Hausner W. 2012. Activation of archaeal transcription mediated by recruitment of transcription factor B. *J Biol Chem* **287**: 18863–18871.

Offre P, Spang A, Schleper C. 2013. Archaea in biogeochemical cycles. *Annu Rev Microbiol* **67**: 437–457.

Ogura T, Wilkinson AJ. 2001. AAA+ superfamily ATPases: common structure--diverse function. *Genes Cells* **6**: 575–597.

Olsen GJ, Woese CR. 1993. Ribosomal RNA: a key to phylogeny. *FASEB J* **7**: 113–123.

Ordonez H, Shuman S. 2013. Mycobacterium smegmatis Lhr Is a DNA-dependent ATPase and a 3'-to-5' DNA translocase and helicase that prefers to unwind 3'-tailed RNA:DNA hybrids. *J Biol Chem* **288**: 14125–14134.

Ouhammouch M, Dewhurst RE, Hausner W, Thomm M, Geiduschek EP. 2003. Activation of archaeal transcription by recruitment of the TATA-binding protein. *Proc Natl Acad Sci USA* **100**: 5097–5102.

Ouhammouch M, Geiduschek EP. 2001. A thermostable platform for transcriptional regulation: the DNA-binding properties of two Lrp homologs from the hyperthermophilic archaeon Methanococcus jannaschii. *EMBO J* **20**: 146–156.

Oxley APA, Lanfranconi MP, Würdemann D, Ott S, Schreiber S, McGenity TJ, Timmis KN, Nogales B. 2010. Halophilic archaea in the human intestinal mucosa. *Environ Microbiol* **12**: 2398–2410.

Oyama T, Ishino S, Shirai T, Yamagami T, Nagata M, Ogino H, Kusunoki M, Ishino Y. 2016. Atomic structure of an archaeal GAN suggests its dual roles as an exonuclease in DNA repair and a CMG component in DNA replication. *Nucleic Acids Res* **44**: 9505–9517.

Pai EF, Sachsenheimer W, Schirmer RH, Schulz GE. 1977. Substrate positions and induced-fit in crystalline adenylate kinase. *J Mol Biol* **114**: 37–45.

Pan M, Kelman LM, Kelman Z. 2011. The archaeal PCNA proteins. *Biochem Soc Trans* **39**: 20–24.

Parsons CA, Tsaneva I, Lloyd RG, West SC. 1992. Interaction of Escherichia coli RuvA and RuvB proteins with synthetic Holliday junctions. *Proc Natl Acad Sci USA* **89**: 5452–5456.

Patel SS, Donmez I. 2006. Mechanisms of helicases. *J Biol Chem* **281**: 18265–18268.

Patel SS, Picha KM. 2000. Structure and function of hexameric helicases. *Annu Rev Biochem* **69**: 651–697.

Paul K, Nonoh JO, Mikulski L, Brune A. 2012. Methanoplasmatales," Thermoplasmatales-related archaea in termite guts and other environments, are the seventh order of methanogens. *Appl Environ Microbiol* **78**: 8245–8253.

Pause A, Sonenberg N. 1992. Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. *EMBO J* **11**: 2643–2654.

Pause A, Sonenberg N. 1993. RNA helicase activity in translation initiation in eukaryotes. In *The Translational Apparatus* (eds. K.H. Nierhaus, F. Franceschi, A.R. Subramanian, V.A. Erdmann, and B. Wittmann-Liebold), pp. 221–228, Springer US, Boston, MA.

Peeters E, Charlier D. 2010. The Lrp family of transcription regulators in archaea. *Archaea* **2010**: 750457.

Peeters E, Driessen RPC, Werner F, Dame RT. 2015. The interplay between nucleoid organization and transcription in archaeal genomes. *Nat Rev Microbiol* **13**: 333–341.

Pefanis E, Wang J, Rothschild G, Lim J, Kazadi D, Sun J, Federation A, Chao J, Elliott O, Liu Z-P, et al. 2015. RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell* **161**: 774–789.

Perevalova AA, Bidzhieva SK, Kublanov IV, Hinrichs K-U, Liu XL, Mardanov AV, Lebedinsky AV, Bonch-Osmolovskaya EA. 2010. Fervidicoccus fontis gen. nov., sp. nov., an anaerobic, thermophilic crenarchaeote from terrestrial hot springs, and proposal of Fervidicoccaceae fam. nov. and Fervidicoccales ord. nov. *Int J Syst Evol Microbiol* **60**: 2082–2088.

Pérez-Rueda E, Janga SC. 2010. Identification and genomic analysis of transcription factors in archaeal genomes exemplifies their functional architecture and evolutionary origin. *Mol Biol Evol* **27**: 1449–1459.

Phung DK, Etienne C, Batista M, Langendijk-Genevaux P, Moalic Y, Laurent S, Liuu S, Morales V, Jebbar M, Fichant G, et al. 2020. RNA processing machineries in Archaea: the 5'-3' exoribonuclease aRNase J of the β-CASP family is engaged specifically with the helicase ASH-Ski2 and the 3'-5' exoribonucleolytic RNA exosome machinery. *Nucleic Acids Res* **48**: 3832–3847.

Phung DK, Rinaldi D, Langendijk-Genevaux PS, Quentin Y, Carpousis AJ, Clouet-d'Orval B. 2013. Archaeal β-CASP ribonucleases of the aCPSF1 family are orthologs of the eukaryal CPSF-73 factor. *Nucleic Acids Res* **41**: 1091–1103.

Pivovarova TA, Kondrat'eva TF, Batrakov SG, Esipov SE, Sheĭchenko VI, Bykova SA, Lysenko AM, Karavaĭko GI. 2002. [Phenotypic features of Ferroplasma acidiphilum strains Yt and Y-2]. *Mikrobiologiia* **71**: 809–818.

Pluchon P-F, Fouqueau T, Crezé C, Laurent S, Briffotaux J, Hogrel G, Palud A, Henneke G, Godfroy A, Hausner W, et al. 2013. An extended network of genomic maintenance in the archaeon Pyrococcus abyssi highlights unexpected associations between eucaryotic homologs. *PLoS One* **8**: e79707.

Pollard TD, Cooper JA. 2009. Actin, a central player in cell shape and movement. *Science* **326**: 1208–1212.

Pomerantz RT, O'Donnell M. 2010. Direct restart of a replication fork stalled by a head-on RNA polymerase. *J Vis Exp*.

Portnoy V, Schuster G. 2006. RNA polyadenylation and degradation in different Archaea; roles of the exosome and RNase R. *Nucleic Acids Res* **34**: 5923–5931.

Prieur D, Erauso G, Geslin C, Lucas S, Gaillard M, Bidault A, Mattenet AC, Rouault K, Flament D, Forterre P, et al. 2004. Genetic elements of Thermococcales. *Biochem Soc Trans* **32**: 184–187.

Probst AJ, Auerbach AK, Moissl-Eichinger C. 2013. Archaea on human skin. *PLoS One* **8**: e65388.

Prokofeva MI, Kostrikina NA, Kolganova TV, Tourova TP, Lysenko AM, Lebedinsky AV, Bonch-Osmolovskaya EA. 2009. Isolation of the anaerobic thermoacidophilic crenarchaeote Acidilobus saccharovorans sp. nov. and proposal of Acidilobales ord. nov., including Acidilobaceae fam. nov. and Caldisphaeraceae fam. nov. *Int J Syst Evol Microbiol* **59**: 3116–3122.

Pugh RA, Honda M, Leesley H, Thomas A, Lin Y, Nilges MJ, Cann IKO, Spies M. 2008. The iron-containing domain is essential in Rad3 helicases for coupling of ATP hydrolysis to DNA translocation and for targeting the helicase to the single-stranded DNA-double-stranded DNA junction. *J Biol Chem* **283**: 1732–1743.

Pühler G, Leffers H, Gropp F, Palm P, Klenk HP, Lottspeich F, Garrett RA, Zillig W. 1989. Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. *Proc Natl Acad Sci USA* **86**: 4569–4573.

Py B, Higgins CF, Krisch HM, Carpousis AJ. 1996. A DEAD-box RNA helicase in the Escherichia coli RNA degradosome. *Nature* **381**: 169–172.

Pyle AM. 2008. Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu Rev Biophys* **37**: 317–336.

Quax TEF, Altegoer F, Rossi F, Li Z, Rodriguez-Franco M, Kraus F, Bange G, Albers S-V. 2018. Structure and function of the archaeal response regulator CheY. *Proc Natl Acad Sci USA* **115**: E1259–E1268.

Qureshi SA, Bell SD, Jackson SP. 1997. Factor requirements for transcription in the Archaeon Sulfolobus shibatae. *EMBO J* **16**: 2927–2936.

Rajagopal V, Patel SS. 2008. Single strand binding proteins increase the processivity of DNA unwinding by the hepatitis C virus helicase. *J Mol Biol* **376**: 69–79.

Rand L, Hinds J, Springer B, Sander P, Buxton RS, Davis EO. 2003. The majority of inducible DNA repair genes in Mycobacterium tuberculosis are induced independently of RecA. *Mol Microbiol* **50**: 1031–1042.

Ray BK, Lawson TG, Kramer JC, Cladaras MH, Grifo JA, Abramson RD, Merrick WC, Thach RE. 1985. ATP-dependent unwinding of messenger RNA structure by eukaryotic initiation factors. *J Biol Chem* **260**: 7651–7658.

Raymann K, Forterre P, Brochier-Armanet C, Gribaldo S. 2014. Global phylogenomic analysis disentangles the complex evolutionary history of DNA replication in archaea. *Genome Biol Evol* **6**: 192–212.

Reno ML, Held NL, Fields CJ, Burke PV, Whitaker RJ. 2009. Biogeography of the Sulfolobus islandicus pan-genome. *Proc Natl Acad Sci USA* **106**: 8605–8610.

Reuven NB, Koonin EV, Rudd KE, Deutscher MP. 1995. The gene for the longest known Escherichia coli protein is a member of helicase superfamily II. *J Bacteriol* **177**: 5393–5400.

Richard P, Feng S, Manley JL. 2013. A SUMO-dependent interaction between Senataxin and the exosome, disrupted in the neurodegenerative disease AOA2, targets the exosome to sites of transcription-induced DNA damage. *Genes Dev* **27**: 2227–2232.

Richards E, Bouché L, Panico M, Arbeloa A, Vinogradov E, Morris H, Wren B, Logan SM, Dell A, Fairweather NF. 2018. The S-layer protein of a Clostridium difficile SLCT-11 strain displays a complex glycan required for normal cell growth and morphology. *J Biol Chem* **293**: 18123–18137.

Richards JD, Cubeddu L, Roberts J, Liu H, White MF. 2008a. The archaeal XPB protein is a ssDNA-dependent ATPase with a novel partner. *J Mol Biol* **376**: 634–644.

Richards JD, Johnson KA, Liu H, McRobbie A-M, McMahon S, Oke M, Carter L, Naismith JH, White MF. 2008b. Structure of the DNA repair helicase hel308 reveals DNA binding and autoinhibitory domains. *J Biol Chem* **283**: 5118–5126.

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling A, Malfatti S, Swan BK, Gies EA, et al. 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**: 431–437.

Rocak S, Linder P. 2004. DEAD-box proteins: the driving forces behind RNA metabolism. *Nat Rev Mol Cell Biol* **5**: 232–241.

Rössler OG, Straka A, Stahl H. 2001. Rearrangement of structured RNA via branch migration structures catalysed by the highly related DEAD-box proteins p68 and p72. *Nucleic Acids Res* **29**: 2088–2096.

Rouillon C, White MF. 2010. The XBP-Bax1 helicase-nuclease complex unwinds and cleaves DNA: implications for eukaryal and archaeal nucleotide excision repair. *J Biol Chem* **285**: 11013–11022.

Rout MP, Field MC. 2017. The evolution of organellar coat complexes and organization of the eukaryotic cell. *Annu Rev Biochem* **86**: 637–657.

Rowlands T, Baumann P, Jackson SP. 1994. The TATA-binding protein: a general transcription factor in eukaryotes and archaebacteria. *Science* **264**: 1326–1329.

Roychowdhury A, Szymanski MR, Jezewska MJ, Bujalowski W. 2009. Mechanism of NTP hydrolysis by the Escherichia coli primary replicative helicase DnaB protein. 2. Nucleotide and nucleic acid specificities. *Biochemistry* **48**: 6730–6746.

Rudolf J, Makrantoni V, Ingledew WJ, Stark MJR, White MF. 2006. The DNA repair helicases XPD and FancJ have essential iron-sulfur domains. *Mol Cell* **23**: 801–808.

Sachsenmaier N, Waldsich C. 2013. Mss116p: a DEAD-box protein facilitates RNA folding. *RNA Biol* **10**: 71–82.

Saikrishnan K, Powell B, Cook NJ, Webb MR, Wigley DB. 2009. Mechanistic basis of 5’-3' translocation in SF1B helicases. *Cell* **137**: 849–859.

Sakai S, Imachi H, Hanada S, Ohashi A, Harada H, Kamagata Y. 2008. Methanocella paludicola gen. nov., sp. nov., a methane-producing archaeon, the first isolate of the lineage “Rice Cluster I”, and proposal of the new archaeal order Methanocellales ord. nov. *Int J Syst Evol Microbiol* **58**: 929–936.

Sakakibara N, Kelman LM, Kelman Z. 2009. How is the archaeal MCM helicase assembled at the origin? Possible mechanisms. *Biochem Soc Trans* **37**: 7–11.

Sanders CM. 2010. Human Pif1 helicase is a G-quadruplex DNA-binding protein with G-quadruplex DNA-unwinding activity. *Biochem J* **430**: 119–128.

Sanders TJ, Wenck BR, Selan JN, Barker MP, Trimmer SA, Walker JE, Santangelo TJ. 2020. FttA is a CPSF73 homologue that terminates transcription in Archaea. *Nat Microbiol* **5**: 545–553.

Schlacht A, Dacks JB. 2015. Unexpected ancient paralogs and an evolutionary model for the COPII coat complex. *Genome Biol Evol* **7**: 1098–1109.

Schleper C, Sousa FL. 2020. Meet the relatives of our cellular ancestor. *Nature* **577**: 478–479.

Schmid SR, Linder P. 1991. Translation initiation factor 4A from Saccharomyces cerevisiae: analysis of residues conserved in the D-E-A-D family of RNA helicases. *Mol Cell Biol* **11**: 3463–3471.

Schmitt E, Coureux P-D, Monestier A, Dubiez E, Mechulam Y. 2019. Start codon recognition in eukaryotic and archaeal translation initiation: A common structural core. *Int J Mol Sci* **20**.

Schönheit P, Buckel W, Martin WF. 2016. On the origin of heterotrophy. *Trends Microbiol* **24**: 12–25.

Scott JF, Eisenberg S, Bertsch LL, Kornberg A. 1977. A mechanism of duplex DNA replication revealed by enzymatic studies of phage phi X174: catalytic strand separation in advance of replication. *Proc Natl Acad Sci USA* **74**: 193–197.

Shahapure R, Driessen RPC, Haurat MF, Albers S-V, Dame RT. 2014. The archaellum: a rotating type IV pilus. *Mol Microbiol* **91**: 716–723.

Shen L, Pelletier J. 2020. General and Target-Specific DExD/H RNA Helicases in Eukaryotic Translation Initiation. *Int J Mol Sci* **21**.

Sheridan PO, Raguideau S, Quince C, Thames Consortium, Williams TA, Gubry-Rangin C. 2020. Gene duplication drives genome expansion in Thaumarchaeota. *BioRxiv*.

Shi Y, Huang Z, Han S, Fan S, Yang H. 2015. Phylogenetic diversity of Archaea in the intestinal tract of termites from different lineages. *J Basic Microbiol* **55**: 1021–1028.

Shibuya T, Tange TØ, Sonenberg N, Moore MJ. 2004. eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay. *Nat Struct Mol Biol* **11**: 346–351.

Shih JW, Tsai TY, Chao CH, Wu Lee YH. 2008. Candidate tumor suppressor DDX3 RNA helicase specifically represses cap-dependent translation by acting as an eIF4E inhibitory protein. *Oncogene* **27**: 700–714.

Shin DS, Pratt AJ, Tainer JA. 2014. Archaeal genome guardians give insights into eukaryotic DNA replication and damage response proteins. *Archaea* **2014**: 206735.

Shin J-H, Jiang Y, Grabowski B, Hurwitz J, Kelman Z. 2003. Substrate requirements for duplex DNA translocation by the eukaryal and archaeal minichromosome maintenance helicases. *J Biol Chem* **278**: 49053–49062.

Shin J-H, Santangelo TJ, Xie Y, Reeve JN, Kelman Z. 2007. Archaeal minichromosome maintenance (MCM) helicase can unwind DNA bound by archaeal histones and transcription factors. *J Biol Chem* **282**: 4908–4915.

Siliakus MF, van der Oost J, Kengen SWM. 2017. Adaptations of archaeal and bacterial membranes to variations in temperature, pH and pressure. *Extremophiles* **21**: 651–670.

Singh SP, Soranno A, Sparks MA, Galletto R. 2019. Branched unwinding mechanism of the Pif1 family of DNA helicases. *Proc Natl Acad Sci USA* **116**: 24533–24541.

Singleton MR, Dillingham MS, Wigley DB. 2007. Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem* **76**: 23–50.

Singleton MR, Scaife S, Wigley DB. 2001. Structural analysis of DNA replication fork reversal by RecG. *Cell* **107**: 79–89.

Singleton MR, Wigley DB. 2002. Modularity and specialization in superfamily 1 and 2 helicases. *J Bacteriol* **184**: 1819–1826.

Singleton MR, Wigley DB. 2003. Multiple roles for ATP hydrolysis in nucleic acid modifying enzymes. *EMBO J* **22**: 4579–4583.

Skourti-Stathaki K, Proudfoot NJ. 2014. A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes Dev* **28**: 1384–1396.

Skourti-Stathaki K, Proudfoot NJ, Gromak N. 2011. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell* **42**: 794–805.

Smollett K, Blombach F, Fouqueau T, Werner F. 2017. A global characterisation of the archaeal transcription machinery. In *RNA metabolism and gene expression in archaea* (ed. B. Clouet-d'Orval), Vol. 32 of *Nucleic acids and molecular biology*, pp. 1–26, Springer International Publishing, Cham.

Snelling TJ, Genç B, McKain N, Watson M, Waters SM, Creevey CJ, Wallace RJ. 2014. Diversity and community composition of methanogenic archaea in the rumen of Scottish upland sheep assessed by different methods. *PLoS One* **9**: e106491.

Song H, Ji X. 2019. The mechanism of RNA duplex recognition and unwinding by DEAD-box helicase DDX3X. *Nat Commun* **10**: 3085.

Song X, Huang Q, Ni J, Yu Y, Shen Y. 2016. Knockout and functional analysis of two DExD/H-box family helicase genes in Sulfolobus islandicus REY15A. *Extremophiles* **20**: 537–546.

Soni RK, Mehra P, Choudhury NR, Mukhopadhyay G, Dhar SK. 2003. Functional characterization of Helicobacter pylori DnaB helicase. *Nucleic Acids Res* **31**: 6828–6840.

Soto-Rifo R, Rubilar PS, Limousin T, de Breyne S, Décimo D, Ohlmann T. 2012. DEAD-box protein DDX3 associates with eIF4F to promote translation of selected mRNAs. *EMBO J* **31**: 3745–3756.

Spang A, Caceres EF, Ettema TJG. 2017. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science* **357**.

Spang A, Eme L, Saw JH, Caceres EF, Zaremba-Niedzwiedzka K, Lombard J, Guy L, Ettema TJG. 2018. Asgard archaea are the closest prokaryotic relatives of eukaryotes. *PLoS Genet* **14**: e1007080.

Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, Lind AE, van Eijk R, Schleper C, Guy L, Ettema TJG. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**: 173–179.

Spies M. 2014. Two steps forward, one step back: determining XPD helicase mechanism by single-molecule fluorescence and high-resolution optical tweezers. *DNA Repair (Amst)* **20**: 58–70.

Sriskanda V, Shuman S. 1998. Mutational analysis of Chlorella virus DNA ligase: catalytic roles of domain I and motif VI. *Nucleic Acids Res* **26**: 4618–4625.

Stahl H, Dröge P, Knippers R. 1986. DNA helicase activity of SV40 large tumor antigen. *EMBO J* **5**: 1939–1944.

Stairs CW, Ettema TJG. 2020. The archaeal roots of the eukaryotic dynamic actin cytoskeleton. *Curr Biol* **30**: R521–R526.

Stein LY, Klotz MG. 2011. Research on nitrification and related processes, Part B. Preface. *Meth Enzymol* **496**: xix–xx.

Stelter M, Acajjaoui S, McSweeney S, Timmins J. 2013. Structural and mechanistic insight into DNA unwinding by Deinococcus radiodurans UvrD. *PLoS One* **8**: e77364.

Story RM, Steitz TA. 1992. Structure of the recA protein-ADP complex. *Nature* **355**: 374–376.

Story RM, Weber IT, Steitz TA. 1992. The structure of the E. coli recA protein monomer and polymer. *Nature* **355**: 318–325.

Studer MK, Ivanović L, Weber ME, Marti S, Jonas S. 2020. Structural basis for DEAH-helicase activation by G-patch proteins. *Proc Natl Acad Sci USA* **117**: 7159–7170.

Subramanya HS, Bird LE, Brannigan JA, Wigley DB. 1996. Crystal structure of a DExx box DNA helicase. *Nature* **384**: 379–383.

Sugino A, Ryu BH, Sugino T, Naumovski L, Friedberg EC. 1986. A new DNA-dependent ATPase which stimulates yeast DNA polymerase I and has DNA-unwinding activity. *J Biol Chem* **261**: 11744–11750.

Takai K, Sugai A, Itoh T, Horikoshi K. 2000. Palaeococcus ferrophilus gen. nov., sp. nov., a barophilic, hyperthermophilic archaeon from a deep-sea hydrothermal vent chimney. *Int J Syst Evol Microbiol* **50 Pt 2**: 489–500.

Tanaka N, Aronova A, Schwer B. 2007. Ntr1 activates the Prp43 helicase to trigger release of lariat-intron from the spliceosome. *Genes Dev* **21**: 2312–2325.

Tanner NK. 2003. The newly identified Q motif of DEAD box helicases is involved in adenine recognition. *Cell Cycle* **2**: 18–19.

Tanner NK, Cordin O, Banroques J, Doère M, Linder P. 2003. The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. *Mol Cell* **11**: 127–138.

Tanner NK, Linder P. 2001. DExD/H box RNA helicases: from generic motors to specific dissociation functions. *Mol Cell* **8**: 251–262.

Tarasov VY, Pyatibratov MG, Tang SL, Dyall-Smith M, Fedorov OV. 2000. Role of flagellins from A and B loci in flagella formation of Halobacterium salinarum. *Mol Microbiol* **35**: 69–78.

Tarn W-Y, Chang T-H. 2009. The current understanding of Ded1p/DDX3 homologs from yeast to human. *RNA Biol* **6**: 17–20.

Teichmann M, Dumay-Odelot H, Fribourg S. 2012. Structural and functional aspects of winged-helix domains at the core of transcription initiation complexes. *Transcription* **3**: 2–7.

Thiel A, Michoud G, Moalic Y, Flament D, Jebbar M. 2014. Genetic manipulations of the hyperthermophilic piezophilic archaeon Thermococcus barophilus. *Appl Environ Microbiol* **80**: 2299–2306.

Thomsen ND, Berger JM. 2009. Running in reverse: the structural basis for translocation polarity in hexameric helicases. *Cell* **139**: 523–534.

Thomsen ND, Berger JM. 2008. Structural frameworks for considering microbial protein- and nucleic acid-dependent motor ATPases. *Mol Microbiol* **69**: 1071–1090.

Tirode F, Busso D, Coin F, Egly JM. 1999. Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Mol Cell* **3**: 87–95.

Tourte M, Schaeffer P, Grossi V, Oger PM. 2020. Functionalized membrane domains: an ancestral feature of archaea? *Front Microbiol* **11**: 526.

Tuteja N, Tuteja R, Rahman K, Kang LY, Falaschi A. 1990. A DNA helicase from human cells. *Nucleic Acids Res* **18**: 6785–6792.

Tuteja R, Malhotra P, Song P, Tuteja N, Chauhan VS. 2002. Isolation and characterization of an eIF-4A homologue from Plasmodium cynomolgi. *Mol Biochem Parasitol* **124**: 79–83.

Valenti A, De Felice M, Perugino G, Bizard A, Nadal M, Rossi M, Ciaramella M. 2012. Synergic and opposing activities of thermophilic RecQ-like helicase and topoisomerase 3 proteins in Holliday junction processing and replication fork stabilization. *J Biol Chem* **287**: 30282–30295.

Van Houten B, Croteau DL, DellaVecchia MJ, Wang H, Kisker C. 2005. Close-fitting sleeves ': DNA damage recognition by the UvrABC nuclease system. *Mutat Res* **577**: 92–117.

Vannier P, Marteinsson VT, Fridjonsson OH, Oger P, Jebbar M. 2011. Complete genome sequence of the hyperthermophilic, piezophilic, heterotrophic, and carboxydotrophic archaeon Thermococcus barophilus MP. *J Bacteriol* **193**: 1481–1482.

Vannier P, Michoud G, Oger P, Marteinsson VÞ, Jebbar M. 2015. Genome expression of Thermococcus barophilus and Thermococcus kodakarensis in response to different hydrostatic pressure conditions. *Res Microbiol* **166**: 717–725.

Velankar SS, Soultanas P, Dillingham MS, Subramanya HS, Wigley DB. 1999. Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* **97**: 75–84.

Venkatesan M, Silver LL, Nossal NG. 1982. Bacteriophage T4 gene 41 protein, required for the synthesis of RNA primers, is also a DNA helicase. *J Biol Chem* **257**: 12426–12434.

Vieille C, Zeikus GJ. 2001. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev* **65**: 1–43.

Villarreal OD, Mersaoui SY, Yu Z, Masson J-Y, Richard S. 2020. Genome-wide R-loop analysis defines unique roles for DDX5, XRN2, and PRMT5 in DNA/RNA hybrid resolution. *Life Sci Alliance* **3**.

Viswanathan M, Dower KW, Lovett ST. 1998. Identification of a potent DNase activity associated with RNase T of Escherichia coli. *J Biol Chem* **273**: 35126–35131.

Wadsworth RI, White MF. 2001. Identification and properties of the crenarchaeal single-stranded DNA binding protein from Sulfolobus solfataricus. *Nucleic Acids Res* **29**: 914–920.

Wagner M, van Wolferen M, Wagner A, Lassak K, Meyer BH, Reimann J, Albers S-V. 2012. Versatile Genetic Tool Box for the Crenarchaeote Sulfolobus acidocaldarius. *Front Microbiol* **3**: 214.

Walker JE, Saraste M, Runswick MJ, Gay NJ. 1982. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J* **1**: 945–951.

Wang H, Peng N, Shah SA, Huang L, She Q. 2015. Archaeal extrachromosomal genetic elements. *Microbiol Mol Biol Rev* **79**: 117–152.

Wang X, Lee H-S, Sugar FJ, Jenney FE, Adams MWW, Prestegard JH. 2007. PF0610, a novel winged helix-turn-helix variant possessing a rubredoxin-like Zn ribbon motif from the hyperthermophilic archaeon, Pyrococcus furiosus. *Biochemistry* **46**: 752–761.

Wassarman DA, Steitz JA. 1991. Alive with DEAD proteins. *Nature* **349**: 463–464.

Webster MPJ, Jukes R, Zamfir VS, Kay CWM, Bagnéris C, Barrett T. 2012. Crystal structure of the UvrB dimer: insights into the nature and functioning of the UvrAB damage engagement and UvrB-DNA complexes. *Nucleic Acids Res* **40**: 8743–8758.

Welander PV, Metcalf WW. 2005. Loss of the mtr operon in Methanosarcina blocks growth on methanol, but not methanogenesis, and reveals an unknown methanogenic pathway. *Proc Natl Acad Sci USA* **102**: 10664–10669.

Werner F. 2012. A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life. *J Mol Biol* **417**: 13–27.

Werner F. 2013. Molecular mechanisms of transcription elongation in archaea. *Chem Rev* **113**: 8331–8349.

Werner F, Weinzierl ROJ. 2002. A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. *Mol Cell* **10**: 635–646.

White MF, Allers T. 2018. DNA repair in the archaea-an emerging picture. *FEMS Microbiol Rev* **42**: 514–526.

Wilkinson SP, Ouhammouch M, Geiduschek EP. 2010. Transcriptional activation in the context of repression mediated by archaeal histones. *Proc Natl Acad Sci USA* **107**: 6777–6781.

Williams TA, Cox CJ, Foster PG, Szöllősi GJ, Embley TM. 2020. Phylogenomics provides robust support for a two-domains tree of life. *Nat Ecol Evol* **4**: 138–147.

Witharana C, Roppelt V, Lochnit G, Klug G, Evguenieva-Hackenberg E. 2012. Heterogeneous complexes of the RNA exosome in Sulfolobus solfataricus. *Biochimie* **94**: 1578–1587.

Woese CR, Fox GE. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci USA* **74**: 5088–5090.

Wold MS. 1997. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu Rev Biochem* **66**: 61–92.

Wong IN, Sayers JR, Sanders CM. 2013. Characterization of an unusual bipolar helicase encoded by bacteriophage T5. *Nucleic Acids Res* **41**: 4587–4600.

Wurtzel O, Sapra R, Chen F, Zhu Y, Simmons BA, Sorek R. 2010. A single-base resolution map of an archaeal transcriptome. *Genome Res* **20**: 133–141.

Xie P. 2016. Dynamics of monomeric and hexameric helicases. *Biophys Chem* **211**: 49–58.

Xu X, Liu Y. 2009. Dual DNA unwinding activities of the Rothmund-Thomson syndrome protein, RECQ4. *EMBO J* **28**: 568–577.

Yang Q, Jankowsky E. 2005. ATP- and ADP-dependent modulation of RNA unwinding and strand annealing activities by the DEAD-box protein DED1. *Biochemistry* **44**: 13591–13601.

Yang Y, McBride KM, Hensley S, Lu Y, Chedin F, Bedford MT. 2014. Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation. *Mol Cell* **53**: 484–497.

Yao NY, O'Donnell ME. 2016. Evolution of replication machines. *Crit Rev Biochem Mol Biol* **51**: 135–149.

Yarranton GT, Gefter ML. 1979. Enzyme-catalyzed DNA unwinding: studies on Escherichia coli rep protein. *Proc Natl Acad Sci USA* **76**: 1658–1662.

Yoshimoto K, Arora K, Brooks CL. 2010. Hexameric helicase deconstructed: interplay of conformational changes and substrate coupling. *Biophys J* **98**: 1449–1457.

Yüce Ö, West SC. 2013. Senataxin, defective in the neurodegenerative disorder ataxia with oculomotor apraxia 2, lies at the interface of transcription and the DNA damage response. *Mol Cell Biol* **33**: 406–417.

Yusufzai T, Kadonaga JT. 2010. Annealing helicase 2 (AH2), a DNA-rewinding motor with an HNH motif. *Proc Natl Acad Sci USA* **107**: 20970–20973.

Zaitsev EN, Kowalczykowski SC. 2000. A novel pairing process promoted by Escherichia coli RecA protein: inverse DNA and RNA strand exchange. *Genes Dev* **14**: 740–749.

Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström D, Juzokaite L, Vancaester E, Seitz KW, Anantharaman K, Starnawski P, Kjeldsen KU, et al. 2017. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**: 353–358.

Zatopek KM, Gardner AF, Kelman Z. 2018. Archaeal DNA replication and repair: new genetic, biophysical and molecular tools for discovering and characterizing enzymes, pathways and mechanisms. *FEMS Microbiol Rev* **42**: 477–488.

Zhai B, DuPrez K, Han X, Yuan Z, Ahmad S, Xu C, Gu L, Ni J, Fan L, Shen Y. 2018. The archaeal ATPase PINA interacts with the helicase Hjm via its carboxyl terminal KH domain remodeling and processing replication fork and Holliday junction. *Nucleic Acids Res* **46**: 6627–6641.

Zhang C, Cooper TE, Krause DJ, Whitaker RJ. 2013a. Augmenting the genetic toolbox for Sulfolobus islandicus with a stringent positive selectable marker for agmatine prototrophy. *Appl Environ Microbiol* **79**: 5539–5549.

Zhang C, She Q, Bi H, Whitaker RJ. 2016. The apt/6-Methylpurine Counterselection System and Its Applications in Genetic Studies of the Hyperthermophilic Archaeon Sulfolobus islandicus. *Appl Environ Microbiol* **82**: 3070–3081.

Zhang H, Zhang Z, Yang J, He Z-G. 2014. Functional characterization of DnaB helicase and its modulation by single-stranded DNA binding protein in Mycobacterium tuberculosis. *FEBS J* **281**: 1256–1266.

Zhang S, Grosse F. 2004. Multiple functions of nuclear DNA helicase II (RNA helicase A) in nucleic acid metabolism. *Acta Biochim Biophys Sin (Shanghai)* **36**: 177–183.

Zhang X-P, Janke R, Kingsley J, Luo J, Fasching C, Ehmsen KT, Heyer W-D. 2013b. A conserved sequence extending motif III of the motor domain in the Snf2-family DNA translocase Rad54 is critical for ATPase activity. *PLoS One* **8**: e82184.

Zheng L, Meng Y, Campbell JL, Shen B. 2019. Multiple roles of DNA2 nuclease/helicase in DNA metabolism, genome stability and human diseases. *Nucleic Acids Res*.

Zhou Z, Pan J, Wang F, Gu J-D, Li M. 2018. Bathyarchaeota: globally distributed metabolic generalists in anoxic environments. *FEMS Microbiol Rev* **42**: 639–655.

Zhu Q, Mai U, Pfeiffer W, Janssen S, Asnicar F, Sanders JG, Belda-Ferre P, Al-Ghalith GA, Kopylova E, McDonald D, et al. 2019. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. *Nat Commun* **10**: 5477.

Zillig W, Holz I, Janekovic D, Schäfer W, Reiter WD. 1983. The Archaebacterium Thermococcus celer Represents, a Novel Genus within the Thermophilic Branch of the Archaebacteria. *Syst Appl Microbiol* **4**: 88–94.

Zinder SH, Dworkin M. 2006. Morphological and physiological diversity. In *The Prokaryotes* (eds. M. Dworkin, S. Falkow, E. Rosenberg, K.-H. Schleifer, and E. Stackebrandt), pp. 185–220, Springer New York, New York, NY.

Zong D, Oberdoerffer P, Batista PJ, Nussenzweig A. 2020. RNA: a double-edged sword in genome maintenance. *Nat Rev Genet*.

Zou Y, Liu Y, Wu X, Shell SM. 2006. Functions of human replication protein A (RPA): from DNA replication to DNA damage and stress responses. *J Cell Physiol* **208**: 267–273.

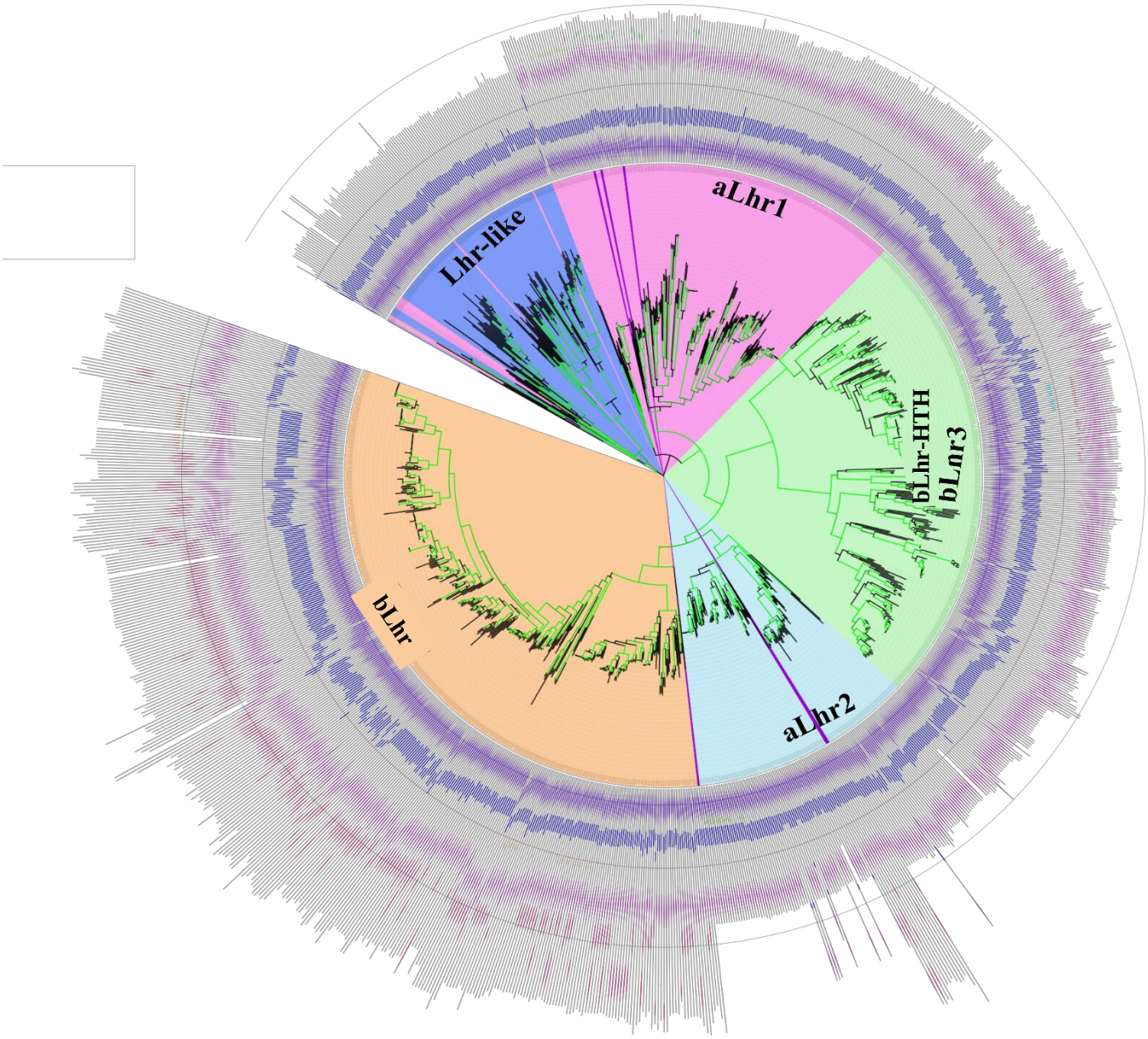
Zuo Y, Steitz TA. 2015. Crystal structures of the E. coli transcription initiation complexes with a complete bubble. *Mol Cell* **58**: 534–540.

Zuo Y, Zheng H, Wang Y, Chruszcz M, Cymborowski M, Skarina T, Savchenko A, Malhotra A, Minor W. 2007. Crystal structure of RNase T, an exoribonuclease involved in tRNA maturation and end turnover. *Structure* **15**: 417–428.

Zvyagintseva, I.S, Tarasov, A.L. 1988. Extreme halophilic bacteria from saline soils.

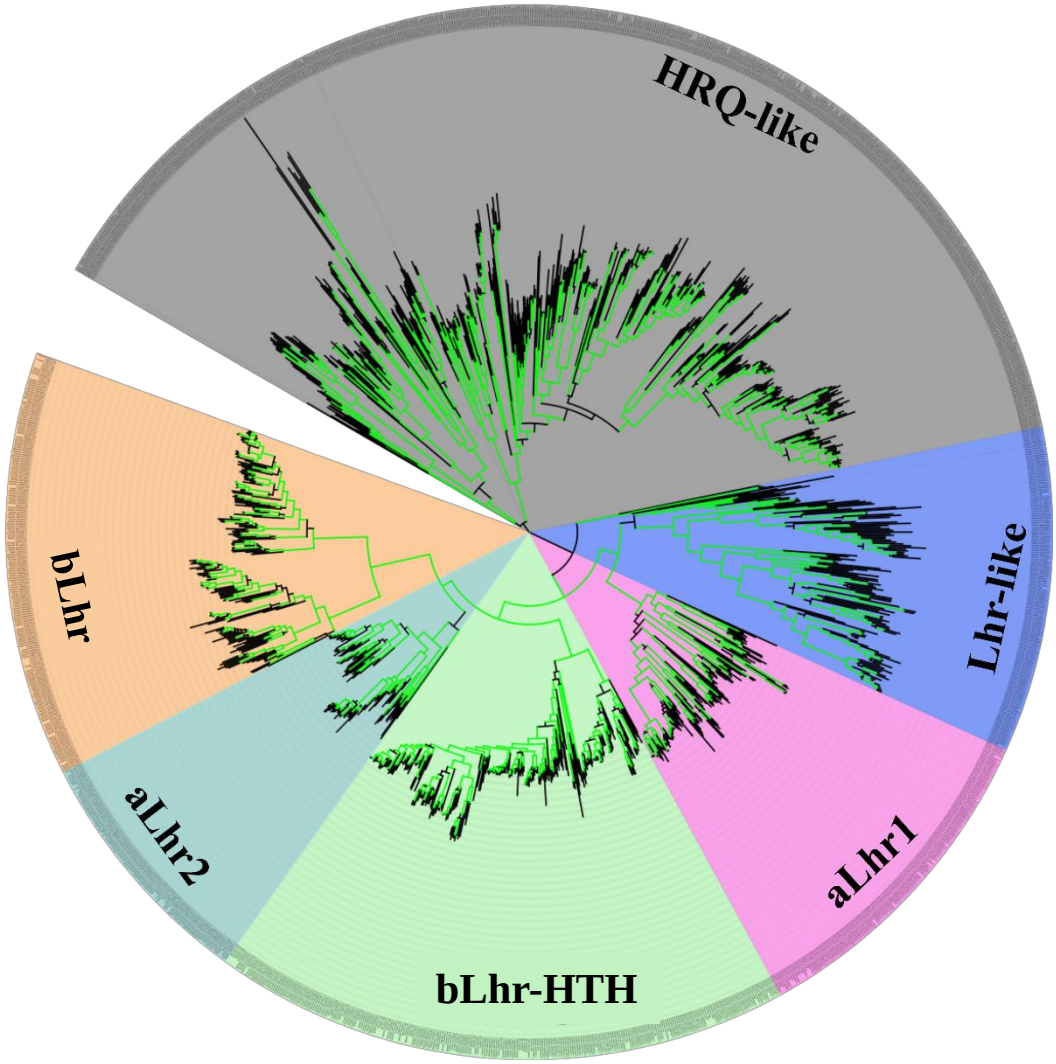
2011. The Genus Archaeoglobus. In *SpringerReference*, Springer-Verlag, Berlin/Heidelberg.

FIGURES ANNEXE



ANNEX 1. Phylogenetic tree of the Lhr-type orthologous groups together with protein sequences retrieved in the Asgard phylum.

The maximum likelihood tree computed using PhyML. A set of the 829 Lhr-type sequences dataset including 6 Asgard sequences was used. The resulted tree shows the segregation of the Asgard Lhr-type into the two clusters aLhr1 and aLhr2. The tree branches are supported by a high bootstrap value (parametric bootstrap supports above 0,7 are represented in green).

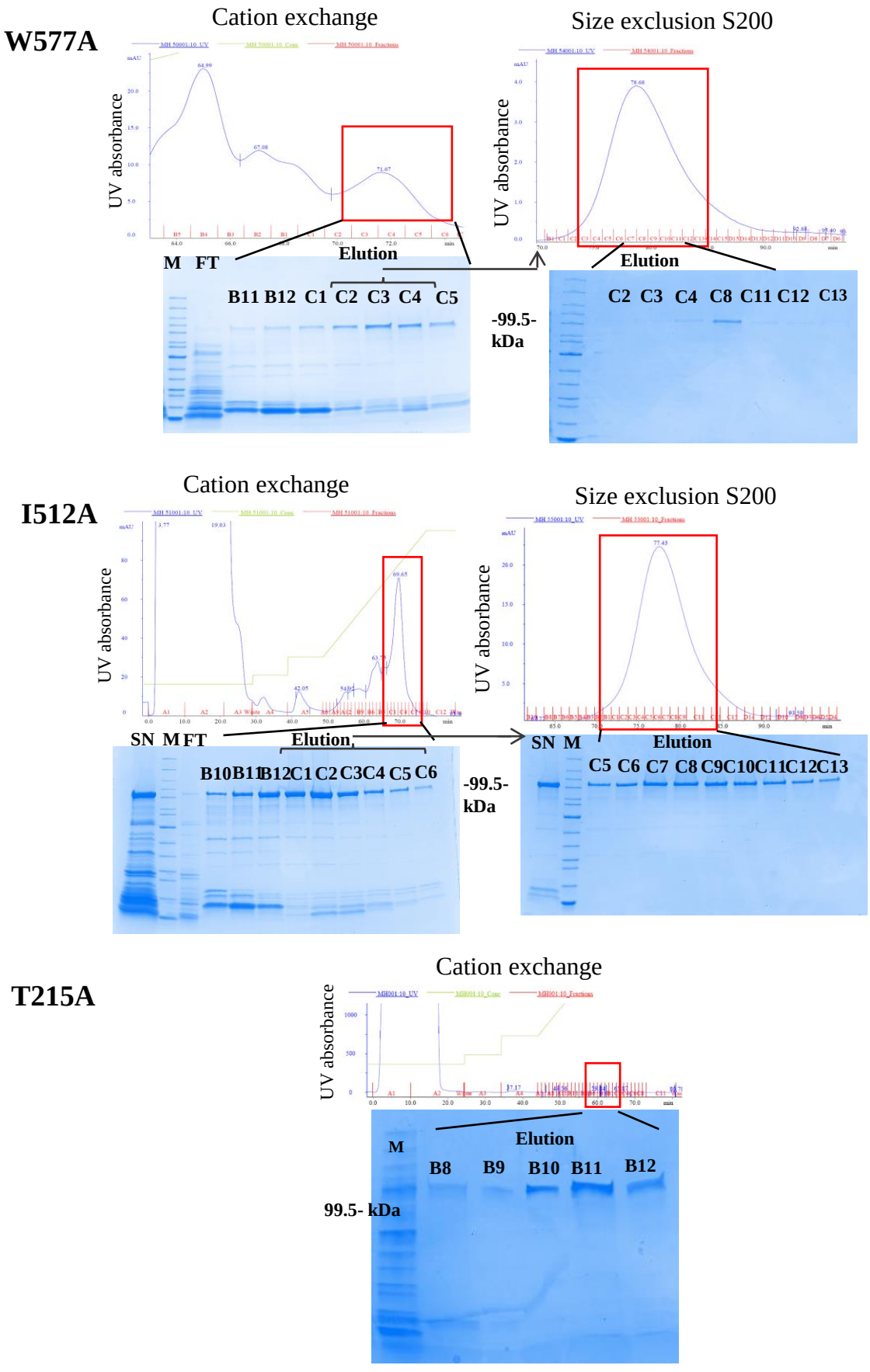


ANNEX 2. Phylogenetic tree of the Lhr-type orthologous groups with HRQ-like helicase as outgroup.

The maximum likelihood tree has been computed using PhyML. A set of 823 Lhr-type and 536 HRQ-like sequences dataset are used. The resulted tree shows the five Lhr-type families and confirmed that the Lhr-like family belongs to the Lhr-type family of helicase despite the uncompleted Lhr core (lacking the Domain 4). The tree branches are supported by a high bootstrap value (parametric bootstrap supports above 0,7 are represented in green).



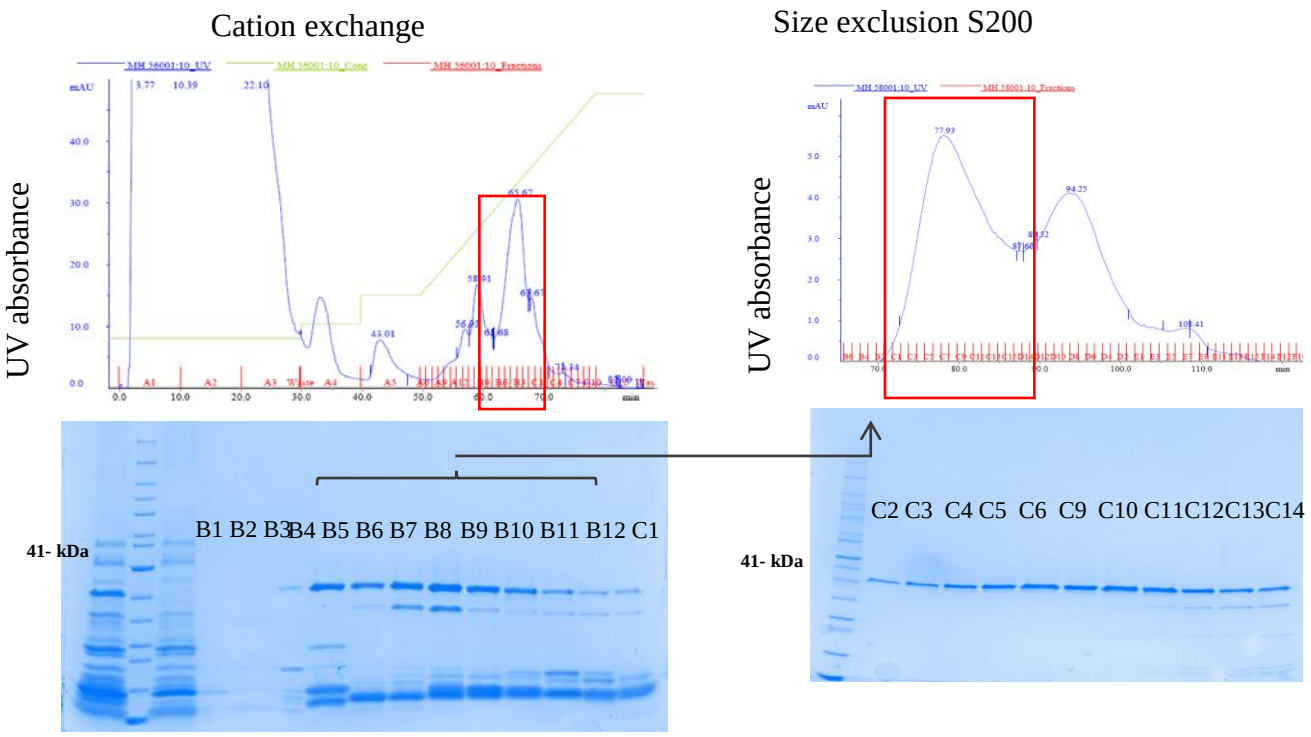
ANNEX 3. Sequence alignment of Lhr-type families.
Clustal omega tool (https://www.ebi.ac.uk/Tools/msa/clustalo/) was used for multiple sequence alignment of *Scoe*-Lhr-like, *Pbay*-aLhr1, *Paby*-aLhr2, *Pput*-bLhr, and *Ecol*-Lhr-HTH. Conserved motifs and domains location are indicated.



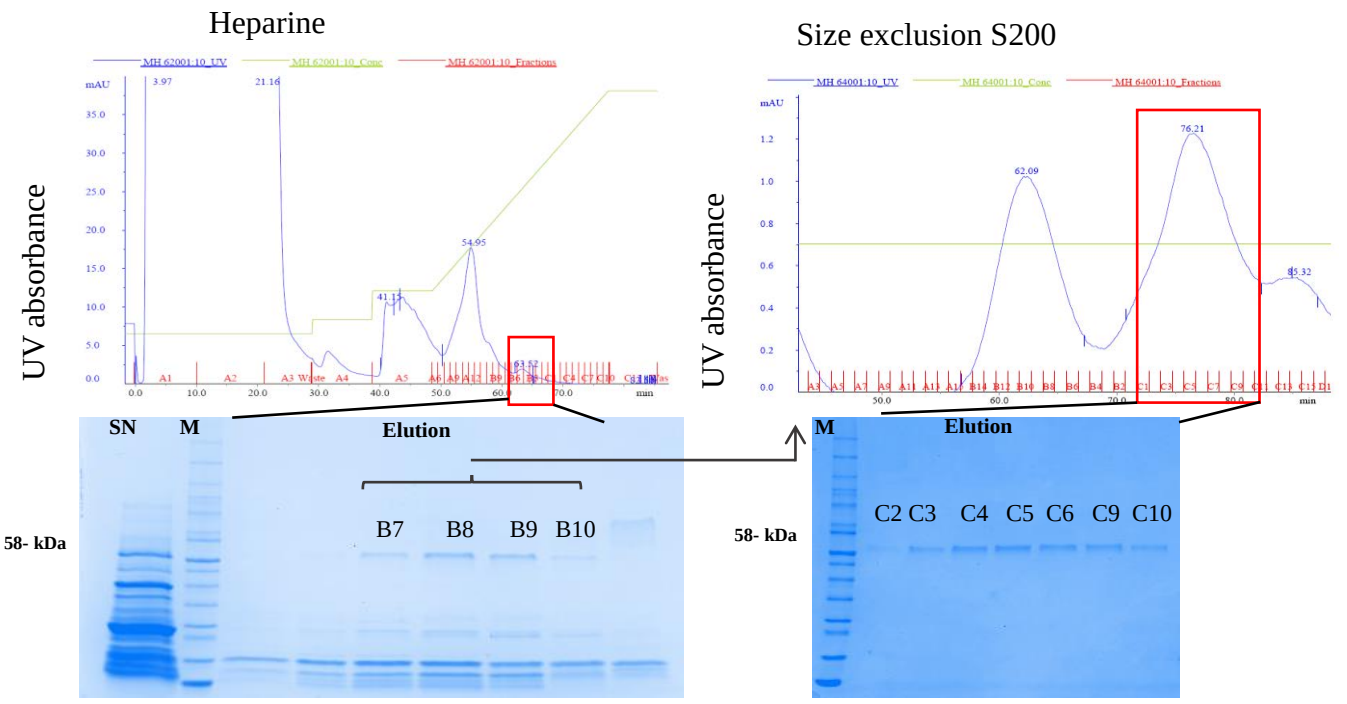
ANNEX 4. Chromatogram and SDS PAGE of purification steps of aLhr2 W557A, I512A and T215A.

Purification steps of aLhr2 W577A, I512A, and T215A by cation exchange & size-exclusion chromatographies. Chromatograms and the corresponding SDS PAGE showing the SN (soluble fraction), FT (Flow through) and the elution fractions visualized by staining with Coomassie blue were shown.

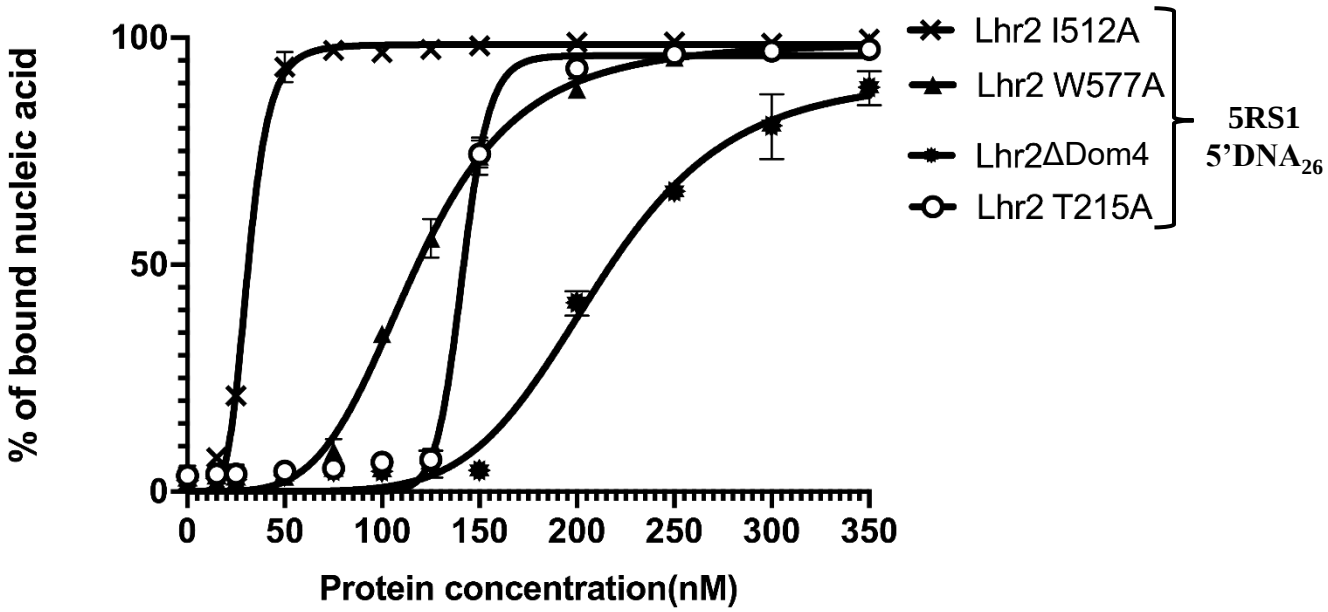
Dom4

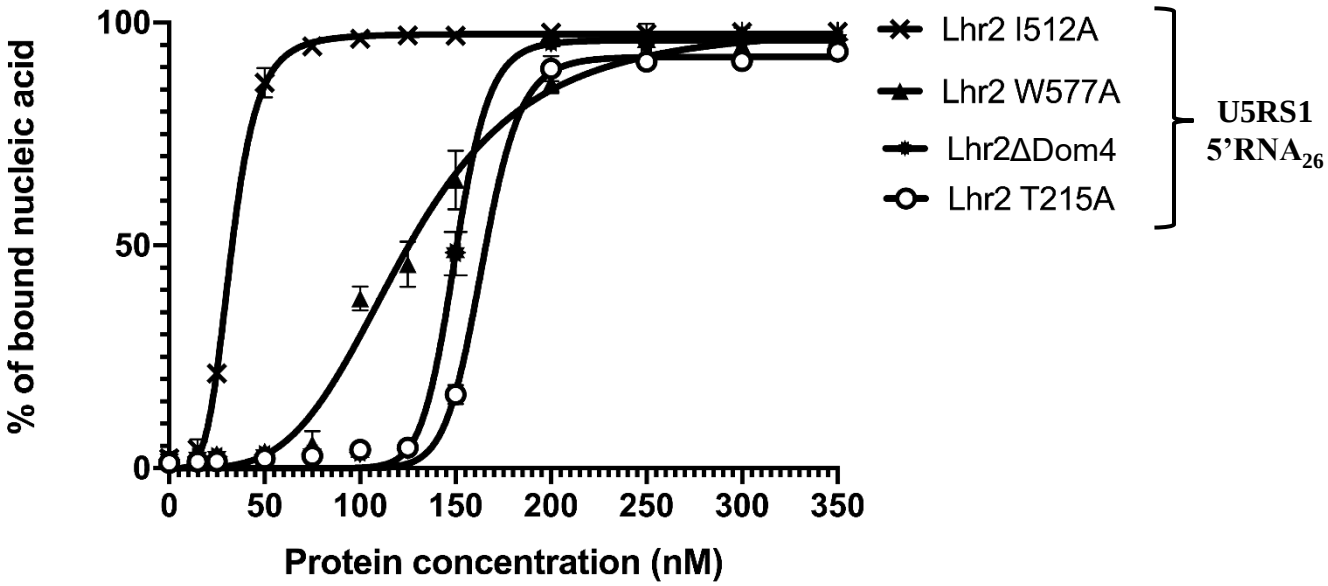


ΔDom4



ANNEX 5. Chromatogram and SDS PAGE of purification steps of aLhr2 Dom4 and ΔDom4. Purification steps of of *Tbar*-aLhr2 dom4 and ΔDom4 by cation exchange/heparin & size-exclusion chromatographies. Chromatograms and the corresponding SDS PAGE showing the SN (soluble fraction), FT (Flow through) and the elution fractions visualized by staining with coomassie blue. Elution fractions after Size-exclusion S200 step were pooled and concentrated.

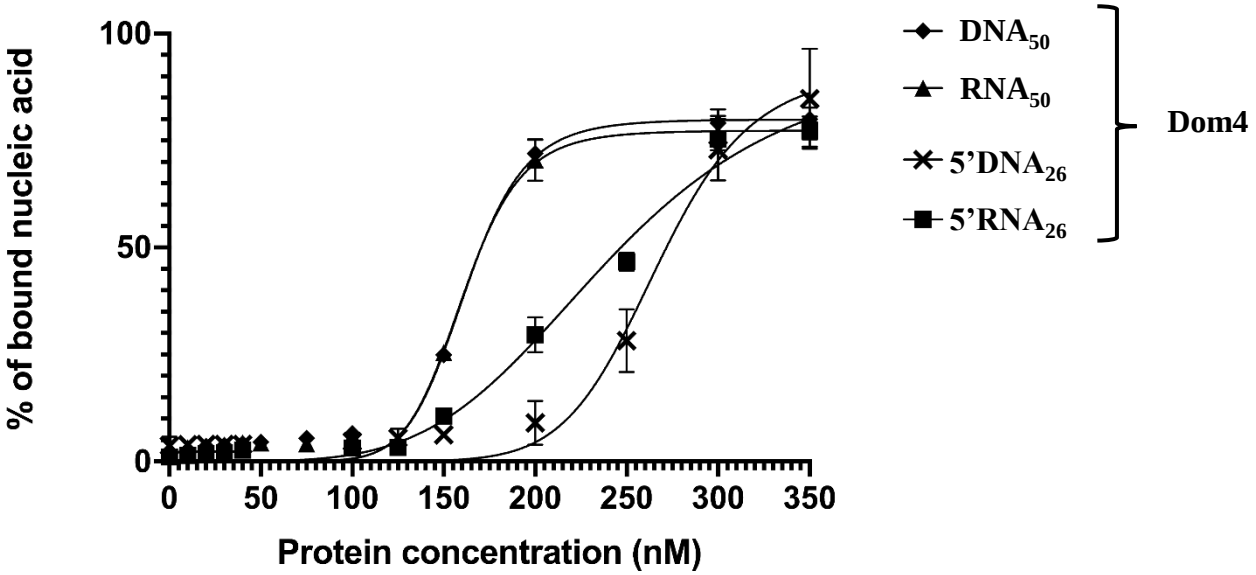




ANNEX 6. Binding affinity of aLhr2 variants with 26nt substrate.

The percentage of RNP complex formation is plotted versus protein concentration to calculate the apparent affinity constant (KD). Up panel: DNA_{26} (5RS1); Lowe panel: RNA_{26} (U5RS1) are used as substrates.

Values are the average of three independent experiments.



ANNEX 7. Binding affinity of aLhr2 Dom4.
The percentage of RNP complex formation is plotted versus protein concentration to calculate the apparent affinity constant (KD).
Values are the average of three independent experiments.