MASTERARBEIT | MASTER'S THESIS

Titel | Title

Chromosome organisation by nucleoid-associated proteins in Ca. Lokiarchaeum ossiferum

verfasst von | submitted by

Grdo. Arnau Sabater i Mezquita

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 830
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Molecular Microbiology, Microbial Ecology and Immunobiology
Betreut von | Supervisor:	Assoz. Prof. Dr. Silvia Bulgheresi Privatdoz.

Table of contents

Zusammenfassung

Die Entstehung der Eukaryoten, die sich durch die strukturelle und funktionelle Komplexität ihrer genetischen Organisation auszeichnen, ist nach wie vor eine der wichtigsten ungelösten Fragen der Evolutionsbiologie. Als die engsten bekannten prokaryotischen Verwandten der Eukaryoten bieten die Mitglieder des Archaeen-Königreichs Promethearchaeati (früher bekannt als Asgard Archaeen) einen einzigartigen Einblick in die frühen Ereignisse der Eukaryogenese. Das Verständnis der Chromatinorganisation in diesen Archaeen ist der Schlüssel zur Aufdeckung der Entwicklung der Komplexität des eukaryotischen Chromatins. In Prokaryonten wird die Chromatinstruktur weitgehend von Nukleoid-assoziierten Proteinen (NAPs) geformt. In *Candidatus* Lokiarchaeum ossiferum, dem einzigen in Europa kultivierten Promethearchaeon, wurden drei NAP-kodierende Gene und Proteine identifiziert: Alba, ein konservierter archäischer NAP, der auch bei der RNA-Regulierung in Eukaryonten eine Rolle spielt, und zwei Homologe von TrmBL2, einem archäenspezifischen DNA-bindenden Protein mit potenziellen chaperonartigen Funktionen. In dieser Arberit präsentiere ich die erste biochemische Charakterisierung von rekombinantem *Ca.* L. ossiferum Alba und TrmBL2b, einschließlich die Analyse der Nukleinsäurebindungseigenschaften durch Massenphotometrie, Mikrokokkalen Nuklease (MN)-Verdau und Mikroskopie. Unsere Ergebnisse deuten darauf hin, dass rekombinantes LoAlba *in vitro* bevorzugt RNA statt DNA bindet und *in vivo* zur Chromatinkompaktion beitragen kann. Rekombinantes LoTrmBL2b hingegen bindet *in vitro* mit hoher Affinität an DNA. Diese Ergebnisse bilden die Grundlage für weitere Studien über die Rolle der NAPs bei der Chromatinarchitektur in Promethearchaea, ihre Wechselwirkungen mit anderen Strukturproteinen, die Bildung von Chromatindomänen höherer Ordnung und ihre mögliche Beteiligung an der Regulierung der Genexpression.

Abstract

The emergence of eukaryotes – characterised by the structural and functional complexity of their genetic organization – remains one of the major unresolved questions in evolutionary biology. As the closest known prokaryotic relatives of eukaryotes, members of the archaeal kingdom Promethearchaeati (formerly known as Asgard archaea) offer a unique window into the early events of eukaryogenesis. Understanding chromatin organization in these archaea is key to uncovering how eukaryotic chromatin complexity evolved. In prokaryotes, chromatin structure is largely shaped by Nucleoid-Associated Proteins (NAPs). Three NAP-encoding genes have been identified in *Candidatus* Lokiarchaeum ossiferum, the only cultivated Promethearchaeon in Europe: Alba, a conserved archaeal NAP also implicated in RNA regulation in eukaryotes, and two homologues of TrmBL2, an archaeal-specific DNA-binding protein with potential chaperone-like functions. In this thesis, I present the first biochemical characterisation of recombinant *Ca.* L. ossiferum Alba and TrmBL2b. This includes the analysis of their nucleic acid-binding properties by Electrophoretic Mobility Shift Assays, Mass Photometry, Micrococcal Nuclease (MN) Digestion, and microscopy. Our findings indicate that recombinant LoAlba preferentially binds RNA over DNA *in vitro* and may contribute to chromatin compaction *in vivo*. Recombinant LoTrmBL2b, by contrast, binds DNA with high affinity *in vitro*. These results lay the groundwork for further studies into the roles of NAPs in chromatin architecture in Promethearchaeati, their interactions with other chromatin proteins, the formation of higher-order chromatin domains, and their potential involvement in the regulation of gene expression.

Acknowledgements

I would like to begin by expressing my sincere gratitude to Prof. Christa Schleper and Prof. Silvia Bulgheresi for giving me the opportunity to be part of the EvoChromo project and for their trust and support throughout the development of my Master's thesis. This thesis would not have been possible without the constant guidance of Dr. Tobias Viehböck, whose supervision was instrumental throughout the project. I am deeply grateful for his patient mentorship and his dedication in teaching me essential laboratory skills and principles of experimental design. I would also like to thank him for carrying out key experimental procedures, including the ChIP-Seq, Mass Photometry, and Mass Spectrometry analyses described in this work. Finally, I would like to thank Prof. Silvia Bulgheresi once again for her trust in my scientific development and for her valuable, constructive feedback during the writing and revision of this thesis, together with her continuous scientific input and support.

My sincere thanks go to Dr. Nika Pende, whose expertise and generosity were invaluable, particularly in teaching me the intricacies of protein purification. Her guidance and hands-on assistance during protein purification were fundamental to the success of this work. I am also grateful to her and MSc. Nicole Krause for their help and guidance with image analysis workflow in Fiji.

I would like to express my appreciation to the entire team working on the maintenance and optimisation of Loki cultures: Dr. Nevena Maslac, MSc. Fatma Baraket (also for providing the genomic DNA extracts used in my PCR reactions), MSc. Karin Hager, Sarah Harrer, Marlene Thalmann, and Sophia Lindorfer. I further acknowledge the contributions of current and former members of the EvoChromo project, especially Dr. Zhen-Hao Luo for his support with bioinformatics and MSc. Emily Grant for her work on SMC complex proteins and FISH experiments.

My sincere thanks also go to Dr. Andrea Tramontano, Ulf Naumann, Mag. Nathalia Jandl, and Hana Chovancová for their essential work in laboratory and departmental organisation within the Archaea Biology and Ecogenomics Unit and the Department of Functional and Evolutionary Ecology. I am also thankful to all members of the Archaea Biology and Ecogenomics Unit for fostering a welcoming and collaborative work environment. Special thanks to the Environmental Cell Biology group for the insightful weekly meetings and discussions, and to Dr. Isabelle Zink for the many casual yet helpful conversations around the lab. I would also like to acknowledge the fellow master's students with whom I had the pleasure of sharing this experience, whether for part or the entirety of the journey.

Lastly, I would like to thank my friends and family for their constant support, patience, and encouragement throughout this journey. Your presence has made this achievement possible.

1. Introduction

1.1 From the Origin of Life to Eukaryotic Complexity

1.1.1. The origin of the eukaryotic cell

The origin and history of life on Earth are probably some of the most important and fascinating topics that humanity has ever investigated; so much so, that they lie at the heart of many, if not all, human societies and cultures. For centuries, questions about the origin and evolution of life were approached from philosophical and theological perspectives, in most cases introducing a divine entity as the catalyst and director of life, human in particular **(Ruse, 2009; Dobzhansky, 1973)**.

At present, the scientific community widely agrees that all cellular life on Earth shares a common ancestor. This is based on molecular evidence (essentially, the shared core metabolism, genetic code, and biopolymer structure among species **(Dobzhansky, 1973; Theobald, 2010)**) and further supported by informatic predictions, which establish a Last Universal Common Ancestor (LUCA) as the most probable source of all extant organisms **(Theobald, 2010)**. As such an ancient (proto)organism, LUCA is considered to have been reminiscent of modern prokaryotes (i.e., bacteria and archaea), anaerobic, and integrated in a complex ecosystem **(Moody et al., 2024; Weiss et al., 2016)**.

A direct consequence of the existence of LUCA is that any differential features present in each extant lineage or species are the outcome of a series of critical events that are traceable all the way to LUCA and to any intermediate ancestor **(Doolittle, 1999)**. Therefore, the identification of major biological transitions, represented by each branching point of the tree of life, and their characterisation at the molecular level, are important goals in the field of Evolutionary Biology **(Ayala, 2009)**.

A most notable evolutionary transition concerns the origin of the eukaryotic cell. Eukaryotic organisms, particularly metazoans, exhibit some of the highest levels of biological complexity known, both at the cellular and organismic levels **(Valentine, 1996)**. The notion that LUCA existed as a prokaryote-like organism implies that the first eukaryotic cell, i.e., the Last Eukaryotic Common Ancestor (LECA), emerged from a bacterial or archaeal cell **(Richards et al., 2024; Vosseberg et al., 2024)**.

Several hypotheses about the origin of LECA have been proposed. According to the most widely accepted, the syntrophic symbiosis between at least two prokaryotes resulted in the engulfment

of one into the other, thereby establishing an endosymbiosis and, ultimately, a eukaryotic cell **(Martin et al., 2015; Gould et al., 2016; Santana-Molina et al., 2025)**. Nowadays, there is sound evidence that mitochondria, one of the hallmarks of eukaryotes alongside the nuclear membrane **(Martin et al., 2015)**, stem from an ancient α-Proteobacterium incorporated into the cytoplasm of another prokaryotic host **(López-García and Moreira, 2015)**. The nature of such host, however, has not been yet resolved, and its identification remains a primary objective for evolutionary biologists **(Santana-Molina et al., 2025; Spang et al., 2019)**.

1.1.2. Archaea and the origin of the eukaryotic cell

The identification and description of archaea by Carl Woese and George Fox during the late 20th century **(Woese and Fox, 1977; Woese et al., 1990)** marked an important milestone in evolutionary biology. Although, until then, archaea were often mistaken for bacteria because of their morphological resemblance, their DNA replication and transcription machineries are strikingly similar to those of eukaryotes **(Decker and Hinton, 2013; Sarmiento et al., 2014)**. This observation, as well as their structural differences with bacteria, led to a redefinition of the evolutionary tree of life, placing archaea not only as a separate domain from bacteria, but also as a sister lineage to eukaryotes, with which they would share a common ancestor **(Woese et al., 1990)**.

Ever since, hundreds of archaeal lineages have been identified, some of which were successfully isolated and studied under controlled laboratory conditions, rapidly expanding the knowledge of archaeal physiology and diversity **(Baker et al., 2020)**. In particular, Thermoproteota belonging to the *Sulfolobus* and *Saccharalobus* genera, as well as Halobacteriota and Methanobacteriota such as, e.g., *Haloferax volcanii* and *Methanococcus jannaschii*, have been established as model organisms in archaeal cell biology and biotechnology **(Leigh et al., 2011; De Lise et al., 2023)**. Despite their amenability to experimentation, these organisms are insufficient to represent the full diversity of archaea and to describe their physiology. *In silico* studies are therefore of critical importance in the exploration of the archaeal domain and its relationship with eukaryotes **(Baker et al., 2020)**.

During the late 2000s and early 2010s, some *in silico* analyses provided evidence supporting the "eocyte hypothesis" for eukaryogenesis **(Cox et al., 2008; Foster et al., 2009; Guy and Ettema, 2011; Williams et al., 2012)**, i.e., that eukaryotes originated from within the archaeal domain rather than constituting their sister taxon. This idea was partially rooted on the identification of molecular components considered to be hallmarks of the eukaryotic domain (such as, e.g., actin-like proteins or proteins of the ESCRT-III membrane remodelling system) in archaea **(Koonin,**

2010; Guy et al., 2014). Some analyses assigned this pre-eukaryotic archeozoan to the TACK superphylum (currently, the kingdom Thermoproteati) **(Guy and Ettema, 2011; Williams et al., 2012; Guy et al., 2014)**, pointing also towards the existence of unexplored clades that would represent the sister lineage to eukaryotes **(Guy and Ettema, 2011; Guy et al., 2014)**.

Indeed, in 2015, a novel archaeal lineage was identified in the vicinity of the Loki's Castle, an active hydrothermal venting site in the Arctic Mid-Ocean Ridge **(Spang et al., 2014)**. Its members were initially recognised as a deeply branching clade within the TACK superphylum in archaeal phylogenies, forming the new "Lokiarchaeota" phylum, and displayed an extensive array of Eukaryotic Signature Proteins (ESPs) encoded in their genomes **(Spang et al., 2015; Rodrigues-Oliveira et al., 2023)**. Accordingly, the inclusion of eukaryotic genomes in those phylogenies revealed that the whole eukaryotic domain emerged from within the "Lokiarchaeota", rather than constituting a sister group to the whole archaeal domain or to other members of the TACK superphylum **(Spang et al., 2015)**.

Because of their close similarity to eukaryotes, significant efforts have been made to identify and characterise archaeal lineages with similar phylogenetic proximity. These novel taxa have been grouped in the kingdom Promethearchaeati (commonly referred to as Asgard archaea or "Asgards") **(Imachi et al., 2024; Baker et al., 2020)**, which includes also the previously identified "Lokiarchaeota" within the phylum Promethearchaeota **(Imachi et al., 2024)**. Among them, the order Hodarchaeales has been most recently proposed as the sister taxon of eukaryotes based on phylogenomic analyses **(Voosseberg et al., 2024; Baker et al., 2020; Eme et al., 2023; Liu et al., 2021)**, albeit this is still being debated **(Zhang et al., 2025; Knopp et al., 2021)**.

1.1.3. Domestication of Asgard archaea

The novelty of Promethearchaea and the scarce knowledge about their physiology have originally hampered any attempt to derive either pure or enrichment cultures. After some years since their discovery, nonetheless, the cultivation conditions of at least four Asgard species were elucidated, enabling the establishment of pure co-cultures with methanogenic archaea and/or sulphate-reducing bacteria. Namely, they are *Ca.* Promethearchaeum syntrophicum **(Imachi et al., 2020)**, *Ca.* Lokiarchaeum ossiferum **(Rodrigues-Oliveira et al., 2023)**, *Ca.* Margulisarchaeum peptidophila, and *Ca.* Flexarchaeum multiprotrusionis **(Imachi et al., 2025)**. From these, *Ca.* Lokiarchaeum ossiferum (henceforth referred to as Loki) is the only Asgard archaeon available in Europe and therefore the organism used in my thesis.

A stable Loki culture was first obtained in 2022 by sequential enrichment from shallow water sediment collected near the coast of Piran, Slovenia, by the group of Prof. Christa Schleper

(Rodrigues-Oliveira et al., 2023). This organism grows anaerobically in a consortium with a sulphate-reducing bacterium of the genus *Desulfovibrio* and methanogenic archaea, and has a generation time of 7 to 14 days. Like the other cultured Promethearchaea, they consist of cocci having a cell body diameter of 0.3-1.0 μm and displaying numerous and heterogeneous protrusions enriched with Loki actin and which may be up to ~6 times their cell body diameter long **(Rodrigues-Oliveira et al., 2023; Imachi et al., 2020; Imachi et al., 2025)**.

The cultivability of Promethearchaea – likely the long sought missing link between prokaryotes and eukaryotes – enables us to characterise their cell biology and physiology, and to further explore the evolutionary relationship between archaea and eukaryotes.

1.1.4. Prokaryotic and eukaryotic chromatin

Chromatin refers to the complex of DNA, proteins, and RNA at a given moment in the life of a cell exposed to a given environment, and therefore represents the "natural state" of the genome **(Gross et al., 2015)**. The main differences in bacterial, archaeal, and eukaryotic chromatin are summarised in **Table 1**. In a prokaryotic cell, chromatin (also referred to as nucleoid) is found in the cytoplasm and is easily accessible by the 30S and 50S ribosome subunits **(Bulgheresi, 2025)**, resulting in the coupling of transcription and translation in some, but not all, prokaryotic species **(Johnson et al., 2020; Irastortza-Olaziregi and Amster-Choder, 2021)**. Eukaryotes, however, possess a lipid bilayer that encases chromatin in its nucleus, spatially separated from the cytoplasm and the translation machinery[1] within **(Bulgheresi, 2025)**. Although some prokaryotic species display similar boundaries between the nucleoid and cytoplasm **(Feijoo-Siota et al., 2017; Sagulenko et al., 2014)**, none have been yet described to possess any homologous structure to the nuclear membrane. Therefore, identifying spatial separation of chromatin and the translation machinery in Loki would confirm the evolutionary relatedness between Promethearchaea and eukaryotes.

No matter whether a nuclear envelope is present or not, all genomes must be heavily condensed to fit within a cell **(Hołówka and Zakrzewska-Czerwińska, 2020)**. Although such compaction is mostly driven by biophysical forces resulting in DNA supercoiling **(Luijsterburg et al., 2008)**, different proteins organise and regulate the dynamic structure of chromatin in accordance with gene expression and the transcriptional machinery **(Luijsterburg et al., 2008, Schwab and Dame, 2025)**. Chromatin proteins are classified into three distinct families: histones, proteins of the Structural Maintenance of Chromosomes (SMC) complex, and Nucleoid-Associated Proteins

[1] Note that translation has been reported to occur to inside the nucleus of some eukaryotic cells, although its full extent and purpose are not fully understood **(Iborra et al., 2001; Fåhraeus, 2024)**.

(NAPs) **(Luijsterburg et al., 2008)**. By binding DNA, they enable the establishment of Chromosome Interaction Domains (CIDs) – the prokaryotic equivalent to eukaryotic Topologically Associated Domains (TADs), genomic regions in close spatial proximity – and compartments, a higher-order structuring of the genome **(Szalay et al., 2024)**. In eukaryotes, compartments are usually referred to as euchromatin, recognised as the transcriptionally active section of the genome, and heterochromatin, the transcriptionally inactive section **(Hildebrand and Dekker, 2020)**.

Capturing hallmarks of eukaryotic chromatin structure and dynamics in Loki, including in its genomic and genic structure, chromosome architectural proteins, and mechanisms of gene expression, is therefore crucial to understand how eukaryotic chromatin evolved.

1.2 Exploration of Loki Chromatin

1.2.1. Loki and promethearchaeal chromatin

Due to their recent identification and even more recent cultivability, the exploration of chromatin in Promethearchaea is still in its infancy. Nevertheless, some *in vitro* and *in silico*-based studies have helped understand some of its features and their similarity to eukaryotes. For instance, the capacity of the Hodarchaeal histone HHoB to form eukaryotic-like structures **(Ranawat et al., 2025)**, the existence of RNA-guided RNA-silencing Argonaut proteins in Promethearchaea **(Bastiaanssen et al., 2024)**, the conservation of DNA-binding proteins of the ESCRT-III protein family **(Nachmias et al., 2023)**, or the identification of group II introns in the genomes of Heimdallarchaeia **(Mattick et al., 2024)**.

As for Loki, it encodes and expresses a total of 11 histones, of which 3 display N-terminal tails similarly to the epigenetically modifiable eukaryotic histones; 7 proteins involved in the formation of the SMC complex (i.e., SMC, ScpA, and ScpB homologues), and 3 NAPs, one Alba and two TrmBL2 homologues (unpublished data). The latter protein group is the focus of my Master's thesis.

NAPs are usually defined as small and abundant proteins with no sequence specificity that bind and structure DNA genome-wide and may regulate gene expression **(Hołówka and Zakrzewska-Czerwińska, 2020; Schwab and Dame, 2025)**. They are, in fact, a very heterogeneous protein family with diverse activities, such as bending, wrapping, or bridging the DNA regions they bind to **(Luijsterburg et al., 2008)**. In the archaeal domain, proteins of the Acetylation Lowers Binding

Affinity (Alba) family are generally considered the main chromosome architectural NAPs, mainly because of their widespread distribution, their structural function to DNA, and their essentiality

BACTERIA	ARCHAEA	EUKARYOTES
Genome size		
· Genomes primarily consist of protein-coding sequences (efficient and compact). · < 15 Mbp long.	· Genomes primarily consist of protein-coding sequences (efficient and compact). · < 10 Mbp long.	· Genomes contain large amounts of non-coding and repetitive DNA. · > 10 and up to ~100,000 Mbp long.
Ploidy and partition		
· Generally, monoploid with a single circular chromosome (some cyanobacteria are polyploid). · Plasmids are common.	· Single circular chromosomes. · Thermoproteota: monoploid. · Methanobacteriota: generally, polyploid. · Plasmids are common.	· Multiple linear chromosomes. · Diploid or polyploid. · Plasmids are not common.
DNA replication (*oriC* number)		
· 1 *oriC* per chromosome. · Carefully regulated.	· Variable. · Halobacteriales and Thermoproteota: ≥ 2 *oriC*. · Methanobacteriota: 1-3 *oriC*.	· Multiple origins of DNA replication (simultaneous DNA replication).
Architectural proteins		
· NAPs (e.g., HU, H-NS). · SMC complex proteins are generally involved in chromosome segregation.	· NAPs (e.g., TrmBL2, Alba) and untailed histones (tailed histones were identified in Asgard archaea).	· Tailed histones susceptible of post-translational modification. · SMC complex proteins.
Compartmentalisation		
· Formation of CIDs.	· Formation of TADs. · Compartments are also present in Thermoproteota.	· Formation of TADs and compartments (eu- and heterochromatin).
Transcription mechanisms		
· 1 RNA polymerase with diverse σ factors. · Transcription and translation are mostly coupled. · Introns are uncommon.	· Complex RNA polymerase. · Promoter structure analogous to eukaryotes. · Transcription and translation are mostly coupled. · Introns are uncommon.	· Complex RNA polymerase. · Transcription and translation are mostly uncoupled. · Introns are common (complex RNA splicing machinery).

Table 1. Comparison between the main features of bacterial, archaeal, and eukaryotic chromatin. Genome size: **Bulgheresi, 2025; Brunk and Martin, 2019**. Ploidy and partition: **Bulgheresi, 2025; Barillà, 2016**. DNA replication (*oriC* number): **Leonard and Méchali, 2013; Sarmiento et al., 2014**. Architectural proteins: **Luijsterburg et al., 2008; Schwab and Dame, 2025**. Compartmentalisation: **Szalay et al., 2024; Hildebrand and Dekker, 2020**. Transcription mechanisms: **Decker and Hinton, 2013; Irastortza-Olaziregi and Amster-Choder, 2021; Novikova and Belfort, 2017; Bulgheresi, 2025**.

in *Sulfolobus* spp. **(Laursen et al., 2021; Zhang et al., 2018)**. Other representative NAPs of the archaeal domain include Cren7 and Sul7d **(Zhang et al., 2020)**, found exclusively in members of the phylum Thermoproteota **(Laursen et al., 2021)**, and TrmBL2, first discovered in hyperthermophilic members of the phylum Methanobacteriota **(Lee et al., 2008; Maruyama et al., 2011)**. To this date, no studies concerning NAPs in Asgard archaea have been reported.

1.2.2. Alba

Alba proteins were first identified and characterised in hyperthermophilic species of the *Saccharolobus* and *Sulfolobus* genera, particularly in *S. solfataricus* by the group of Steve D. Bell **(Wardleworth et al., 2002)**. In this organism, Alba was found to bind throughout the genome without any specificity, confirming its nature as a NAP. Since then, *S. solfataricus* has been used as the main model organism to study archaeal Alba.

Two Alba homologues with slightly different function were identified in *S. solfataricus*, Alba1 and Alba2. While the first can form both homo- and heterodimers with Alba2, the latter can only form heterodimers with Alba1 **(Laurens et al., 2012; Jelinska et al., 2005)**. A closer characterisation of these proteins revealed that Alba1 homodimers can multimerise on the DNA double helix by establishing lateral dimer-dimer interactions mediated by a phenylalanine residue at position 60 **(Laurens et al., 2012)**. This behaviour allows the formation of a rigid protein cover on DNA *in vitro*, reducing the overall flexibility of the nucleic acid **(Laurens et al., 2012; Maruyama et al., 2020)**. In addition, both Alba homo- and heterodimers function as DNA bridgers, i.e., they bring distant regions of DNA in sequence to close spatial proximity **(Luijsterburg et al., 2008)**. This function is considered crucial in maintaining a proper tridimensional organisation of the genome and may also contribute to the regulation of expression in certain genes, for example, by allowing the interaction between promoter and enhancer regions **(Luijsterburg et al., 2008; Szalay et al., 2024)**.

Furthermore, Alba was reported to interact with deacetylases of the Sir2 family *in vitro*, which can acetylate the protein at specific lysine residues (Lys16 in *S. solfataricus*) and mediate thereby the repression of transcriptional activity **(Bell et al., 2002; Wardleworth et al., 2002)**. Although post-translational modification (PTM) of Alba is a well-established phenomenon, recent data showed that *S. solfataricus* Alba is subject to (tri)methylation *in vivo* rather than to acetylation **(Cao et al., 2018)**. Moreover, the same study demonstrated that the *S. islandicus* homologue of Alba can be also methylated, and such PTM would not alter gene expression but, instead, the turnover and stability of the protein. However, proteins mediating such PTMs are still unknown.

Besides in *Sulfolobus* and *Saccharolobus* spp., Alba homologues have been identified in almost all archaeal lineages, although their function is considered less relevant in the members of the kingdom Methanobacteriati **(Goyal et al., 2016)**. For example, Alba is not essential in the mesophilic archaeon *Methanococcus voltae*, and its deletion leads to changes in gene expression comparable to those after deletion of a histone gene **(Heinicke et al., 2004)**. In fact, they are completely absent in Halobacteria and the genus Methanosarcina **(Laursen et al., 2021)**, in which the function of Alba is thought to be executed by other NAPs **(Goyal et al., 2016)**. Alba proteins are also found in most Promethearchaeati candidate species, which might encode up to 5 different Alba homologues in a single genome (see **Section 3.1**). However, the differences and similarities between promethearchaeal and conventional Alba remain largely unexplored.

Besides archaea, proteins of the Alba-like domain superfamily, i.e., containing the so-called Alba domain **(Aravind et al., 2003)**, are also found in bacteria and, most frequently, in eukaryotes **(Jagadeesh and Vembar, 2024)**. Bacterial Alba-like proteins are very scarce and are considered to have been acquired by horizontal gene transfer from the archaeal domain and form a unified subfamily with all archaeal Albas **(Jagadeesh and Vembar, 2024)**. Eukaryotic Albas, on the other hand, are outstandingly diverse. They have been mostly studied in unicellular pathogenic eukaryotes, such as, e.g., *Plasmodium falciparum* and *Toxoplasma gondii*, in which they modulate gene expression at either transcriptional and translational levels by binding DNA and RNA **(Chêne et al., 2012; Gissot et al., 2013)**. In other unicellular and multicellular eukaryotes, such as the yeast *Saccharomyces cerevisiae* or the flowering plant *Arabidopsis thaliana*, Alba proteins have been reported to regulate tRNA maturation, RNA stability, and gene expression **(Chan et al., 2018; Náprstková et al., 2021)**.

The recent analysis of over 15,000 Alba domain-containing proteins by Jagadeesh and Vembar classified them into 1 archaeal and 12 eukaryotic clusters, including the previously identified Rpp20/Pop7 and Rpp25/Pop6 subfamilies, which most likely originated from an archaeal Alba protein ancestor **(Aravind et al., 2003)**. Promethearchaeal Alba proteins, including Loki Alba (LoAlba), cluster within the archaeal Alba subfamily **(Jagadeesh and Vembar, 2024)**.

Despite the extreme diversity of eukaryotic Alba-like proteins, they tend to be much larger than their archaeal counterparts, maintaining only the size of their Alba domain **(Aravind et al., 2003)**, and to operate as RNA-binding proteins rather than NAPs **(Aravind et al., 2003; Jagadeesh and Vembar, 2024)**. Even though these features might appear as unique innovations of the eukaryotic domain, RNA-binding and processing properties have also been reported in archaeal Alba, including in the *Sulfolobus* and *Saccharolobus* genera **(Marsh et al., 2005; Guo et al., 2003;**

Zhang et al., 2020). Among these, *S. solfataricus*, *S. shibatae*, and *S. islandicus* Alba were reported to bind RNA *in vitro* and *in vivo*, particularly in association with ribosomal RNA or exhibiting single-stranded RNA chaperone activity. Interestingly, *S. shibatae* Alba does not bind DNA *in vivo* but only RNA, challenging the overall definition of Alba as a NAP **(Guo et al., 2003)**.

Alba thus constitutes an outstandingly diverse DNA/RNA-binding protein superfamily, whose members exhibit specific DNA and/or RNA-related activity which may vary even among members of the same genera. Because of this, it is considered that Alba proteins diversified from an ancient RNA-binding archaeal protein, while their function as NAPs would be the result of secondary acquisitions in specific archaeal lineages, particularly in Thermoproteota **(Aravind et al., 2003; Jagadeesh and Vembar, 2024)**.

Because of the phylogenetic proximity of Promethearchaea to eukaryotes, promethearchaeal Alba proteins are predicted to display RNA-related activity *in vivo*. However, the close similarity between LoAlba and *S. solfataricus* Alba1 suggests that LoAlba may also have DNA-binding activity and a prominent function in chromatin organisation in Loki. Therefore, the exploration of the physiochemical properties of LoAlba will clarify whether it is functionally similar to eukaryotic Alba and, consequently, will shed light on the evolution of the Alba superfamily.

1.2.3. TrmBL2

TrmBL2 was first identified as a NAP in 2011, in the hyperthermophilic archaeon *Thermococcus kodakarensis*, of the phylum Methanobacteriota **(Maruyama et al., 2011)**. This protein exhibited both chromosome organisation and transcriptional repression activity, in coordination with histones and Alba **(Maruyama et al., 2011; Maruyama et al., 2020; Efremov et al., 2015)**. In fact, TrmBL2 belongs to the larger TrmB protein family of transcriptional regulators and sugar transporters of archaea **(Lee et al., 2008)** and is considered its only NAP representative **(Kim et al., 2016)**. To date, TrmBL2 has not been reported in the eukaryotic domain.

Unlike Alba, the physiochemical properties of TrmBL2 remain largely unexplored, while most studies have been conducted in *T. kodakarensis* and *Pyrococcus furiosus* **(Maruyama et al., 2011; Ahmad et al., 2015)**. In these, TrmBL2 has been shown to bind DNA non-specifically by oligomerising in pairs of TrmBL2 homodimers on opposite sides of the DNA double helix, covering it with their structure **(Ahmad et al., 2015)**. In *P. furiosus*, TrmBL2 binds both double-stranded DNA and single-stranded DNA with similar affinity and contributes to their stabilisation by polymerising into a thick and fibrous filament **(Maruyama et al., 2011; Wierer et al., 2016)**. Unlike Alba, however, it does not establish bridges between non-contiguous DNA regions and is thus thought to act purely as a stabiliser of DNA secondary structure **(Wierer et al., 2016)**.

Although this function is thought to be crucial in preventing DNA denaturation in hyperthermophiles, in mesophilic archaea, including Loki, TrmBL2 might be involved in different activities. Like other members of the TrmB family, TrmBL2 acts as a global transcriptional repressor in *P. furiosus* and *T. kodakarensis*, although a certain degree of specificity to promoters might be influenced by regulating its concentration **(Maruyama et al., 2011; Kim et al., 2016)**. Similarly, TrmBL2 is involved in the regulation of expression of a halolysin in the mesophilic haloarchaeon *Natrinema gari* **(Yin et al., 2024)**. More interestingly, TrmBL2 appears to regulate the formation of topologically associating domains (TADs) by acting as a roadblock during DNA loop extrusion by SMC in *T. kodakarensis* **(Yamaura et al., 2025)**.

All in all, TrmBL2 is regarded as an important chromatin architectural NAP with critical functions in the regulation of gene expression in archaea. Examination of physiochemical properties of both Loki TrmBL2 paralogues (LoTrmBL2a and LoTrmBL2b) will therefore contribute to the overall characterisation of this remarkably unknown NAP, as well as in understanding the structure and dynamics of Loki chromatin.

1.3 Aims of my Master's Thesis

In October 2024, I joined the EvoChromo project under the coordination of Prof. Christa Schleper and under the joint supervision of Prof. Silvia Burghelesi and Dr. Tobias Viehböck (Universität Wien) to conduct my Master's thesis. Overall, we seek to understand whether Loki chromatin exhibits features that, up to now, are exclusively eukaryotic. If discovered, such features would support the archaeal origin of eukaryotes beyond bioinformatic predictions.

In my thesis, I focus on finding some of this evidence in the characterisation of LoAlba, LoTrmBL2a, and LoTrmBL2b. In particular, I aim at expressing and purifying recombinant LoAlba and conducting biochemical assays to elucidate its functions. The purification and biochemical characterisation of LoTrmBL2a and LoTrmBL2b are included only as secondary goals.

1.3.1. Research hypotheses

1. LoAlba, LoTrmBL2a, and LoTrmBL2b are key chromatin organisers in Loki alongside proteins of the SMC complex and histones. This hypothesis is subdivided into the following:
    a. The three proteins bind DNA throughout the Loki genome and are enriched at boundaries between genomic domains and/or compartments.
    b. The three proteins exhibit at least one of DNA bending, bridging, or wrapping activities and may thereby be considered NAPs in Loki.

c. LoAlba interacts with histones, LoTrmBL2a and/or LoTrmBL2b.
d. LoTrmBL2a and LoTrmBL2b interact with proteins of the SMC complex present in Loki.

2. LoAlba binds RNA and influences its folding and stability.

3. The three proteins dimerise and/or oligomerise upon binding to DNA. Oligomerisation of LoAlba is mediated by a phenylalanine residue at position 74.

4. LoAlba is post-translationally modified in a lysine residue at position 31 and therefore interacts with methyltransferases and demethylases and/or acetyltransferases and deacetylases present in Loki.

5. Tyr52 and Asp53 of LoTrmBL2a, and Tyr56 and Glu57 of LoTrmBL2b mediate interaction with DNA.

This thesis presents the first characterization of the three NAPs identified to date in Loki. The findings contribute to testing the proposed hypotheses and lay the groundwork for future studies on Loki NAPs and the organization of Loki chromatin.

2. Methods

2.1. Establishment of *E. coli* Expression Strains

2.1.1 Alba, TrmBL2a, and TrmBL2b Expression Strains

Loki *albA* was identified by Dr. Tobias Viehböck by homology-based search on the Loki-B35 GCA_025839675.1 genome assembly **(NCBI Datasets, O'Leary et al., 2024)** using the hidden Markov models (HMM) from the Pfam database corresponding to Alba. For the Loki *trmBL2a* and *trmBL2b* genes, a similar HMM-based search for TrmB-related proteins was conducted.

albA, *trmBL2a*, and *trmBL2b* were amplified from *Ca.* L. ossiferum B-35 strain genomic DNA extracts by Polymerase Chain Reaction (PCR) and cloned into a pT7 vector in frame with a 6x Histidine and SUMO tag by Gibson assembly at an insert-to-vector molar ratio of 3:1. The pT7 vector has a high copy-number origin of replication and kanamycin (Kan) resistance cassette and allows for controlled expression of recombinant genes upon induction with lactose or isopropyl β-D-1-thiogalactopyranoside (IPTG) **(Tabor, 2011)**. pT7 plasmids containing Loki *albA*, *trmBL2a* or *trmBL2b* were transformed into chemically competent *Escherichia coli* Top10 by heat shock as reported in **Froger and Hall, 2007** and plated in Luria-Bertani (LB)-agar plates containing 50 μg mL^{-1} Kan. Successful transformation was confirmed by Sanger sequencing (Eurofins Genomics).

Plasmids of transformed colonies were purified with the Monarch® Plasmid Miniprep Kit and transformed into the expression strains *E. coli* BL21(DE3) or *E. coli* C41(DE3) as described above. Viable colonies were inoculated into LB + Kan and incubated overnight (o/n) at 37ºC to derive cell stocks stable at -70ºC in 25% glycerol.

2.1.2. Mutagenesis of Alba, TrmBL2a, and TrmBL2b

pT7 vectors for mutant Alba, TrmBL2a, and TrmBL2b proteins were established by PCR. Primers were designed to misalign at specific sites on the pT7 vectors cloned with the wild-type protein versions, thereby inserting mutated codons upon amplification. Mutated sites are specified in the **Supplementary Material**. PCR products were circularised by incubation with a Kinase, Ligase, and DpnI Enzyme Mix (New England Biolabs) for 5 minutes at room temperature (RT) and 1 hour at 37ºC, also digesting the plasmid used as template. Mutagenised plasmids were then transformed into *E. coli* BL21(DE3) as described above.

Successful transformation of all expression strains was confirmed by whole plasmid sequencing by Plasmidsaurus using Oxford Nanopore Technology. All primer and plasmid sequences cited,

including those for the mutant gene variants, as well as all PCR reagents and programs used, are specified in the **Supplementary Material**.

2.2 Purification of Recombinant Proteins

2.2.1 LoAlba, LoTrmBL2a, and LoTrmBL2b Expression Tests

Expression of recombinant LoAlba, LoTrmBL2a, and LoTrmBL2b was first tested at three different temperatures. BL21(DE3) or C41(DE3) transformed with genes encoding for wild-type LoAlba, LoTrmBL2a, or LoTrmBL2b were grown until saturation at 37ºC in 2YT broth (Carl Roth GmbH) supplemented with NPS-Mg (25 mM $(NH_4)_2SO_4$, 50 mM KH_2PO_4, 50 mM Na_2HPO_4, 1 mM $MgSO_4$) and 0.8% glucose, and containing 50 µg mL^{-1} Kan. Cells were then diluted to 1:1,000 in 10 mL fresh 2YT + Kan supplemented with NPS-Mg and an Autoinduction Supplement (AIS; 0.4% glycerol (v/v), 0.5 g L^{-1} glucose, 2 g L^{-1} α-lactose; see **Studier, 2005**) and incubated for 4 hours at 37ºC at 180 or 120 rpm agitation, after which their growth temperature was adjusted to RT, 27ºC, or 37ºC. After ~16h of incubation, 2 mL of each culture were harvested and cells obtained by centrifugation (10 minutes, 2,300 rcf, 4ºC).

Cell pellets were resuspended in 1.5 mL pre-cooled Lysis Buffer (50 mM Tris-HCl pH 8, 300 mM NaCl, 0.25 mM DTT, 1 mM $MgCl_2$) further containing a Pierce™ EDTA-free Protease Inhibitor Mini Tablet (Thermo Scientific), lysozyme, benzonase, and Bugbuster® Protein Extraction Reagent (Sigma-Aldrich); and incubated on a rotating wheel (45 minutes, 20 rpm, 4ºC). After lysis, 20 µL of cell lysate were harvested and mixed with 5 µL NuPAGE 4X LDS Sample Buffer (LD; Invitrogen). To obtain the soluble protein fraction, the remaining lysate was centrifuged (10 minutes, 13,000 rcf, 4ºC), 20 µL of supernatant were harvested, and mixed with 5 µL LD.

The remaining supernatant for each condition tested was loaded onto 100 µL washed HisPur™ Ni-NTA Resin (Thermo Scientific) and incubated on a rotating wheel (30 minutes, 20 rpm, 4ºC). Washes were performed once with 1 mL MiliQ-H_2O (MiliQ) and thrice with 1 mL Lysis Buffer by centrifugation (the first two washes were 1-minute long, the last was 2-minute long; 3,500 rcf, 4ºC). After incubation, Ni-NTA-bound protein pellets were recovered, washed thrice with Lysis Buffer by centrifugation (2 minutes, 3,500 rcf, 4ºC) and mixed with 10 µL LD. For proteins with low solubility (LoTrmBL2a and LoTrmBL2b), pelleted cell lysates were resuspended in 1.5 mL Urea Buffer (8 M urea, 50 mM Tris-HCl pH 8, 300 mM NaCl, 0.25 mM DTT, 1 mM $MgCl_2$) and incubated on a rotating wheel (30 minutes, 20 rpm, 4ºC). Solubilised proteins were then

separated from insoluble complexes and cell debris by centrifugation (10 minutes, 13,000 rcf, RT) and 20 µL of the resulting supernatant were mixed with 5 µL LD.

6 µL from each sample were loaded on a 15% acrylamide SDS-PAGE or a ProteinEle® Precast 4-20% Tris-Glycine gel (TransGen Biotech Co.) and electrophoresed at 200 V for 50 minutes. Gels were stained with InstantBlue® Coomassie Protein Stain (Abcam Ltd.), washed successively with deionised H_2O (dH_2O), and imaged with an iBright Imaging System (Thermo Fisher Scientific).

In addition to molecular weight-based identification of recombinant proteins on Coomassie stained PAGE gels, *E. coli* protein extracts were blotted onto a 0.2 µm nitrocellulose membrane by wet transfer in Transfer Buffer (25mM Tris, 192 mM glycine, 20% methanol (v/v), 0.01% SDS, pH 8.3), 100 V, 1 hour. To confirm that proteins did transfer properly, membranes were briefly stained with Ponceau S dye and rinsed with Phosphate-buffered Saline solution (PBS; 8 g L^{-1} NaCl, 1.42 g L^{-1} Na_2HPO_4, 0.2 g L^{-1} KH_2PO_4, 0.2 g L^{-1} KCl). Membranes were then blocked with 5% milk powder in PBS with 0.05% Tween (PBT) for 45 minutes at RT under agitation. After blocking, membranes were incubated with a rabbit anti-His tag antibody (Cell Signalling Technology®) in a 1:2,000 dilution in milk-PBT at 4ºC o/n and under gentle agitation. Membranes were then washed for 5 minutes thrice with milk-PBT and incubated with an HRP-conjugated goat anti-rabbit antibody (Abcam Ltd) in a 1:5,000 dilution in milk-PBT at RT for 45 minutes under gentle agitation. Membranes were washed for 5 minutes thrice with PBT, briefly once with PBS, and kept in PBS before imaging. Signal was revealed with either the NeoPRO Nano or Pico Ultra Western ECL Substrate (Neo Biotech) kits, depending on the resolution desired for each recombinant protein.

2.2.2. Large-scale Expression of LoAlba and LoTrmBL2b and Protein Extraction

Expression of target proteins was scaled up by reproducing the conditions yielding the highest protein amount in the expression tests (see **Section 2.2.1**). Expression strains for LoAlba and LoTrmBL2b were grown in 10 mL 2YT + Kan supplemented with NPS-Mg and 0.8% glucose at 37ºC until saturation. Cells were then diluted to 1:1,000 in expression medium (1 L 2YT + Kan supplemented with AIS), incubated at 37ºC and 180 rpm agitation for 4 hours, and continued to grow at 37ºC until saturation. After ~16 hours, cultures were centrifuged (20 minutes, 4,500 rpm, 4ºC; Sorvall Lynx 4000 Centrifuge, Thermo Scientific) to recover cell pellets. These were resuspended in 5 mL PBS, recentrifuged (20 minutes, 1,900 rcf, 4ºC), and stored at -20ºC.

Cells were thawed on ice and lysed in 50 mL pre-cooled Lysis Buffer (300 mM NaCl, 50 mM HEPES pH 7 (LoAlba) or pH 8 (LoTrmBL2b), 5% glycerol (v/v), 15 mM imidazole, 0.25 mM DTT, benzonase, lysozyme, and a Pierce™ EDTA-free Protease Inhibitor Mini Tablet) on a rotating

wheel (1 hour, 4ºC). To ensure all cellular components were released, cells were placed on ice and sonicated at 35% of the maximum amplitude (Sonic Dismembrator Model 705 with FB-4420 microtip probe, Fisher Scientific) with alternating pulses and breaks of 1 second during a total time 1 minute, until no opaque clumps could be observed in the cell lysate (~ after 5 sonication cycles). 20 µL of the total lysate were then harvested and mixed with 5 µL LD. To extract the soluble protein fraction, the remaining lysate was centrifuged at a high speed (20 minutes, 20,000 rcf, 4ºC) and 20 µL of the supernatant were collected and mixed with 5 µL LD.

2.2.3. Purification of LoAlba and LoTrmBL2b

Purification of LoAlba and LoTrmBL2b was achieved by sequential separation, first, by affinity chromatography in a HisTrap™ HP 5 mL His-tag protein purification column, and secondly by a Size Exclusion Chromatography (SEC) column, using an Äkta Pure HPLC system (Cytiva), as described in **Pende et al., 2021**. Before the start, the system and the HisTrap column were washed with degassed MiliQ to replace their storage solution (20% EtOH) and the column was subsequently equilibrated with 5 column volumes (CV) of Immobilized Metal Affinity Chromatography (IMAC) A (300 mM NaCl, 50 mM HEPES pH 7 or 8, 5% glycerol (v/v), 15 mM imidazole). After the system preparation and column equilibration, the soluble lysate was loaded onto the column at a flow rate of 3 mL min^{-1}. The column was washed with 10 CV IMAC A to remove all proteins that did not bind it, and 20 µL of the flowthrough were collected and mixed with 5 µL LD for analysis.

Elution started thereafter by gradually injecting IMAC B (300 mM NaCl, 50 mM HEPES pH 7 or 8, 5% glycerol (v/v), 500 mM imidazole) into the column, creating a linear gradient of imidazole for 20 CV at a flow rate of 5 mL min^{-1} until reaching 100% of IMAC B injection. The eluate was collected in 1 mL fractions and analysed in 20 µL aliquots mixed with 5 µL LD. During elution, light absorbance at 280 nm was measured to select fractions likely containing recombinant proteins. In some experiments, absorbance at 260 nm was also measured to identify fractions containing recombinant proteins bound to nucleic acids. After elution, the HisTrap column was washed sequentially with 5 CV of IMAC A, MiliQ, and 20% EtOH, and stored at 4ºC for reuse. A separate HisTrap column was used for each of the proteins purified.

All samples in LD were analysed by gel electrophoresis as described in **Section 2.2.1**. For the total cell extract, soluble protein fraction, and column flowthrough samples, 5 µL were loaded on the gel, while 20 µL were used for each eluted fraction.

All fractions containing the protein of interest, as derived from the gel electrophoresis results, were pooled together to be further purified by size exclusion chromatography. The SUMO

protease Ulp1 (~ 200 to 350 µg, depending on the number of pooled fractions) was added to the fractions and the whole volume was transferred to a SnakeSkin™ Dialysis Tubing (3.5K Molecular Weight Cut-off (MW CO), 16 mm diameter dry; Thermo Scientific) and placed on 2 L SEC Buffer (150 mM NaCl, 50 mM HEPES pH 7 or 8, 5% glycerol (v/v)) at 4ºC under constant magnetic stir to remove the imidazole, o/n. Some fractions were pooled and dialysed separately without Ulp1 incubation to obtain pure protein samples retaining their 6x-His-SUMO tag. To increase the efficiency of the SEC column separation, DTT was added to the SEC buffer during dialysis of LoTrmBL2b in the second purification attempt, to help remove any protein complex that hindered the recovery of pure protein. 20 µL of the pooled fractions were collected and mixed with 5 µL LD both before the addition of Ulp1 and after dialysis, to confirm the proper cleavage of the 6x-His-SUMO tag.

Only during the first purification round of LoAlba, all cleavage products were loaded into a spin chromatography column (Bio-Rad Laboratories) to separate the 6x-His-SUMO tag from the target protein in a so-called "reverse IMAC" (rIMAC) protocol. Prior to purification, the HisPur™ Ni-NTA resin was washed once with MiliQ and twice with SEC buffer by centrifugation and stored at 4ºC. The quantity of Ni-NTA beads was adjusted according to protein concentration so that 1 mL of crude Ni-NTA resin was used per each 50 mg of protein, calculated prior to cleavage by Ulp1. The chromatography column was previously equilibrated with MiliQ and SEC Buffer.

Dialysed proteins were then incubated with the washed Ni-NTA resin on a rotating wheel (1 hour, 20 rpm, 4ºC) and subsequently loaded onto the chromatography column. The flowthrough was recovered and 20 µL were mixed with 5 µL LD for analysis by electrophoresis.

Since cleaved LoAlba stayed bound to the resin and no protein was detected in the flowthrough (see **Results 3.1.2**), all resin-bound proteins were eluted by saturating the column with IMAC B buffer and recovering the flowthrough. Imidazole was removed from the eluate by repeating the dialysis as described above. Given that this step strongly affected recombinant protein yield, it was omitted in all subsequent runs.

Proteins recovered were concentrated to a variable volume between 2 and 5 mL using a Vivaspin® concentrator column (Cytiva) with a MW CO of 3 kDa (LoAlba) or 10 kDa (LoTrmBL2b) by repeated centrifugation (10 minutes, 3,620 rcf, 4ºC). The membrane filters in each column were periodically cleared of precipitates that unclogged and limited concentration efficiency by resuspending the protein mix on the concentrator column. In addition, the concentrating protein solution was transferred to a new recipient and spun down (10 minutes, 3,620 rcf, 4ºC) to remove any insoluble complexes blocking the membrane whenever deemed appropriate.

Before SEC separation, a HiLoad™ 16/600 Superdex 75 pg SEC column (Cytiva) was washed in 1 CV degassed MiliQ and equilibrated with 1 CV degassed SEC Buffer. A slow flow rate (0.5 mL min^{-1}) was used during column preparation to avoid the creation of overpressure which could hinder the size resolution during the elution step. The concentrated protein was then injected into the Äkta Pure HPLC machine with special care not to introduce air bubbles into the system. Proteins were loaded into a loop prior to entrance in the SEC column, ensuring a controlled flow rate upon column charging. Proteins were eluted isocratically in 1.1 CV in fractions of 0.8 mL volume at a slower flow rate (0.3 to 0.5 mL min^{-1}) to increase size resolution. Light absorbance at 280 nm, and 260 nm only in some purification runs, was recorded. 20 µL of each fraction of interest were harvested and mixed with 5 µL LD for analysis. After elution, the SEC column and Äkta system were sequentially washed with MiliQ and 20% EtOH for long-term storage.

All collected samples during purification were analysed by gel electrophoresis as described in **Section 2.2.1**. Protein concentration was measured with a NanoPhotometer® N60 (Implen) and adjusted according to the extinction coefficient $\varepsilon_{0.1\%}$ (also, $Abs_{0.1\%}$) of each of the target proteins, as calculated with the ProtParam online tool from Expasy **(Gasteiger et al., 2005)**. The final concentration value results by dividing the output value from the NanoPhotometer machine, which is calculated assuming an $\varepsilon_{0.1\%} = 1$, with each protein's $\varepsilon_{0.1\%}$ value.

2.2.4. Detection of DNA in Purified Protein Fractions by Acrylamide Gel Electrophoresis

Fractions of interest were further tested for their nucleic acid content as follows. Proteinase K (Carl Roth GmbH) was added to 20 µL of each analysed fraction at a concentration of 0.5 mg mL^{-1} to digest all proteins and liberate the nucleic acids present in the sample. The mix was incubated for 30 minutes at 65ºC and 1X Tris-Borate-EDTA (TBE) Sample Buffer (0.01% bromo-phenol blue (w/v), 10% glycerol (v/v)) was added. Samples were loaded on a 12% acrylamide gel and electrophoresed at 150 V for 50 minutes and the gel was stained with SYBR™ Safe DNA Gel Stain (Invitrogen) and imaged using an iBright Imaging System. In addition, the UV absorbance spectrum of each fraction of interest was measured with a Nano-Photometer, as described above.

2.3. DNA staining of *E. coli* expressing LoAlba or LoTrmBL2b

2.3.1 Induction of LoAlba and LoTrmBL2b expression with IPTG

E. coli BL21(DE3) transformed with a pT7 vector encoding either LoAlba or LoAlba(p.Lys31Ala; Phe74Ala) (Alba_dm) or carrying no insert; and *E. coli* C41(DE3) transformed with a pT7 vector

encoding either LoTrmBL2b or LoTrmBL2b(p.Tyr56Asn;Glu57Gln) (TrmBL2b_dm), or carrying no insert, were grown until saturation at 37ºC in LB + Kan. Each culture was diluted to 1:1,000 in triplicates of 50 mL LB + Kan and grown at 37ºC at 180 or 120 rpm agitation. OD_{600} was recorded upon inoculation, 2 hours after, and subsequently after every hour.

Upon reaching an OD_{600} between 0.3 and 0.6 (roughly after 3.5 hours after inoculation), 1mM IPTG was added to half of the cultures to induce the expression of recombinant genes. After 3 hours of expression, 1 mL from each uninduced and induced strain were harvested and fixed with 3% formaldehyde for 30 minutes at RT. Excess formaldehyde was removed by sequential centrifugation (10 minutes, 3,300 rcf, 4ºC) and resuspension in 1 mL PBS thrice.

To verify whether target proteins were indeed expressed, 2 mL of each uninduced and induced strain were harvested, and cells were collected by centrifugation (10 minutes, 3,300 rcf, 4ºC). These were then resuspended in 500 µL, of which 60 µL were mixed with 20 µL LD and boiled for 5 minutes at 99ºC. 10 µL of each sample were used for Western Blot analysis as described in section 2.2.1**.**

2.3.2 DNA Staining and Imaging

Hoechst stain was added to fixed *E. coli* cells in PBS to a concentration of 1:2,000 and incubated for 10 minutes at RT in the dark. Stained cells were washed twice by centrifugation (10 minutes, 3,300 rcf, RT) and resuspended in 70 µL PBS. Microscopy slides were prepared with a 1% (w/v) agarose pad to provide a uniform, even surface suitable for sample mounting. 5 µL cells and 1 µL Vecta-Shield Antifade Mounting Medium (Biozol Diagnostica Vertrieb GmbH) were placed on the slides and evenly spread using a glass cover.

Slides were imaged using a Nikon Eclipse Ni-U upright microscope (Model 941490; Nikon Corporation) equipped with a Nikon Intensilight C-HGFI mercury fibre illuminator. Images were captured using a Nikon Plan Apo 100X/1.45 OFN Ph3 DM oil immersion objective and a ProgRes® MFCool CCD camera with the Capture Pro 2.8.8 software (Jenoptik).

Images were analysed with Fiji **(Schindelin et al., 2012)** and the MicrobeJ plugin **(Ducret et al., 2016)**. Cell shape parameters were fine-tuned to identify single *E. coli* cells, including those undergoing cell division. Cell aggregates were manually excluded or separated into single cells to be individually analysed. A minimum of 300 cells per strain and treatment were considered.

2.4. Digestion of *E. coli* Chromatin upon LoAlba Expression by Micrococcal Nuclease (MNase)

2.4.1. MNase digestion of *E. coli* chromatin

E. coli BL21(DE3) cell pellets were obtained after LoAlba induction with IPTG as described in **Section 2.3.1**. These were fixed with 10% formaldehyde in PBS (v/v) and incubated on the rotating wheel for (10 minutes, 10 rpm, RT). Fixation was quenched by the addition of 140 mM glycine to the mix, which was further incubated on the rotating wheel (5 minutes, 10 rpm, 4ºC). Cells were washed twice with PBS by centrifugation (5 minutes, 4,000 rcf, 4ºC) and resuspended in 4 mg mL^{-1} lysozyme in PBS. The mixture was incubated for 20 minutes on ice and the resulting lysate was centrifuged (3 minutes, 16,000 rcf, RT). The remaining pellet was resuspended in 1 mL PBS containing 1% NP-40 Alternative, Protein Grade ® Detergent (Sigma Aldrich) to remove all lipids still present and incubated on ice for 10 minutes. Genomic DNA was pelleted and washed with PBS by centrifugation (3 minutes, 16,000 rcf, RT) and finally resuspended in 320 µL MNase Reaction Buffer (New England Biolabs).

100 µL of cell lysate were diluted in MNase Reaction Buffer containing an increasing value of MNase activity units in a final volume of 225 µL and incubated for 10 minutes at RT. Digestion was stopped by adding 100 mM EDTA and 10 mM EGTA in each reaction tube and waiting for 1 minute. 100 µg mL^{-1} RNase A were subsequently added and the mix was incubated for 10 minutes at 37ºC. Finally, 1% SDS and 1 mg mL^{-1} proteinase K were added to the mix, which was incubated o/n at 65ºC with constant agitation at 500 rpm.

2.4.2. DNA extraction and analysis

Digested DNA was isolated by phenol:chloroform extraction. The total volume of each sample was adjusted to 450 µL with PBS and subsequently mixed with an equal volume of a 25:24:1 mixture of phenol, chloroform, and isoamyl alcohol (Carl Roth GmbH) by vortexing for 5 minutes. The aqueous phase was recovered by centrifugation (5 minutes, 16,000 rcf, RT) and DNA was precipitated by adding 40 µL 3M NaOAc and 1 mL EtOH at -20ºC, followed by an incubation of 1 hour at -70ºC. DNA was recovered by centrifugation and washed with cold 80% EtOH (1 hour pelleting, 30 minutes washing; 16,000 rcf, 4ºC). The supernatant was discarded and all remaining EtOH was removed by speed-vacuum drying (5 minutes, 30ºC). DNA was rehydrated in 20 µL 5 mM Tris-HCl at 37ºC for 30 minutes and mixed with TriTrack DNA Loading Dye (Thermo Scientific). Samples were run by electrophoresis in a 2.5% agarose 1X TAE gel (Carl Roth, GmbH) containing SYBR Safe for 50 minutes at 100 V and imaged using an iBright system.

2.5. Electrophoretic Mobility Shift Assay (EMSA)

2.5.1. DNA and RNA probe preparation

A 615 base-pair (bp) DNA fragment was amplified from the pT7 plasmid by PCR using Cyanine 5 (Cy5)-labelled DNA primers, thereby adding the fluorophore to the amplified DNA molecule. To derive a complementary RNA molecule, a T7 class III phi6.5 promoter sequence was cloned upstream of the aforementioned 615 bp-long DNA fragment. The RNA molecule was produced by *in vitro* transcription and labelled with Cyanine 3 (Cy3) with a High-Yield T7 Cy3 RNA Labeling Kit (Jena Bioscience). All necessary primers and the sequences of the aforementioned DNA and RNA fragments are specified in the **Supplementary Material**.

2.5.2. EMSAs

EMSAs were executed by combining DNA and/or RNA with LoAlba or LoTrmBL2b at varying ratios in a final volume of 20 µL and incubating the mix for 30 minutes at RT. To estimate the binding affinities of Alba and TrmBL2b to nucleic acids, increasing concentrations of protein, DNA, or RNA were used. When held constant, the concentrations of DNA and RNA were adjusted to 40 µM, based on the intensity of the observed fluorescent signal. In contrast, the concentrations of LoAlba and LoTrmBL2b were selected based on the minimum amount required to produce a complete shift of DNA or RNA in the EMSA at fixed nucleic acid concentrations. For some experiments, the DNA-binding protein unspecific competitor Poly(dI·dC) LightShift™ (Thermo Scientific) was added at increasing concentrations to hinder DNA binding by Alba or TrmBL2b.

Heat-inactivated (HI) negative controls were generated by boiling the test protein at 99ºC for 10 minutes, followed by immediate cooling on ice. Additionally, BSA was included as a negative control for DNA binding, applied at the highest concentration of the test protein used in each experiment. Protein dilutions in SEC Buffer were prepared when necessary. LoAlba and LoTrmBL2b as well as DNA and RNA concentrations were measured with a NanoPhotometer prior to each experiment.

After incubation, samples were mixed with 3 µL of custom DNA loading dye (6X: 10 mM Tris-HCl pH 7.6, 1.5 mg mL^{-1} orange G, 60% glycerol (v/v)). Orange G was used as the only tracking dye to avoid background fluorescence by bromophenol blue and xylene cyanole, commonly present in commercial DNA loading dyes. EMSA reaction samples were loaded on a 1% agarose 1x TAE gel, run by electrophoresis for 45 minutes at 100 V, and stained using an iBright system.

3. Results

3.1. Identification of LoAlba

We identified the gene coding for *Ca.* L. ossiferum Alba (*albA*) in the Loki-B35 GCA_025839675.1 genome assembly by homology search with the Alba domain. The predicted LoAlba consists of a globular protein of 118 amino acids that is highly enriched in the *Ca.* L. ossiferum proteome (data not shown). Moreover, it is slightly acidic (isoelectric point, pI = 6.58) and has a molecular weight (MW) of 13.65 kDa. LoAlba is longer but structurally homologous to Alba1 of *S. solfataricus* (**Figure 1**) and therefore conserves its DNA-binding (antiparallel β-hairpin loop between strands β3 and β4) and homodimerisation (helix α2 and strands β3 and β4) domains **(Wardleworth et al., 2002)**. *S. solfataricus* Alba1 Lys16 and Phe60 are involved in DNA-Alba1 duplex stabilisation, post-translational modification, and lateral stacking and correspond to LoAlba Lys31 and Phe74, respectively.

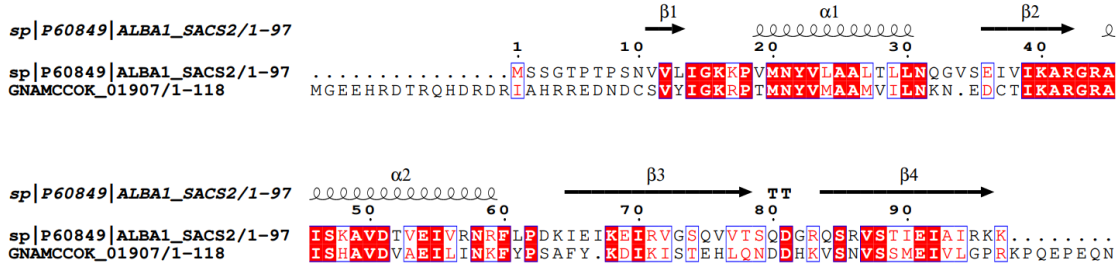


Figure 1. Structural alignment between *S. solfataricus* Alba1 (top) and LoAlba (bottom). Residues forming β-sheet and α-helical secondary structures are indicated on top. Conserved and analogous residues are shown in red. Provided by Dr. Zhen-Hao Luo (Universität Wien).

To assess the phylogenetic placement of LoAlba within Promethearchaeati, we constructed a phylogenetic tree of promethearchaeal Alba-like proteins and quantified the number of paralogues in each analysed genome (**Figure 2**). We found that the evolutionary diversification of Alba proteins matches the established phylogenetic classification of Promethearchaeati, with individual genomes encoding from zero to five distinct Alba paralogues.

3.2. LoAlba binds DNA *in vitro* and *in vivo*

3.2.1. Expression of LoAlba in *E. coli* BL21(DE3)

According to its definition as a NAP and its homology to archaeal Alba proteins, we wanted to test whether LoAlba can bind DNA *in vitro* and *in vivo*. Therefore, we constructed *E. coli* BL21 (DE3) recombinant strains expressing LoAlba under the control of a *lac promoter* for expression

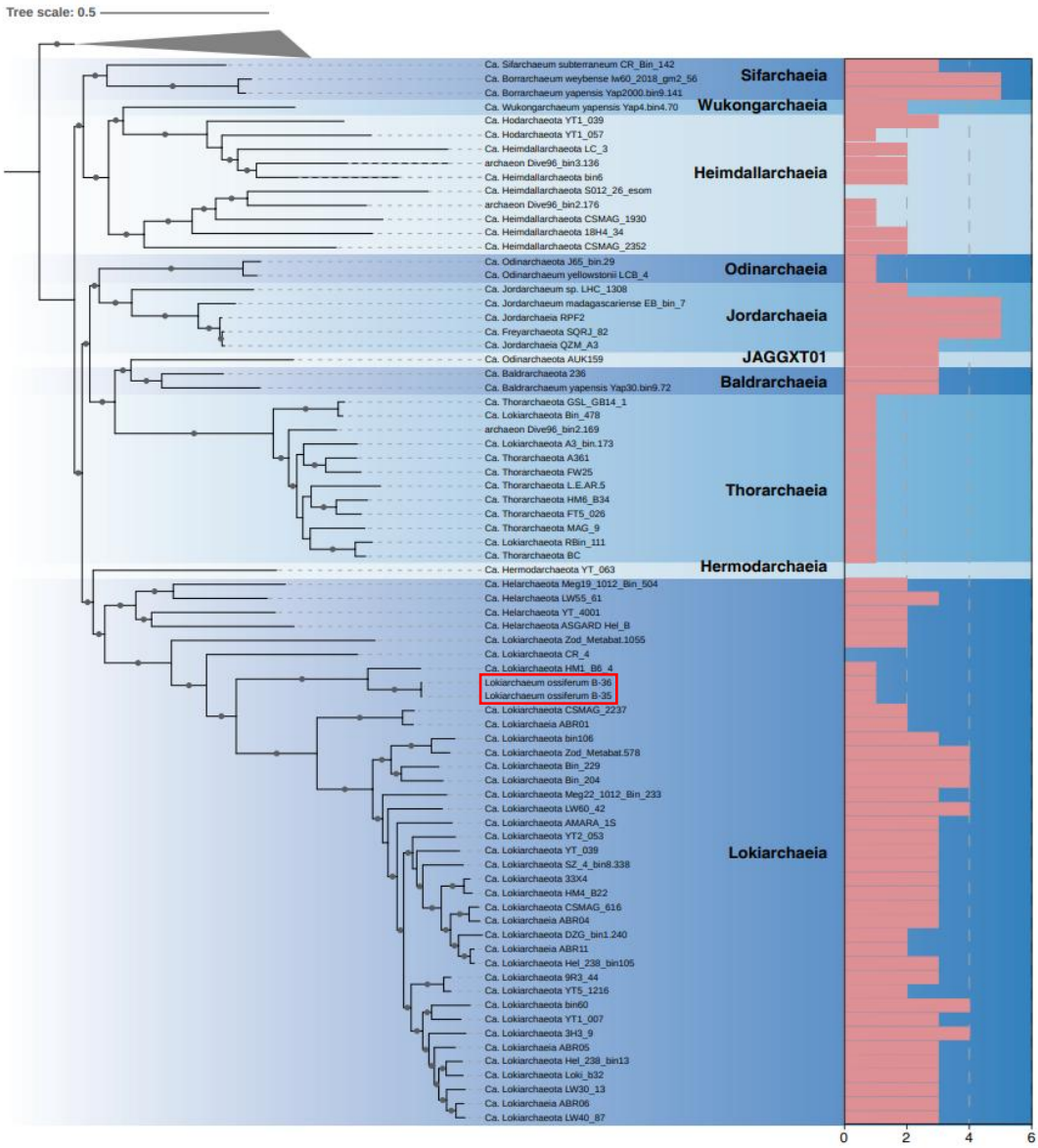


Figure 2. Distribution of Alba homologues in Promethearchaeati. *Ca.* L. ossiferum genomes (strains B35 and B36) are boxed (in red.) The figure was created and provided by Dr. Zhen-Hao Luo.

and purification. To increase the solubility of the protein and enable its purification by affinity chromatography, we fused a 6x His and SUMO tag (12.2 kDa) to the N'-terminal end of the protein.

First, we examined the expression levels of LoAlba. Since Loki thrives at 20ºC, we assessed Alba expression at three different conditions, from the optimal growth temperature for *Ca.* L. ossiferum to that of *E. coli* (20ºC, 27ºC, and 37ºC). Our results indicate that Alba is expressed at all tested temperatures, but optimally at either 27ºC or 37ºC (**Figure 3**). The latter was chosen to express Alba at a larger scale because of convenience (i.e., a higher availability of incubators).

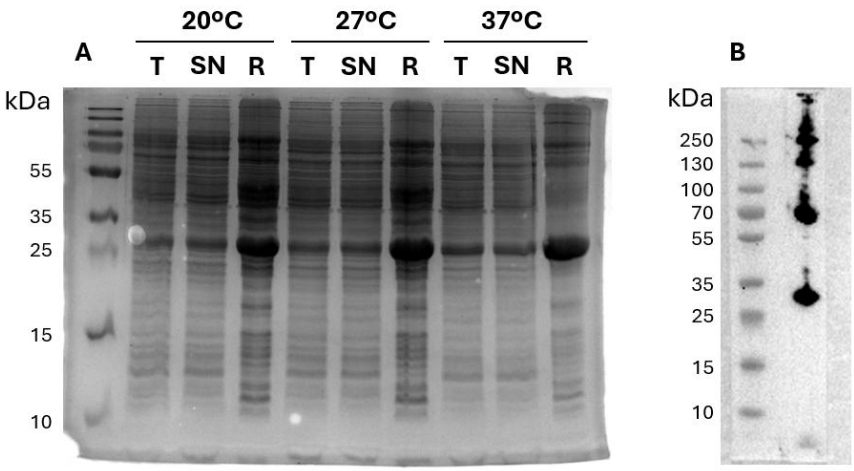
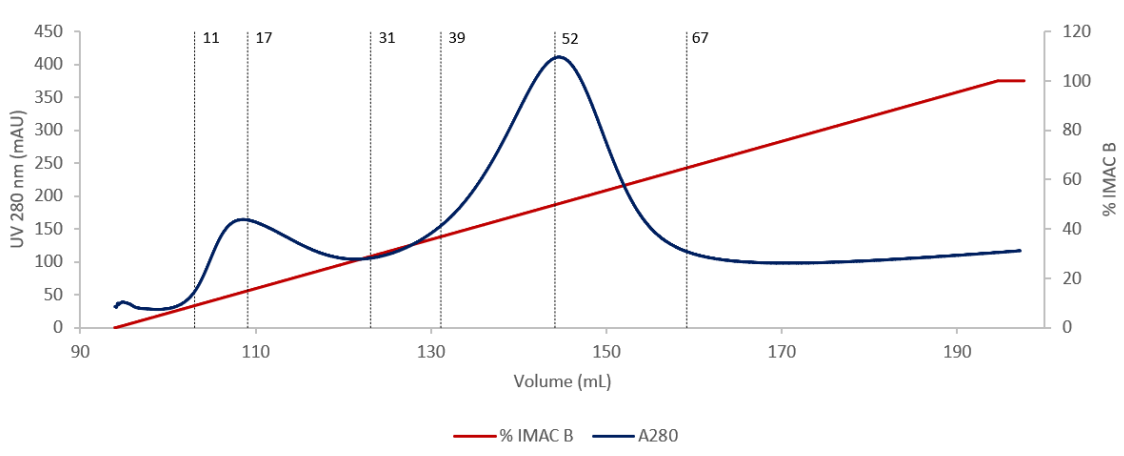


Figure 3. Detection of recombinant 6x-His-SUMO-LoAlba in *E. coli* BL21(DE3) extracts after expression with an autoinduction supplement. **A.** 15% polyacrylamide-SDS gel displaying the total cell lysate (T), soluble fraction (SN), and HisTrap resin-bound fraction (R) of cell extracts after growth at 20ºC, 27ºC, or 37ºC. **B.** Western Blot of fraction R upon growth at 37ºC probed with an anti-6x-His tag antibody. MW 6x-His-SUMO-LoAlba: 25.8 kDa; LoAlba: 13.6 kDa; 6x-His-SUMO: 12.2 kDa.

3.2.2. Purification of LoAlba

E. coli BL21(DE3) carrying pT7-6x-His-SUMO-LoAlba were grown at 37ºC in preparation for purification of LoAlba. Protein extracts were obtained, loaded on a HisTrap column, and eluted with a linear imidazole gradient. For each eluted fraction, the UV light absorbance spectrum at 280 nm (A280) served as an indicator of protein presence (**Figure 4**). Based on the spectrum, aliquots of protein-rich fractions were sampled and analysed by gel electrophoresis. We detected two absorption peaks centred in fractions 17 and 52. However, only the latter fraction correlated with the detection of a band of the expected molecular weight by gel electrophoresis likely corresponding to 6x-His-SUMO-LoAlba (25.8 kDa; concentration = 10.68 mg mL^{-1}). Fractions containing LoAlba (31 to 67) were therefore pooled to proceed with its purification.

A



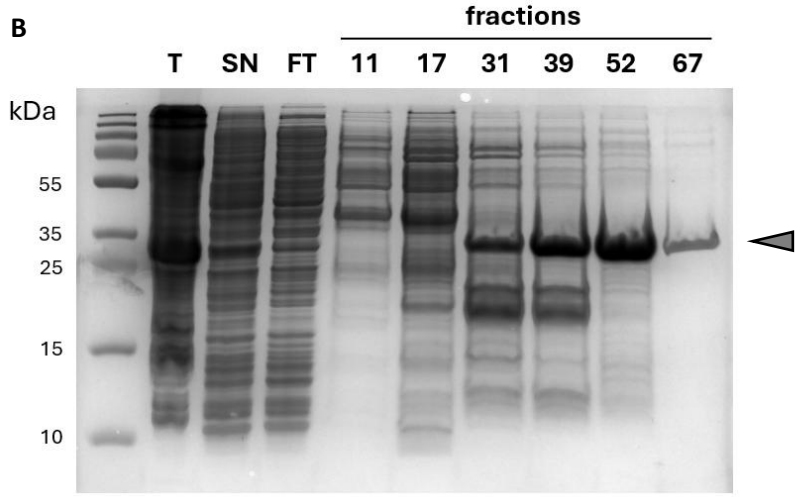


Figure 4. Separation of *E. coli* BL21(DE3) extracts via imidazole gradient after expression with an auto-induction supplement. **A.** A280 spectrum during HisTrap column elution at increasing concentrations of IMAC B. Dashed lines indicate the fractions analysed by gel electrophoresis. **B.** 15% polyacrylamide-SDS gel loaded with selected pre-purification samples and eluted fractions. T: total cell lysate. SN: soluble fraction of the lysate. FT: HisTrap column flowthrough after washing with IMAC A. Grey arrowhead points to putative 6x-His-SUMO-LoAlba.

Fractions 53 and 64 to 67, however, were kept for further biochemical analysis. After dialysis of these selected fractions, we obtained a total of 6.3 mg 6x-His-SUMO-LoAlba.

O/n incubation with Ulp1 resulted in the cleavage of 6x-His-SUMO tag from most LoAlba, although a small portion remained uncleaved (**Figure 5**). When the protein solution was loaded in a chromatography column containing a Ni-NTA resin to separate the cleaved 6x-His-SUMO tag from LoAlba, however, both remained attached to the resin (**Figure 5**). LoAlba was ultimately recovered from the column by saturating the Ni-NTA resin with imidazole, sweeping along the 6x-His-SUMO tag, too. A small fraction of Alba could not be eluted and remained bound to the resin (**Figure 6B**).

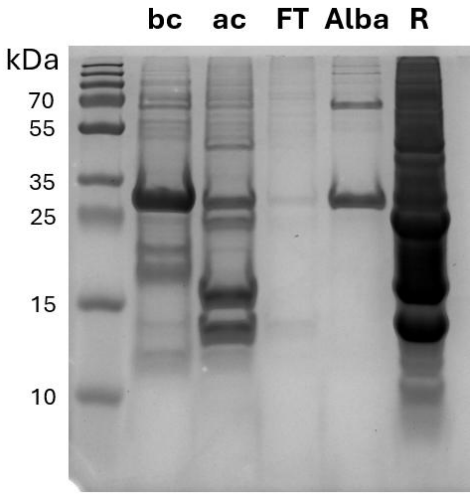


Figure 5. 15% polyacrylamide-SDS gel loaded with samples and controls extracted during rIMAC separation. bc: pooled fractions after IMAC separation and before cleavage by Ulp1. ac: pooled IMAC fractions after dialysis and cleavage by Ulp1. FT: rIMAC chromatography column flowthrough. Alba: pooled IMAC fractions containing uncleaved 6x-His-SUMO-LoAlba after dialysis. R: Ni-NTA resin-bound proteins during rIMAC chromatography.

LoAlba was ultimately isolated from the 6x-His-SUMO tag and the proteins that remained bound to the Ni-NTA resin through SEC chromatography (**Figure 6**). Interestingly, we identified two non-consecutive elution peaks at fractions 17 and 37. We therefore pooled fractions 15 to 22, containing a LoAlba variant henceforth referred to as LoAlba_heavy, and 35 to 43, containing the second variant (LoAlba_light) separately for further characterisation. Mass Spectrometry analysis confirmed that both species corresponded to LoAlba (see **Supplementary Material**). A total of 2.8 mg LoAlba_heavy and 1.7 mg LoAlba_light were obtained. The elution peak centred at fraction 51 corresponded to the cleaved 6x-His-SUMO tag (data not shown).

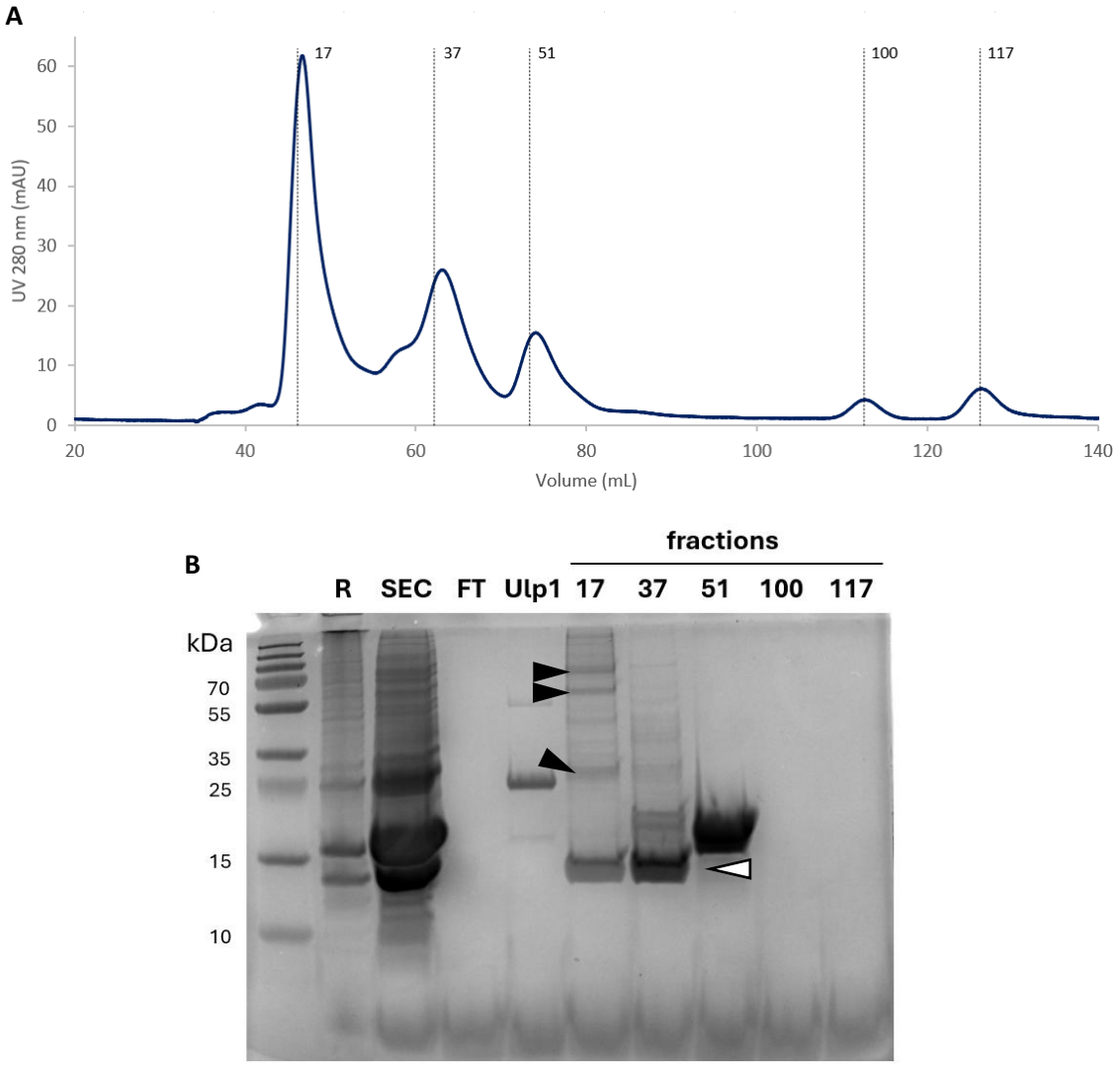


Figure 6. Isolation of LoAlba in a SEC column after elution from a rIMAC column. **A.** A280 spectrum during SEC column elution. Dashed lines indicate the fractions analysed by gel electrophoresis (13, 37, 51, 100, 117). **B.** 15% polyacrylamide-SDS gel loaded with selected pre-purification samples and eluted fractions. R: resin-bound proteins after rIMAC column stripped with imidazole. SEC: SEC column input. FT: flowthrough of the Vivaspin® concentrator column. Ulp1: Ulp1 protein used for 6x-His-SUMO-tag cleavage. Black arrowheads point to LoAlba_heavy and white arrowhead points to LoAlba_light.

We then conducted an additional round of LoAlba purification, incorporating two modifications to improve yield. Based on the DNA-binding behaviour of Alba proteins in *Sulfolobus* spp. **(Laurens et al., 2012)**, we hypothesised that the elution of LoAlba as two distinct variants after SEC could be a consequence of the formation of oligomers that would co-elute with DNA. Therefore, we quadrupled the concentration of benzonase in the Lysis Buffer compared to that of the first round. Given that benzonase is a nuclease used in biochemistry to eliminate nucleic acids (DNA and RNA) from cell lysates, we reasoned that increasing its concentration would reduce nucleic acid contamination and prevent Alba oligomerisation. Furthermore, given that Alba was almost completely bound to the Ni-NTA resin during rIMAC and could not be separated from the 6x-His-SUMO tag after cleavage by Ulp1, this step was omitted and the eluate from the IMAC column was loaded directly on the SEC column, instead.

The elution A280 spectrum from the HisTrap with an imidazole gradient and subsequent PAGE analysis were comparable to those of the former purification round (**Figure 7**). We thus pooled fractions 33 to 66 and then cleaved, dialysed, and loaded them onto the SEC column. Light absorbance at 260 nm (A260), which correlates with the nucleic acid concentration in each fraction, was recorded alongside A280 during the SEC column elution to detect any contamination by nucleic acids (**Figure 8A**). We obtained a spectrum very similar to the one ob-

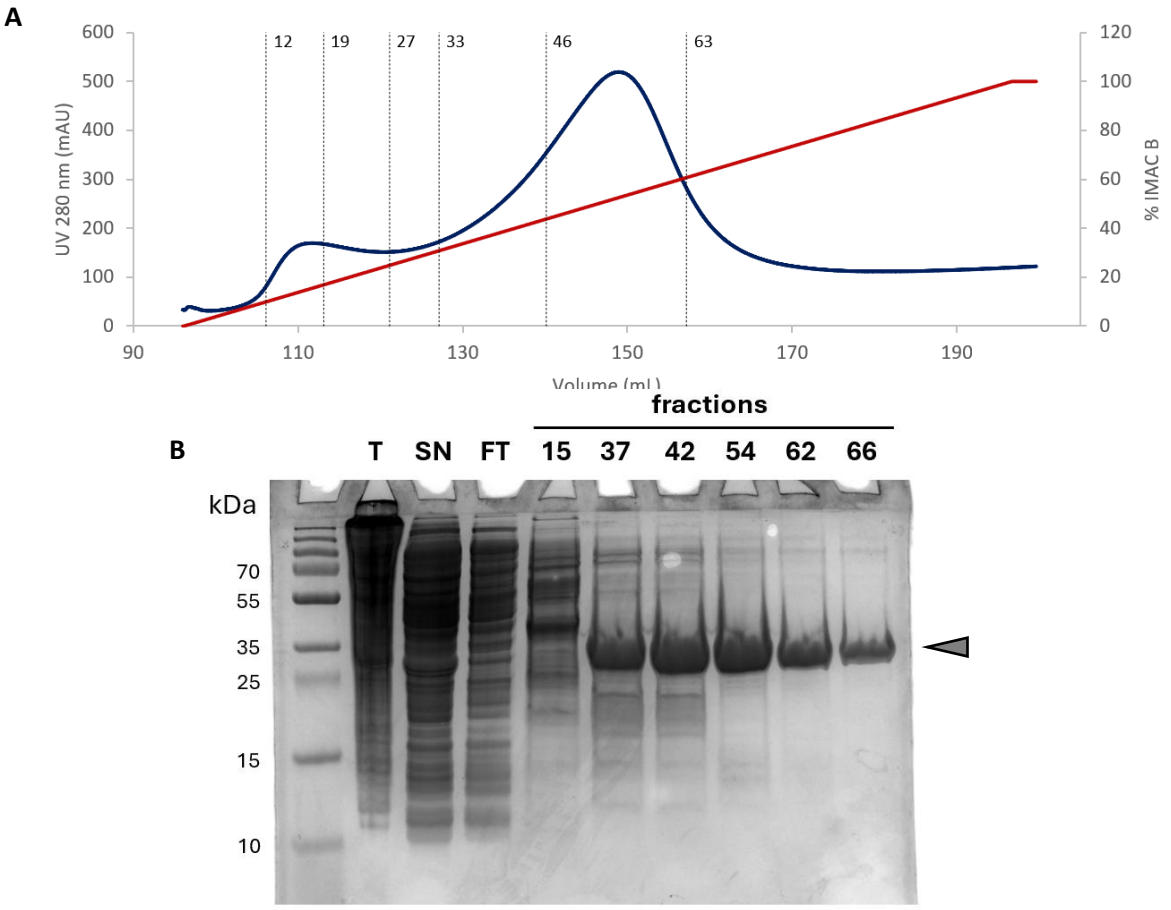


Figure 7. Separation of *E. coli* BL21(DE3) extracts via imidazole gradient after expression with an auto-induction supplement (second purification round). **A.** A280 spectrum during HisTrap column elution at increasing concentrations of IMAC B. Dashed lines indicate the fractions analysed by gel electrophoresis. **B.** 15% polyacrylamide-SDS gel loaded with selected samples pre-purification and eluted fractions. T: total cell lysate. SN: soluble fraction of the lysate. FT: HisTrap column flowthrough after washing with IMAC A. Grey arrowhead points to 6x-His-SUMO-LoAlba.

tained in the first run, although the second one yielded more LoAlba_light and less LoAlba_heavy (**Figure 8A**). We analysed fractions predicted to contain LoAlba_light (36 to 42) by gel electrophoresis (**Figure 8B**) and, after pooling them, we obtained a total of 6.5 mg LoAlba_light. Similarly to the previous run, the elution peak centred at fraction 56 corresponded to the cleaved 6x-His-SUMO tag (data not shown).

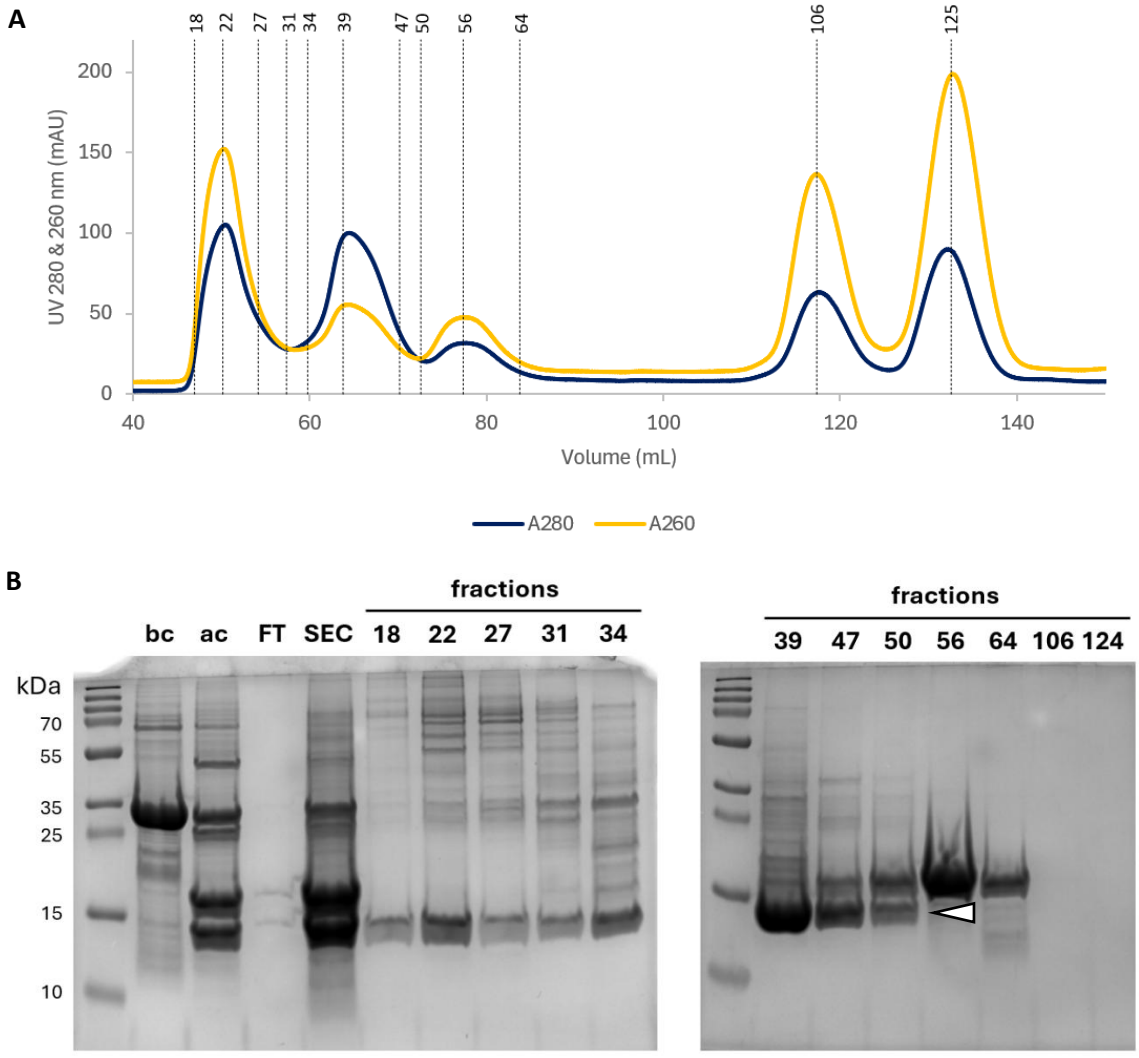


Figure 8. Purification of LoAlba by size-exclusion chromatography after affinity purification on a HisTrap column. **A.** A280 and A260 spectra of SEC column eluates. Dashed lines indicate the fractions analysed by gel electrophoresis. **B.** 15% polyacrylamide-SDS gels loaded with selected pre-purification samples and

eluted fractions. bc: pooled fractions before dialysis and cleavage by Ulp1. ac: pooled fractions after dialysis and cleavage by Ulp1. FT: Vivaspin® concentrator column flowthrough. SEC: protein concentrate loaded onto the SEC column. White arrowhead points to LoAlba_light.

To further elucidate whether the elution of LoAlba as LoAlba_heavy or LoAlba_light is a consequence of the formation of DNA-associated LoAlba oligomers by the former, the ratio between A260 and A280 values was measured in each of the pooled fractions. As expected, A260/A280 was higher in fractions containing LoAlba_heavy (1.401), while it was lower in those containing LoAlba_light (0.424) (**Figures 8A and 9A-B**). In addition, nucleic acids were detected by gel electrophoresis only in fractions containing LoAlba_heavy (**Figure 9C**), suggesting that LoAlba_heavy may consist of nucleic acid-associated Alba oligomers.

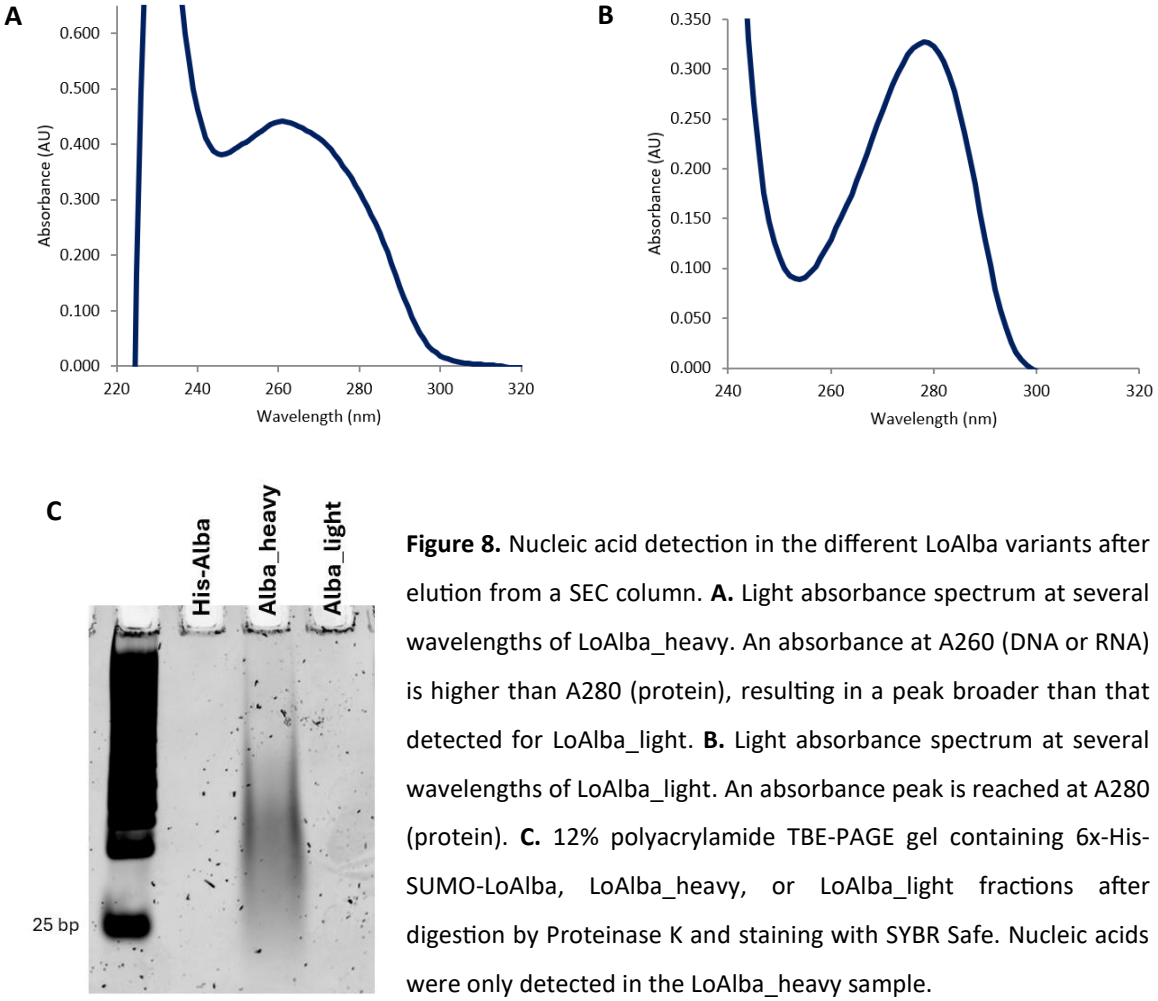


Figure 8. Nucleic acid detection in the different LoAlba variants after elution from a SEC column. **A.** Light absorbance spectrum at several wavelengths of LoAlba_heavy. An absorbance at A260 (DNA or RNA) is higher than A280 (protein), resulting in a peak broader than that detected for LoAlba_light. **B.** Light absorbance spectrum at several wavelengths of LoAlba_light. An absorbance peak is reached at A280 (protein). **C.** 12% polyacrylamide TBE-PAGE gel containing 6x-His-SUMO-LoAlba, LoAlba_heavy, or LoAlba_light fractions after digestion by Proteinase K and staining with SYBR Safe. Nucleic acids were only detected in the LoAlba_heavy sample.

In summary, we successfully isolated recombinant LoAlba as two distinct variants, a heavier variant bound to nucleic acids, likely oligomeric, and a free, 13.6 kDa variant.

3.2.3. LoAlba binds DNA non-specifically *in vitro*

We then tested the DNA binding capacity of LoAlba by EMSA. Namely, we incubated increasing concentrations of LoAlba_heavy, LoAlba_light, or 6x-His-SUMO-LoAlba with a fragment of fluorescently labelled DNA at the optimal growth temperature for Loki (20ºC) and observed its migration in a 1% agarose gel. All LoAlba species tested retarded DNA (**Figure 10A-B**). However, LoAlba_light bound DNA at lower concentration than LoAlba_heavy (**Figure 10A**), indicating a greater affinity for DNA. 6x-His-SUMO-LoAlba displayed the lowest affinity for DNA, retarding it only at the highest concentration tested (**Figure 10B**). In addition, when the competitor for unspecific DNA binding poly(dI·dC) was included in the reaction, the affinity of LoAlba_light for DNA decreased in inverse relation to the concentration of competitor used (i.e., the more competitor we added, the less LoAlba_light retarded the DNA fragment) (**Figure 10C**).

Interestingly, heat inactivation of LoAlba_heavy and LoAlba_light was insufficient to inhibit their DNA binding capacity (**Figure 10A**). Therefore, in the subsequent EMSAs, denaturing agents were added to the reaction to inhibit DNA-binding by LoAlba (see **Section 2.5.2**).

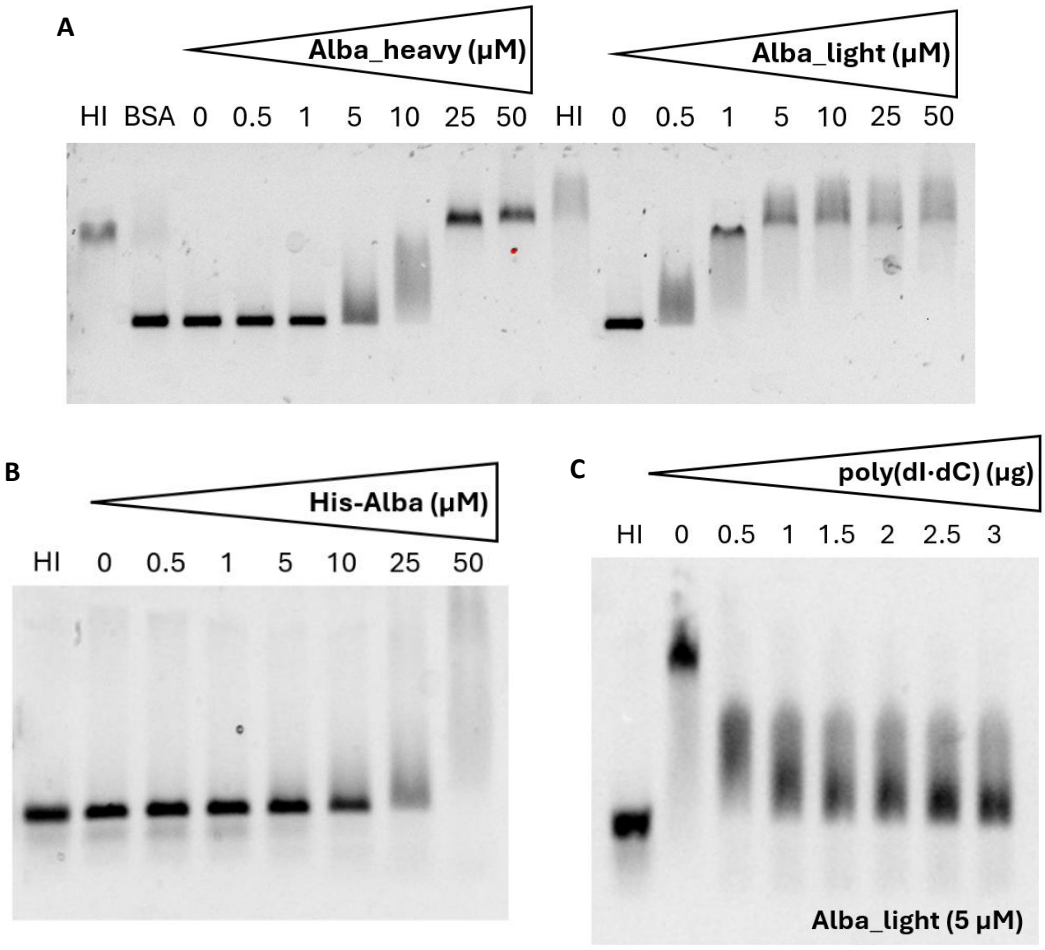


Figure 10. DNA EMSAs of different Alba species with or without the presence of the competitor poly(dI·dC). Images display the Cy5 fluorescence signal (λ = 665 nm). **A.** EMSAs at increasing concentrations of LoAlba_heavy and LoAlba_light and constant DNA concentration (1.5 ng μL^{-1}). **B.** EMSA at increasing concentrations of 6x-His-SUMO-LoAlba and constant DNA concentration (1.5 ng μL^{-1}). **C.** Competitive EMSA at constant DNA (1.5 ng μL^{-1}) and Alba_free concentration and increasing poly(dI·dC) concentrations. BSA: negative control reaction using BSA instead of LoAlba. HI: negative control with a heat-inactivated and/or denatured LoAlba. In both controls, protein concentration corresponds to the maximum concentration of LoAlba used in each assay.

To further characterise the DNA-binding capacity of LoAlba_heavy, LoAlba_light, and 6x-His-SUMO-LoAlba, we analysed them by mass photometry in the absence or presence of DNA. Unlike LoAlba_light or 6x-His-SUMO-LoAlba, LoAlba_heavy appeared in multiple oligomeric states of different size (see three yellow peaks in **Figure 11A**). In addition, the presence of DNA appeared to favour multimerisation (**Figure 11B**). Whether the smallest LoAlba_heavy species, LoAlba_light or 6x-His-SUMO-LoAlba corresponded to protein monomers or dimers is unclear, due to the detection threshold of mass photometry being > 30 kDa. Finally, even upon addition of DNA, we identified neither LoAlba_light nor 6x-His-SUMO-LoAlba oligomers (data not shown).

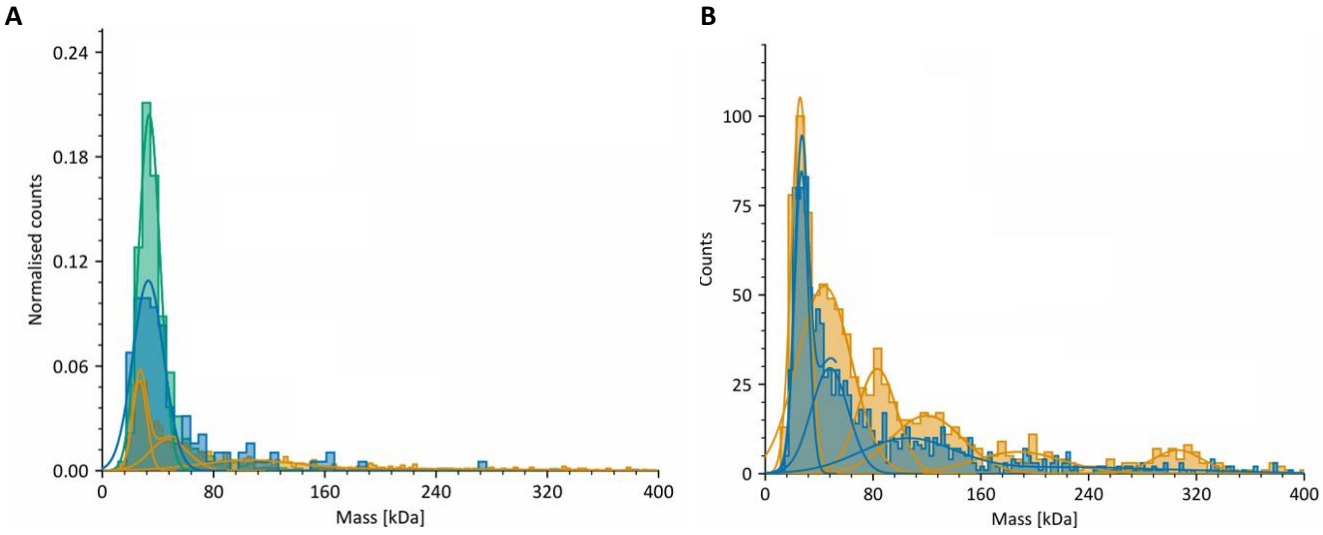


Figure 11. Mass photometry analysis of LoAlba_heavy, LoAlba_light, and 6x-His-SUMO-LoAlba. **A.** Comparison between the masses of 6x-His-SUMO-LoAlba (200 nM, dark blue), LoAlba_heavy (50 nM, orange), and LoAlba_light (50 nM, green). **B.** LoAlba_heavy mass shift in the presence of an 80 bp-long DNA fragment (blue: no DNA; orange: DNA added).

All in all, purified recombinant *Ca.* L. ossiferum Alba bound DNA *in vitro* whereas its 6x-His-SUMO tagged counterpart did not. Moreover, we found that LoAlba_heavy – a variant of recombinant *Ca.* L. ossiferum Alba co-purified with nucleic acids – had less affinity for DNA in EMSA than a seemingly nucleic acid-free variant, LoAlba_light. Conversely, in mass photometric assays,

LoAlba_heavy appeared more prone to multimerize in the presence of DNA. Fearing that the presence of complex nucleic acids might confound the functional characterization of recombinant *Ca*. L. ossiferum Alba, we exclusively used the LoAlba_light variant in all ensuing assays involving purified recombinant LoAlba.

3.2.4. LoAlba binds DNA *in vivo*

Following the observation that recombinant LoAlba binds DNA *in vitro*, we tested whether it would bind DNA *in vivo*. Given that no antibody against LoAlba is available yet, nor can we edit the *Ca.* L. ossiferum genome, we tested LoAlba-DNA binding in *E. coli* BL21(DE3).

NAPs are generally considered to partly protect DNA from nucleolytic cleavage by limiting its accessibility by nucleases **(Farrants, 2018; Carone, 2025)**. MNase is commonly used to identify regions of the genome which are protected from nucleolytic cleavage because they are associated with DNA-binding proteins, such as NAPs or histones **(Tolstorukov et al., 2016; Chereji et al., 2017)**. We hypothesised that, if recombinant LoAlba bound the genome of *E. coli*, it would protect it from cleavage. Thus, we compared the digestion pattern of *E. coli* BL21(DE3) genomic DNA in an Alba-IPTG-induced and non-induced strains after incubation of cell extracts with increasing MNase concentrations. Contrary to our expectations, *E. coli* DNA was more efficiently digested when LoAlba was expressed than when it was not (**Figure 12**). Therefore, MNase digestion does not support the Loki Alba binding to *E. coli* DNA *in vivo*.

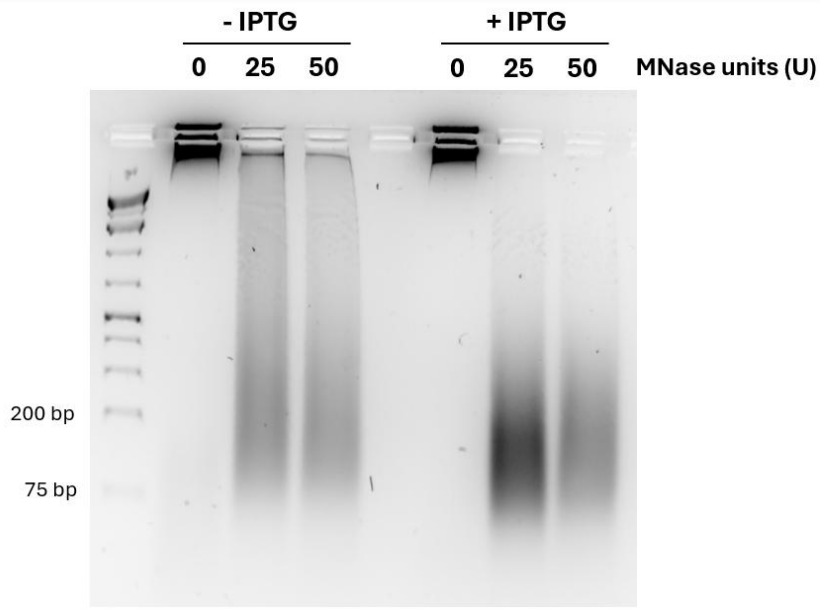


Figure 12. MNase digestion assay of *E. coli* BL21(DE3) genomic DNA after LoAlba induction with IPTG, or from an uninduced control, at increasing MNase concentrations.

Therefore, we decided to conduct a Chromatin Immunoprecipitation and Sequencing (ChIP-Seq) experiment using an anti-6x-His tag antibody on *E. coli* BL21(DE3) after expression of LoAlba. Hits were compared with an *E. coli* BL21(DE3) strain carrying an empty pT7-SUMO vector, i.e., without any insert for recombinant LoAlba expression. DNA sequencing revealed the enrichment of LoAlba in certain genomic regions in *E. coli* compared to the control, but these did not correlate with high or low GC content or with the presence or absence of specific genomic regions, like promoters or intergenic regions (**Figure 13**).

Collectively, our results indicate that recombinant LoAlba binds *E. coli* BL21(DE3) DNA *in vivo* non-specifically but that it does not protect it from digestion by MNase.

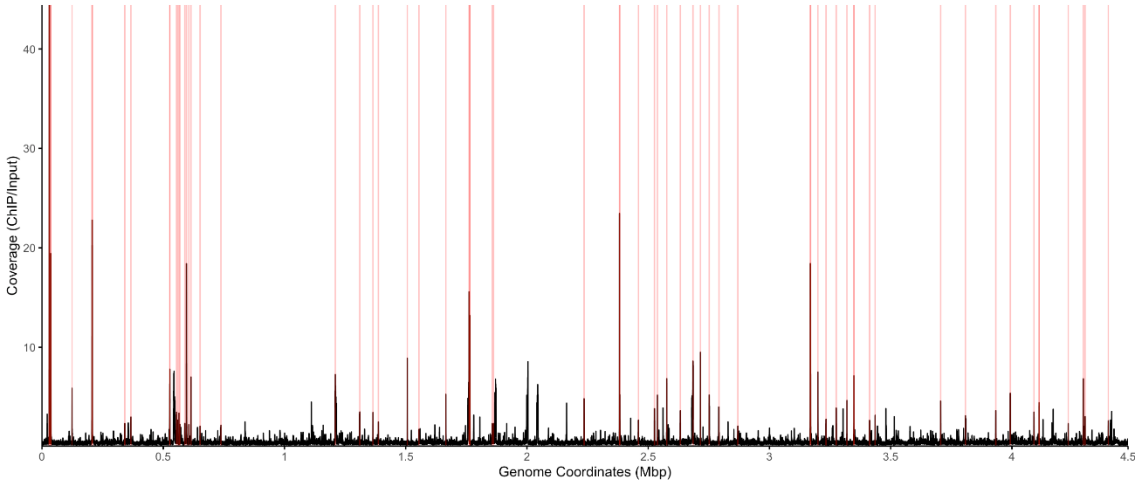


Figure 13. Peak profile of ChIP-Seq results after capturing Alba. Red lines indicate genomic locations significantly enriched with Loki Alba.

3.3. Expression of Loki Alba leads to DNA compaction in *E. coli* BL21(DE3)

After testing the DNA-binding activity of *Ca.* L. ossiferum Alba *in vitro* and *in vivo*, we decided to visually examine whether its expression impacted chromatin structure in *E. coli* BL21(DE3). *S. solfataricus* Alba1 homodimers compact DNA by bridging distant chromatin regions and placing them in close spatial proximity **(Luijsterburg et al., 2008; Laurens et al., 2012)**. Considering the great structural similarity between LoAlba and Alba1 (see **Section 3.1**), we hypothesised that LoAlba could compact the DNA in a similar way.

E. coli BL21(DE3) + pT7-6x-His-SUMO-LoAlba were grown until mid-late exponential phase in the absence or presence of IPTG. These were then harvested and stained with Hoechst dye for DNA visualisation and microscopic analysis. Only in cells recombinantly expressing LoAlba, we

observed a significant reduction in their nucleocytoplasmic (NC) ratio ($p < 0.001$) (**Figure 14A-B, D, F**). To validate that this effect is specifically caused by the expression of LoAlba but not by the expression of other DNA-binding or non-DNA-binding recombinant proteins, as reported in the literature **(Yilmaz and Schnetz, 2022; Thompson et al., 2022)**, we also examined the DNA signal of *E. coli* BL21(DE3) recombinantly expressing *Methanobrevibacter smithii* SepF or *Simonsiella muelleri* ParB. We observed a slight decrease in NC ratio upon expression of *M. smithii* SepF, although significantly less pronounced than that seen upon expression of LoAlba ($p < 0.001$). As for *S. muelleri* ParB, its expression led the opposite, a NC ratio increase rather than a decrease ($p < 0.001$) (**Figure 14A, C, E-F**).

In addition, proteomic analysis of strains actively expression LoAlba revealed no upregulation of proteins involved in general stress responses (data not shown). A retardation in growth was observed in all strains upon induction of LoAlba variants with IPTG when grown at identical temperature (**Figure 14G**).

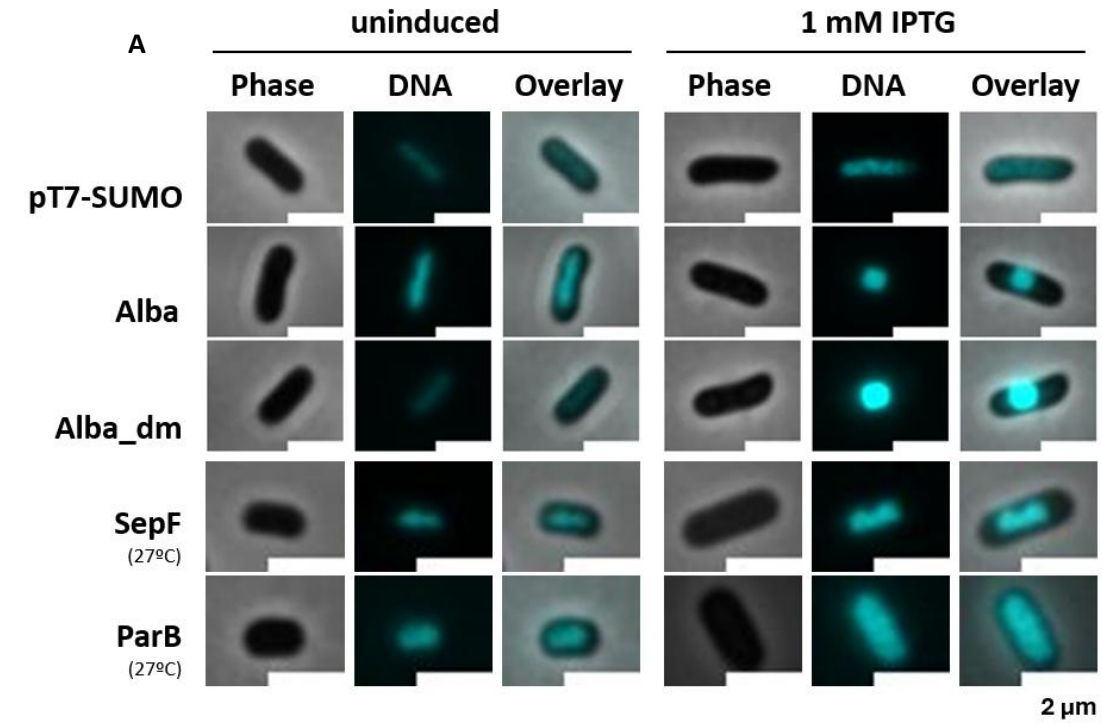
Unexpectedly, expression of LoAlba(p.Lys31Ala;Phe74Ala), a mutant derivative of LoAlba containing a substitution in two residues presumably involved in DNA binding, bridging, and oligomerisation (see **Section 3.1**), henceforth referred to as LoAlba_dm for LoAlba (double mutant), compacted DNA as much as LoAlba did ($p < 0.001$) (**Figure 14A-B, D, F**).

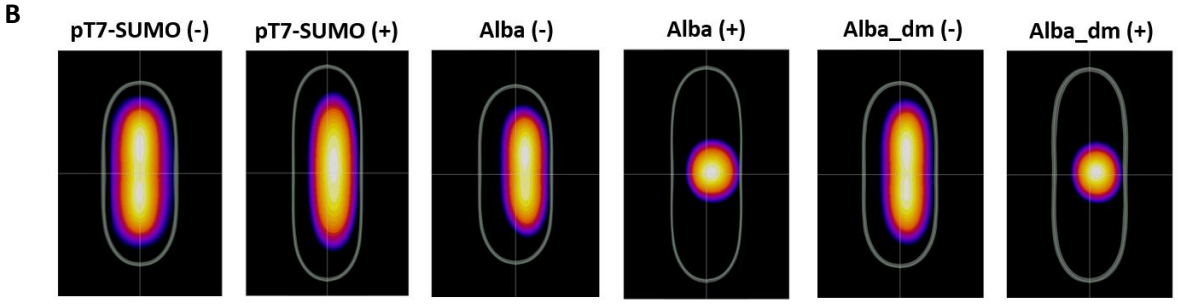
We therefore conclude that expression of recombinant LoAlba leads to chromatin compaction in *E. coli* BL21(DE3) without eliciting a stress response and through a mechanism independent of residues Lys31 and Phe74.

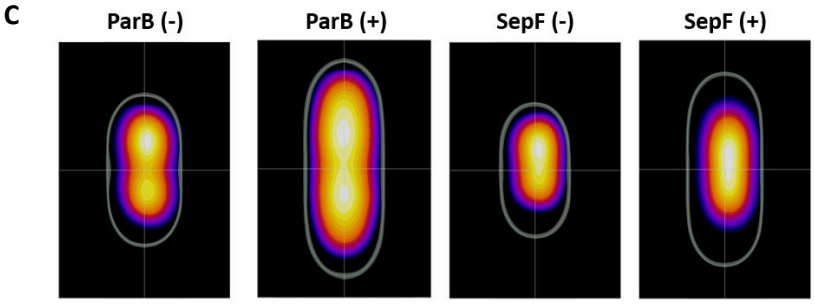
3.4. Loki Alba binds RNA *in vitro*

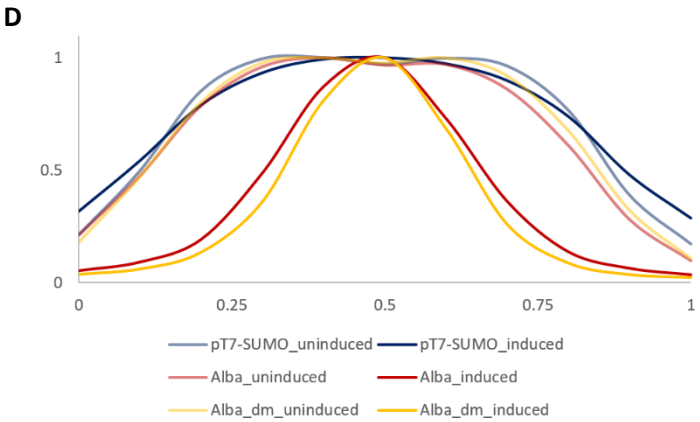
Not only are Alba proteins important DNA-binding NAPs in archaea, but RNA-binding properties have also been reported in both the archaeal domain and, most predominantly, in eukaryotes **(Jagadeesh and Vembar, 2024)**. Because of the critical evolutionary junction between both domains, I decided to examine whether LoAlba also exhibited RNA-binding activity.

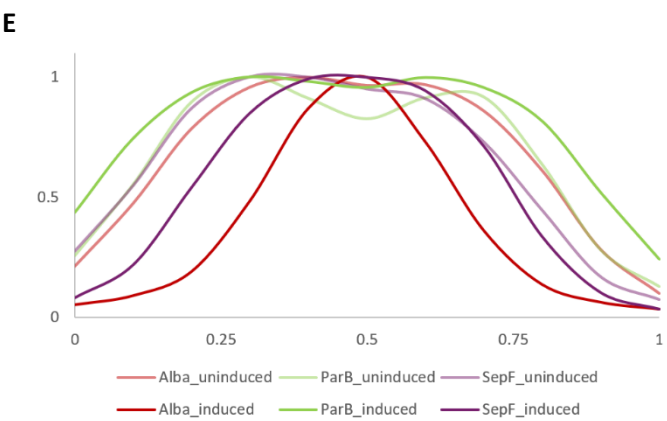
An EMSA was conducted by mixing the complementary RNA to the DNA sequence used in other EMSAs at an equivalent molarity (see **Section 3.2.3**) with increasing concentrations of LoAlba_light. A complete shift was observed from a concentration of LoAlba_light as low as 0.5 µM, suggesting a greater affinity for RNA than DNA (**Figure 15A-B**). This behaviour was consistent throughout different competitive EMSAs, in which LoAlba_light was incubated with


A
uninduced
1 mM IPTG
Phase
DNA
Overlay
Phase
DNA
Overlay
pT7-SUMO
Alba
Alba_dm
SepF
(27ºC)
ParB
(27ºC)
2 µm


B
pT7-SUMO (-)
pT7-SUMO (+)
Alba (-)
Alba (+)
Alba_dm (-)
Alba_dm (+)


C
ParB (-)
ParB (+)
SepF (-)
SepF (+)


D
1
0.5
0
0
0.25
0.5
0.75
1
pT7-SUMO_uninduced
pT7-SUMO_induced
Alba_uninduced
Alba_induced
Alba_dm_uninduced
Alba_dm_induced


E
1
0.5
0
0
0.25
0.5
0.75
1
Alba_uninduced
ParB_uninduced
SepF_uninduced
Alba_induced
ParB_induced
SepF_induced

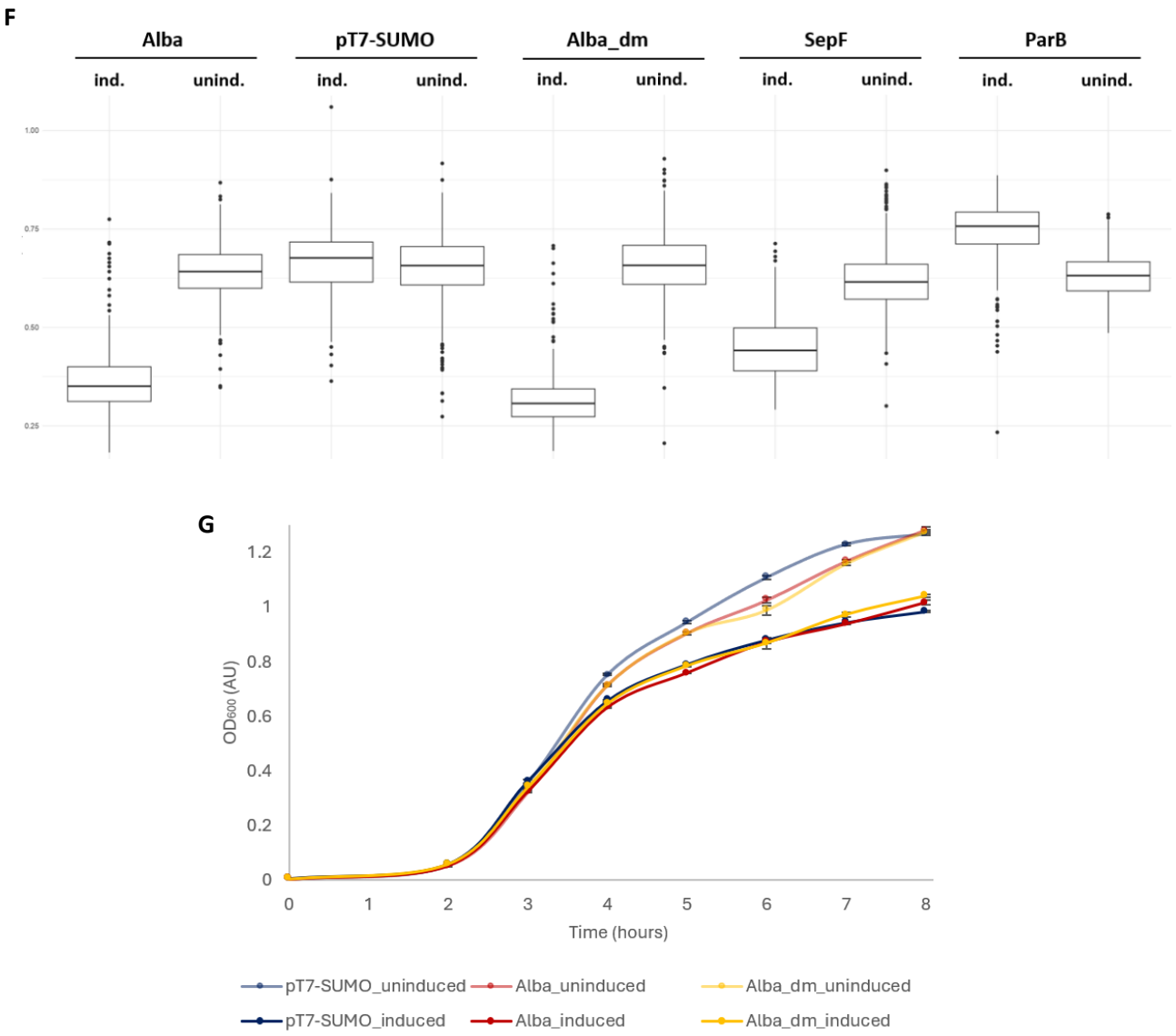
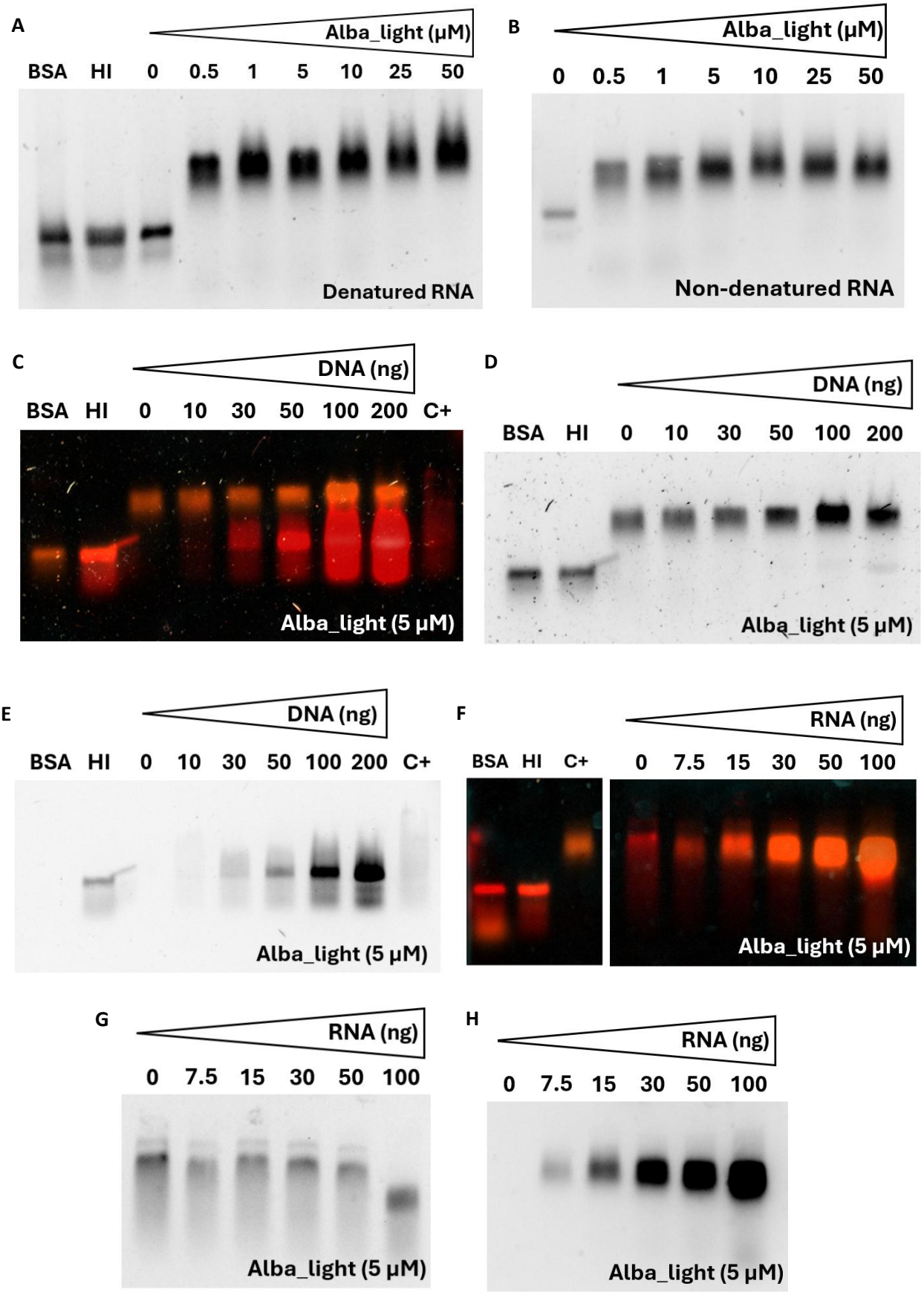


Figure 14. Analysis of the effects on DNA localisation and growth in *E. coli* BL21(DE3) upon expression of recombinant LoAlba. **A.** Representative images of all IPTG-induced and uninduced expression strains. **B, C.** Representation of the mean fluorescence intensity of DAPI-stained DNA along the normalised cell shape, averaged across all cells analysed. Horizontal and vertical axes correspond to a total length of 2.2 and 3 µm. Brighter colours indicate a higher mean fluorescence intensity, while darker colours indicate a lower value. The signal depiction is scaled to fit within the normalised cell outline. Number of cells analysed per strain: pT7-SUMO (-): 1,580; pT7-SUMO (+): 471; Alba (-): 739; Alba (+): 806; Alba_dm (-): 1,298; Alba_dm (+): 390; ParB (-): 303; ParB (+): 465; SepF (-): 835; SepF (+): 350. **D, E.** Normalised mean fluorescence intensity along the normalised cell length, averaged across all cells analysed. X: normalised cell length. Y: normalised mean fluorescence intensity. **F.** Cell area covered by fluorescent signal per cell per strain and treatment, represented as the normalised area under the curve (AUC) in figure D. **G.** Growth curves of expression strains for the empty pT7 vector, LoAlba, and Alba_dm, with or without IPTG induction. Error bars indicate the standard deviation of the mean OD_{600} value.

both DNA and RNA at different concentration ratios (**Figure 15C-K**). All in all, LoAlba preferentially bound RNA to DNA *in vitro*.



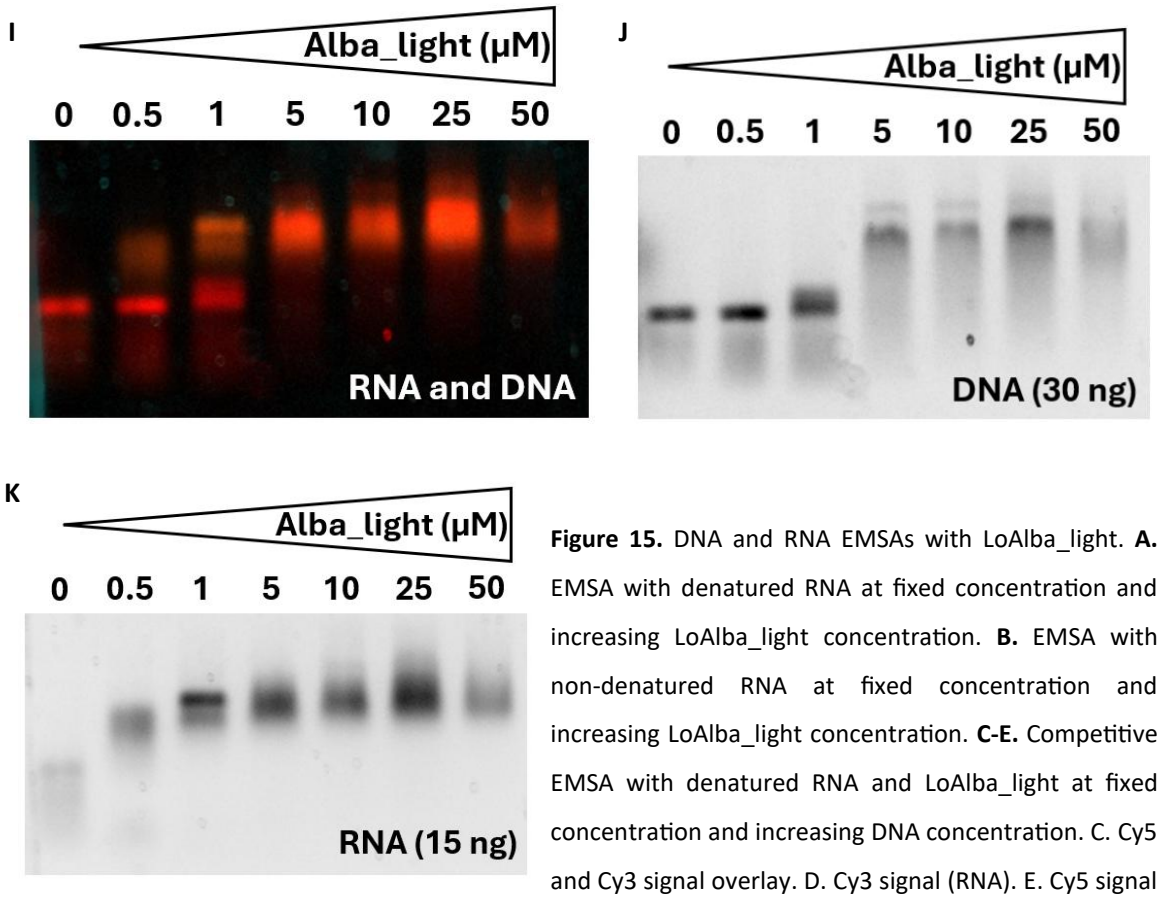


Figure 15. DNA and RNA EMSAs with LoAlba_light. **A.** EMSA with denatured RNA at fixed concentration and increasing LoAlba_light concentration. **B.** EMSA with non-denatured RNA at fixed concentration and increasing LoAlba_light concentration. **C-E.** Competitive EMSA with denatured RNA and LoAlba_light at fixed concentration and increasing DNA concentration. C. Cy5 and Cy3 signal overlay. D. Cy3 signal (RNA). E. Cy5 signal (DNA). **F-H.** Competitive EMSA with DNA and LoAlba_light at fixed concentrations and increasing concentration of denatured RNA. F. Cy5 and Cy3 signal overlay. G. Cy5 signal (DNA). H. Cy3 signal (RNA). **I-K.** Competitive EMSA with denatured RNA and DNA at fixed concentrations and increasing LoAlba concentration. I. Cy5 and Cy3 signal overlay. J. Cy5 signal (DNA). K. Cy3 signal (RNA). BSA: negative control reaction using BSA instead of LoAlba_light. HI: negative control with a heat-inactivated and denatured LoAlba_light. In both controls, protein concentration corresponds to the maximum LoAlba concentration used in each assay. C+: positive controls of DNA or RNA shift caused by binding of LoAlba_light.

3.5. Identification of LoTrmBL2a and LoTrmBL2b

We identified the genes for LoTrmBL2a (*trmBL2a*) and LoTrmBL2b *(trmBL2b*) on the Loki-B35 GCA_025839675.1 genome assembly by homology with TrmB-related proteins. We identified 8 TrmB-like homologues in *Ca.* L. ossiferum, of which only LoTrmBLa and LoTrmBL2b belonged to the TrmBL2 protein subfamily, displaying high sequence similarity to *Natrialba magadii* TrmBL2 (data not shown). LoTrmBL2a is a globular protein of 265 amino acids with a predicted pI of 9.08 and predicted MW of 30.77 kDa. Similarly, LoTrmBL2b is a globular protein of 262 amino acids with a predicted pI of 9.07 and predicted MW of 30.23 kDa. While LoTrmBL2b is abundant in the proteome of *Ca.* L. ossiferum (top 250), expression of LoTrmBL2a is lower (top 1,200; data not

shown). Moreover, both LoTrmBL2a and TrmBL2b display residues reportedly involved in DNA binding (Tyr 52 and Asp53 in LoTrmBL2a and Tyr56 and Glu57 in LoTrmBL2b), conserved among archaeal TrmBL2 proteins **(Ahmad et al., 2015)**.

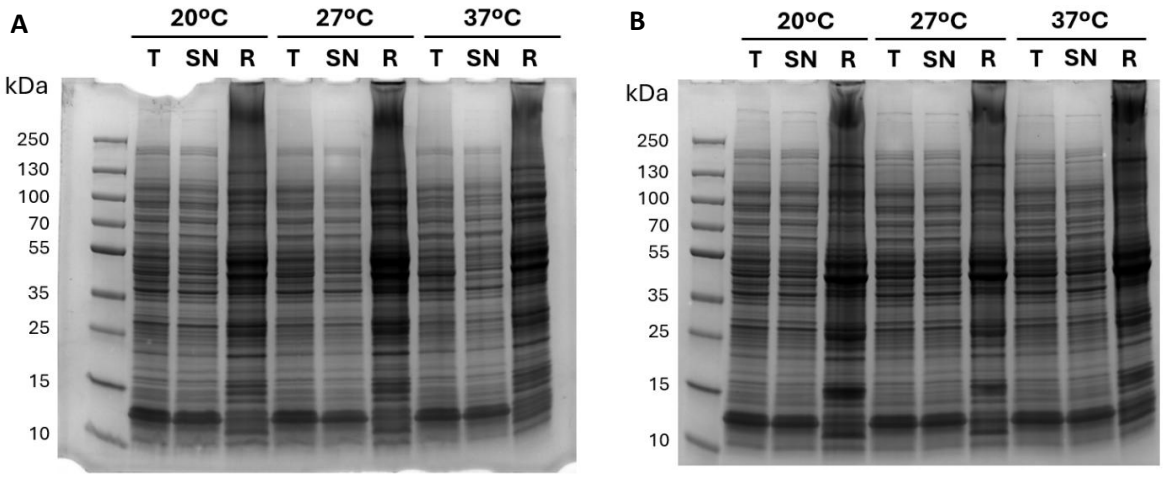
3.6. Loki TrmBL2b binds DNA *in vitro*

3.6.1. Expression of LoTrmBL2a and LoTrmBL2b in *E. coli* BL21(DE3) and C41(DE3)

Like Alba, TrmBL2 proteins in archaea have been described as nonspecific DNA-binding proteins **(Ahmad et al., 2015)**. To determine whether the two TrmBL2 homologues in Loki exhibit similar properties, we first sought to isolate LoTrmBL2a and LoTrmBL2b from a recombinant expression strain.

LoTrmBL2a and LoTrmBL2b were expressed in *E. coli* BL21(DE3) at 20ºC, 27ºC, and 37ºC. As described in **Section 3.2.1**, proteins were fused to a 6x-His-SUMO tag at their N'-terminal ends and were expressed in an autoinduction medium under the control of a *lac* promoter. Both proteins were expressed optimally at 27ºC and 37ºC, although overall yields were lower than those observed for LoAlba. In addition, we could not detect LoTrmBL2a in the Ni-NTA resin-bound fraction (data not shown).

We hypothesised that the low expression levels of LoTrmBL2a and LoTrmBL2b might be the consequence of a toxic effect when overexpressed in *E. coli*. Thus, we constructed new expression strains for both proteins based on *E. coli* C41(DE3), an *E. coli* cell line optimised for the expression of toxic proteins **(Dumon-Seignovert et al., 2004)** and examined their expression levels. Expression of both proteins was indeed enhanced in the new strains at all temperatures tested, and was optimal, again, at 27ºC and 37ºC (**Figure 16A-B**). However, most LoTrmBL2a was insoluble and therefore could not be recovered in the Ni-NTA resin (**Figure 16C-D**). LoTrmBL2b,



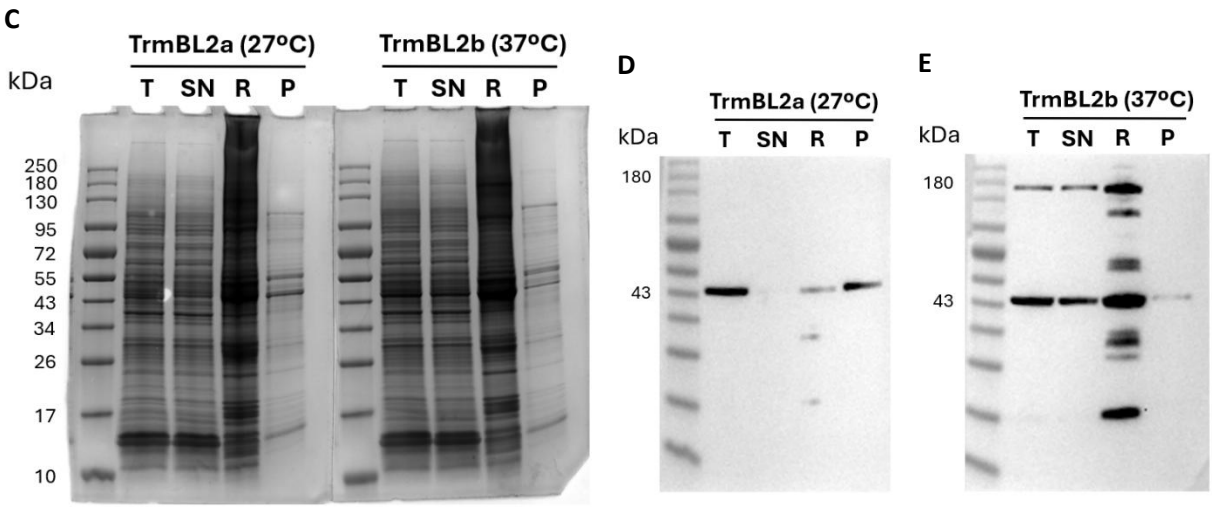


Figure 16. Recombinant expression of 6x-His-SUMO-LoTrmBL2a and LoTrmBL2b in *E. coli* C41(DE3) in the presence of an autoinduction supplement. **A, B.** Coomassie stained ProteinEle® Precast 4-20% Tris-Glycine gels displaying the total cell lysate (T), soluble fraction (SN), and HisTrap resin-bound fraction (R) of cell extracts after growth at 20ºC, 27ºC, or 37ºC. **C.** Coomassie stained ProteinEle® Precast 4-20% Tris-Glycine gel displaying T, SN, R, and the pelleted lysate fraction (P) of cell extracts after growth at 27ºC or 37ºC. **D, E.** Western Blot of T, SN, R, and P probed with an anti-6x-His tag antibody. MW 6x-His-SUMO-LoTrmBL2a: 42.9 kDa; LoTrmBL2a: 30.8 kDa 6x-His-SUMO-LoTrmBL2b: 42.3 kDa; LoTrmBL2b: 30.2 kDa.

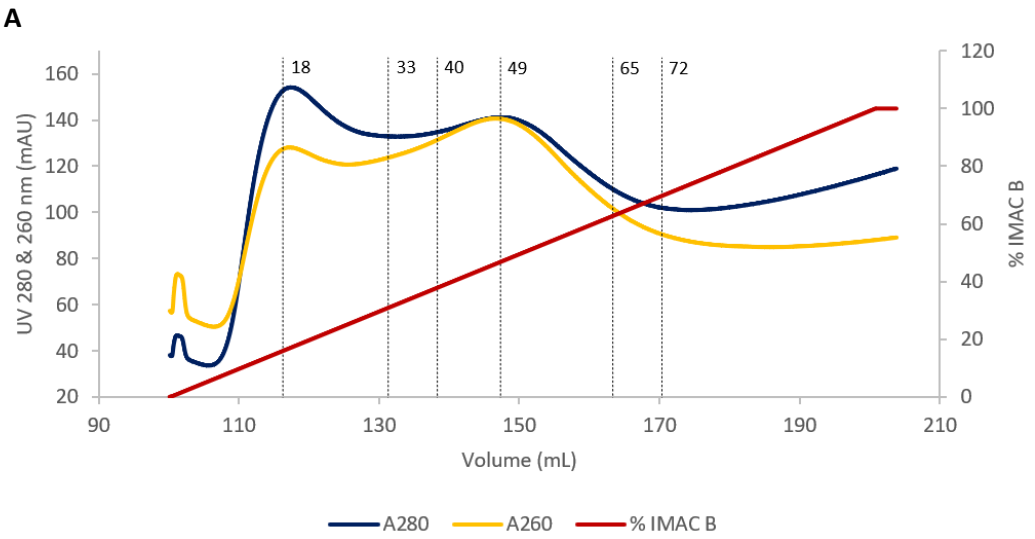
on the contrary, was soluble and effectively bound the resin (**Figure 16E**). Interestingly, a protein band was detected reproducibly at a size of ~180 kDa when probing Western Blots with an antibody against the 6x-His tag. Such band could correspond to a 6x-His-SUMO-LoTrmBL2b tetramer.

3.6.2. Purification of LoTrmBL2b

We first attempted to purify recombinant LoTrmBL2b from *E. coli* BL21(DE3). Although the protein was recovered, purity and yield were insufficient for functional characterisation or the production of antibodies against LoTrmBL2b (data not shown). Thus, we expressed LoTrmBL2b in *E. coli* C41(DE3) at 37ºC in an autoinduction medium until saturation. Like for LoAlba, we loaded protein extracts onto a HisTrap column, applied a linear imidazole gradient, and analysed A280 to determine the protein concentration in the eluted fractions. Two distinct absorption peaks were detected, centred in fractions 18 and 49, but only the latter correlated with the presence of a band of the expected MW on PAGE predicted to be 6x-His-SUMO-LoTrmBL2b (**Figure 17**). Fractions 33 to 70 were therefore collected and pooled for dialysis. We then pooled fractions 71 and 72 separately for further biochemical analysis of the untagged 6x-His-SUMO-LoTrmBL2b. 0.7 mg of tagged protein were obtained.

As performed for LoAlba, o/n incubation with Ulp1 resulted in the cleavage of the 6x-His-SUMO tag from most LoTrmBL2b, although a small portion of the protein remained uncleaved (**Figure 18B**). During elution from a SEC column, the presence of LoTrmBl2b did not correlate with a maximum value of the A280 profile, but rather with the shoulder that followed the main absorbance peak (**Figure 18A-B**). We pooled fractions 21 to 29 based on the detection of TrmBL2b in a polyacrylamide gel and concentrated them into a volume of 1.5 mL. 1.3 mg protein were obtained. Pooled fractions were checked by Mass Spectrometry, revealing that TrmBL2b was only obtained at a low purity (see **Supplementary Material**).

Throughout elution, A260 remained lower than A280 (A260/A280 = 0.603) in fractions containing LoTrmBL2b, indicating that, unlike LoAlba, LoTrmBL2b did not co-elute with nucleic acids. Accordingly, we did not detect nucleic acids upon SYBR-Safe staining of fractions after digestion with proteinase K (data not shown).



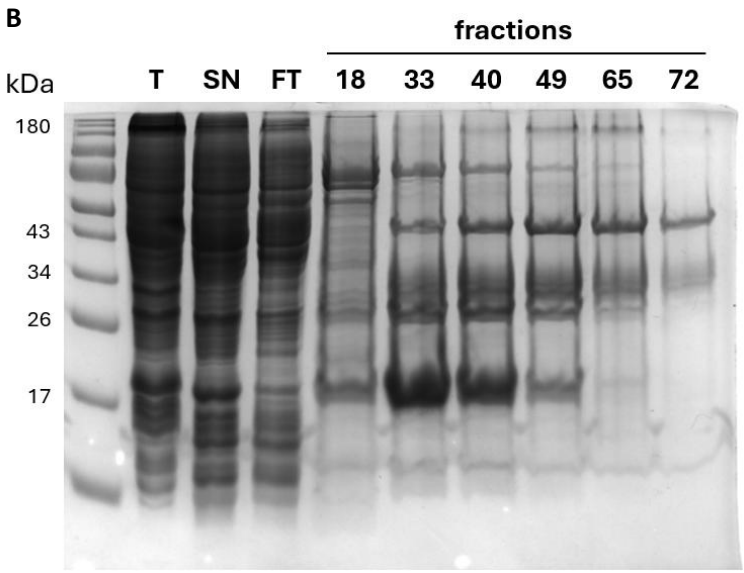


Figure 17. Separation of *E. coli* C41(DE3) extracts via imidazole gradient after expression with an auto-induction supplement. **A.** A280 and A260 spectra during HisTrap column elution at increasing concentrations of IMAC B. Dashed lines indicate the fractions analysed by gel electrophoresis. **B.** 15% polyacrylamide-SDS gel loaded with selected samples pre-purification and eluted fractions. T: total cell lysate. SN: soluble fraction of the lysate. FT: HisTrap column flowthrough after washing with IMAC A.

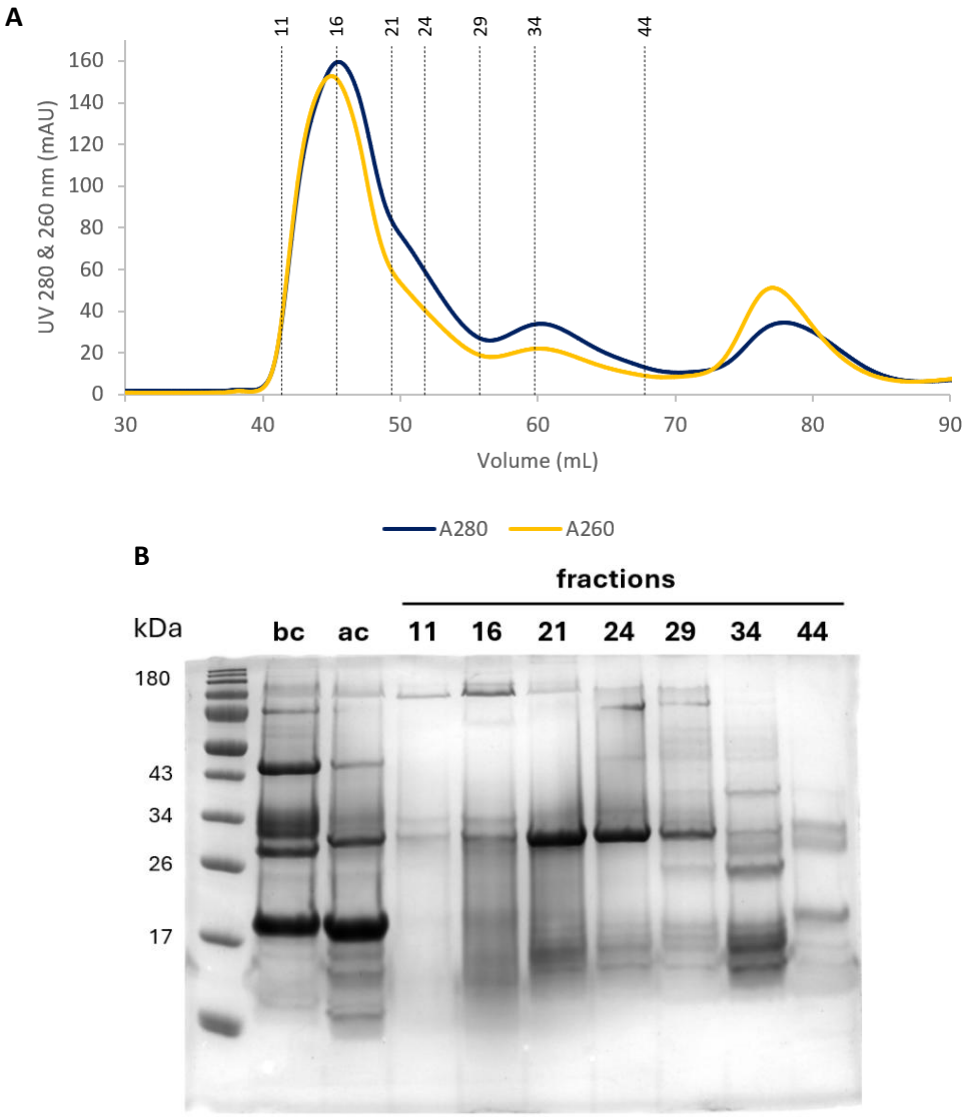


Figure 18. Isolation of LoTrmBL2b in a SEC column after elution from a HisTrap column. **A.** A280 and A260 spectra during SEC column elution. Dashed lines indicate the fractions analysed by gel electrophoresis. **B.** 15% polyacrylamide-SDS gels loaded with selected samples pre-purification and eluted fractions. bc: pooled fractions before dialysis and cleavage by Ulp1. ac: pooled fractions after dialysis and cleavage by Ulp1.

3.6.3. LoTrmBL2b binds DNA *in vitro*

Although we could not purify LoTrmBL2b as efficiently as LoAlba, we proceeded with the examination of its DNA-binding capacity *in vitro* by EMSA. In our initial setup, LoTrmBL2b bound the DNA probe at the lowest concentration tested (0.5 µM) and caused a complete shift at 2 µM. Further examination revealed that LoTrmBL2b can bind DNA at a concentration as low as 0.1 µM (**Figure 19**) When the competitor for non-specific DNA-binding proteins poly(dI·dC) was used, probe retardation was partially aborted, with only a fraction of DNA exhibiting a complete shift. Increasing the concentration of poly(dI·dC) also caused the accumulation of DNA at the lower end of the gel (**Figure 19**).

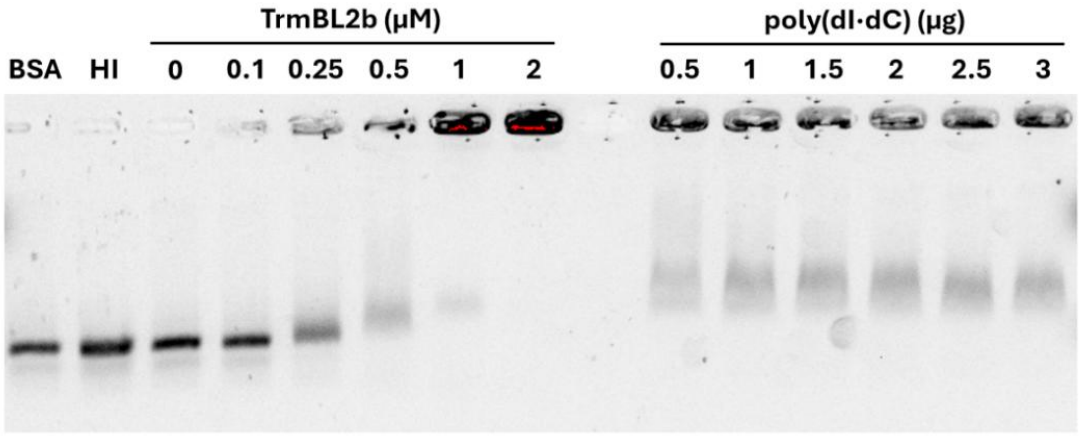
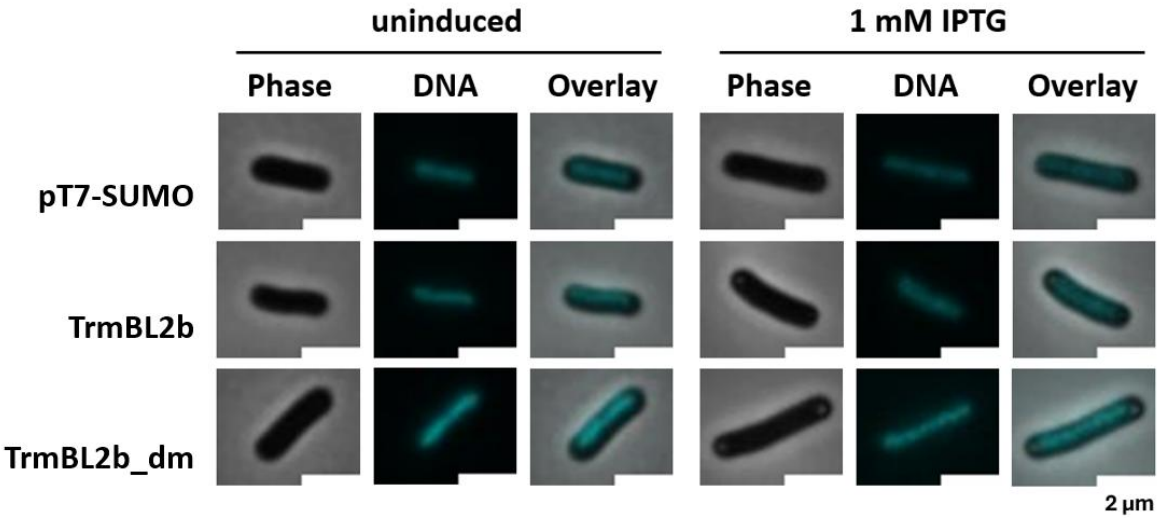


Figure 19. DNA EMSAs of LoTrmBL2b with or without the presence of poly(dI·dC). Images display the Cy5 fluorescence signal (λ = 665 nm). DNA concentration = 1.5 ng μL^{-1}. BSA: negative control reaction using BSA instead of TrmBL2b. HI: negative control with a heat-inactivated and denatured Alba. In both controls, as well as in the competitive assay with poly(dI·dC), protein concentration corresponds to the maximum concentration of LoTrmBL2b tested (2 µM).

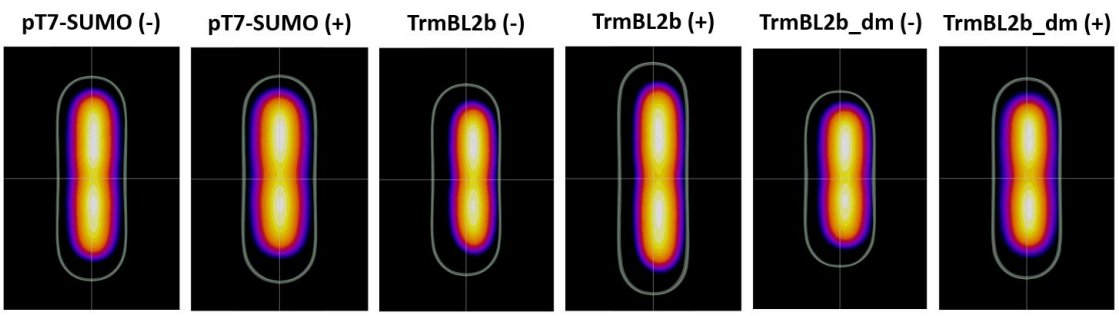
3.6.4. Expression of LoTrmBL2b does not lead to chromatin compaction in *E. coli* C41(DE3)

We induced LoTrmBL2b expression in *E. coli* C41(DE3) with IPTG to test for differences in the NC ratios between induced and uninduced strains under the microscope. The mutant variant TrmBL2b(p.Tyr56Asn;Glu57Gln), henceforth referred to as TrmBL2b_dm, was included in the analysis. No effect on DNA structure or NC ratio could be observed in any of the induced strains (**Figure 20A-D**). However, inclusion bodies were detected only in strains expressing either of the recombinant proteins. No significant differences in growth were observed between induced and uninduced strains (**Figure 20E**).

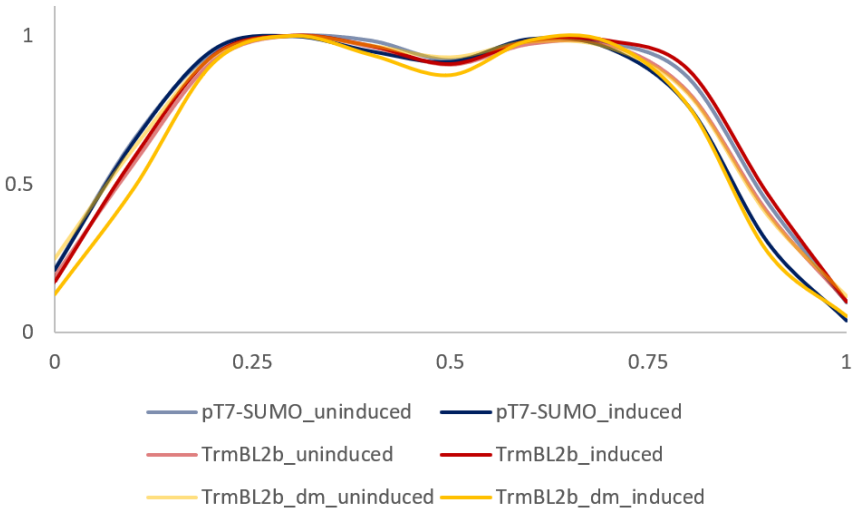
A


uninduced
1 mM IPTG
Phase
DNA
Overlay
Phase
DNA
Overlay
pT7-SUMO
TrmBL2b
TrmBL2b_dm
2 µm

B


pT7-SUMO (-)
pT7-SUMO (+)
TrmBL2b (-)
TrmBL2b (+)
TrmBL2b_dm (-)
TrmBL2b_dm (+)

C


1
0.5
0
0
0.25
0.5
0.75
1
pT7-SUMO_uninduced
pT7-SUMO_induced
TrmBL2b_uninduced
TrmBL2b_induced
TrmBL2b_dm_uninduced
TrmBL2b_dm_induced

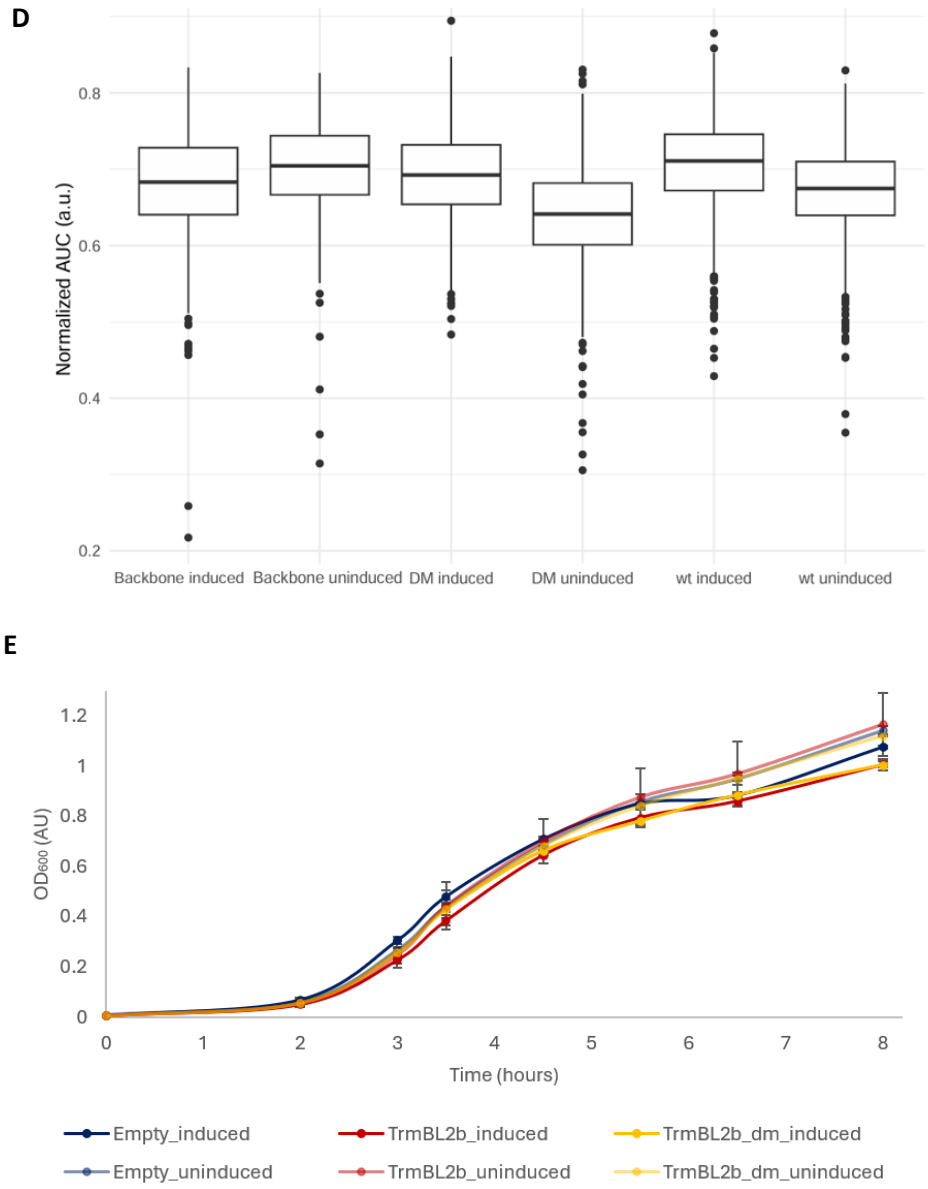


Figure 20. Analysis of the effects of the expression of recombinant LoTrmBL2b on DNA compaction and growth in *E. coli* C41(DE3). **A.** Representative images of all expression strains either IPTG-induced or uninduced. **B.** Representation of the mean fluorescence intensity of DAPI-stained DNA along the normalised cell shape, averaged across all cells analysed. Horizontal and vertical axes correspond to a total length of 2.2 and 4 µm. Brighter colours indicate a higher mean fluorescence intensity, while darker colours indicate a lower value. The signal depiction is scaled to fit within the normalised cell outline. Number of cells analysed per strain: pT7-SUMO (-): 545; pT7-SUMO (+): 555; TrmBL2b (-): 926; TrmBL2b (+): 839; TrmBL2b_dm (-): 693; TrmBL2b_dm (+): 1,249. **C.** Normalised mean fluorescence intensity along the normalised cell length, averaged across all cells analysed. X: normalised cell length. Y: normalised mean fluorescence intensity. **D.** Cell area covered by fluorescent signal per cell per strain and treatment, represented as the normalised area under the curve (AUC) in panel C. **F.** Growth curves of expression strains for the empty pT7 vector, TrmBL2b, and TrmBL2b_dm, with or without IPTG induction. Error bars indicate the standard deviation of the mean OD_{600} value.

4. Discussion

4.1. Identification and Characterisation of recombinant LoAlba

4.1.1. Expression and purification of recombinant LoAlba

LoAlba was recombinantly expressed in *E. coli* BL21(DE3) and purified by sequential affinity (IMAC) and size-exclusion (SEC) chromatography. Affinity purification via a polyhistidine tag is widely used for isolation of recombinant proteins from *E. coli* **(Kimple et al., 2013; Structural Genomics Consortium et al., 2008)**. We therefore fused a 6x-His tag to the N'-terminal end of the recombinant LoAlba to enable its purification, together with a SUMO tag to increase its solubility.

Despite the efficiency of this system, contamination by DNA-binding and/or histidine-rich proteins from *E. coli* have been reported to often co-elute with histidine-tagged target proteins and associated chaperones **(Structural Genomics Consortium et al., 2008; Bolanos-Garcia and Davies, 2006)**, thereby limiting their purification. Even after cleavage of the polyhistidine tag, Alba can bind the Ni-NTA resin, possibly due to the presence of surface-exposed histidine, cysteine, or methionine residues (see **Supplementary Material** and **Section 3.1.1.**) that may coordinate with the Ni^{2+} ions present in the resin. Therefore, omitting the rIMAC step in our protocol highly improved the overall purification yield. Nonetheless, the use of different tags (such as, e.g., streptavidin or biotin-based tags) might be useful when higher protein purity is necessary **(Kimple et al., 2013)**.

LoAlba consistently eluted as two separate species of different apparent molecular weight. This behaviour may be a consequence of the oligomerisation of LoAlba upon binding DNA. This is consistent with the observation that contamination by nucleic acids was detected in fractions containing LoAlba_heavy but not in fractions containing LoAlba_light. Accordingly, increasing benzonase concentration in the Lysis Buffer to enhance nucleolytic cleavage promoted the elution of LoAlba_light at the expense of LoAlba_heavy. Consistent with LoAlba_heavy being complexed with nucleic acids, its affinity for DNA molecules appeared lower than that of LoAlba_light in EMSAs. However, in mass photometry assays, nucleic acid-bound LoAlba_heavy appeared to be more prone to DNA-induced multimerization than DNA-free LoAlba_light. It will be necessary to ascertain whether differences in folding and/or post-translational modifications are responsible for the different behaviours observed in EMSAs and mass photometric assays. Examination with more sensitive techniques, such as, e.g., protein crystallography, NMR spectroscopy, and Size-Exclusion Chromatography with Multi-Angle Light Scattering (SEC-MALS),

would help to elucidate the structure and oligomerisation state of each LoAlba species **(Chetri et al., 2022; Some et al., 2019)** and explain both their separate existence and distinctive DNA-binding properties.

Finally, concerning Western Blot bands corresponding to multiples of the size of 6x-His-SUMO-LoAlba, we ascribe them to the formation of oligomers that could resist the denaturing conditions imposed by SDS-PAGE **(Jelinska et al., 2010)**.

4.1.2. Recombinant LoAlba binds both RNA and DNA *in vitro*

Despite being often defined as a NAPs, the Alba protein superfamily exhibits a diverse array of functions involving both RNA and/or DNA-binding activity. Our results show that recombinant LoAlba can bind both DNA and RNA *in vitro*, and *E. coli* DNA *in vivo*. In fact, LoAlba exhibits higher affinity for RNA than for DNA: in separate *in vitro* reactions, complete RNA shifting occurs at concentrations ~10 times lower than those required for DNA. Furthermore, the reduced DNA-binding affinity observed in the presence of RNA suggests that both nucleic acids likely compete for the same binding site (see **Section 3.1.1**and **Figure 15J**).

In addition, the gradual (rather than ladder-like) increase in DNA retardation observed in EMSA at increasing LoAlba concentrations suggests that retardation is due to the formation of oligomers or multimers of different length. This is consistent with the non-specific and cooperative DNA-binding model of Alba **(Molloy, 2000)**.

Despite its high homology to Alba1 of *S. solfataricus*, LoAlba displayed a remarkable RNA-binding activity. This forces us to broaden the definition of "NAP" and suggests an involvement of LoAlba in processes beyond chromatin organisation. While archaeal Alba proteins are considered key chromosome architectural proteins, the function of their eukaryotic homologues has been mostly related to the modulation of gene expression and RNA processing. However, the existence of RNA-exclusive and DNA/RNA-binding Alba proteins in *Sulfolobus* and *Saccharolobus* clearly opposes such divide and therefore renders the classification of LoAlba as either archaeal or eukaryotic-like meaningless. Although this work represents the first biochemical characterisation of a recombinant promethearchaeal Alba, the *in vivo* functions of LoAlba remain unclear. It is even possible that native LoAlba does not bind DNA in *Ca.* L. ossiferum, but performs functions exclusively related to RNA (e.g., folding, protection, localisation). Approaches targeting the role of native LoAlba, such as ChIP and cross-linking immunoprecipitation (CLIP-Seq), together with immunostaining with a specific anti-Alba antibody (ongoing), will be instrumental in uncovering the role of Alba in Loki.

Furthermore, the characterisation of LoAlba might be insufficient to represent the whole diversity of Alba proteins within Promethearchaeati, whose members may encode none or up to five different Alba-like proteins in their genome, potentially specialised in distinct functions.

4.1.3. Recombinant LoAlba might compact chromatin *in vivo*

The chromatin of *E. coli* expressing recombinant LoAlba was significantly more compacted than the chromatin of *E. coli* expressing other NAPs or cytoskeletal proteins. Therefore, compaction may not merely be ascribed to a general stress response triggered by the expression of a recombinant protein. Although LoAlba may act as a key structural protein in *Ca.* L. ossiferum chromatin and although we observed chromatin compaction in *E. coli* cells, it is unclear whether this was caused by LoAlba binding or whether it was an indirect effect. DNA bridging experiments, as described in **Shwab and Dame, 2025**, together with ChIP-Seq analysis of LoAlba-bound chromatin regions, will hopefully reveal whether DNA compaction is directly caused by LoAlba. Crucially, DNA compaction may impact the biophysical properties of the cytoplasm, as well as the accessibility of chromatin by transcription factors, transcription and translation coupling, or polysome stability **(Gray et al., 2019)**.

Therefore, chromatin compaction by LoAlba, if proven, could be a rudimentary mechanism of cellular compartmentalisation, as observed in bacteria with low NC ratio, such as *E. coli*[2]. In such a case, the emergence of a nuclear envelope, during eukaryogenesis, could represent the enhancement of a pre-existing chromatin-cytoplasm separation.

That being said, we observed chromatin compaction in *E. coli* expressing LoAlba independently of the presence of a lysine and phenylalanine residues at positions 31 and 74, respectively. These residues are presumed to be critical for the DNA-binding activity of LoAlba, as well as its capacity to oligomerise. It is thus possible that LoAlba does not compact DNA by directly binding to DNA, but rather indirectly by interacting with other DNA-binding proteins (such as, e.g., other NAPs, or proteins of the SMC complex) present in *E. coli*.

During my Master's thesis, I performed a pull-down assay to identify putative interaction partners of LoAlba (e.g., deacetylases of the Sir2 family, methylases, or demethylases **(Bell et al., 2002; Cao et al., 2018)**) but results were inconclusive (data not shown).

[2] While bacterial and eukaryotic cells with either high or low NC ratios have been reported **(Gray et al., 2019)**, knowledge on NC ratios across archaea is lacking. As for eukaryotes, the existence of the eukaryotic nuclear membrane leads to chromatin confinement in a specific cellular compartment, thereby physically limiting the maximum NC ratio values they can achieve **(Lemière et al., 2022)**.

4.2. Identification and Characterisation of LoTrmBL2b

4.2.1. Expression and purification of recombinant LoTrmBL2b

LoTrmBL2b was purified following the same strategy as for LoAlba. Expression in *E. coli* BL21 (DE3) cells was, however, suboptimal, and recombinant LoTrmBL2b was therefore purified from *E. coli* C41(DE3). In comparison to LoAlba, purification of LoTrmBL2b was challenging, both in terms of quantity and purity. The formation of inclusion bodies in the LoTrmBL2b expression strain suggests toxicity and likely contributed to its reduced yield.

Based on our Mass Spectrometry results, the DnaK was the major contaminant. This protein is a common contaminant in recombinant protein purification from *E. coli*, as DnaK helps protein folding and thereby commonly co-elutes with recombinant proteins. It does not, however, display any DNA-binding activity **(Christensen et al., 2022)**.

The purity of LoTrmBL2b after SEC was lower than that immediately after IMAC. Perhaps this is because a lot of LoTrmBL2b was lost during cleavage of the 6x-His-SUMO tag, concentration, and SEC. It is possible that the IMAC or SEC buffers used (adjusted to a pH of 8) could not have enough buffering capacity to maintain both 6x-His-SUMO-TrmBL2b (pI = 6.89) and untagged LoTrmBL2b (pI = 9.07) soluble throughout purification. Thus, we hypothesise that protein yield could increase if different pH values were used in the IMAC and SEC buffers, each being closer to the pI values of either 6x-His-SUMO-TrmBL2b or untagged LoTrmBL2b.

4.2.2. Recombinant LoTrmBL2b binds DNA *in vitro*

Despite its low purity, we still subjected TrmBL2b to EMSAs. Here, LoTrmBL2b displayed unspecific DNA-binding activity, consistently with the activity of TrmBL2 homologues in other archaeal species. Its affinity for DNA appeared ~20 times higher compared to that of LoAlba and is therefore a strong NAP candidate in Loki. Unlike LoAlba, LoTrmBL2b seems to bind DNA non-cooperatively, given that either a complete shift in DNA mobility, or no shift at all, is observed when LoTrmBL2b is added. It is likely that LoTrmBL2b binds DNA as a tetramer, following the DNA-binding pattern of TrmBL2 homologues in archaea and the detection of bands corresponding to the apparent MW of 6x-His-SUMO-LoTrmBL2b tetramers in Western Blot analysis against the 6x-His tag.

We found no evidence that recombinant LoTrmBL2b compacts or alters chromatin structure in *E. coli* cells. However, as reported in **Wierer et al., 2016**, archaeal TrmBL2 proteins cover the DNA double helix in a thick and rigid filament, resulting in a more extended DNA conformation than the random coil configuration established in the absence of NAPs. Compaction of DNA by

LoTrmBL2b is therefore unlikely and structural changes in chromatin may only be observable with techniques such as circular dichroism or atomic force microscopy **(Garbett et al., 2007; Beckwitt et al., 2018)**. Verifying DNA-binding by LoTrmBL2b *in vivo* by ChIP-Seq remains essential to continue exploring its function in *Ca.* L. ossiferum.

4.3. Conclusions and outlook

The discovery of archaea, particularly of Promethearchaea, represented a turning point in the study of eukaryogenesis, introducing compelling evidence of the archaeal origin of eukaryotes **(Spang et al., 2015; Rodrigues-Oliveira et al., 2023; Imachi et al., 2020)**. Therefore, Promethearchaeal chromatin, which contains a combination of NAPs and histone-like proteins, may represent an evolutionary intermediate between NAP-based compaction and regulation (as in bacteria and archaea) and nucleosome-based chromatin (as in eukaryotes). However, it is unclear whether Promethearchaeal chromatin is "more bacterial" or "more archaeal" or whether it already has the main hallmarks of eukaryotic chromatin (e.g., complete uncoupling of transcription and translation or chromatin remodeling to finely control gene expression). The characterisation of chromatin in Promethearchaeota such as *Ca.* L. ossiferum is thus imperative to understanding how eukaryotic chromatin evolved and, more specifically, to address the following questions:

1. What is the function of known non-histone chromatin proteins (i.e., NAPs) and can they be the precursors of eukaryotic chromatin remodelers or co-regulators?
2. Are there NAPs unique to Promethearchaeota which are the progenitors of eukaryotic chromatin remodelling complexes?
3. Do Promethearchaeal NAPs interact with promethearchaeal histones or SMC complexes and is there a functional division between NAP-based chromatin domains and histone-based chromatin domains?

Here, I focused on the first question and provided the first biochemical analysis of two NAPs, Alba and TrmbL2b, of *Ca*. Lokiarchaeum ossiferum, the only culturable Promethearchaeon in Europe. Based on our biochemical characterisation of LoAlba and LoTrmBL2b, we conclude that:

- Recombinant LoAlba binds both RNA and DNA *in vitro*, although it displays a higher affinity for RNA.
- Recombinant LoAlba binds *E. coli* DNA *in vivo* with no apparent correlation between GC content or genomic features like promoter regions or intergenic regions.

- Expression of recombinant LoAlba leads to chromatin compaction in *E. coli* without an ensuing stress response.
- Recombinant LoTrmBL2b binds DNA with high affinity *in vitro*.
- Expression of recombinant LoTrmBL2 does not lead to chromatin compaction in *E. coli*.

Further investigation of promethearchaeal NAPs is crucial to fully understand their role in chromatin compartmentalisation, remodelling, and gene expression, and how they interact with histones and SMC complexes. More generally, this work may help us to understand how the epigenetic fine-tuning of gene expression emerged, arguably the most distinctive feature of eukaryotic chromatin. In this direction, future work will include the following:

- The identification of LoAlba interaction partners. Relevant putative interactors include other chromosome architectural proteins (i.e., histones, proteins of the SMC complex, and other NAPs) and proteins suspected to mediate post-translational modification of LoAlba, such as acetylases and deacetylases or methylases and demethylases.
- The characterisation of *Ca.* L. ossiferum chromatin domains and, if existing, compartments by chromosome conformation capture techniques and the role of NAPs, SMC complex proteins, and histones in the creation of chromatin domains / compartments.
- The identification of novel chromatin-binding proteins, including architectural proteins, chromatin remodelers, or enzymes involved in post-translational modification by agnostic approaches such as ChIP-X (ongoing analysis in collaboration with the Sebé-Pedrós Lab – Centre for Genomic Regulation, Barcelona, Spain).

5. Supplementary Material

5.1. Establishment of *E. coli* Expression Strains and Probe Synthesis

5.1.1. Target *Ca.* L. ossiferum gene sequences and expected translation products

- ***Lo albA* (GNAMCCOK_01907):**

5’-atgggtgaagaacatagagatactcgacaacacgatcgagatagaatagcgcacagacgggaggataatgattgtagtgtgtatatagga**aag**
cgtccgaccatgaactacgttatggcggcgatggtaatcttaaacaagaacgaagattgtacaattaaagcacgcggcagagcaatttcccatgcagt
cgatgtggcagaaattctcatcaataaa**ttt**tatccgagtgctttttataaagacataaagatttcaacggaacacttacagaatgatgaccataaagtt
tctaatgtcagttcaatggaaattgttttaggaccacgcaaaccacaagaaccagaacaaaattag-3’

N’-MGEEHRDTRQHDRDRIAHRREDNDCSVYIG**K**RPTMNYVMAAMVILNKNEDCTIKARGRAISHAVDVAEILINK**F**Y
PSAFYKDIKISTEHLQNDDHKVSNVSSMEIVLGPRKPQEPEQN-C’

- ***Lo trmBL2a* (GNAMCCOK_03675):**

5’-atgagctcgaatataccgcaagaagtttatgcagcattgaaagaatttggcttaacggactatgaaacccgagcctttattgccttaattaccaatg
gcatatcctcggccaaagaattatcggatcggactaaaattccttatagtcgaatt**tacgat**gtattggtaaatttggaaaatattggctgggttaaagt
gatcagcgggcgtccgatgaaatataaagcagaacgccccgcatttgttgcaaaattagcaaagaaacaaattgaagctaaatactcccgtattgaa
accgctcttattgaaaaactcgaacctttgtatggacaagatgaacaagtcgaatccacacctatatggatcctgaacggtaatgttcaacaaaaagtt
attgatatggtaaaagctaccaataaaacccttacgatctttctgaaaaacccaaatgaagggcttcttgctgatatgtttgaacatttcttatatttaacc
gagaaaaaggtcgatattaatattttggttaatgagagtaatttaaacgtaacaaataaaaacatttggcgaaagattgcatcgatttcaaaggttcag
accgttaaaaatgtattattcgatgctttcttatttgataataatgaaatgttaattttcttaacaacctttttccgaatttctattaaggatgagaacatggt
ctttttcattcaagaaagccgtttgatcaattatacaaaaacttacttcaaaatgttatggtcaatgggtgagaagtttttgttaaaaaattattcttaa-3’

N’-MSSNIPQEVYAALKEFGLTDYETRAFIALITNGISSAKELSDRTKIPYSRI**YD**VLVNLENIGWVKVISGRPMKYKAERPA
FVAKLAKKQIEAKYSRIETALIEKLEPLYGQDEQVESTPIWILNGNVQQKVIDMVKATNKTLTIFLKNPNEGLLADMFEHF
LYLTEKKVDINILVNESNLNVTNKNIWRKIASISKVQTVKNVLFDAFLFDNNEMLIFLTTFFRISIKDENMVFFIQESRLINYT
KTYFKMLWSMGEKFLLKNYS-C’

- ***Lo trmBL2b* (GNAMCCOK_03810):**

5’-ttgaaattagggaaaaatattaaaattgattctgatgttcgtggtgccttgaaaaatttaggatttacagattattttaccaacatttacattgctctttt
agattctggagaaatgaacgcccacgagttaagtaaattaacaagtgttccttattcccgaatt**tatgaa**gtgttaaatgatatggttaagcgaaggat
aatcacaaaattagaaggacgcccaagtacatttattggaaatgatccttccgaagtttttggaaatataaaaaaaacccaagaaactgaatttcaac
aaaatatggattgttccctccccttttctaaaacaattgtttggggaaaaacatccagcaaaaaaagaacaattgacgatttatgaaggcaatagagcat
ctacggatcactttcggaatgttctgaatggaactaatcgatcaattaaagcgtttattaaaaacatgaatgaaatatttccaattattaaaatgaatctt
gattttcttcgtgctaaaggaattgaaacccatttaatcattgaagaacgatttcgagaacgatcatttatcagtattctgaagaaatatggcaccatcaa
attctatccaaatgtagaacaaggagtgctcatttctgatgataacgttggaattcagctgatcaagggtaattttaacattgcaaaaccaaaagaatta
gaatactccatctttagctccaattcaacttcttatgtcacgtttctaaccgaattattttcaagaatgtgggaaaaagctatagaatag-3’

N’-LKLGKNIKIDSDVRGALKNLGFTDYFTNIYIALLDSGEMNAHELSKLTSVPYSRI**YE**VLNDMVKRRIITKLEGRPSTFIGN
DPSEVFGNIKKTQETEFQQNMDCSLPFLKQLFGEKHPAKKEQLTIYEGNRASTDHFRNVLNGTNRSIKAFIKNMNEIFPII
KMNLDFLRAKGIETHLIIEERFRERSFISILKKYGTIKFYPNVEQGVLISDDNVGIQLIKGNFNIAKPKELEYSIFSSNSTSYVTF
LTELFSRMWEKAIE-C’

5.1.2. Plasmid vector sequences

- **6x-His-SUMO pT7 vector** (the 6x-His-SUMO tag sequence is highlighted)**:**

5’-ttgagatccttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactct
ttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccg
cctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggata
aggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctat
gagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagg
gggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaa
aacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattac
cgcctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtatttt
ctccttacgcatctgtgcggtatttcacaccgcaatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtatacactccgctatcgc
tacgtgactgggtcatggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctg
tgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagctgcggtaaagctcatcagcgtggtcgtgaagc
gattcacagatgtctgcctgttcatccgcgtccagctcgttgagtttctccagaagcgttaatgtctggcttctgataaagcgggccatgttaagggcggt
tttttcctgtttggtcactgatgcctccgtgtaagggggatttctgttcatgggggtaatgataccgatgaaacgagagaggatgctcacgatacgggtta
ctgatgatgaacatgcccggttactggaacgttgtgagggtaaacaactggcggtatggatgcggcgggaccagagaaaaatcactcagggtcaatg
ccagcgcttcgttaatacagatgtaggtgttccacagggtagccagcagcatcctgcgatgcagatccggaacataatggtgcagggcgctgacttcc
gcgtttccagactttacgaaacacggaaaccgaagaccattcatgttgttgctcaggtcgcagacgttttgcagcagcagtcgcttcacgttcgctcgcg
tatcggtgattcattctgctaaccagtaaggcaaccccgccagcctagccgggtcctcaacgacaggagcacgatcatgcgcacccgtggggccgcca
tgccggcgataatggcctgcttctcgccgaaacgtttggtggcgggaccagtgacgaaggcttgagcgagggcgtgcaagattccgaataccgcaag
cgacaggccgatcatcgtcgcgctccagcgaaagcggtcctcgccgaaaatgacccagagcgctgccggcacctgtcctacgagttgcatgataaag
aagacagtcataagtgcggcgacgatagtcatgccccgcgcccaccggaaggagctgactgggttgaaggctctcaagggcatcggtcgagatcccg
gtgcctaatgagtgagctaacttacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcggcca
acgcgcggggagaggcggtttgcgtattgggcgccagggtggtttttcttttcaccagtgagacgggcaacagctgattgcccttcaccgcctggccctg
agagagttgcagcaagcggtccacgctggtttgccccagcaggcgaaaatcctgtttgatggtggttaacggcgggatataacatgagctgtcttcggt
atcgtcgtatcccactaccgagatatccgcaccaacgcgcagcccggactcggtaatggcgcgcattgcgcccagcgccatctgatcgttggcaacca
gcatcgcagtgggaacgatgccctcattcagcatttgcatggtttgttgaaaaccggacatggcactccagtcgccttcccgttccgctatcggctgaatt
tgattgcgagtgagatatttatgccagccagccagacgcagacgcgccgagacagaacttaatgggcccgctaacagcgcgatttgctggtgacccaa
tgcgaccagatgctccacgcccagtcgcgtaccgtcttcatgggagaaaataatactgttgatgggtgtctggtcagagacatcaagaaataacgccg
gaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggatagttaatgatcagcccactgacgcgttgcgcgagaagattgtgc
accgccgctttacaggcttcgacgccgcttcgttctaccatcgacaccaccacgctggcacccagttgatcggcgcgagatttaatcgccgcgacaattt
gcgacggcgcgtgcagggccagactggaggtggcaacgccaatcagcaacgacttttgcccgccagttgttgtgccacgcggttgggaatgtaattca
gctccgccatcgccgcttccactttttcccgcgttttcgcagaaacgtggctggcctggttcaccacgcgggaaacggtctgataagagacaccggcata

ctctgcgacatcgtataacgttactggtttcacattcaccaccctgaattgactctcttccgggcgctatcatgccataccgcgaaaggttttgcgccattc
gatggtgtccgggatctcgacgctctcccttatgcgactcctgcattaggaagcagcccagtagtaggttgaggccgttgagcaccgccgccgcaagga
atggtgcatgcaaggagatggcgcccaacagtcccccggccacggggcctgccaccatacccacgccgaaacaagcgctcatgagcccgaagtggc
gagcccgatcttccccatcggtgatgtcggcgatataggcgccagcaaccgcacctgtggcgccggtgatgccggccacgatgcgtccggcgtagag
gatcgagatctcgatcccgcgaaattaatacgactcactataggggaattgtgagcggataacaattcccctctagaaataattttgtttaactttaaga
aggagatataccatgggt**catcaccatcatcatcacgggtcggactcagaagtcaatcaagaagctaagccagaggtcaagccagaagtcaagcc**
tgagactcacatcaatttaaaggtgtccgatggatcttcagagatcttcttcaagatcaaaaagaccactcctttaagaaggctgatggaagcgttcg
ctaaaagacagggtaaggaaatggactccttaagattcttgtacgacggtattagaattcaagctgatcaggcccctgaagatttggacatggagg
ataacgatattattgaggctcaccgcgaacagattggtggcagatccggctgctaacaaagcccgaaaggaagctgagttggctgctgccaccgctg
agcaataactagcataacccctttggggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccggattggcgaatgggacgcgcc
ctgtagcggcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccctt
cctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgaccccaaaaaa
cttgattagggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttcc
aaactggaacaacactcaaccctatctcggtctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaa
aatttaacgcgaattttaacaaaatattaacgcttacaatttaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacat
tcaaatatgtatccgctcatgaattaattcttagaaaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggattatcaataccatatttttg
aaaaagccgtttctgtaatgaaggagaaaactcaccgaggcagttccataggatggcaagatcctggtatcggtctgcgattccgactcgtccaacat
caatacaacctattaatttcccctcgtcaaaaataaggttatcaagtgagaaatcaccatgagtgacgactgaatccggtgagaatggcaaaagtttat
gcatttctttccagacttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatcaaccaaaccgttattcattcgtgattgcgcctgagcg
agacgaaatacgcgatcgctgttaaaaggacaattacaaacaggaatcgaatgcaaccggcgcaggaacactgccagcgcatcaacaatattttcac
ctgaatcaggatattcttctaatacctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtacggataaaatgcttgat
ggtcggaagaggcataaattccgtcagccagtttagtctgaccatctcatctgtaacatcattggcaacgctacctttgccatgtttcagaaacaactct
ggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagcccatttatacccatataaatcagcatccatgttgg
aatttaatcgcggcctagagcaagacgtttcccgttgaatatggctcataacacccttgtattactgtttatgtaagcagacagttttattgttcatgacc
aaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttc-3’

5.1.3. DNA and RNA sequences for EMSA

- **DNA Sequence** (the T7 promotor sequence required for RNA production is highlighted)**:**

5’-**taatacgactcactatagcg**taatcgcggcctagagcaagacgtttcccgttgaatatggctcataacacccttgtattactgtttatgtaagcaga
cagttttattgttcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttt
tctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaact
ggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctc
tgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcga
gctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgc
ttcccgaagggagaaaggcggacaggtatccggtaagcg-3’

- **RNA Sequence:**

5’-gcguaaucgcggccuagagcaagacguuucccguugaauauggcucauaacaccccuuguauuacuguuuauguaagcagacaguu
uuauuguucaugaccaaaaucccuuaacgugaguuuucguuccacugagcgucagaccccguagaaaagaucaaaggaucuucuugag
auccuuuuuuuucugcgcguaaucugcugcuugcaaacaaaaaaaccaccgcuaccagcgguggguuuguuugccggaucaagagcuacca
acucuuuuuccgaagguaacuggcuucagcagagcgcagauaccaaauacuguccuucuaguguagccguaguuaggccaccacuucaa
gaacucuguagcaccgccuacauaccucgcucugcuaauccuguuaccaguggcugcugccaguggcgauaagucgugucuuaccgggu
uggacucaagacgauaguuaccggauaaggcgcagcggucgagcugaacgggggguucgugcacacagcccagcuuggagcgaacgaccu
acaccgaacugagauaccuacagcgugagcuaugagaaagcgccacgcuucccgaagggagaaaggcggacagguauccgguaagcg-3’

5.1.4. Design of LoAlba, TrmBL2a, and TrmBL2b mutants

albA, *trmBL2a*, and *trmBL2b* sequences were targeted at specific codons to generate protein mutant versions. Mutated amino acids were selected based on the functionally critical residues of the corresponding homologues of LoAlba, LoTrmBL2a, and LoTrmBL2b in archaea. Targeted and introduced codons, as well as the corresponding amino acids and their proposed function, are highlighted in the previous sequences and listed in the table below.

Gene	Original codon	New codon	Mutation	Presumed function
albA	AAG (c.91_93)	GCG	Lys31Ala	PTM and DNA/RNA binding **(Wardleworth et al., 2002; Bell et al., 2002; Guo et al., 2014)**
	TTT (c.220_222)	GCG	Phe74Ala	Dimer-dimer interaction **(Laurens et al., 2012)**
trmBL2a	TAC (c.154_156)	AAC	Tyr52Asn	DNA binding **(Ahmad et al., 2015)**
	GAT (c.157_159)	AAC	Asp53Asn	DNA binding **(Ahmad et al., 2015)**
trmBL2b	TAT (c.166_168)	AAC	Tyr56Asn	DNA binding **(Ahmad et al., 2015)**
	GAA (c.169_171)	CAG	Glu57Gln	DNA binding **(Ahmad et al., 2015)**

5.1.5. PCR primers and programs

- **pT7-6x-His-SUMO vector:**

The plasmid was amplified from an *E. coli* colony stored in 25% glycerol containing the pT7-6x-His-SUMO vector transformed with SepF from *Methanopyrus kandleri* (personal transfer by Dr. Nika Pende) and diluted 1:50 in 50 µL Phusion™ High-Fidelity DNA Polymerase Reaction Buffer (1X Phusion HF Buffer, 200 µM dNTP mix, 0.5 µM forward primer, 0.5 µM reverse primer, 1 U Phusion DNA polymerase).

- Forward primer (F'): 5'-AGATCCGGCTGCTAACAAAGCC-3'
- Reverse primer (R'): 5'-GCCACCAATCTGTTCGCGGTG-3'

Initial Denaturing		98ºC	3 min
35 cycles	Denaturing	98ºC	15 s
	Annealing	62ºC	20 s
	Extension	72ºC	2 min 40 s
Final Extension		72ºC	5 min
Cooling		4ºC	∞

- ***Lo albA*:**

albA was amplified from genomic extracts of *Ca.* Lokiarchaeum ossiferum (personal transfer by Fatma Baraket, Univeristät Wien) and diluted 1:50 in 50 µL Phusion™ High-Fidelity DNA Polymerase Reaction Buffer.

- F': 5'-CCGCGAACAGATTGGTGGCATGGGTGAAGAACATAGAGA-3'
- R': 5'-CTTTGTTAGCAGCCGGATCTCTAATTTTGTTCTGGTTCTT-3'

Initial Denaturing		98ºC	3 min
35 cycles	Denaturing	98ºC	15 s
	Annealing	60ºC	20 s
	Extension	72ºC	20 s
Final Extension		72ºC	5 min
Cooling		4ºC	∞

- ***Lo trmBL2a*:**

trmBL2a was amplified from genomic extracts of *Ca.* Lokiarchaeum ossiferum and diluted 1:50 in 50 µL Phusion™ High-Fidelity DNA Polymerase Reaction Buffer.

- F': 5'-ACCGCGAACAGATTGGTGGCATGAGCTCGAATATACCGC-3'
- R': 5'-CTTTGTTAGCAGCCGGATCTTTAAGAATAATTTTTTAACA-3'

Initial Denaturing		98ºC	3 min
35 cycles	Denaturing	98ºC	15 s

	Annealing	64ºC	20 s
	Extension	72ºC	30 s
Final Extension		72ºC	5 min
Cooling		4ºC	∞

- ***Lo trmBL2b*:**

trmBL2b was amplified from genomic extracts of *Ca.* Lokiarchaeum ossiferum and diluted 1:50 in 50 µL GoTaq® G2 Flexi DNA Polymerase Reaction Buffer (1X GoTaq® Flexi Buffer, 2mM $MgCl_2$, 200 µM dNTP mix, 0.1 µM F', 0.1 µM R', 1.25 U GoTaq® G2 Flexi DNA Polymerase).

- F': 5'-ACCGCGAACAGATTGGTGGCATGAGCTCGAATATACCGC-3'
- R': 5'-CTTTGTTAGCAGCCGGATCTTTAAGAATAATTTTTTAACA-3'

Initial Denaturing		95ºC	2 min
35 cycles	Denaturing	95ºC	30 s
	Annealing	46ºC	30 s
	Extension	72ºC	1 min
Final Extension		72ºC	5 min
Cooling		4ºC	∞

- **Sequencing reactions of pT7-6x-His-SUMO-LoAlba, LoTrmBL2a, and LoTrmBL2b:**

To confirm the proper integration of *Lo albA*, *Lo trmBL2a*, and *Lo trmBL2b* into the pT7-6x-His-SUMO vector, DNA containing each of the inserted genes were amplified from a transformed *E. coli* Top10 colony diluted in 25 µL OneTaq® DNA Polymerase Reaction Buffer (1X OneTaq Standard Reaction Buffer, 0.2 µM F', 0.2 µM R'). Amplified PCR fragments were then purified and prepared for San-ger sequencing (Eurofins Genomics GmbH).

- F': 5'-CCCGCGAAATTAATACGACTCAC-3'
- R': 5'-CCTCAAGACCCGTTTAGAGGCC-3'

Initial Denaturing		94ºC	3 min
30 cycles	Denaturing	94ºC	30 s
	Annealing	53ºC	30 s
	Extension	68ºC	1 min (Alba)
			1 min 30 s (TrmBL2a/b)

Final Extension	68ºC	10 min
Cooling	4ºC	∞

- ***Lo albA* mutants:**

All pT7-6x-His-SUMO-LoAlba single mutant variants were synthesised from 1 ng pure pT7-6x-His-SUMO-LoAlba in 50 µL Q5® High-Fidelity DNA Polymerase Buffer (1X Q5® Reaction Buffer, 200 µM dNTPs, 0.5 µM F', 0.5 µM R', 1 U Q5® High-Fidelity DNA Polymerase). The pT7-6x-His-SUMO-LoAlba(K31A) plasmid was used as template for the synthesis of the double mutant variant, ins-tead. The following primers were used:

- *albA*(c.91_93delinsGCG):
    - F': 5'-GTATATAGGAGCGCGTCCGACCATG-3'
    - R': 5'-ACACTACAATCATTATCCTC-3'

- *albA*(c.220_222delinsGCG and double mutant):
    - F': 5'-CATCAATAAAGCGTATCCGAGTGCTTTTTATAAAG-3'
    - R': 5'- AGAATTTCTGCCACATC-3'

Initial Denaturing		98ºC	30 s
25 cycles	Denaturing	98ºC	10 s
	Annealing	52ºC	20 s
	Extension	72ºC	2 min 30 s
Final Extension		72ºC	2 min
Cooling		4ºC	∞

- ***Lo trmBL2a*(c.157_159delinsAAC):**

The pT7-6x-His-SUMO-TrmBL2a(D53N) single mutant variant was synthesised from 1 ng pure pT7-6x-His-SUMO-TrmBL2a in 50 µL Q5® High-Fidelity DNA Polymerase Buffer.

- F': 5'-TCGAATTTACAACGTATTGGTAAATTTGG-3'
- R': 5'-CTATAAGGAATTTTAGTCCGATCC-3'

Initial Denaturing		98ºC	30 s
25 cycles	Denaturing	98ºC	10 s
	Annealing	50ºC	20 s

	Extension	72ºC	3 min 20 s
Final Extension		72ºC	5 min
Cooling		4ºC	∞

- ***Lo trmBL2a*(c.154_156delinsAAC and double mutant):**

The pT7-6x-His-SUMO-TrmBL2a(Y52N) single mutant variant was synthesised from 1 ng pure pT7-6x-His-SUMO-TrmBL2a in 50 µL Phusion™ High-Fidelity DNA Polymerase Reaction Buffer. The pT7-6x-His-SUMO-LoAlba(D53N) plasmid was used as template for the synthesis of the double mutant variant, instead. The following primers were used:

- *trmBL2a*(c.154_156delinsAAC):
  - F': 5'-TAGTCGAATTAACGATGTATTGGTAAATTTGG-3'
  - R': 5'-TAAGGAATTTTAGTCCGATCC-3'

- *trmBL2a* (double mutant):
  - F': 5'-TAGTCGAATTAACAACGTATTGGTAAATTTGG-3'
  - R': 5'-TAAGGAATTTTAGTCCGATCC-3'

Initial Denaturing		98ºC	3 min
25 cycles	Denaturing	98ºC	15 s
	Annealing	54ºC	20 s
	Extension	72ºC	3 min 20 s
Final Extension		72ºC	10 min
Cooling		4ºC	∞

- ***Lo trmBL2b* mutants:**

All pT7-6x-His-SUMO-TrmBL2b single mutant variants were synthesised from 1 ng pure pT7-6x-His-SUMO-TrmBL2b in 50 µL Phusion™ High-Fidelity DNA Polymerase Reaction Buffer. The pT7-6x-His-SUMO-TrmBL2b(E57Q) plasmid was used as template for the synthesis of the double mutant variant, instead. The following primers were used:

- *trmBL2b*(c.166_168delinsAAC):
  - F': 5'-TTCCCGAATTAACGAAGTGTTAAATG-3'
  - R': 5'-TAAGGAACACTTGTTAATTTAC-3'

- *trmBL2b*(c.169_171delinsCAG):
    - F': 5'-CCGAATTTATCAGGTGTTAAATGATATGG-3'
    - R': 5'-GAATAAGGAACACTTGTTAATTTAC-3'

- *trmBL2b* (double mutant):
    - F': 5'-TTCCCGAATTAACCAGGTGTTAAATG-3'
    - R': 5'-TAAGGAACACTTGTTAATTTAC-3'

Initial Denaturing		98ºC	3 min
25 cycles	Denaturing	98ºC	15 s
	Annealing	54ºC	20 s
	Extension	72ºC	3 min 20 s
Final Extension		72ºC	10 min
Cooling		4ºC	∞

- **EMSA DNA fragments:**

The DNA fragments for in EMSAs and for *in vitro* transcription were synthesised from 40 ng pure pT7-6x-His-SUMO-LoAlba (K31A) in 25 µL OneTaq® DNA Polymerase Reaction Buffer. The following primers were used:

- DNA fragment for EMSA:
    - F': 5'-Cy5-TAATCGCGGCCTAGAGCAAG-3'
    - R': 5'-Cy5-CGCTTACCGGATACCTGTCC-3'
- DNA fragment for *in vitro* transcription:
    - F': 5'-TAATACGACTCACTATAGCGTAATCGCGGCCTAGAGCAAG-3'
    - R': 5'-CGCTTACCGGATACCTGTCC-3'

Initial Denaturing		94ºC	3 min
30 cycles	Denaturing	94ºC	30 s
	Annealing	58ºC	30 s
	Extension	68ºC	40 s
Final Extension		68ºC	40 s
Cooling		4ºC	∞

5.3. Mass Spectrometry Results

5.3.1. Mass Spectrometric analysis of recombinant LoAlba (1st purification round)

Protein	Sequence coverage (%)	rel. iBAQ (%)	Other relevant hits
6x-His-SUMO-LoAlba	67.4	65.62	Ribosomal proteins (<6%)
Alba_heavy	93.2	50.73	Termination factor Rho (<3%)
Alba_light	89.9	69.57	Peptidylpropyl isomerase SylD (<7%)

Table S1. Mass Spectrometric analysis of the first purification round of LoAlba. Sequence coverage: the percentage of the protein sequence that was identified by matching overlapping peptide sequences. rel. iBAQ: the relative Intensity-based Absolute Quantification without contaminants, corresponding to all peptide ion intensities corresponding to the target protein normalised by the number of theoretically observable peptides of that protein, omitting common contaminants such as human keratin. Other relevant hits: the most abundant contaminating proteins identified in each fraction.

These values were considered appropriate for antibody production and the characterisation of its DNA-binding properties by EMSA.

5.3.2. Mass Spectrometric analysis of recombinant LoTrmBL2b (2nd purification round)

Fraction nº	Sequence coverage (%)	rel. iBAQ (%)	Other relevant hits
33-70 (IMAC)	80.1	56.8	30S ribosomal protein S15 (6%)
16 (SEC)	74	16.3	Murein lipoprotein Lpp (9.0%)
21 (SEC)	81.4	22.43	Molecular chaperone DnaK (15.0%)
35 (SEC)	75.7	22.27	Peptidylpropyl isomerase (31.8%)

Table S2. Mass Spectrometric analysis of recombinant LoTrmBL2b (fractions 21-29).

Although the purity of recombinant LoTrmBL2b in fraction 21 was low, we still used the pooled fractions 21-29 in EMSAs.

6. References

Ahmad MU, Waege I, Hausner W et al. Structural Insights into Nonspecific Binding of DNA by TrmBL2, an Archaeal Chromatin Protein. *J Mol Biol* 2015; 427(20): 3216-29.

Aravind L, Iyer LM, and Anantharaman V. The two faces of Alba: the evolutionary connection between proteins participating in chromatin structure and RNA metabolism. *Genome Biol* 2003; 4: R64.

Ayala FJ. Darwin and the scientific method. *PNAS USA* 2009; 106: 10033-9.

Baker BJ, De Anda V, Seitz KW et al. Diversity, ecology and evolution of Archaea. *Nat Microbiol* 2020; 5: 887-900.

Baker BJ, De Anda V, Seitz KW et al. Diversity, ecology and evolution of Archaea. *Nat Microbiol* 2020; 5: 887-900.

Barillà D. Driving Apart and Segregating Genomes in Archaea. *Trends Microbiol* 2016; 24(12): 957-67.

Bastiaanssen C, Bobadilla Ugarte P, Kim K et al. RNA-guided RNA silencing by an Asgard archaeal Argonaute. *Nat Commun* 2024; 15: 5499 (2024).

Beckwitt EC, Kong M, and Van Houten B. Studying protein-DNA interactions using atomic force microscopy. *Semin Cell Dev Biol* 2018; 73: 220-30.

Bell SD, Botting CH, Wardleworth BN et al. The Interaction of Alba, a Conserved Archaeal Chromatin Protein, with Sir2 and Its Regulation by Acetylation. *Science* 2002; 296: 148-51.

Bolanos-Garcia VM and Davies OR. Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. *Biochim Biophys Acta* 2006; 1760(9): 1304-13.

Brunk CF and Martin WF. Archaeal Histone Contributions to the Origin of Eukaryotes. *Trends Microbiol* 2019; 27(8): 703-14.

Bulgheresi S. 15 DNA facts. In: Fisher M, editor. Brave Genomes – Microbial Genome Plasticity in the Face of Environmental Challenge. Academic Press; 2025. p. 19-20.

Bulgheresi S. DNA very fun facts. In: Fisher M, editor. Brave Genomes – Microbial Genome Plasticity in the Face of Environmental Challenge. Academic Press; 2025. p. 181.

Bulgheresi S. On Top of the DNA – Nucleoid-associated Proteins. In: Fisher M, editor. Brave Genomes – Microbial Genome Plasticity in the Face of Environmental Challenge. Academic Press; 2025. p. 103-129.

Cao J, Wang Q, Liu T et al. Insights into the post-translational modifications of archaeal Sis10b (Alba): lysine-16 is methylated, not acetylated, and this does not regulate transcription or growth. *Mol Microbiol* 2018; 109: 192-208.

Carone, BR. MNase-seq to Identify Genome-Wide DNA-Protein Interactions. In: Carabetta VJ and Akintunde O, editors. High Throughput Gene Screening. Methods in Molecular Biology. New York, New Yorl: Humana Press; 2025. vol. 2866.

Chan CW, Kiesel BR, and Mondragón A. Crystal Structure of Human Rpp20/Rpp25 Reveals Quaternary Level Adaptation of the Alba Scaffold as Structural Basis for Single-stranded RNA Binding. *J Mol Biol* 2018; 430(10): 1403-16.

Chêne A, Vembar SS, Rivière L et al. PfAlbas constitute a new eukaryotic DNA/RNA-binding protein family in malaria parasites. *Nucleic Acids Res* 2012; 40(7): 3066-77.

Chereji RV, Ocampo J, Clark DJ. MNase-Sensitive Complexes in Yeast: Nucleosomes and Non-histone Barriers. *Mol Cell* 2017; 65(3): 565-77.e3.

Chetri PB, Khan H, and Tripathi T. Chapter 5 - Methods to determine the oligomeric structure of proteins. In: Tripathi T and Dubey VK, editors. Advances in Protein Molecular and Structural Biology Methods. Academic Press; 2022. p. 49-76.

Christensen S, Rämisch S, and André I. DnaK response to expression of protein mutants is dependent on translation rate and stability. *Commun Biol* 2022; 5: 597.

Cox CJ, Foster PG, Hirt RP et al. The archaebacterial origin of eukaryotes. *Proc Natl Acad Sci USA* 2008; 105(51): 20356-61.

De Lise F, Iacono R, Moracci M et al. Archaea as a Model System for Molecular Biology and Biotechnology. *Biomolecules* 2023; 13(1): 114.

Decker KB and Hinton DM. Transcription regulation at the core: similarities among bacterial, archaeal, and eukaryotic RNA polymerases. *Annu Rev Microbiol* 2013; 67(1): 113-39.

Dobzhansky T. Nothing in Biology Makes Sense except in the Light of Evolution. *Am Biol Teach* 1973; 35(3): 125-9.

Doolittle WF. Phylogenetic Classification and the Universal Tree. *Science* 1999; 284: 2124-8.

Ducret A, Quardokus E, and Brun Y. MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis. *Nat Microbiol* 2016; 1: 16077.

Dumon-Seignovert L, Cariot G, and Vuillard L. The toxicity of recombinant proteins in Escherichia coli: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expr Purif* 2004; 37(1):203-6.

Efremov AK, Qu Y, Maruyama H et al. Transcriptional Repressor TrmBL2 from Thermococcus kodakarensis Forms Filamentous Nucleoprotein Structures and Competes with Histones for DNA Binding in a Salt- and DNA Supercoiling-dependent Manner. *J Biol Chem* 2015 19; 290(25): 15770-84.

Eme L, Tamarit D, Caceres EF et al. Inference and reconstruction of the heimdallarchaeial ancestry of eukaryotes. *Nature* 2023; 618: 992-9.

Fåhraeus, R. Has translation in the nucleus found its purpose? *Nat Rev Mol Cell Biol* 2024; 25: 1-2.

Farrants AK. DNA Accessibility by MNase Digestions. In: Visa N and Jordán-Pla A, editors. Chromatin Immunoprecipitation. Methods in Molecular Biology. New York, New York: Humana Press; 2018. vol. 1689.

Feijoo-Siota L, Rama JL, Sánchez-Pérez A et al. Considerations on bacterial nucleoids. *Appl Microbiol Biotechnol* 2017; 101: 5591-5602.

Foster PG, Cox CJ, and Embley TM. The primary divisions of life: a phylogenomic approach employing composition-heterogeneous methods. *Philos Trans R Soc Lond B Biol Sci* 2009; 364(1527): 2197-207.

Froger A, Hall JE. Transformation of plasmid DNA into E. coli using the heat shock method. *J Vis Exp* 2007; 6: 253.

Garbett N, Ragazzon P, and Chaires J. Circular dichroism to determine binding mode and affinity of ligand–DNA interactions. *Nat Protoc* 2007; 2: 3166-72.

Gasteiger E, Hoogland C, Gattiker A et al. Protein Identification and Analysis Tools on the Expasy Server. In: Walker JM, editor. The Proteomics Protocols Handbook. Humana Press; 2005. p. 571-607.

Gissot M, Walker R, Delhaye S, et al. Toxoplasma gondii Alba proteins are involved in translational control of gene expression. *J Mol Biol* 2013; 425(8): 1287-301.

Gould SB, Garg SG, and Martin WF. Bacterial Vesicle Secretion and the Evolutionary Origin of the Eukaryotic Endomembrane System. *Trends Microbiol* 2016; 24(7): 525-34.

Goyal M, Banerjee C, Nag S et al. The Alba protein family: Structure and function. *Biochim Biophys Acta* 2016; 1864(5): 570-83.

Gray WT, Govers SK, Xiang Y et al. Nucleoid Size Scaling and Intracellular Organization of Translation across Bacteria. *Cell* 2019; 177(6): 1632-48.e20.

Gross DS, Chowdhary S, Anandhakumar J et al. Chromatin. *Current Biology* 2015; 25(24): R1158-63.

Guo L, Ding J, Guo R et al. Biochemical and Structural Insights into RNA Binding by Ssh10b, a Member of the Highly Conserved Sac10b Protein Family in Archaea. *J Biol Chem* 2014; 289(3): 1478-90.

Guo R, Xue H, and Huang L. Ssh10b, a conserved thermophilic archaeal protein, binds RNA in vivo. *Mol Microbiol* 2003; 50(5): 1605-15.

Guy L and Ettema TJ. The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol* 2011; 19(12): 580-7.

Guy L, Saw JH, and Ettema TJ. The archaeal legacy of eukaryotes: a phylogenomic perspective. *Cold Spring Harb Perspect Biol.* 2014; 6(10): a016022.

Heinicke I, Müller J, Pittelkow M et al. Mutational analysis of genes encoding chromatin proteins in the archaeon Methanococcus voltae indicates their involvement in the regulation of gene expression. *Mol Genet Genomics* 2004; 272: 76-87.

Hildebrand EM and Dekker J. Mechanisms and Functions of Chromosome Compartmentalization. *Trends Biochem Sci* 2020; 45(5): 385-96.

Hołówka J and Zakrzewska-Czerwińska J. Nucleoid Associated Proteins: The Small Organizers That Help to Cope with Stress. *Front Microbiol* 2020; 11: 590.

Iborra FJ, Jackson DA, and Cook PR. Coupled transcription and translation within nuclei of mammalian cells. *Science* 2001; 293(5532): 1139-42.

Imachi H, Nobu MK, Ishii S et al. Eukaryotes' closest relatives are internally simple syntrophic archaea. *bioRxiv* 2025; 640444.

Imachi H, Nobu MK, Kato S et al. Promethearchaeum syntrophicum gen. nov., sp. nov., an anaerobic, obligately syntrophic archaeon, the first isolate of the lineage 'Asgard' archaea, and proposal of the new archaeal phylum Promethearchaeota phyl. nov. and kingdom Promethearchaeati regn. nov. *Int J Syst Evol Microbiol* 2024; 74(7): 006435.

Imachi H, Nobu MK, Nakahara N et al. Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature* 2020; 577: 519-25.

Irastortza-Olaziregi M and Amster-Choder O. Coupled Transcription-Translation in Prokaryotes: An Old Couple with New Surprises. *Front Microbiol* 2021; 11: 624830.

Jagadeesh J and Vembar SS. Evolution of sequence, structural and functional diversity of the ubiquitous DNA/RNA-binding Alba domain. *Sci Rep* 2024; 14, 30363.

Jelinska C, Conroy MJ, Craven CJ et al. Obligate heterodimerization of the archaeal Alba2 protein with Alba1 provides a mechanism for control of DNA packaging. *Structure* 2005; 13(7): 963-71.

Jelinska C, Petrovic-Stojanovska B, Ingledew WJ et al. Dimer-dimer stacking interactions are important for nucleic acid binding by the archaeal chromatin protein Alba. *Biochem J* 2010; 427(1): 49-55.

Johnson GE, Lalanne JB, Peters ML et al. Functionally uncoupled transcription–translation in Bacillus subtilis. *Nature* 2020; 585: 124-8.

Kim M, Park S, and Lee SJ. Global transcriptional regulator TrmB family members in prokaryotes. *J Microbiol* 2016; 54(10): 639-45.

Kimple ME, Brill AL, Pasker RL. Overview of affinity tags for protein purification. *Curr Protoc Protein Sci* 2013;73: 9.9.1-9.9.23.

Knopp M, Stockhorst S, van der Giezen M et al. The Asgard Archaeal-Unique Contribution to Protein Families of the Eukaryotic Common Ancestor Was 0.3. *Genome Biol Evol* 2021; 13(6): evab085.

Koonin EV. The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome Biol* 2010; 11: 209.

Laurens N, Driessen R, Heller I et al. Alba shapes the archaeal genome using a delicate balance of bridging and stiffening the DNA. *Nat Commun* 2012; 3: 1328.

Laursen SP, Bowerman S, and Luger K. Archaea: The Final Frontier of Chromatin. *J Mol Biol* 2021; 433(6): 166791.

Lee SJ, Surma M, Hausner W et al. The role of TrmB and TrmB-like transcriptional regulators for sugar transport and metabolism in the hyperthermophilic archaeon Pyrococcus furiosus. *Arch Microbiol* 2008; 190(3):247-56.

Leigh JA, Albers SV, Atomi H et al. Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfolobales. *FEMS Microbiol Rev* 2011; 35(4): 577-608.

Lemière J, Real-Calderon P, Holt LJ et al. Control of nuclear size by osmotic forces in Schizosaccharomyces pombe. *eLife* 2022; 11: e76075.

Leonard AC and Méchali M. DNA replication origins. *Cold Spring Harb Perspect Biol* 2013; 5(10): a010116.

Liu Y, Makarova KS, Huang WC et al. Expanded diversity of Asgard archaea and their relationships with eukaryotes. *Nature* 2021; 593: 553-7.

López-García P and Moreira D. Open Questions on the Origin of Eukaryotes. *Trends Ecol Evol* 2015; 30(11): 697-708.

Luijsterburg MS, White MF, van Driel R et al. The Major Architects of Chromatin: Architectural Proteins in Bacteria, Archaea and Eukaryotes. *Crit Rev Biochem Mol Biol* 2008; 43(6): 393-418.

Manning M and Colón W. Structural Basis of Protein Kinetic Stability: Resistance to Sodium Dodecyl Sulfate Suggests a Central Role for Rigidity and a Bias Toward β-Sheet Structure. *Biochem* 43(35): 11248-54.

Marsh VL, Peak-Chew SY, and Bell SD. Sir2 and the acetyltransferase, Pat, regulate the archaeal chromatin protein, Alba. *J Biol Chem* 2005; 280(22): 21122-8.

Martin WF, Garg S, and Zimorski V. Endosymbiotic theories for eukaryote origin. *Philos Trans R Soc Lond B Biol Sci* 2015; 370(1678): 20140330.

Maruyama H, Prieto EI, Nambu T et al. Different Proteins Mediate Stepwise Chromosome Architectures in Thermoplasma acidophilum and Pyrobaculum calidifontis. *Front Microbiol* 2020; 11: 1247.

Maruyama H, Shin M, Oda T et al. Histone and TK0471/TrmBL2 form a novel heterogeneous genome architecture in the hyperthermophilic archaeon Thermococcus kodakarensis. *Mol Biol Cell* 2011; 22(3): 386-98.

Mattick JS, Malik SB, and Delwiche CF. Group II Introns in Archaeal Genomes and the Evolutionary Origin of Eukaryotic Spliceosomal Introns. *bioRxiv* 2024: 627823.

Molloy PL. Electrophoretic Mobility Shift Assays. In: Tymms MJ, editor. Transcription Factor Protocols. Methods in Molecular Biology. Humana Press; 2000. vol. 130.

Moody, ER, Álvarez-Carretero S, Mahendrarajah TA et al. The nature of the last universal common ancestor and its impact on the early Earth system. *Nat Ecol and Evol* 2024; 8: 1654-66.

Nachmias D, Melnikov N, Zorea A et al. Asgard ESCRT-III and VPS4 reveal conserved chromatin binding properties of the ESCRT machinery. *ISME J* 2023; 17(1): 117-29.

Náprstková A, Malínská K, Záveská Drábková L et al. Characterization of ALBA Family Expression and Localization in Arabidopsis thaliana Generative Organs. *Int J Mol Sci* 2021; 22(4): 1652.

Novikova O and Belfort M. Mobile Group II Introns as Ancestral Eukaryotic Elements. *Trends Genet* 2017; 33(11): 773-83.

O'Leary NA, Cox E, Holmes JB et al. Exploring and retrieving sequence and metadata for species across the tree of life with NCBI Datasets. *Sci Data* 2024; 11: 732.

Peak Proteins. Protein concentration determination by UV absorbance at 280 nm [Internet]. Macclesfield (UK): Peak Proteins; 2022 [cited 2025 Jul 10]. Available from: https://peakproteins.com/wp-content/uploads/2022/06/PPP_Protein-Conc-by-UV.pdf

Pende N, Sogues A, Megrian D et al. SepF is the FtsZ anchor in archaea, with features of an ancestral cell division system. *Nat Commun* 2021; 12: 3214.

Ranawat HM, Cajili MK, Quail T et al. Cryo-EM reveals open and closed Asgard chromatin assemblies. *bioRxiv* 2025; 653377.

Richards TA, Eme L, Archibald JM et al. Reconstructing the last common ancestor of all eukaryotes. *PLoS Biol* 2024; 22(11): e3002917.

Rodrigues-Oliveira T, Wollweber F, Ponce-Toledo RI et al. Actin cytoskeleton and complex cell architecture in an Asgard archaeon. *Nature* 2023; 613: 332-9.

Ruse M. The History of Evolutionary Thought. In: Ruse M and Travis J, editors. Evolution – The First Four Billion Years. Cambridge, Massachusetts: The Belknap Press of Harvard University Press; 2009. p. 1-48.

Sagulenko E, Morgan GP, Webb RI et al. Structural studies of planctomycete Gemmata obscuriglobus support cell compartmentalisation in a bacterium. *PLoS One* 2014; 9(3): e91344.

Santana-Molina C, Williams TA, Snel B et al. Chimeric origins and dynamic evolution of central carbon metabolism in eukaryotes. *Nat Ecol Evol* 2025; 9: 613-27.

Sarmiento F, Long F, Cann I et al. Diversity of the DNA replication system in the Archaea domain. *Archaea* 2014; 2014: 675946.

Schindelin J, Arganda-Carreras I, Frise E et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods* 2012; 9: 676–82.

Schwab S and Dame RT. Identification, characterization and classification of prokaryotic nucleoid-associated proteins. *Mol Microbiol* 2025; 123(3): 206-217.

Some D, Amartely H, Tsadok A et al. Characterization of Proteins by Size-Exclusion Chromatography Coupled to Multi-Angle Light Scattering (SEC-MALS). *J Vis Exp* 2019; 148: e59615.

Spang A, Saw J, Jørgensen S et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 2015; 521: 173-9.

Spang A, Stairs CW, Dombrowski N et al. Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. *Nat Microbiol* 2019; 4: 1138-48.

Structural Genomics Consortium, Architecture et Fonction des Macromolécules Biologiques, Berkeley Structural Genomics Center et al. Protein production and purification. *Nat Methods* 2008; 5: 135-46.

Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 2005; 41(1): 207-34.

Szalay MF, Majchrzycka B, Jerković I et al. Evolution and function of chromatin domains across the tree of life. *Nat Struct Mol Biol* 2024; 31: 1824-37.

Tabor S. Expression using the T7 RNA polymerase/promoter system. *Curr Protoc Mol Biol* 2001; 16: 16.2.

Theobald D. A formal test of the theory of universal common ancestry. *Nature* 2010; 465: 219-22.

Thompson MK, Nocedal I, Culviner PH et al. Escherichia coli SymE is a DNA-binding protein that can condense the nucleoid. *Mol Microbiol* 2022; 117(4): 851-70.

Tolstorukov MY, Virnik K, Zhurkin VB et al. Organization of DNA in a bacterial nucleoid. *BMC Microbiol* 2016; 16: 22.

Valentine JW. The Evolution of Complexity in Metazoans. In: Riste T, Sherrington D, editors. Physics of Biomaterials: Fluctuations, Self-assembly and Evolution. NATO ASI Series, vol 322. Dordrecht: Springer; 1996. p. 327-62.

Vosseberg J, van Hooff JE, Köstlbacher S et al. The emerging view on the origin and early evolution of eukaryotic cells. *Nature* 2024; 633: 295-305.

Wardleworth BN, Russell RJ, Bell SD et al. Structure of Alba: an archaeal chromatin protein modulated by acetylation. *EMBO J* 2002; 21(17): 4654-62.

Weiss M, Sousa F, Mrnjavac N et al. The physiology and habitat of the last universal common ancestor. *Nat Microbiol* 2016; 1: 16116.

Wierer S, Daldrop P, Ahmad MU et al. TrmBL2 from Pyrococcus furiosus Interacts Both with Double Stranded and Single-Stranded DNA. *PLoS ONE* 2016; 11(5): e0156098.

Williams TA, Foster PG, Nye TM et al. A congruent phylogenomic signal places eukaryotes within the Archaea. *Proc Biol Sci.* 2012; 279(1749): 4870-9.

Woese CR and Fox GE. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc Natl Acad Sci USA* 1977: 74(11): 5088-90.

Woese CR, Kandler O, and Wheelis ML. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* 1990; 87: 4576-9.

Yamaura K, Takemata N, Kariya M et al. Chromosomal domain formation by archaeal SMC, a roadblock protein, and DNA structure. *Nat Commun* 2025; 16: 1312.

Yilmaz C and Schnetz K. High Abundance of Transcription Regulators Compacts the Nucleoid in Escherichia coli. *J Bacteriol* 2022; 204: e00026-22.

Yin J, Liu Y, He D et al. A TrmBL2-like transcription factor mediates the growth phase-dependent expression of halolysin SptA in a concentration-dependent manner in Natrinema gari J7-2. *Appl Environ Microbiol* 2024; 90(7): e0074124.

Zhang C, Phillips AP, Wipfler RL et al. The essential genome of the crenarchaeal model Sulfolobus islandicus. *Nat Commun* 2018; 9(1): 4908.

Zhang J, Feng X, Li M et al. Deep origin of eukaryotes outside Heimdallarchaeia within Asgardarchaeota. *Nature* 2025; 642: 990-8.

Zhang N, Guo L and Huang L. The Sac10b homolog from Sulfolobus islandicus is an RNA chaperone. *Nucleic Acids Res* 2020; 48(16): 9273-84.

Zhang Z, Zhan Z, Wang B et al. Archaeal Chromatin Proteins Cren7 and Sul7d Compact DNA by Bending and Bridging. *mBio* 2020; 11(3): e00804-20.