X-ray crystallographic analyses of archaeal DNA binding proteins TrmBL2 and TrmB

Dissertation submitted for the degree of

Doctor of Natural Sciences

Presented by

Misbha Ud Din Ahmad

at the

Universität Konstanz

Faculty of Sciences

Department of Biology

Date of the oral examination: 11-11-2015

First referee: Prof. Dr. Wolfram Welte

Second referee: Prof. Dr. Andreas Marx

Third referee: Prof. Dr. Michaeal Thomm

CONTENTS

PART II: TrmB

LIST OF ABBREVIATIONS

ADAR1	**A**denosine **d**eaminase **a**cting on **R**NA
Alba	**A**cetylation **l**owers **b**inding **a**ffinity
ArgA	**A**mino-acid **a**cetyltransferase
AT	**A**denine **T**hymine
ATP	**A**denosine **t**ri**p**hosphate
bp	**b**ase **p**air
BSA	**B**uried **S**urface **A**rea
CC	**C**oiled **c**oil
CHES	*N*-**C**yclo**h**exyl-2-amino**e**thane**s**ulfonic acid
CSL	**C**BF1, **S**uppressor of Hairless, **L**ag-1
CTD	**C**-**t**erminal **d**omain
CV	**C**olumn **v**olume
DMSO	**D**i**m**ethyl **s**ulf**o**xide
DNA	**D**eoxyribose **n**ucleic **a**cid
ecTrmBL2	TrmBL2 expressed in ***E**scherichia **c**oli*
E-SET	**E**RG-associated protein with **SET** domain
ETS	**E**-**t**wenty **s**ix
ewHTH	**E**xtended **w**inged **h**elix **t**urn **h**elix
FOX	**Fo**rkhead bo**x**
GC	**G**uanine **C**ytosine
HEPES	4-(2-**h**ydroxy**e**thyl)-1-**p**iperazine**e**thane**s**ulfonic acid
H-NS	**H**istone like **n**ucleoid **s**tructuring protein
Hox-Exd	**Ho**meobo**x-Ex**tra**d**enticle
HySS	**Hy**brid **S**ubstructure **S**earch
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IEX	**I**on **ex**change
kDa	**k**ilo **Da**lton
MD	**M**alto**d**extrins
MPD	2-**M**ethyl-2,4-**p**entane**d**iol
MR	**M**olecular **R**eplacement

NAP	**N**ucleoid **a**ssociated **p**rotein
NCS	**N**on **c**rystallographic **s**ymmetry
PDB	**P**rotein **d**ata **b**ank
PEG	**P**oly**e**thylene **g**lycol
pfTrmBL2	TrmBL2 expressed in ***P**yrococcus **f**uriosus*
PLD	**P**hospho**l**ipase **D**
PTM	**P**ost **t**ranslational **m**odification
RFX1	**R**egulatory **f**actor **X**
RMSD	**R**oot **M**ean **S**quare **D**eviation
RNA	**R**ibose **n**ucleic **a**cid
SAD	**S**ingle wavelength **a**nomalous **d**ispersion
Scr	**SC**AREC**R**OW
SDS-PAGE	**S**odium **d**odecyl **s**ulphate - **P**oly**a**crylamide **g**el **e**lectrophoresis
SelB	**Sel**enocystine tRNA-specific elongation factor
SeMet	**Se**leno**met**hionine
SUMO	**S**mall **U**biquitin like **Mo**difier
TGM	**T**hermococcales **g**lycolytic **m**otif
TLS	**T**ranslation **L**ibration **S**crew
TM	**T**rehalose **M**altose
TrmB	**T**ranscriptional **r**egulator of the **m**altose operon
TrmBL2	**TrmB L**ike **2**
ULP1	**U**biquitin-**l**ike-specific **p**rotease 1
WRN	**W**e**rn**er syndrome ATP-dependent helicase
XDS	**X**-ray **D**etector **S**oftware
YpR	P**y**rimidine **p**hosphate pu**r**ine

SUMMARY

In archaea, packaging of DNA into chromatin and transcriptional regulation are closely related processes and the proteins involved therein often exhibit overlapping roles. In case of extremophilic archaea, the challenge to protect the DNA from detrimental effects of the harsh environments has led many of these chromatin proteins to evolve the additional function of DNA protection. The multiple roles these proteins play in DNA metabolism makes them interesting candidates for structural studies.

The first and major part of this work deals with the structural elucidation of TrmBL2, a member of the TrmB family of transcriptional regulators. Recent studies have to a great extent established the role of TrmBL2 in both chromatin shaping and transcriptional regulation. Upon association with DNA, TrmBL2 has been shown to form thick fibrous structures. Deletion of TrmBL2, in addition to the loss of fibrous structure leads to upregulation of many unrelated genes thereby providing evidence for its role as a dual functional protein.

For structure determination, TrmBL2 was heterologously expressed in *E.coli*, subjected to ion-exchange and size exclusion chromatographic purification and crystallized with 19 or
17 bp TGM (Thermococcales Glycolytic Motif) dsDNA. In the absence of a suitable molecular replacement model, phases were determined by the Selenium Single Wavelength Anomalous Dispersion (Se-SAD) method. The structure of DNA-free TrmBL2 expressed in *Pyrococcus furiosus* was subsequently determined by molecular replacement.

TrmBL2 crystallizes as a tetramer in an asymmetric unit, both in the DNA-bound and DNA-free forms. The structure reveals an extended winged Helix Turn Helix (ewHTH) domain at the N-terminus followed by a coiled coil dimerization domain and a Phospholipase D (PLD) like domain at the C-terminus. While the electron density of the sugar-phosphate backbone of the bound TGM dsDNA is clearly distinguishable, the density for the nucleobases is averaged and represents a superposition of three binding modes with a 3 bp shift around the central 19 bp DNA at the 5’ and 3’ ends thereby explaining the observed 25 bp density. For the 17 bp DNA, the observed 21 bp density could be explained by a similar 2 bp shift at the 5’ and 3’ ends. During refinement, the occupancy of the nucleotides was adjusted so that the overall occupancy sums to the actual number of the base pairs used in crystallization. Given the non-specific binding of TrmBL2 to the DNA, the observed multiple binding modes and the resultant averaging out of the nucleobase density is not surprising.

The structure of the DNA-free TrmBL2 does not show any major differences from the DNA-bound structure.

The TrmBL2-DNA complex structure described in this work shows a hitherto unknown mode of tetramerization and DNA binding. The analyses of the crystal structures provide a basis for the reported non-specific binding of TrmBL2 to the DNA and also provides an explanation for its observed roles in chromatin structuring and transcriptional regulation.

The second part of this work details the efforts to devise a protocol for overexpressing TrmB and to overcome the low solubility issues of this protein with the ultimate aim of solving the structure of TrmB in complex with TM and MD promoters. Towards this end several constructs were tried but the problem of TrmB proteolysis proved to be a major hindrance in the realization of these goals.

ZUSAMMENFASSUNG

DNA Kondensation in Chromatin und Transkriptionsregulation sind in Archaeen eng zusammenhängende Prozesse und die beteiligten Proteine besitzen oft überlappende Funktionen. Bei den extremophilen Archaeen hat die Notwendigkeit, die DNA vor Schäden durch die harte Lebensumgebung zu schützen, dazu geführt, dass viele der Chromatin Proteine zusätzliche Schutzfunktionen entwickelt haben. Die multiplen Funktionen, die diese Proteine im DNA Stoffwechsel spielen, machen sie zu interessanten Kandidaten für Strukturuntersuchungen.

Der erste und Hauptteil dieser Doktorarbeit behandelt die Strukturaufklärung von TrmBL2, einem Mitglied der TrmB Familie der Transkriptionsregulatoren. Neuere Studien haben die Rolle von TrmBL2 in der Chromatinbildung und Transkriptionsregulation deutlich nachgewiesen. Bei der Anlagerung von TrmBL2 an DNA entsteht eine dicke Faserstruktur. Nach dem Ablösen von TrmBL2 wird zusätzlich zum Ausbleiben der dicken Faserstruktur die Hochregulierung vieler scheinbar nicht zusammenhängender Gene beobachtet, was eine Doppelfunktion klar werden lässt.

Für die Strukturanalyse wurde TrmBL2 in *E.coli* Zellen exprimiert, mit Ionenaustauscher- und Gelpermeationschromatographie gereinigt und mit einer 19- oder 17- Basenpaare langen TGM (Thermococcales Glycolytic Motif) Doppelstrang-DNA kokristallisiert. Wegen des Fehlens eines geeigneten Strukturmodells wurden die Phasen experimentell mit der Methode des Selenomethioninersatzes (Single Wavelength Anomalous Dispersion) bestimmt. Anschliessend wurde die Struktur von TrmBL2 ohne DNA mit TrmBL2 aus *Pyrococcus furiosus* Zellen mit Molecular Replacement bestimmt.

Die asymmetrische Einheitszelle der Kristalle enthält sowohl im Fall des DNA Komplexes als auch bei dem DNA-freien TrmBL2 ein Tetramer. Am N-Terminus bildet das Protein eine extended winged Helix-Turn-Helix (ewHTH) Domäne aus. Darauf folgt eine Dimerisierungshelix, die mit einem zweiten Protomer ein Coiled-Coil bildet. Am C-Terminus bildet das Protein eine Phospholipase D (PLD) ähnliche Domäne aus.

Während die Elektronendichte für das Zucker-Phosphatrückgrat der gebundenen TGM-DNA deutlich erkannt werden kann, lässt sich die Dichte der Nukleobasen nur als Überlagerung dreier gegeneinander jeweils um 3 Basenpaare verschobener Kopien des 10 Basenpaare langen TGM Motifs erklären, was eine scheinbar 25 Basenpaare lange Elektronendichte ergibt. Im Kristall mit der 17 Basenpaare langen DNA ergibt sich eine ähnliche Erklärung durch

Überlagerung von drei DNA-Kopien mit gegenseitigem Versatz von jeweils 2 Basenpaaren. Die Besetzung der Nukleotide wurde während der Verfeinerung so eingestellt, dass die Gesamtbesetzung die tatsächliche Anzahl an Basenpaaren ergibt, die während der Kristallisation verwendet wurden.

In Anbetracht der bekannten nichtspezifischen Bindung von TrmBL2 an DNA sind die beobachtete multiple Bindung und die Überlagerung mehrerer gegeneinander verschobener DNA Kopieen nicht überraschend.

Die Struktur von TrmBL2 ohne gebundene DNA zeigt keine wesentlichen Unterschiede zur Struktur mit gebundener DNA.

Die hier beschriebene Struktur des TrmBL2-Komplexes mit DNA zeigt einen bislang unbekannten Modus der Tetramerisierung und der DNA Bindung. Die Strukturanalyse erklärt die bekannte nichtspezifische DNA Bindung von TrmBL2 und seine beobachtete Rolle in der Strukturbildung von Chromatin und der Transkriptionsregulation.

Der zweite Teil der Doktorarbeit beschreibt die Versuche, ein Überexpressions- und Reinigungsverfahren für TrmB zu entwickeln und die schlechte Löslichkeit des Proteins mit dem Ziel zu verbessern, eine Kristallstruktur des Komplexes von TrmB mit dem TM und dem MD Promotor zu bestimmen. Zu diesem Zweck wurden mehrere Expressionskonstrukte hergestellt, aber das Problem der Proteolyse des Proteins erwies sich als unüberwindliches Hindernis.

PART I: TrmBL2

1. INTRODUCTION

1.1 Archaea: Life at the extremes

Archaea, earlier classified together with bacteria, were later recognized as a separate domain of life [1]. This recognition was prompted by observation of features that are unique to this domain. Prominent among these include unique cell membrane constituents [2], and production of methane as a metabolic end product [3]. Although lacking a nucleus and thereby being classified as prokaryotes, archaeal transcriptional and replication machineries and the proteins involved therein are strikingly similar to eukaryotes [4, 5]. Features shared with bacteria include the presence of a single circular chromosome, clustering of a large number of genes into operons, an evolutionary conserved cell division apparatus and similar types of transcriptional factors [6-8].

The most recent classification categorizes archaea into three main phyla: Euryarchaeota, Crenarchaeota and Thaumarchaeota [9, 10]. Several other phyla are also reported in the literature but their recognition as separate phyla is not yet confirmed [11].

1.2 Major archaeal phyla:

Euryarchaeota

The phylum Euryarchaeota comprises methanogens, extreme halophiles, hyperthermophiles and a majority of acidophilic-thermophilic prokaryotes. Production of methane by a large number of Euryarchaeota is an essential feature of their metabolism.

The extreme halophilic archaea live in highly saline environments and a majority of species require high salt concentrations ranging from 2-4 M for optimal growth.

Acidophiles flourish in low pH environments, usually at a pH < 3, with species of genus *Picrophilus* capable of growing in sub-zero pH conditions.

Hyperthermophiles, as the name indicates, thrive at extreme temperatures generally above 80°C. *Pyrococcus furiosus*, for example, can grow at temperatures ranging from 70 - 105°C with a growth optimum at 100°C.

Crenarchaeota

Phylum Crenarchaeota comprises a single class, *Thermoprotei*, and includes both hyperthermophilic archaea as well as prokaryotes found in cold environments. A majority of hyperthermophilic Crenarchaeota grow in geothermal heated soils, elemental sulfur and sulfide containing waters or hydrothermal vents. *Pyrodyctum fumarii*, an extreme hyperthermophile can grow at 113°C[12].

Thaumarchaeota

This phylum which includes the genus *Cenarchaeum* is represented by ammonia-oxidizing archaea which make a significant contribution to the global Nitrogen and Carbon cycles. Making up 20% of all bacteria and archaea present in the ocean and having a very low substrate threshold for total ammonium, Thaumarchaeota are among the dominant species that oxidize ammonia in the open ocean[13].

1.3 Chromatin organization in different forms of life

Eukaryota

Across all the domains of life, genomes need to be compacted inside the cell in such a way that its effective volume is considerably reduced and at the same time is made available to a myriad of DNA metabolic processes associated with it like transcription, replication, repair and recombination[14]. Eukaryotes, owing to their structural and functional complexity have evolved the most advanced mechanisms to deal with this problem. Wrapping of the DNA around histone octamers represents the primary mechanism which eukaryotes employ to pack the DNA inside the nucleus[15]. The structures thus formed are called nucleosomes[16]. A host of dedicated ATP-dependent proteins, depending on the metabolic requirements of the cell, actively alter the structure of the nucleosome in a process called as chromatin remodelling[17]. At the global level, among other factors, it has been hypothesized that the posttranslational modifications (PTMs) of the histone tails, which constitute the so called “histone code”, determine whether specific genes are activated or repressed[18]. Therefore histones apart from packaging the eukaryotic DNA also act as global determinants of gene expression.

Bacteria

Bacteria in contrast lack histone proteins. Instead a group of chromatin binding proteins called Nucleoid-Associated Proteins or NAPs take up the function of DNA condensation. Some of the *Escherchia coli* NAPs which have been characterized in considerable detail include H-NS, HU, Fis, IHF and StpA [19]. These NAPs serve to both organize the bacterial chromosome and regulate gene expression [20]. In particular, H-NS has been shown to bind preferentially to AT-rich gene sequences and act as a global transcriptional silencer [21-27]. Oligomerization of H-NS has been shown to assist in the formation of higher order chromosome structures [28] possibly through DNA looping, bridging and stiffening or a combination of these [29-31]. Biophysical characterization of H-NS and other NAPs coupled with the evidence that they bind to numerous sites on the DNA has led to the hypothesis that these proteins function as chromosome organizing centers [32]. Whether these in vitro observations about NAPs hold true in vivo, and exactly how the organization of bacterial chromatin is affected globally by these proteins remains largely unknown [33].

Archaea

In archaea, a large number of chromatin binding proteins with variable phylogenetic distribution have been characterized. Every archaeal species encodes more than one type of chromatin protein with differing effects on DNA structure. In many cases, these proteins are expressed as functionally overlapping multiple paralogues. In Euryarchaeota, histones are the major chromatin proteins. Homologues of eukaryotic core histones H3 and H4 are encoded by a majority of the species belonging to this phylum. As tetramers, these histones wrap approximately 60bp of DNA and form “beads-on-a-string” kind of structures, analogous to the ones found in eukaryotes [34-37].

Crenarchaeota usually lack histones and instead possess a group of small, usually basic, proteins that bear function homology to bacterial NAPs [38].Some of the well-known examples include Sul7 family of proteins, Sac10a, Cren7 and CC1.

The Sul7 family of proteins derives its name from the genus *Sulfolobus*, species of which express a number of highly conserved, 7kDa, architectural proteins. These highly expressed basic proteins bind to DNA without any apparent sequence specificity. Besides their role in

chromosome shaping, these proteins impart stability to DNA against thermal denaturation[39].

Sac10a homologues apart from being widely distributed in crenarchaeota, are also found in some euryarchaeal species. Electron microscopy studies of Sac10a protein show that this protein assists in DNA bridging [40].

Cren7, an abundantly expressed chromatin protein is present in almost all of the crenarchaeal species. The few species which lack this protein instead encode histone [41, 42]. Although this protein hardly shares any amino acid sequence similarity with Sul7, both proteins have similar overall structure and biochemical properties.

A small, 6kDa protein, CC1 (crenarchaeal chromatin protein 1) which binds both ssDNA and dsDNA with same affinity is found in a few crenarchaeal species. CC1 is believed to be an architectural chromatin protein but exactly how it influences chromatin structure is not yet fully understood [43, 44].

Another group of chromatin proteins belonging to the Alba (acetylation lowers binding affinity) superfamily is present throughout the archaeal kingdom and has been extensively characterized. In addition to dsDNA binding, Alba proteins also interact with RNA and are thought to be involved in RNA metabolism [45-47]. The expression profiles and DNA binding properties of Alba proteins differ across archaeal phyla. In methanogenic euryarchaeal species, Alba is a low-abundance protein which binds in a sequence specific manner to dsDNA [43] whereas in crenarchaea Alba proteins are highly expressed and bind dsDNA in a non-specific manner [48, 49]. Depending on whether Alba paralogues are expressed or not, Alba forms homo or heterodimers and binds DNA in a cis or trans-manner resulting in the formation of stiff filamentous structures in the former case and looped structures in the latter [50, 51].

Table 1 details the various chromatin proteins and transcription factors found in different archaeal phyla.

			Chromatin proteins											Transcription regulators			
Phylum	Order	Genus	Alba	Sul7	CC1	HU	Cren7	Sso10a	Histone	Sso7c	MC1	7kMK	TrK	Lrp	TrmB	ArsR	HTH_3
Thaumarchaeota		*Cenarchaeum*	✓	–	–	–	–	–	✓	–	–	–	–	✓	✓	✓	✓
Aigarchaeota		*'Candidatus Caldiarchaeum'*	✓	–	–	–	–	–	✓	–	–	–	–	✓	✓	✓	✓
Crenarchaeota	Sulfolobales	*Sulfolobus*	✓	✓	–	–	✓	✓	–	✓	–	–	–	✓	✓	✓	✓
		Metallosphaera	✓	✓	–	–	✓	✓	–	✓	–	–	–	✓	✓	✓	✓
		Acidianus	✓	✓	–	–	✓	✓	–	✓	–	–	–	✓	✓	✓	✓
	Desulfurococcales	*Ignicoccus*	✓	–	–	–	✓	–	–	–	–	–	–	✓	✓	✓	✓
		Desulfurococcus	✓	–	✓	–	✓	–	–	–	–	–	–	✓	✓	✓	✓
		Aeropyrum	✓	–	✓	–	✓	✓	–	–	–	–	–	✓	✓	✓	✓
		Ignisphaera	✓	–	–	–	✓	✓	–	–	–	–	–	✓	✓	✓	✓
	Thermoproteales	*Acidilobus*	✓	–	–	–	✓	–	–	–	–	–	–	✓	✓	✓	✓
		Pyrobaculum	✓	–	✓	–	✓	✓	–	–	–	–	–	✓	✓	✓	✓
		Thermoproteus	✓	–	✓	–	✓	✓	–	–	–	–	–	✓	✓	✓	✓
		Vulcanisaeta	✓	–	–	–	✓	✓	✓	–	–	–	–	✓	✓	✓	✓
		Caldivirga	✓	–	–	–	✓	✓	✓	–	–	–	–	✓	✓	✓	✓
		Thermofilum	✓	–	–	–	–	–	✓	–	–	–	–	✓	✓	✓	✓
Korarchaeota		*'Candidatus Korarchaeum'*	✓	–	–	–	–	–	✓	–	–	–	–	✓	✓	✓	✓
Euryarchaeota		*Archaeoglobus*	✓	–	–	–	–	✓	✓	–	–	–	–	✓	✓	✓	✓
		Methanosarcina	–	–	–	–	–	✓	✓	–	✓	–	–	✓	✓	✓	✓
		Halobacterium	–	–	–	–	–	–	✓	–	✓	–	–	✓	✓	✓	✓
		Methanobacterium	✓	–	–	–	–	–	✓	–	–	–	–	✓	✓	✓	✓
		Methanococcus	✓	–	–	–	–	–	✓	–	–	–	–	✓	✓	✓	✓
		Methanosphaerula	✓	–	–	–	–	–	✓	–	–	–	–	✓	✓	✓	✓
		Methanopyrus	✓	–	–	–	–	–	✓	–	–	✓	–	✓	✓	✓	✓
		Pyrococcus	✓	–	–	–	–	–	✓	–	–	–	✓	✓	✓	✓	✓
		Thermococcus	✓	–	–	–	–	–	✓	–	–	–	✓	✓	✓	✓	✓
		Thermoplasma	✓	–	–	✓	–	–	–	–	–	–	–	✓	✓	✓	✓
Nanoarchaeota		*Nanoarchaeum*	✓	–	–	–	–	–	✓	–	–	–	✓	✓	✓	–	–

*Protein distributions were obtained by carrying out standard protein BLAST[100].

Table 1. Distribution of chromatin proteins and transcriptional regulators across various archaeal phyla.

Table reproduced from [52] with permission from the publisher vide license number 3654660066282 of the Copyright Clearance Center.

1.4 Molecular properties of DNA

A, B and Z-DNA: variation in surface and electrostatic properties

Most of the observed biochemical properties of DNA are explained by the double helical model proposed by Watson and Crick [53]. The most commonly occurring form of dsDNA under physiological conditions is B-DNA [54]. A right handed helical structure with base pairs oriented almost perpendicular to the helical axis, B-DNA features a wide, shallow major groove and a narrow, deep minor groove (Figure 1, b, c) [55] with the minor groove being more electronegative than the major groove (Figure 1, f, g). Depending on where either of the groove is enriched in AT or GC base pairs, the electrostatic potentials vary accordingly. The electronegativity of minor groove is enhanced in case of AT rich sequences compared to GC rich ones. This difference in electronegativity of groove surfaces between AT rich and GC rich sequences is due to the different orientation of the polar groups present at the base edges.

Under dehydrating condition, dsDNA assumes the A-form which differs from the B-form in featuring a narrow, deep major groove and a shallow, wide minor groove (Figure 1, a). GC sequences have a higher propensity of forming A-DNA structures. A-DNA forms a right handed helical structure with the base pairs tilted 20° with respect to the helical axis and also shifted towards the major groove. These variations from B-DNA result in the different major and minor groove features (Figure 1, a, e) [55-57].

Z-DNA is formed when alternating purine-pyrimidine sequences, under high salt conditions, form a left handed double helical structure [58, 59]. This type of structures is thought to be formed in negatively supercoiled DNA. Z-DNA does not feature a recognizable major groove (Figure 1, d) and the minor groove is similar in dimensions to the B-DNA minor groove.

DNA structure: deviations from ideal B-form

A global change frequently observed in crystal structures of protein-DNA complexes is DNA bending. Remo Rohs et al. define DNA bending “as a curvature distributed over a stretch of several base pairs, leading to a different orientation of the regions on both sides of the curvature” [60]. DNA bending, as in the case of other structural variations in dsDNA, is

sequence dependent and is commonly observed in sequences that contain A-tracts, regions which contain A:T base pairs including ApA[1] and TpT but not TpA steps [61-63].

Variations in DNA shape which occur on a local scale include DNA kinks and minor groove narrowing. DNA kinks are defined as local distortions of an otherwise linear helix [60]. A partial or complete disruption of base stacking interactions at a single base pair step results in the generation of kinks. Stabilization by base stacking interactions is weakest for Pyrimidine-purine (YpR)[2] base steps and minimal for TpA step [55, 64] which is therefore also referred to as the "hinge step" [65]. In many cases, the loss of the base pair stacking interactions is compensated by the intercalation of amino acid side chains of DNA binding proteins which stabilize these local deformations in DNA structure.

Another local structural variation observed in DNA is minor groove narrowing. Differences in base stacking interactions and hydrogen bonding patterns, which depend on the nucleotide sequence, exhibit their effect on the shape of the minor groove. Optimizing these interactions involves varying three rotational parameters which are:

1. Roll: Relative rotation between adjacent base pairs with respect to the base pairing axis.

2. Helical twist: Relative rotation between adjacent base pairs with respect to the helical axis and

3. Propeller twist: Relative rotation between bases within a base pair with respect to base pairing axis. [66]

Variations in the base pair steps and the three rotational parameters lead either to the narrowing or widening of the minor groove [63, 67].

[1] p stands for phosphate and is introduced to discriminate a base pair step from base pair.
[2] **Y** : P**y**rimidine, **R** : Pu**r**ine

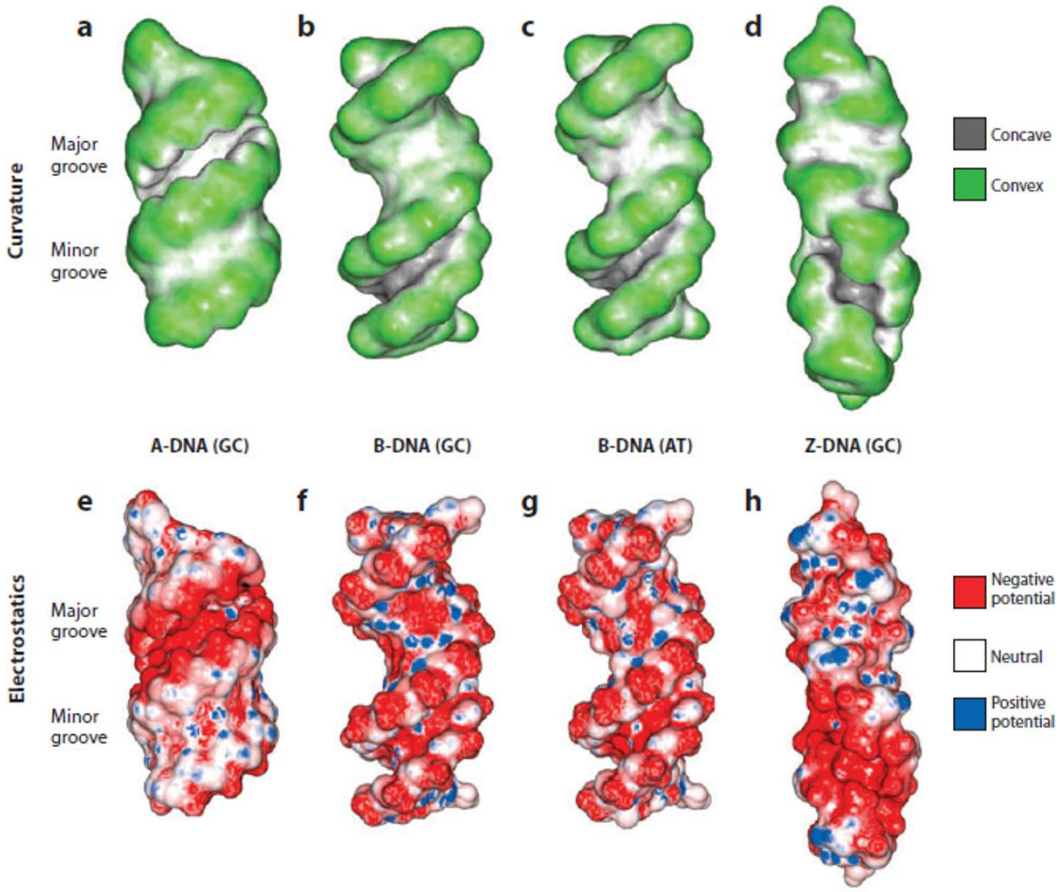


Figure 1. Molecular shape and electrostatic potential representations of three main forms of DNA
Top panel (a-d): GRASP2 [68] generated images of 14-mer dsDNA. Dark grey and green colors represent concave and convex surfaces respectively.
a) $d(GC)_7$ A-DNA model with a deep (9.5Å), narrow (2.2 Å) major groove and a wide (10.9Å), shallow (no depth) minor groove.
b) $d(GC)_7$ B-DNA model with a wide (11.4Å), shallow (4Å) major groove and a deep (5.5Å), narrow (5.9Å) minor groove.
c) $d(AT)_7$ B-DNA model. Because the models are built based on fiber diffraction data, a sequence dependent difference of molecular shape is not observed between AT and GC B-DNA models.
d) $d(GC)_7$ model of Z-DNA lacking a major groove (13.2 Å wide, no defined depth) and a deep (5Å), narrow(2.4Å) minor groove.
Bottom panel: DelPhi [69, 70] calculated electrostatic surface representation of 14-mer dsDNA. Positive and negative potentials are represented by blue and red colors respectively.
e) A-DNA displays a strong electronegative major groove and a hydrophobic minor groove which is partly a consequence of the exposed C3´ endo sugar groups.
f) $d(GC)_7$ B-DNA exhibits a strong electronegative minor groove and less electronegative minor groove.
g) $d(AT)_7$ B-DNA.
Difference in electrostatic potential between **f)** and **g)** arise because of the different functional groups that are exposed in the major and minor grooves. Guanine amino groups contribute to the positive potential in the major groove of GC rich sequences whereas in case of AT rich sequences, the methyl groups of thymine contribute to neutral potential in the major groove.

Figure adapted from [60] with permission from the publisher vide confirmation number 11360641 of the Copyright Clearance Center**.**

1.5 DNA binding proteins

DNA binding proteins, based on secondary structural elements which form the DNA-binding domain belong to three main classes viz:

1. Predominantly α

2. Predominantly β and

3. Mixed α/ β

1. Predominantly α

Proteins belonging to this class have DNA binding domains mainly composed of α-helices. Examples of this class include λ-repressor-like proteins, homeodomains and leucine zippers. α-helix is by far the most common secondary structural element used for base specific recognition in the major groove of the DNA. The orientation of the α-helix with respect to the helical axis of the DNA differs among protein families and also shows variations between proteins belonging to the same family [71]. Although mostly binding in the major groove, interaction of α-helix with the minor groove of the DNA is observed in the case of Lac repressor [72, 73] and intron endonucleases [74-76]. Some of the common motifs found in this class of proteins include the helix-turn-helix, helix-loop-helix and leucine zipper motifs.

Helix turn helix motif (HTH)

HTH motif along with its variations is one of the most common motifs found in DNA-binding proteins. The basic HTH motif consists of three α-helices which pack in a right handed fashion to form a three-helical bundle [77, 78]. In majority of cases, proteins containing the HTH motif insert the H3 helix, also called the “recognition helix”, into the major groove where its presents amino acid side chains for base specific recognition of DNA. (Figure 2, a). Despite being a predominantly major groove binding motif, HTH motif has also been shown to bind to the minor groove of DNA in case of O^6-alkylyguanine-DNA alkyltransferase [79].

Winged Helix turn Helix motif (wHTH)

A versatile DNA-binding motif, wHTH represents a variation of the HTH motif in which a β-hairpin is appended to its C-terminus (Figure 2, b). Proteins containing the wHTH domain are present across all forms of life [80].First discovered in hepatocyte nuclear factor-3γ (HNF-

3γ)/forkhead box [81], the wHTH domain is involved in almost all the aspects of nucleic acid metabolism. The recognition helix H3, like in the canonical HTH domain, is involved in the base specific interactions when presented to the major groove of DNA. Examples of proteins where such a mode of DNA binding is observed include the FOX family [82], E2F-DP2 transcription factor [83], lambda excision A (LexA) repressor [84] and ETS family [85]. In many of the cases where H3 is presented to the recognition helix, the wing1 (w1) has been observed to form minor groove contacts that are not well conserved.

A reversal of roles for H3 helix and wing1 is observed in RFX1, where the former interacts with the minor groove and the latter with the major groove [86].

R.DpnI restriction endonuclease [87] from *Streptococcus pneumoniae*, represents an example of wHTH domain where the H3 helix is used to recognize methylated targets, specifically, methylated adenines, in the major groove of DNA.

The use of wHTH domain in an enzymatic function is demonstrated in the case of O^6-alkylguanine-DNA methyltransferase (AGT) where the wHTH domain after recognizing specific structural features of alkylated DNA, removes alkyl groups from damaged guanine bases in a nonreversible reaction. In this case, the H3 helix of wHTH domain interacts weakly with the minor groove and the majority of interactions are with the sugar-phosphate backbone of DNA. Amino acids from helices other than H3 form these backbone contacts [79].

wHTH proteins have also been shown to bind Z-DNA, which are transitory structures formed at particular transcriptional sites. ADAR1 [88, 89], DAI [90], vaccinia virus E3L protein [91] and zebrafish PKZ protein kinase [92] represent examples where the H3 and w1 of wHTH domains are involved in stabilizing interactions with the sugar-phosphate backbone of Z-DNA.

Apart from interacting with DNA, wHTH proteins are also involved in sequence and conformation dependent recognition of RNA. SelB and La proteins represent examples where wHTH domains recognize unique RNA features and employ a combination of specific and non-specific interactions to carry out the required roles respectively, of selenocysteine incorporation into nascent polypeptides and binding to ssRNA [93-97].

Helicases are enzymes which expend energy from NTP (nucleoside triphosphate) hydrolysis to separate nucleic acid strands. In many of these enzymes it is the wHTH domain which

carries out the function of DNA unzipping. In the case of human WRN, residues present in w1 are involved in the unwinding of dsDNA [98].

wHTH proteins in some cases can switch roles from being primarily nucleic acid binding proteins to motifs that mediate protein-protein interactions. In case of Fok1, a restriction endonuclease, one of the wHTH domains interacts weakly with the DNA and is supposed to be playing a larger role of mediating protein-protein interactions. Exclusive use of wHTH domains for protein-protein interaction is seen in case of the heterotrimeric human replication protein A (hRPA) where this domain does not interact with DNA but acts as a scaffold for binding of other proteins [99].

Extended winged helix turn helix domain (ewHTH)

A variation of wHTH where extra helices α0 and α5 are appended to the N- and C -terminus respectively is observed in archaeal TFE [100] and RecQ helicase WRN [98] forming the extended helix turn helix commonly abbreviated as ewHTH domain (Figure 2, c).

Basic-region leucine zipper (bZIP) domain

DNA binding proteins belonging to this class are composed of long α-helices, of about 60 residues, and dimerize at the C-terminus via hydrophobic interactions of the leucine residues present at the dimer interface. The N-terminal DNA binding domains bind to the major grooves on either side of the DNA forming an open-scissor or chopstick like structure [101] (Figure 2, d) with the DNA binding region structured only when bound to DNA [102].

Basic-region helix loop helix (bHLH) domain

The bHLH domains share a very similar mode of DNA binding with bZIP proteins. However in the case of bHLH proteins the connection between the C-terminal dimerization helices and the N-terminal DNA binding domains is intervened by loop regions (Figure 2, e).

2. Predominantly β

DNA binding domains consisting mainly of β-strands are less frequent than their α-helical counterparts. A brief description of representative classes is presented below:

TATA box-binding protein

TATA binding proteins (TBPs) bind to the minor groove of DNA by presenting an extensive β-sheet surface and cause significant distortion of the dsDNA helix (Figure 2, f). This distortion facilitates base specific contacts in the otherwise less accessible minor groove. [103, 104].

Immunoglobulin-like-β-sandwich

DNA binding proteins which use an Immunoglobulin like fold include p53-like transcription factors [105], E-SET domains [106, 107] and Runt domains [108]. Loops which connect the β-strands in this type of sandwich fold are the main regions which are involved in DNA binding (Figure 2, g).

β-trefoil

An example of the β-trefoil fold is found in the case of effector CSL where it is present as an insertion between the N and C-terminal domains. Along with the N-terminal domain, it has been shown to be involved in DNA binding. A capped β-barrel, the β-trefoil fold displays an approximate 3-fold symmetry, with the strands 1 and 4 forming the wall of the β-barrel, and strands 2 and 3 capping the 12-stranded structure [109] (Figure 2, h).

β-β-β-sandwich

This type of DNA-binding domain as observed in the crystal structure of the *Staphylococcus aureus* AgrA LytTR Domain [110] consists of 10 β-strands and 2 short α-helices with an N-C secondary structure arrangement of 2β-3β-α-2β-3β-α. The β-strands are arranged in an antiparallel manner and DNA binding is mainly achieved through residues located in the loop regions (Figure 2, i).

3. Mixed α/β

Mixed α/β domains are present in a vast majority of DNA-binding proteins. DNA-recognizing elements can be either individual secondary structural elements or any combination of these. Below is a short description of two representative examples:

Zinc finger proteins

A rather small, 30 residue-DNA-binding domain composed of an α-helix and a two stranded antiparallel β-sheet is found in many DNA-binding proteins (Figure 2, j). Depending on the number and type of residues that coordinate zinc, the domain is classified into Cys_2His_2, Cys_4, and Cys_6. The α-helix presents the residues involved in the base specific recognition in the major groove of the DNA [111].

Ribbon-helix-helix (RHH) motif

RHH DNA binding fold is observed in bacterial repressors Met and Arc [112-114] and consists of a two α-helices preceded by an antiparallel β-ribbon. Base specific recognition in the major groove is carried out by residues from the β-ribbon whereas the α-helices pack together to form a hydrophobic core and assist in dimerization (Figure 2, k).

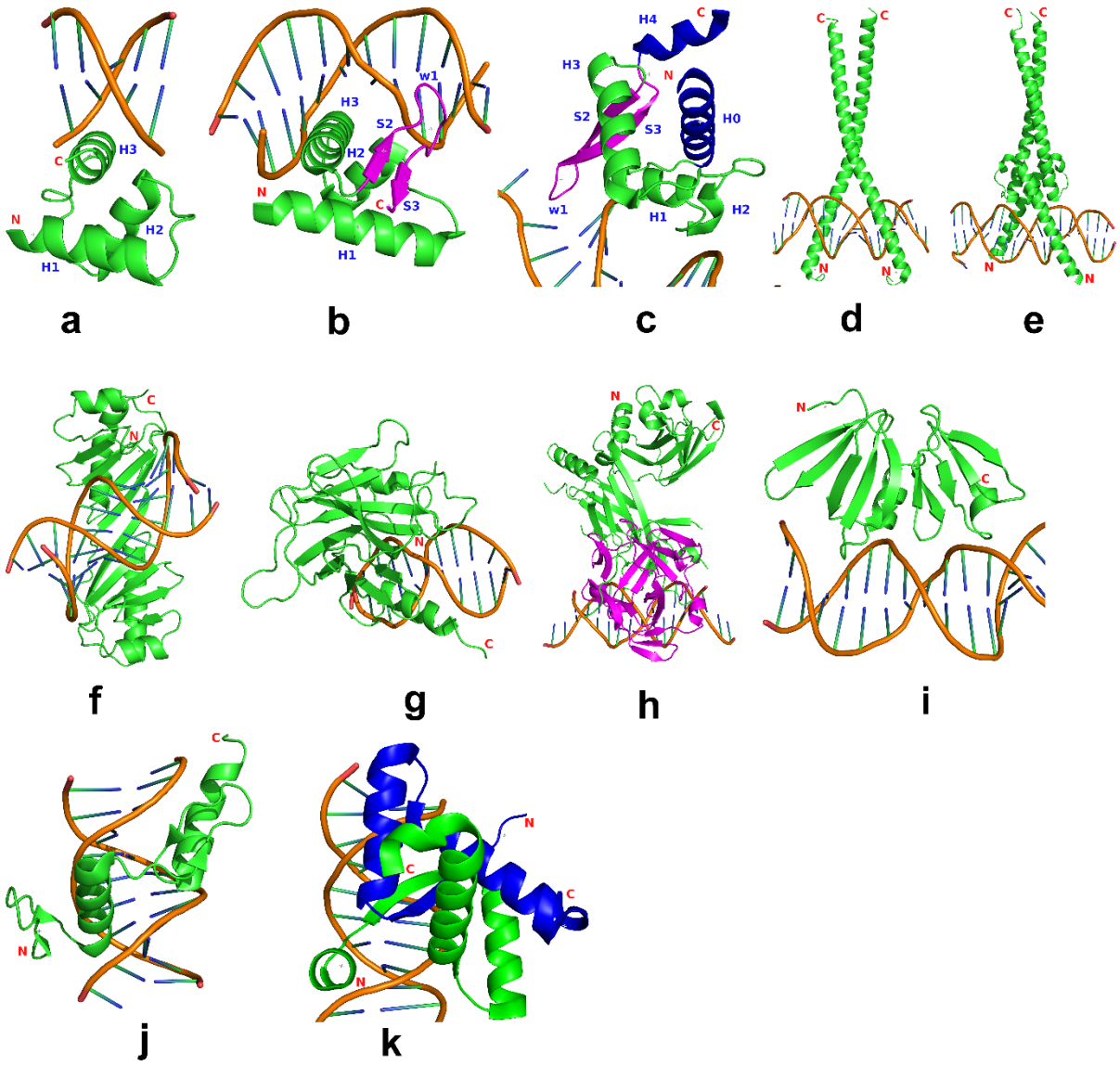


Figure 2. Representative examples of predominantly α, predominantly β and mixed α/β type transcription factors.

Top panel; predominantly α type: **(a)** Typical HTH motif (PDB:1K78) **(b)** wHTH motif (PDB:3JSO), the additional antiparallel β sheet (S2 and S3) and wing (w1) is colored in pink **(c)** ewHTH motif (PDB:2WWY), additional N-terminal (H0) and C-terminal (H4) helices are colored in blue. **(d)** bZIP domain (PDB:2E42) **(e)** bHLH domain (1NKP)

Middle panel; predominantly β type: **(f)** TATA-binding protein (PDB:1CDW), **(g)** Immunoglobulin like β sandwich domain illustrated by p53 core domain (PDB:4IBU), **(h)** β-trefoil domain, colored in pink, represented by nuclear effector CSL (PDB:3BRD), **(i)** AgrA LytTR domain (PDB:3BS1) exhibiting the β-β-β sandwich fold.

Bottom panel; mixed α/β type: **(j)** Zinc finger domain illustrated by mouse ZFP57 (PDB:4GZN), **(k)** A dimeric RHH motif represented by omega repressor (PDB:2BNW); monomers comprising the dimer are colour green and blue.

1.6 Factors determining specificity in protein-DNA recognition

Non-coding DNA sequences that are recognized by DNA-binding proteins are commonly viewed as linear strings composed of four types of nucleotides, A, G, C, T and their modifications rather than thinking of them as molecules that can form complex three dimensional structures. The properties of these three dimensional structures vary based on the composition, arrangement and positioning of these four nucleotides along the helical DNA structure. Although the specific interactions of DNA-binding proteins with nucleotide base-pairs are an important parameter in conferring specificity, they represent only a part of a much larger interaction surface which is defined by the three dimensional structure of both protein as well as DNA. When analysing protein-DNA interactions, the two interacting molecules should be considered as equally essential contributing partners [60]. For long, protein-DNA interactions have been divided into two categories; direct readout and indirect readout. Direct readout is defined as the specific hydrogen bonds which the amino acid side chains of the proteins make with the base pair edges in the DNA major groove. This lead to the assumption that the amino acid sequence is related to the base sequence in a one-to-one manner. However later studies [115] found that there is no such simple recognition code and therefore direct readout only cannot determine the sequence specificity of DNA-binding proteins.

Indirect readout refers to the protein-DNA interactions which do not involve direct contacts of amino acid side chains with the nucleotide bases but instead depend on the sequence of the nucleotides that have a propensity to deviate from the ideal B-DNA. This deviation results in a deformity such as a bend, which then facilitates the formation of hydrogen bonds or non-polar contacts. In the absence of such a deformation, there is a lesser likelihood of such interactions taking place. Indirect readout also encompasses water-mediated base specific interaction like the ones observed in high resolution Trp-operator complex [60].

With the increasing number of protein-DNA complexes whose crystal structures are available now, it is becoming apparent that direct readout and indirect readout are not isolated mechanisms. Rather a combination of these is used to achieve specificity in protein-DNA recognition. It has therefore been suggested to replace the terms direct readout and indirect

readout respectively with more descriptive terms as base readout and shape readout, respectively [60].

Base readout

Base specific interactions with the amino acid side chains usually provide more specificity in the major groove than in the minor groove. The polar groups that are displayed on the base edges have a unique pattern and can be clearly distinguished for A:T, T:A, C:G or G:C patterns in the major groove whereas such a distinction of base pair specific hydrogen bond acceptors and donors is not very clear in the minor groove (Figure 3). Commonly used domains that are involved in base specific interactions include HTH domains (homeodomains, 434, λ and Trp repressors), zinc fingers (TFIIIA), immunoglobulin folds (NF-κB) and bZip proteins (Max transcription factor) which present the N-terminal DNA binding helix to the major groove.

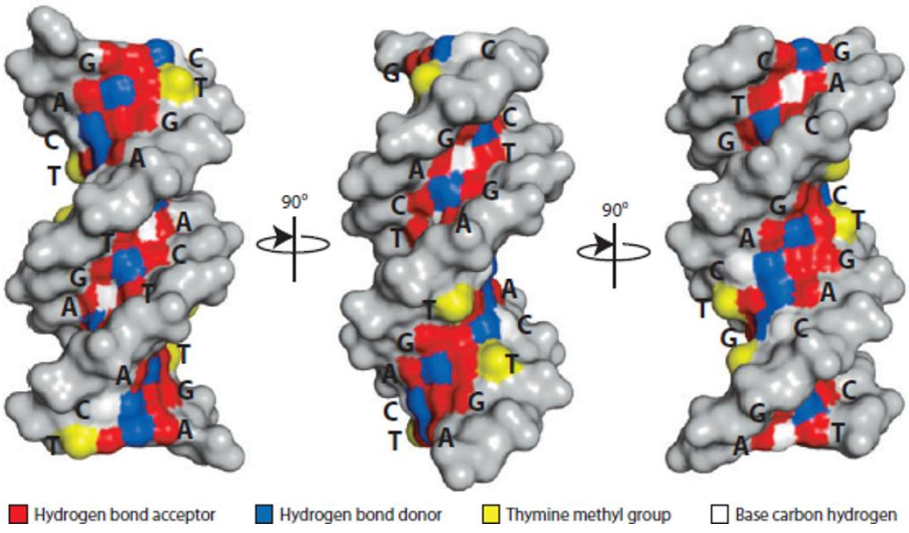


Figure 3. Disposition of the functional groups of bases in the major and minor grooves

Model of a dodecamer $(GATC)_3$ built by 3DNA based on fiber diffraction data [66]. Hydrogen bond acceptors and hydrogen bond donors are shown in red and blue color respectively. Yellow color represents thymine methyl group and white indicates a base carbon hydrogen. A distinction between A:T and T:A or G:C and C:G base pairs is not clear in the minor groove. In contrast the unique features arising from these interactions are clearly distinguishable in the major groove.

Figure adapted from [60] with permission from the publisher vide confirmation number 11360641 of the Copyright Clearance Center.

Interactions involving base readout

Hydrogen bonds

Specificity of interactions is determined not only on the number of hydrogen bonds but also on the type of hydrogen bonds formed. Bidentate hydrogen bonds, two hydrogen bonds involving different donor and acceptor atoms, provide for the greatest specificity followed by bifurcated hydrogen bonds, two hydrogen bonds from a single donor, and single hydrogen bonds. Bidentate hydrogen bonds which provide for exceptional specificity can involve a single base, a base pair, two adjacent bases on the same strand or two diagonally positioned bases on opposite strands.

Highly ordered water molecules have been observed in many structures to act as hydrogen bonding bridges. A prominent example is the Trp repressor where majority of base specific interactions are water mediated [115, 116]. RXR-DNA complex represents another example where the role of water molecules in base specific readout is observed [117].

Hydrophobic interactions

While hydrogen bonding interactions mostly involve purines, hydrophobic interactions in the major groove involve pyrimidines and are used to distinguish thymine (which carries a methyl group) from cytosine. Examples include bacteriophage 434 repressor and lambdoid bacteriophage P22 c2 repressor-operator complex [118-120]. In the latter example, a valine residue specifically recognizes an indentation created by four thymine methyl groups.

A second example where hydrophobic interactions can provide for specificity is observed in bacterial cold shock proteins where phenylalanine and histidine residues are involved in stacking interactions with thymines populated on polythymine strands, and discriminate them from cytosine through hydrogen bonding [121, 122].

Hydrogen bonding with bases in the minor groove, although observed in many proteins [103] hardly provides for specificity. It is usually the shape and plasticity of the minor groove which are used as discriminating factors. Architectural proteins bind mostly to the minor groove and make extensive hydrophobic contacts with the bases resulting in the widening of the minor groove in many observed cases [123].

Shape readout

Base recognition is not the only determinant of specificity. It has been observed that a vast majority of DNA-binding proteins recognize local or global changes in DNA structure to achieve specificity. One of the structural features which many DNA binding proteins utilize to enhance specificity is the minor groove shape which in turn is sequence dependent. Local sequence variations that influence the shape of the minor groove have been shown to act as discriminators in case of Scr-specific and Hox-Exd consensus site recognition. The difference in these consensus sites result in the creation of unique, sequence dependent, electrostatic potentials which contribute to their specificity [124, 125].

AT rich sequences have a propensity of forming narrow minor grooves with enhanced negative electrostatic potential. Many proteins including the DNA binding arm of Scr use arginine side chains to recognize this property. It has been proposed that such a feature may represent a general mode of sequence specific shape recognition of DNA [126].

Examples of minor groove shape recognition is also observed in case of architectural proteins which bind exclusively to the minor groove where base specific contacts are used to stabilize DNA shape [123].

Contribution of major groove shape to binding specificity has been observed in case of hRFX1, a wHTH protein, in which the recognition helix H3 instead of binding the major groove contacts the minor groove. This results in a widening of the minor groove with concomitant narrowing of the major groove which enhances its complementarity to be specifically recognized by wing w1 of hRFX1 [86, 127].

Kinks represent deviations from ideal B-DNA structure at a local scale. These local, sequence dependent deformities in DNA structure facilitate the amino acid- nucleobase interactions and are often stabilized by intercalation of hydrophobic residues which make for the loss of base stacking interactions [123, 128].

DNA bending, a global deviation from ideal B-DNA structure is observed in many crystal structures. E2 protein from papillomavirus binds as a dimer to two half sites separated by a flexible linker ($ACCGN_4CGGT$). The linker in this case causes the bending of the DNA which in turn increases the interactions between the two monomers which form the dimer.

Bending of DNA sequences is either an intrinsic property or a forced effect which occurs upon protein binding. Recent studies suggest that high-affinity binding sites represent the former and low-affinity binding sites the latter case [129, 130].

An example of DNA bending playing a role in determining specificity is observed in the case of phage 434 repressor where the bending of the operator sequence increase the interactions of the bases with the recognition helix. GC base pairs when present in the operator sequence reduce bending as they offer more resistance to bending than AT base pairs [131, 132].

Nucleosomes regulate gene expression as well as the compaction of the genomic DNA by utilizing the sequence dependent variations that affect DNA shape at both local and global levels. The narrowing of minor grooves in TATA boxes as a consequence of DNA wrapping around the histone core obstructs the binding of TBP (TATA binding protein) [133]. On the other hand, it has been suggested that the bending of DNA in nucleosomes aides in the binding of p53 by forming an accessible surface away from the histone cores [134]. The proclivity towards bending, of short A-tracts, is made use of at DNA positons which face the histone core [126, 135] where a certain degree of bendability is required for efficient wrapping. In yeast genomic DNA it has been observed that short A-tracts are repeated at periodic intervals which coincide with the wrapping of DNA around the histone core [126].The wrapping of DNA around the histone core is further facilitated by kinks generated by CpA steps adjoining the A-tracts. A second consequence of the A-tract periodicity is the creation of narrow minor grooves at periodic intervals which act as sites for interaction with histone proteins mainly via their arginine side chains [126]. These examples further illustrate the role sequence dependent variations in DNA shape play in determining specificity of protein-DNA interactions.

2. MATERIALS AND METHODS

2.1 TrmBL2 expression in *Pyrococcus furiosus* (pfTrmBL2) and purification

Pyrococcus cells expressing N-terminal His_6 tagged TrmBL2 were obtained from our collaboration partners at the Universität Regensburg. The details of the N-terminal His tag introduction into TrmBL2 and subsequent transformation into *Pyrococcus furiosus* resulting in the generation of the MURPf10 strain are detailed in [136, 137].

30g of *Pyrococcus furiosus* cell pellet was resuspended in 50ml of buffer consisting of 40mM HEPES pH 7.5, 1M NaCl, 20% glycerol, 10µg/ml of DNase (Roche) and 1/3rd of a protease inhibitor tablet (Roche). Cell lysis was carried out by passing the cells several times through cell disruptor (Constant Systems Limited, UK). The lysate was centrifuged at 185500g for 60 minutes. Supernatant containing soluble pfTrmBL2 was filtered by passing through a 0.45µm filter before being loaded onto a 5ml Ni HisTrap FF column (GE Healthcare) pre-equilibrated with Buffer A (40mM HEPES pH 7.5, 20mM imidazole, 1M NaCl and 20% glycerol). After washing the column with 10CV of Buffer A, protein elution was carried out by applying a linear imidazole gradient with Buffer B (40mM HEPES pH 7.5, 0.5M imidazole, 1M NaCl, and 20% glycerol). Peak fractions were subjected to SDS-PAGE gel electrophoresis to determine the presence of pfTrmBL2. Fractions containing pfTrmBL2 were pooled and concentrated to 500µl in a 30kDa centrifugal ultrafiltration device (Vivaspin, Sartorius). To remove any precipitates or large aggregates, the protein was filtered through a 0.1µm centrifugal filter (Ultrafree®, Millipore) before being loaded onto a 60ml Superdex 200 column pre-equilibrated with 40mM HEPES pH 7.5, 150mM NaCl and 20% glycerol. The column was run at a flowrate of 0.2ml/min during the course of elution.

2.2 Crystallization of pfTrmBL2

For crystallization, pfTrmBL2 was concentrated in a 30kDa cut-off Vivaspin concentrator and the concentration was determined by measuring the UV absorbance of the protein at 280nm (Eppendorf Biophotometer). After filtering the protein through a 0.1µm centrifugal filter, initial screening was carried out in 96-well sitting drop plates with a starting protein concentration of 16mg/ml. Fine screening was done in Qiagen 15-well hanging drop plates.

2.3 TrmBL2 expression in *E.coli* (ecTrmBL2) and purification

TrmBL2 transformed BL21 (DE3) *E.coli* cells were grown overnight in 100ml of LB medium containing 0.1% v/v Kanamycin (50 mg/ml) at 30°C. 10 ml of overnight culture was used to inoculate 1L of fresh LB medium and cells were grown at 37°C. Protein expression was induced by adding 1ml of 1mM IPTG at OD_{600} of 0.6. Cells were harvested 4 hours post induction and stored at -80°C.

6g of *E.Coli* cell pellet was resuspended in 40ml of lysis buffer containing 40mM HEPES pH 7.5, 150mM NaCl, 20% glycerol and 1/3rd of a protease inhibitor tablet (Roche). Cell lysis was done at room temperature by passing the resuspended cells four times through French Press at 16000 psi. The lysate was heated for 25 minutes at 80°C prior to centrifugation at 185500g for 60 minutes. The supernatant containing ecTrmBL2 was diluted to 50mM NaCl by adding appropriate volume of 40mM HEPES pH 7.5. Glycerol was used to only stabilize the protein during heat treatment and was not required during the purification process. The supernatant was filtered to remove any particulate matter and loaded onto as 30ml Q-Sepharose anion exchange column. After washing with 10CV of Buffer A, the protein was eluted by applying a linear gradient of NaCl with Buffer B (40mM HEPES pH 7.5, 1M NaCl). After analysis of the peak fractions by SDS-PAGE, fractions containing ecTrmBL2 were pooled and concentrated in a 10kDa filter to 500µl. The protein was filtered through a 0.1µm filter before loading onto a 60ml Superdex 200 column pre-equilibrated with 40mM HEPES pH 7.5 and 150mM NaCl. For elution, the column was run at 0.2ml/min.

2.4 Crystallization of ecTrmBL2

For the crystallization of ecTrmBL2, purified protein was concentrated in a 30kDa cut-off Vivaspin concentrator to 10mg/ml. For preparing the protein-DNA complexes, the protein was mixed with either 19bp (5'GTATCACTATCGATGATAC3') or 17bp (5'TATCACTATCGATGATA3') TGM (*Thermococalles* glycolytic motif) sequence in a 1(ecTrmBL2):3.4(TGM) molar ratio, incubated at 37°C for 30 minutes and filtered through a 0.1µm centrifugal filter before screening for crystals. Initial high throughput screening was carried out in 96-well sitting drop plates using the Phoenix ARI crystallization robot. Initial hits were optimized by fine screening in Qiagen 15-well hanging drop plates.

2.5 Selenomethionine (SeMet) incorporation into ecTrmBL2

For SeMet incorporation into TrmBL2, protein expression was carried out in M9 minimal medium which was prepared in two steps as described below:

1. **5x M9 salts preparation**

   In 800ml ddH_2O, dissolve

   64g $Na_2HPO_4.7H_20$

   15g KH_2PO_4

   2.5g NaCl

   5g NH_4Cl

   Adjust pH to 7.2

   Adjust volume to 1000ml by adding dH_2O and autoclave.

2. **M9 minimal media preparation**

   To 780ml of autoclaved dH_2O add

   200ml of 5x M9 salts

   20ml of 20% sterile filtered glucose

   2ml of 1M $MgSO_4$

   100µl of 1M $CaCl_2$

ecTrmBL2 transformed *E.Coli* cells grown in LB medium (Kan^+) for approximately 10 hours were used to inoculate M9 minimal medium which was then grown overnight at 37°C. 1L of fresh M9 minimal medium (Kan^+) was inoculated with 10ml of overnight culture and grown at 37°C till an OD_{600} of 0.5 was reached. At this point, an amino acid mix consisting of 100mg Lys, Phe and Thr; 50mg of Ile, Leu and Val; and 60mg SeMet per litre (I used to dissolve the amino acid mix in minimum amount of water before adding it to the culture) was added to the cell culture and cells were grown for additional 15 minutes. Protein expression was induced by adding 1ml of 1mM IPTG and cells were grown at 18°C for approximately 18 hours before being harvested and stored at -80°C.

The above mentioned protocol for SeMet labelling was obtained from UCLA-DOE (http://work.doe-mbi.ucla.edu/local/protocols/). Original protocol is described in [138].

2.6 Data collection, phasing, structure determination and refinement

For data collection, crystals were flash cooled in liquid N_2 and data were collected at beamlines X06SA or X06DA of Swiss Light Source (SLS) Villigen, Switzerland. Data processing was done with XDS [139] and for SeMet data anomalous scatterers were found from a single data set (Se-SAD) by the HySS submodule [140] of the Phenix package [141]. A partial model was built by Autobuild module of Phenix and further model building was done in Coot [142]. Refinement was carried in Phenix with Chain A being NCS retrained to Chain B and Chain C to Chain D. TLS groups used in the refinement were selected from the output of TLS Motion Determination server [143]. Subsequent structures of ecTrmBL2-TGM17 and ecTrmBL2-TGM19 complexes were solved by Molecular Replacement [144].

The DNA-free pfTrmBL2 structure was solved by Molecular Replacement using the ecTrmBL2-TGM19 structure as the model. The MR solution had a Translation function Z score (TFZ) of 30.3 and a Log likelihood gain (LLG) of 5989. Initial model building was aided by Rosetta [145] followed by refinement in Phenix.

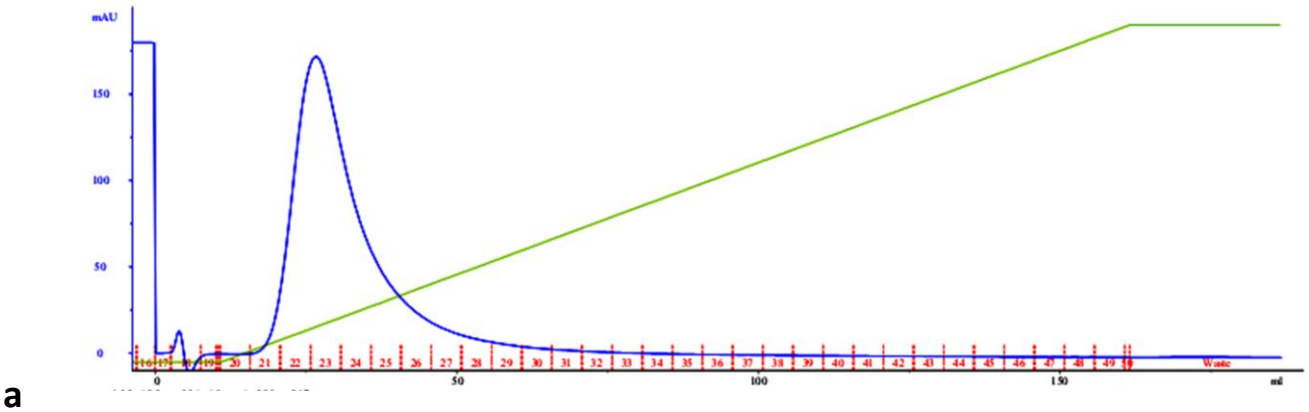
3. RESULTS

For a detailed description of purification, crystallization and structure determination procedures refer to Materials and Methods section.

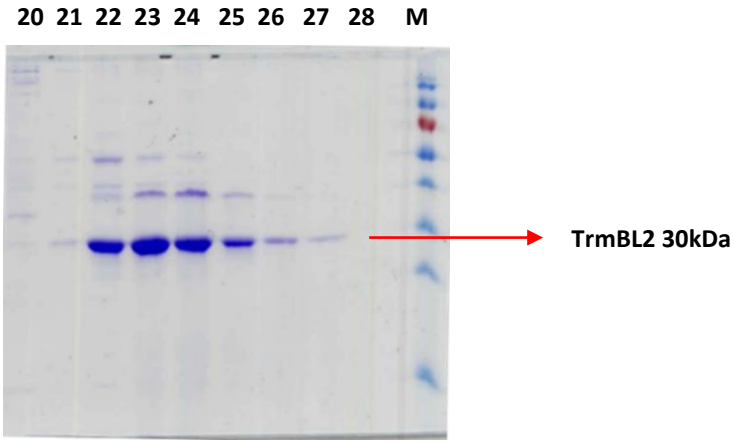
3.1 pfTrmBL2: Purification and crystallization

Nickel affinity purification of pfTrmBL2

As a first purification step, His-tagged pfTrmBL2 was passed over a 5ml Ni HisTrap column and after application of a linear Imidazole gradient eluted as a single peak at 70mM imidazole concentration. The reason for eluting at such a low imidazole concentration could be the presence of 1M NaCl in the buffers.



a



b

Figure 4.

a: Elution profile of pfTrmBl2 from a Ni HisTrap column.
X-axis: Volume; Y-axis: UV-absorbance

b: SDS-PAGE analysis of the peak fractions.

Gel filtration chromatography of pfTrmBL2

To obtain a homogenous pfTrmBL2, a prerequisite for successful crystallization, the protein was purified over a Superdex 200 gel filtration column. The main peak corresponding to fractions 16-18 was estimated to be a tetramer by comparing it with known molecular weight calibration markers.

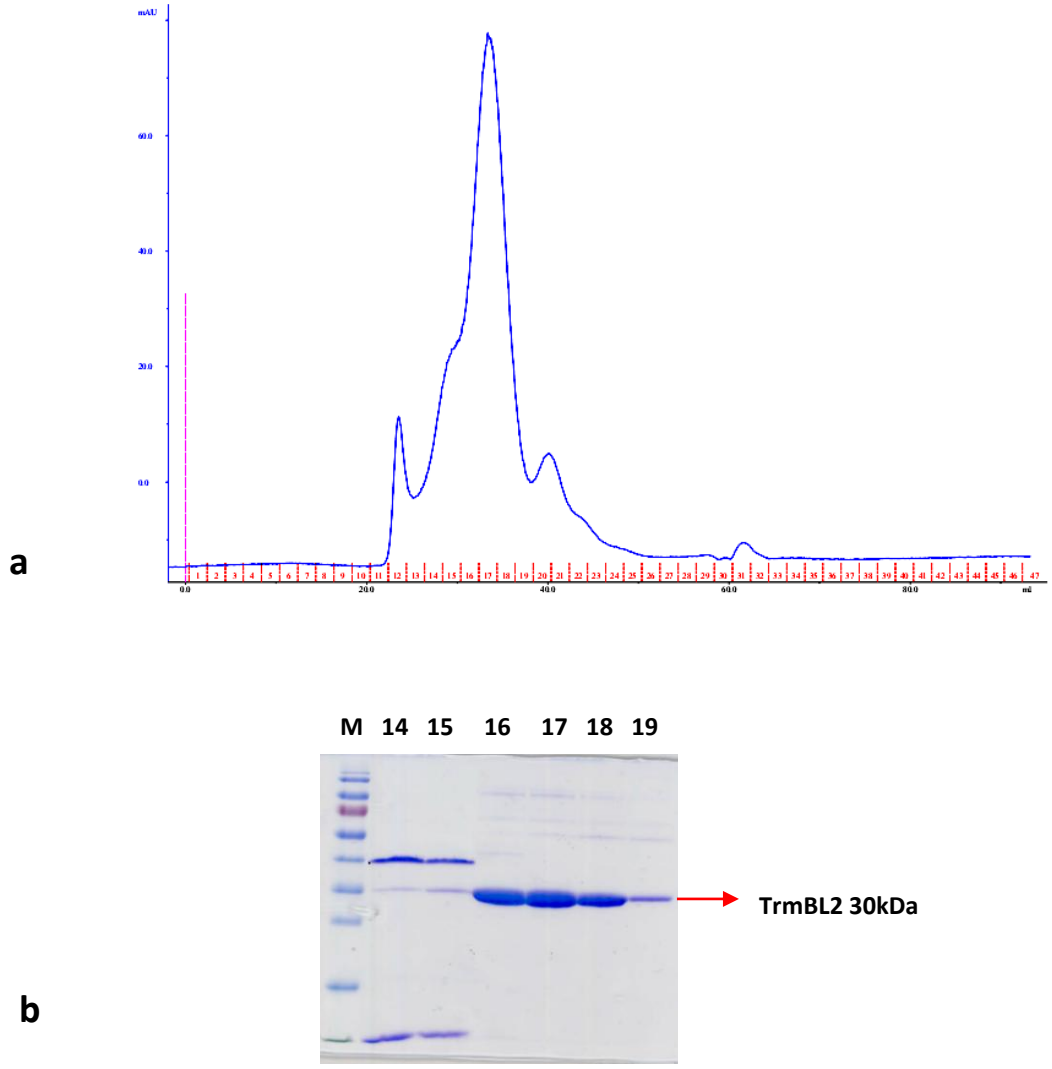


Figure 5.

a: Elution profile from a Superdex200 column and

b: SDS-PAGE analysis of the peak fractions.

Crystallization

Prior to crystallization, the protein was concentrated to 16mg/ml and after extensive fine screening of the initial hits best diffracting crystals with 2D plate morphology could be grown out of the following two conditions in Qiagen 15 well hanging drop plates by using a protein to reservoir ratio of 1:2:

0.2M $Ca(CH_3COO)_2$, 21% PEG 3350 and

0.2M $CaCl_2$, 21% PEG 3350

Crystals grew to a maximum size of 200µm after 1 week of incubation at 18°C.

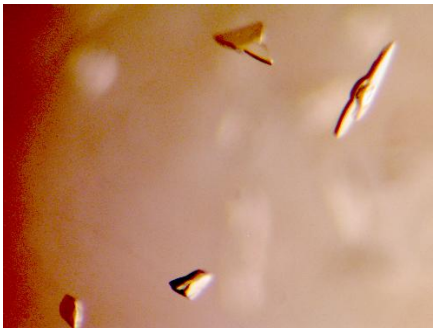


Figure 6. pfTrmBL2 crystals

.

3.2 ecTrmBL2: Purification and crystallization

Ion exchange chromatography of ecTrmBL2

The extreme thermostability of TrmBL2 allowed for the removal of most protein impurities by heating the cell lysate at 80°C for 25 minutes prior to ion exchange chromatography.

For Ion exchange chromatography, ecTrmBL2 was subjected to a linear gradient elution on a 30ml Q-Sepharose anion exchange column. The protein eluted as multiple peaks that were analysed on SDS-PAGE for selecting the most pure fractions which were determined to be eluting between 18% and 20% Buffer B (Figure 7).

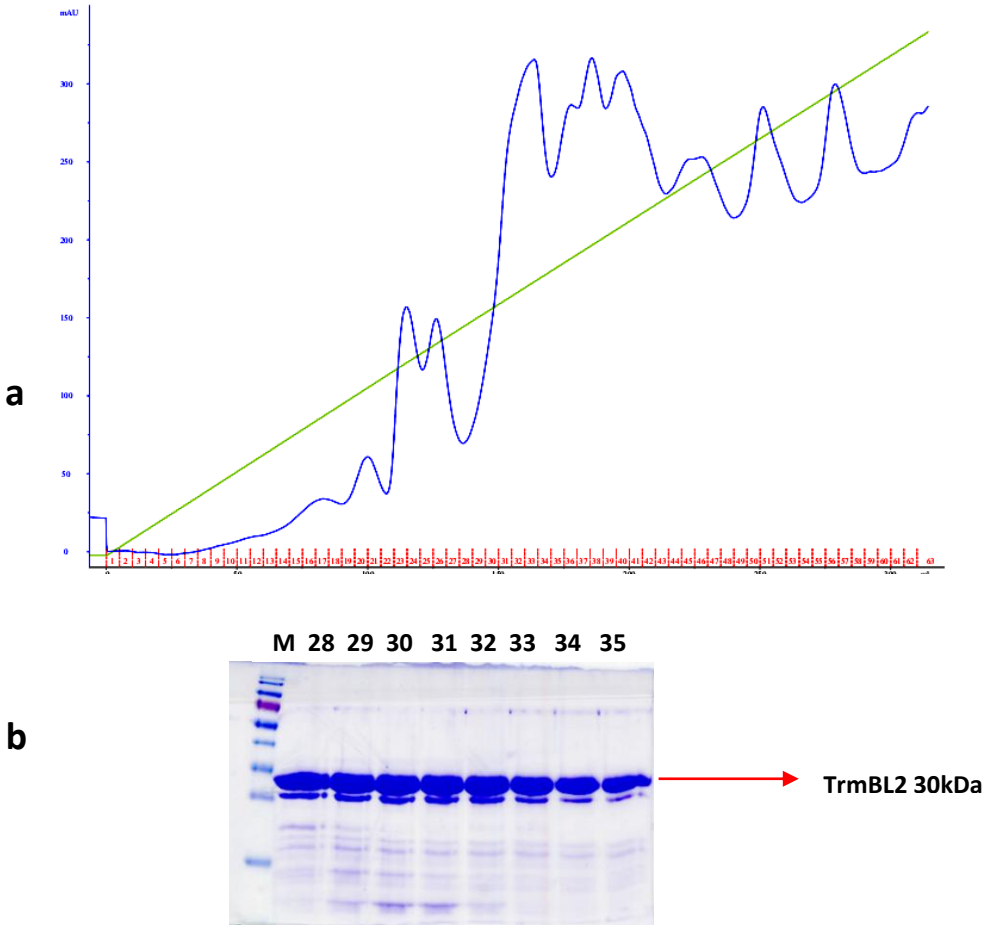


Figure 7.

a: Elution profile from a 30ml Q-Sepharose column.

b: SDS-PAGE analysis of fractions corresponding to Buffer B concentration of 15-20%.

Remark: Fractions were always chosen based on the salt concentration they eluted at and not on the UV absorbance. Although 85% of the protein elutes beyond 20% Buffer B as UV peaks of very high absorbance, all of it consists of higher oligomers and aggregates.

Gel filtration chromatography of ecTrmBL2

As a final polishing step, ecTrmBL2 was loaded onto a Superdex 200 column and eluted as a single sharp peak that corresponded to its dimeric form after comparison with known molecular weight calibration markers (Figure 8).

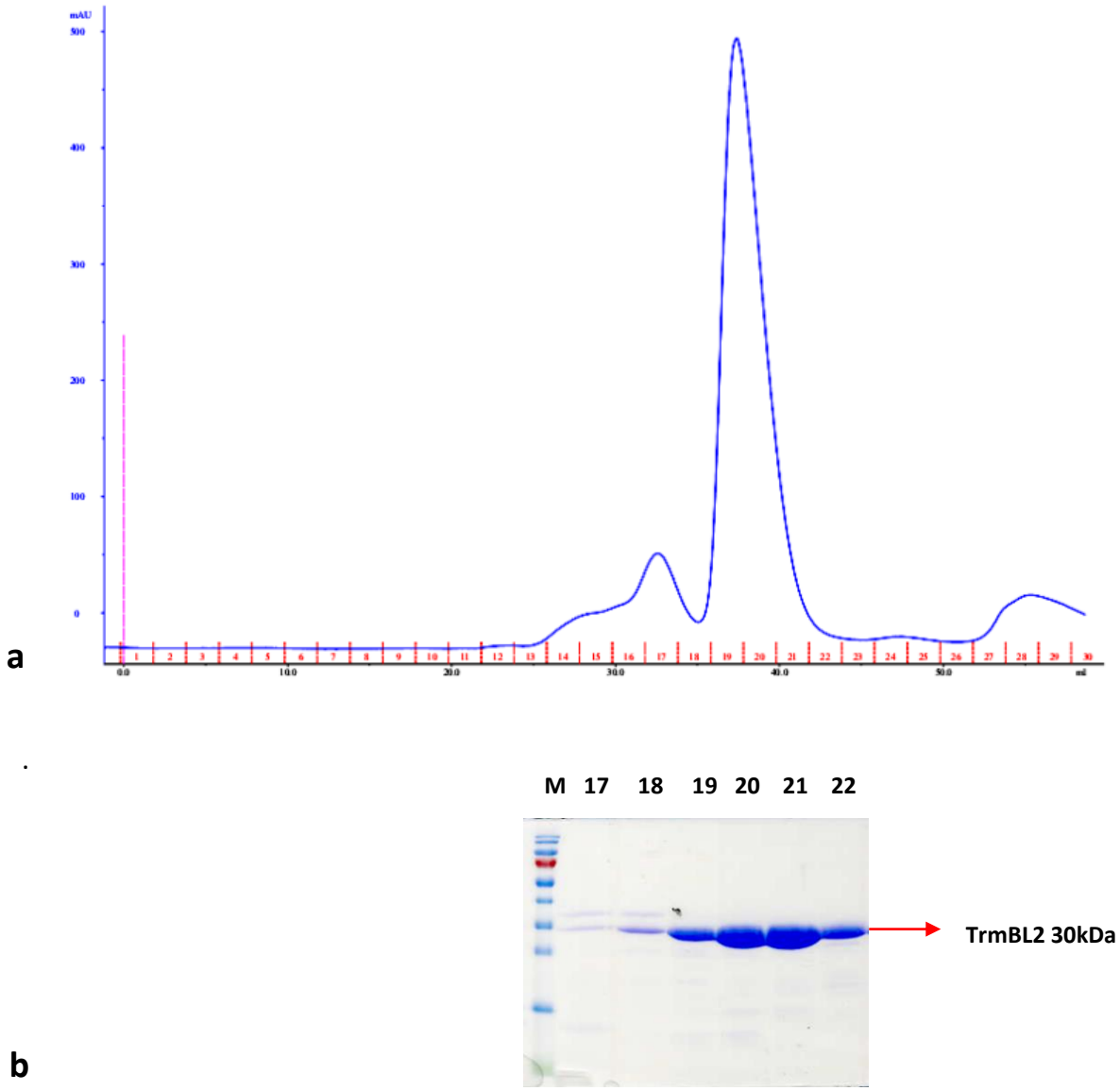


Figure 8.

a: Elution profile from a Superdex200 column.

b: SDS-PAGE analysis of peak fractions.

Crystallization

Prior to crystallization, the protein was concentrated to 10mg/ml and mixed with either TGM-17 or TGM-19 dsDNA in a 1:3.4 molar ratio. After extensive fine screening of the initial hits, best diffracting ecTrmBL2-TGM19 complex crystals appeared in the following condition:

60% (v/v) MPD and 0.1M HEPES pH 6.5

For ecTrmBL2-TGM17 complex, crystals were obtained in the following conditions:

0.1M HEPES pH 6.5 and 60% MPD

0.1M HEPES pH 7.5 and 60% MPD

0.1M imidazole HCl pH 8.0, 30% MPD and 10% PEG 4000

All the crystals grew as stacks of thin plates and reached a maximum size of 150µm in 4 weeks when grown at 18°C (Figure 9).

Figure 9. ecTrmBL2 crystals growing as stacked thin plates.

Structure solution

The structure of ecTrmBL2-TGM17 complex was solved by Se-SAD (Selenium-Single wavelength Anomalous Dispersion) and subsequent structures of ecTrmBL2-TGM19, ecTrmBL2-TGM17 and Apo-pfTrmBL2 were solved by Molecular Replacement.

3.3 Structure of the ecTrmBL2-TGM19 complex

Full length, wild type TrmBL2, expressed in *E.coli* crystallizes as a tetramer bound to a single copy of dsTGM19 DNA in the asymmetric unit. The same tetrameric assembly is found in ecTrmBL2-TGM17 complex and DNA-free pfTrmBL2 crystals. Table 2 details the data collection and refinement statistics.

The TrmBL2 tetramer is composed of two dimers, AC and BD, related by a twofold non-crystallographic symmetry, which also relates the two strands of the dsTGM19 DNA (Figure 10, a). Binding of TrmBL2 results in the bending of DNA which is clearly seen when the complex is viewed along a direction perpendicular to the DNA helical axis (Figure 10, b).

TrmBL2 shows a modular arrangement of domains with an N-terminal DNA-binding domain followed by a dimerization helix and a C-terminal domain.

The N-terminal domain of TrmBL2 (residues 1-74) belongs to the extended winged Helix Turn Helix (ewHTH) family of DNA binding proteins which are characterized by additional helices α0 and α4 at the N and C-terminus respectively, of the canonical winged Helix turn Helix (wHTH) domain [100, 127] (Figure 11, a).

An amphipathic dimerization helix (residues 78-109) contributes critically to dimer stability by forming an extensive hydrophobic interface with an oppositely arranged helix in a dimer (Figure 12, c).

Seven β-strands sandwiched between a long α-helix on one side and three smaller α-helices on the other, form the C-terminal domain (residues 125-264) (Figure 11, a).

Within a dimer, the N-terminal domains and the dimerization helices are related by a twofold NCS which does not extend to the whole dimer because of the different orientation of C-terminal domains. As a consequence of this asymmetry, one of the ewHTH domains in the dimer forms an extensive interface with the CTD of its dimeric partner (BSA 1383Å^2) whereas for the other ewHTH domain this interface is much smaller (BSA 851Å^2) (Figure 12, a). Because the two dimers are related by a twofold NCS, the same holds true for the other dimer. The overall buried surface areas for the two dimers AC and BD are 2844Å^2 and 2804Å^2 respectively.

In the TrmBL2 tetramer, the ewHTH α3 helices of the two dimers are shifted by two base pairs with respect to each other which results in a rotation of about 80° (with α3 helices as the

reference) of the two dimers with respect to an approximate DNA helical axis (Figure 10, c). This offset places two of the ewHTH domains (A and B) at the distal end of the tetramer and the other two (C and D) occupy proximal positons (Figure 12, a). This arrangement of the dimers on the dsDNA is promoted by the formation of an extensive interaction interface between the C-terminal domains of Chains A and D and their NCS related partners, Chains B and C. The strength of these interactions is reflected by the large buried surface areas (1385Å^2 for AD and 1371 Å^2 for BC) and extensive polar interactions (Figure 13).

Surface potential representation of TrmBL2 tetramer (Figure 14) shows that the two ewHTH domains of a dimer have differently charged environments, a consequence of the dimer asymmetry. In addition to the hydrophobic interactions between the oppositely arranged dimerization helices, the opposite surface charges of the ewHTH domains and the adjacent CTDs also contribute to dimer stability. Moreover, the differently arranged CTDs show opposite surface potentials and contribute to tetramer stability.

	ecTrmBL2-TGM19	ecTrmBL2-TGM17	ecTrmBL2-TGM17	Apo pfTrmBL2	ecTrmBL2-TGM17 Peak
Data collection					
Wavelength (Å)	1.00	1.00	1.00	1.00	0.97942
Resolution range (Å)	50-2.5 (2.59-2.5)*	50-2.4 (2.49-2.4)*	50-3.2 (3.31-3.2)*	50-3.0 (3.11-3.0)*	50-2.7 (2.78-2.7)*
Space group	$P2_1$	$P2_1$	$P2_1\,2_1\,2_1$	$P2\,2_1\,2_1$	$P2_1$
Unit cell parameters					
a (Å)	83.69	95.79	58.76	63.51	97.18
b (Å)	105.79	58.67	154.37	83.29	59.50
c (Å)	93.23	143.07	176.12	235.14	143.87
β (°)	96.75	92.82	90	90	92.45
Multiplicity	3.4 (3.4)	6.5 (5.5)	13.1 (12.8)	6.3 (6.6)	12.4 (10.6)
Completeness (%)	100 (100)	99 (96)	100 (98)	99 (100)	99 (88)
Mean I/sigma	7.84 (0.41)	8.18 (0.69)	7.8 (0.99)	9.96 (0.6)	9.54 (0.62)
R_{meas}	0.178 (3.63)	0.15 (2.29)	0.39 (2.96)	0.1346 (2.89)	0.2547 (3.55)
$CC_{1/2}$	0.995 (0.14)	0.996 (0.369)	0.995 (0.408)	0.999 (0.312)	0.996 (0.234)
Refinement					
Resolution range (Å)	50-2.5 (2.53-2.5)*	50-2.4 (2.44-2.4)*	50-3.2 (3.26-3.2)*	50-3.0 (3.06-3.0)*	-
Number of reflections	55705	62270	27265	25446	-
R_{work}	0.2242	0.2336	0.2546	0.2326	-
R_{free}	0.2720	0.2828	0.2988	0.2888	-
No. of amino acid residues	1052	1045	1039	1038	-
No. of nucleic acid residues	50	42	42	-	-
No. of heteroatoms	66	28	6	2	-
RMS bonds (Å)	0.006	0.002	0.005	0.005	-
RMS angles (°)	0.99	0.55	0.63	1.06	-
Ramachandran favoured (%)	94.5	96.6	95.7	94	-
Ramachandran allowed (%)	5	3.3	4	5.6	-
Ramachandran outliers (%)	0.5	0.1	0.3	0.4	-
Average B-factor (Å ^{2})					-
Protein	89	79	82	134	-
DNA	98	96	103	-	-

*** Values in parenthesis are for the highest resolution shell.**

Table 2. Data collection and refinement statistics

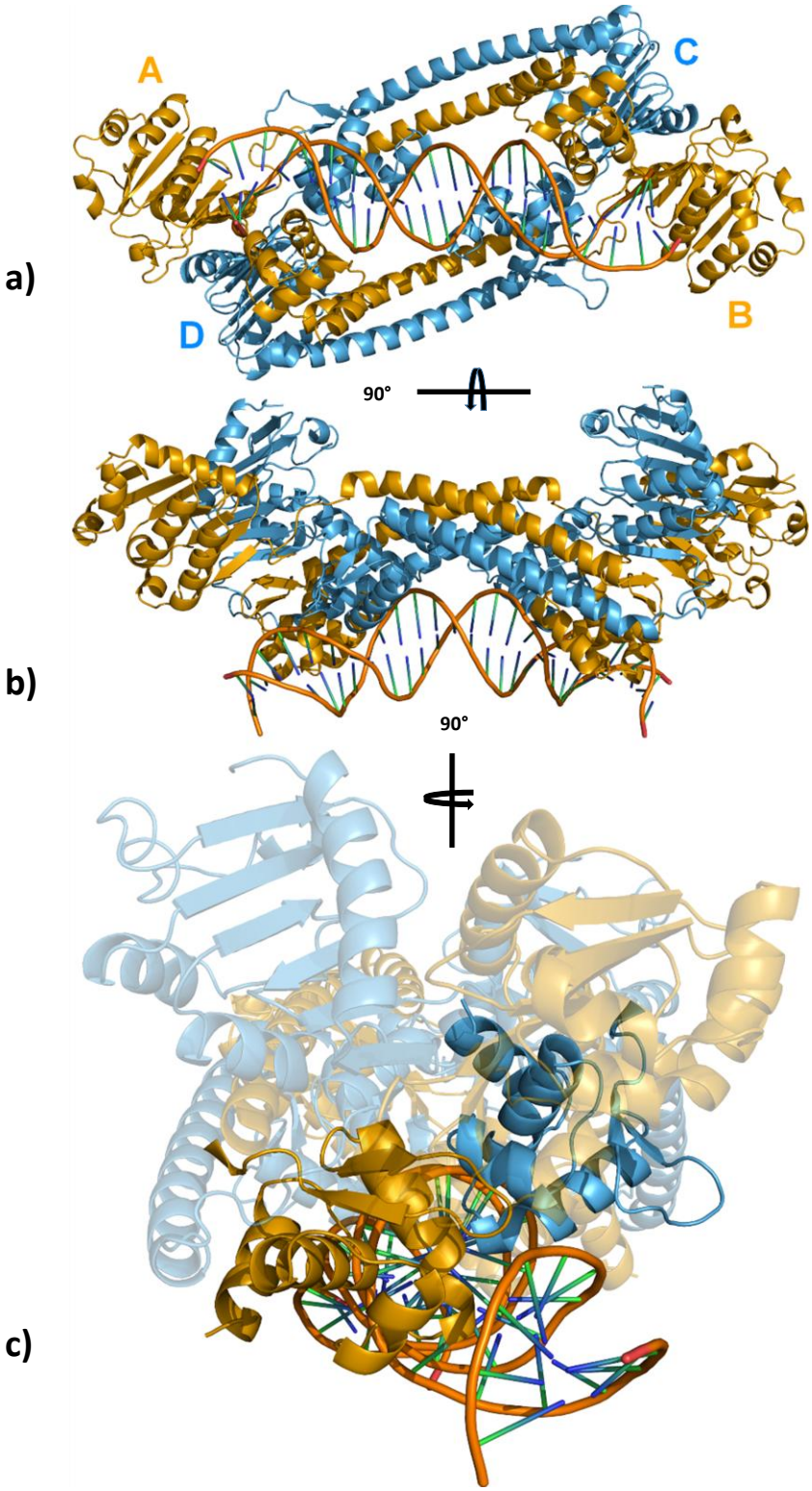


Figure 10. ecTrmBL2 tetramer in cartoon representation with bound TGM19 DNA. NCS related chains are shown in the same colour.

a: View along the twofold symmetry axis. From top to bottom: Chains C (blue), A (orange), B (orange) and D (blue). Chains A and C form one dimer which is related by a twofold NCS to the other dimer formed by chains B and D. The two ewHTH domains of chains A and B are close to the tetramer

ends and are termed “distal” hereafter. The others from chains C and D are termed “proximal” hereafter.

b: View perpendicular to the bent DNA with twofold symmetry axis oriented vertically.

c: View on one end of the complex along the bent DNA with the twofold symmetry axis oriented vertically. This panel is enlarged compared to the other two and all except the distal ewHTH domain of chain A and the proximal ewHTH domain of chain D are set in a higher transparency to allow a better view of the two ewHTH domains.

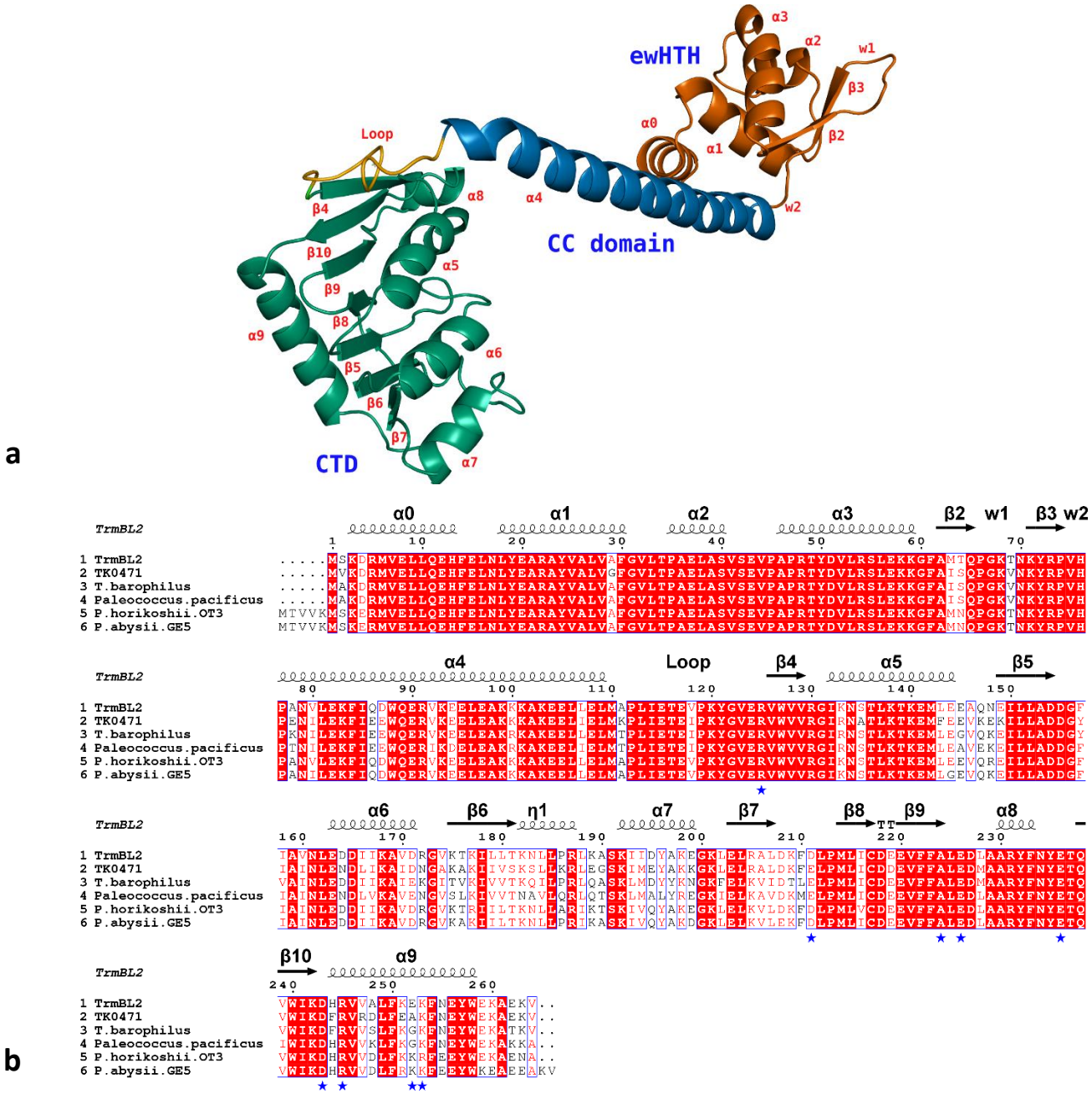


Figure 11.

a: TrmBL2 monomer depicted as cartoon representation of chain C. Secondary structural elements are labelled. ewHTH, CC (Coiled coil) domain and C-terminal domain are labelled and coloured in red, green and blue respectively. The flexible loop connecting the CC helix to the CTD is coloured in orange. Secondary structural elements of the ewHTH domain are numbered to match the conventional labelling of wHTH domains. Secondary structural elements are assigned according to DSSP [146].

b: Sequence alignment of TrmBL2 homologs in hyperthermophilic archaea labelled with secondary structure elements from the TrmBL2 structure. Strictly conserved residues are highlighted in red. Turns, 3_{10} helices and wings are denoted by TT, η and w respectively. Residues forming polar contacts at the interface between the CTDs are labelled with a blue asterisk. The alignment was generated using the ENDscript server [147].

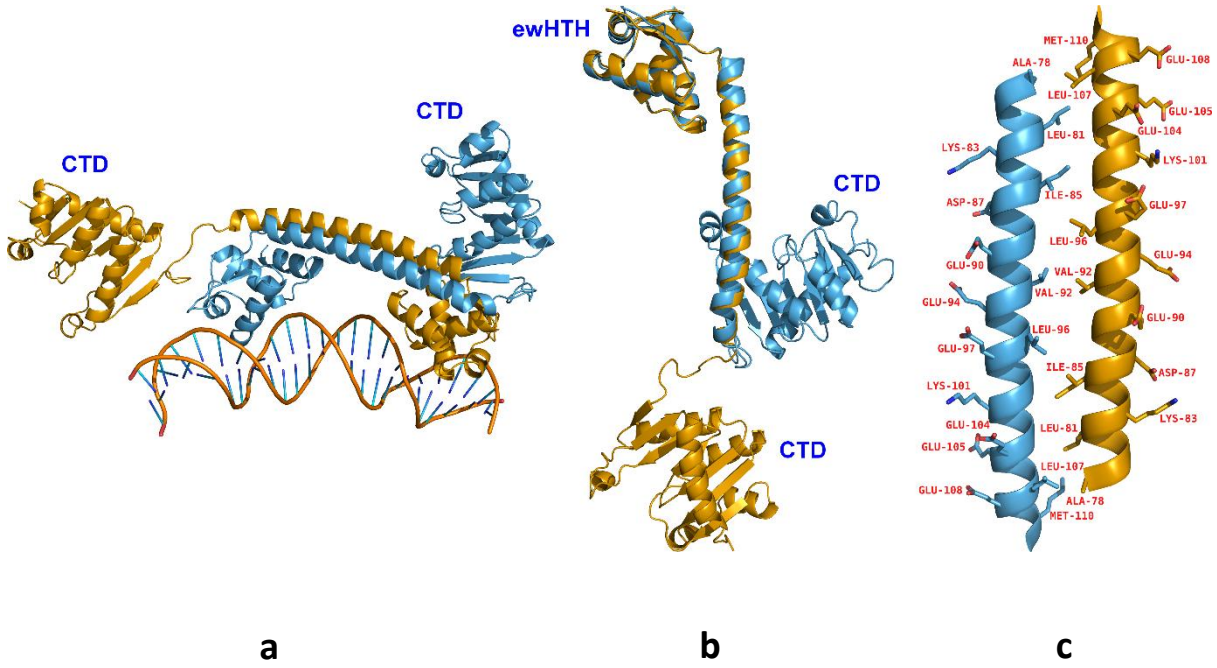


Figure 12.

a: Broken twofold symmetry of TrmBL2 dimer. View of the TrmBL2 dimer perpendicular to the symmetry axis relating the ewHTH domains and the dimerization helices. The symmetry is broken due to different conformations of the CTDs. While the distal ewHTH domain (orange) forms a large buried surface with the CTD (blue), the same is not the case with the proximal ewHTH domain (blue).

b: Superposition of the two TrmBL2 monomers of the dimer. The difference in the conformation of the CTDs are obvious.

c: The coiled-coil formed by two oppositely arranged α4 helices in the TrmBL2 dimer. Solvent exposed and core hydrophobic residues are labelled. The helices are almost parallel with a crossing angle of 178°.

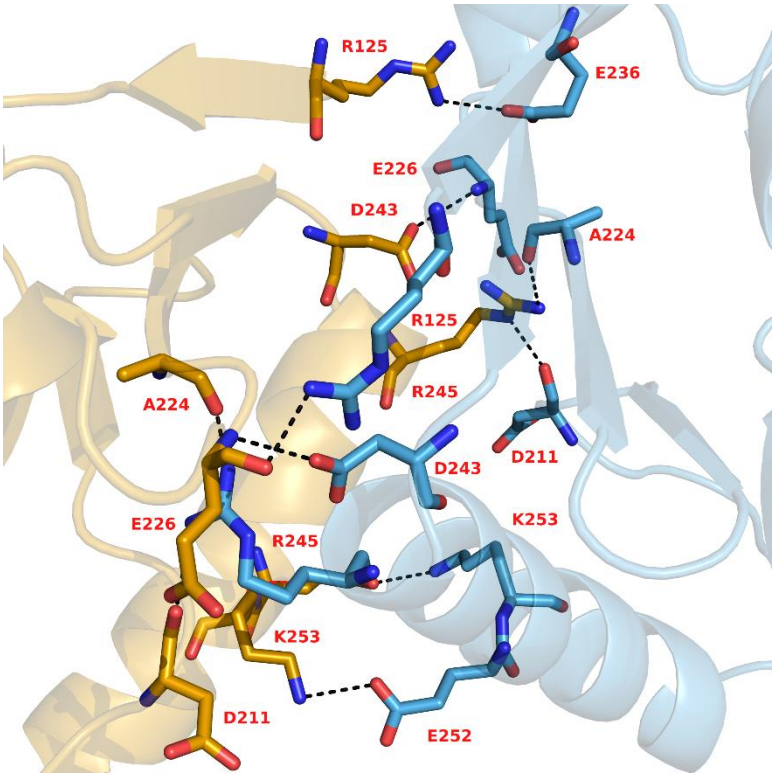


Figure 13. Interface of the CTDs from Chains A (orange) and D (blue) with densely spaced polar interactions. Residues involved in polar contacts are labelled. Because of the twofold symmetry that relates the CTDs, the same interface is formed at the other end of the tetramer (not shown).

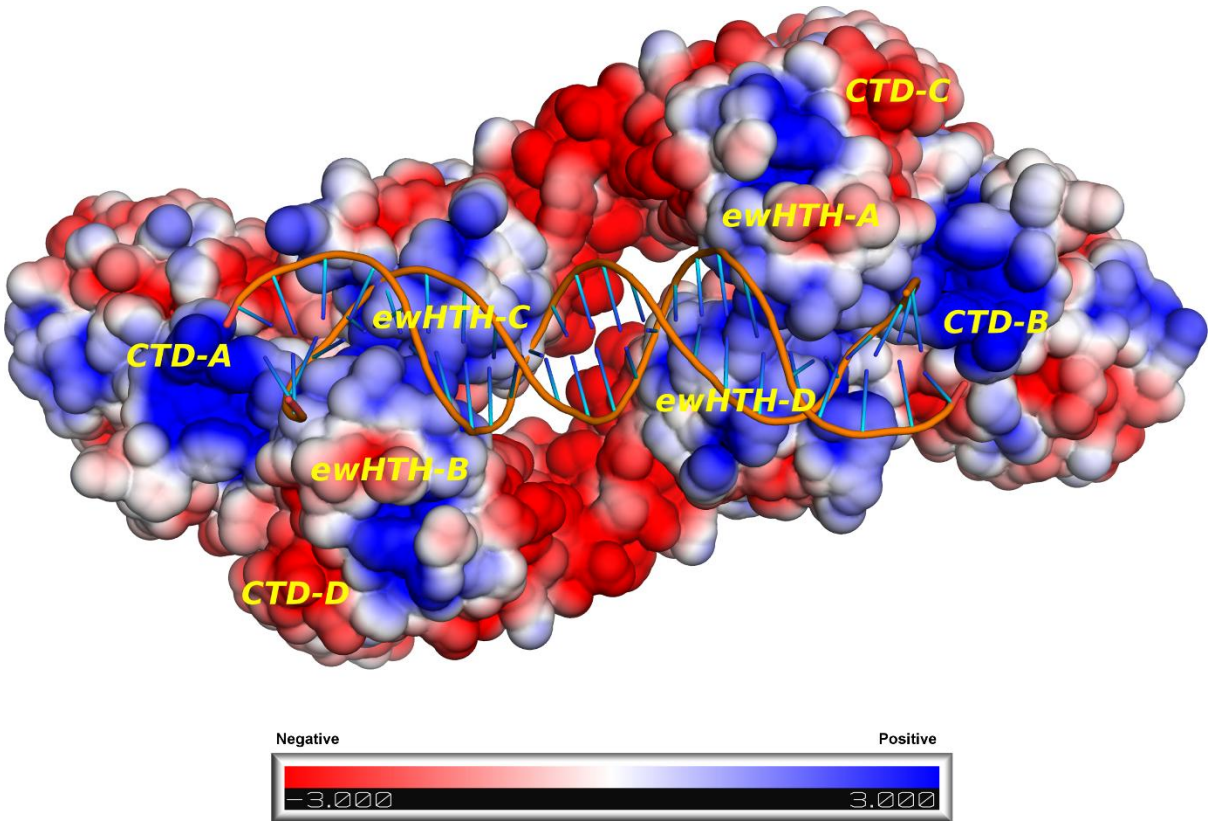


Figure 14. Surface potential representation of TrmBL2 tetramer bound to TGM19 DNA shown in cartoon representation. The tetramer is oriented as in Figure 1, top panel. The two ewHTH domains of a dimer display different electrostatic surfaces as they are rotated by 80° with respect to each other. The differently charged environments of the two ewHTH domains which are a consequence of dimer asymmetry can be seen in the figure. Electrostatic surface potential was calculated by APBS [148-150].

3.4 Refinement of the ecTrmBL2-TGM19 complex

When modelling the structure of the TGM19 DNA bound to ecTrmBL2, there was density for 25 base pairs and the density for the nucleobases was ambiguous at most of the positions prohibiting clear assignment as one of the four bases. Apart from that, the density for the phosphates and the deoxyribose rings was clearly seen. Similarly in both TGM17 structures a density for 21 base pairs was seen. Moreover this extra three base pair density in case of TGM19 and two-base pair density in case of TGM17 at both DNA ends was significantly weaker but because these parts of the model showed up as significant difference map peaks during subsequent cycles of refinement, it was decided to place extra base pairs at these positions and to assume that the DNA is bound to TrmBL2 along the ewHTH domains in various positions. To take this into account, several approaches were tried during the refinement. A grouped occupancy refinement with 3 copies of TGM-19, each positioned along the visible density of the phosphate and ribose backbone with an offset of three base pairs against each other (Figure 15 a) resulted in an increase in R_{free} by 5 % and was therefore discarded. To reduce the number of refinement parameters, the parsimonious model shown in Figure 15 b was constructed. This model has the known 19bp segment as its centerpiece, but duplicates the first and last three base pairs, thus obtaining 25 base pairs. To account for the fact that the total number of base pairs must be 19, the occupancy of the first and last three residues was adjusted to 1/3, base pairs 4-6 and 20-22 to 2/3 and base pairs 7-19 to 1.0. We conclude from inspection of the electron density map and the refinement that in the ecTrmBL2 tetramer the TGM19 DNA is bent and likely bound at least at 3 different positions along the four ewHTH domains. Likewise, for the TGM 17 there are also three binding positions. This suggests that the four ewHTH domains offer at least three binding sites of roughly the same affinity to the dsTGM DNA which are approximately related to each other by screw axis operations corresponding to a shift of three base pairs in case of TGM19 and two base pairs in case of TGM17.

3.5 TrmBL2-DNA interactions

As with the case for most non-specific DNA binding proteins, the majority of TrmBL2 interactions with DNA are confined to the phosphoribosyl backbone. From the TrmBL2-TGM19 interaction scheme (Figure 15 c), it can be seen that the proximal ewHTH domains

present more residues for interaction with the DNA than the distal ones. The extra residues presented by the proximal ewHTH domains are R54 from α3 and N70 from w1.

Binding of the distal ewHTH domains causes distortion of the major grooves with the result that the surface area which the α3 helices from these domains bury with the DNA is smaller (238Å^2) as compared to the one buried by α3 helices of the proximal ewHTH domains (441Å^2) [151].

The majority of HTH proteins which bind DNA in the major groove form sequence specific contacts with the dsDNA by polar interactions of the residues from the recognition helix α3 with the nucleobases. In the case of TrmBL2, the only residues from the recognition helix which interact with the nucleobases are R48 from the distal ewHTH domains (Chains A and B) and P47 from all four ewHTH domains with the latter being a nonpolar residue (Figure 16).

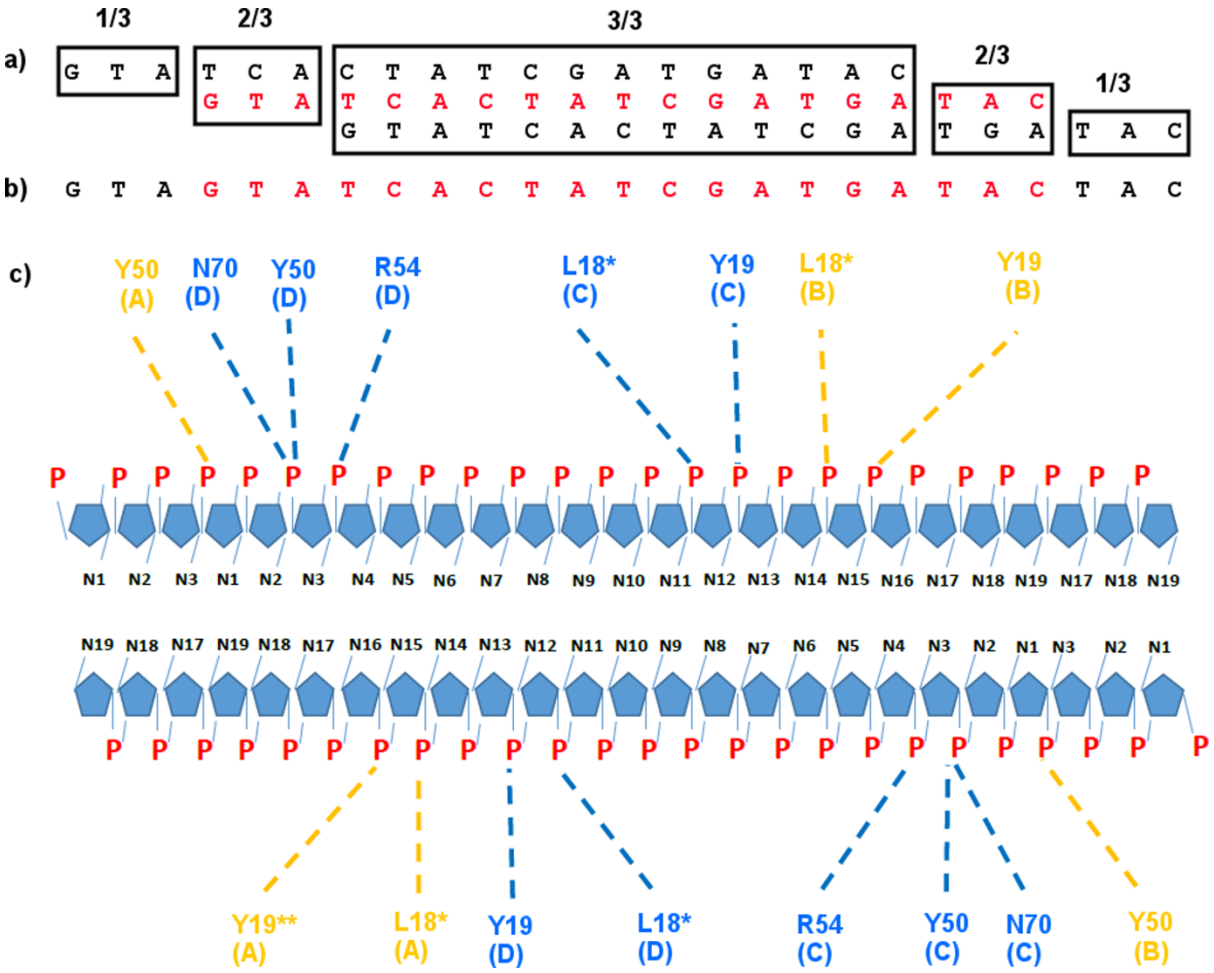


Figure 15. Occupancy refinement and TrmBL2-TGM19 interactions.

The TGM sequence is shown in red. Amino acids interacting with the sugar-phosphate backbone are shown in the same colour as the chains they belong to (Figure 1), which are indicated below each residue.

a) Occupancy refinement with three copies of TGM19. The sequences above and below the TGM sequence (coloured in red) are shifted by 3 base pairs at the 5' and 3' ends respectively. The occupancies of the base pairs are shown on top. Grouped refinement with 3 copies of TGM19 sequence resulted in an increase in R_{free}.
b) 25 bp sequence which was fitted in the density. The TGM19 sequence (red) is shown in the middle with the 3 base pairs extensions at the 5' and 3' ends shown in black. Refinement got better by placing only one nucleotide at each position with adjusted occupancy along the 25 bp density stretch. The occupancy of the residues is adjusted so that the total number of base pairs adds to 19. It is important to note that two base pairs with equal probability occupy positions 4-6 and 20-22, and 3 base pairs with equal probability occupy positions 7-19.
c) Symmetric interactions of the amino acids from the four ewHTH domains with the sugar-phosphate backbone of TGM19. Residues interacting with the DNA backbone within 3.5 Å are shown.
* Main chain contacts with the TGM19 sugar-phosphate backbone
** 3.67Å distance from the TGM19 sugar-phosphate backbone

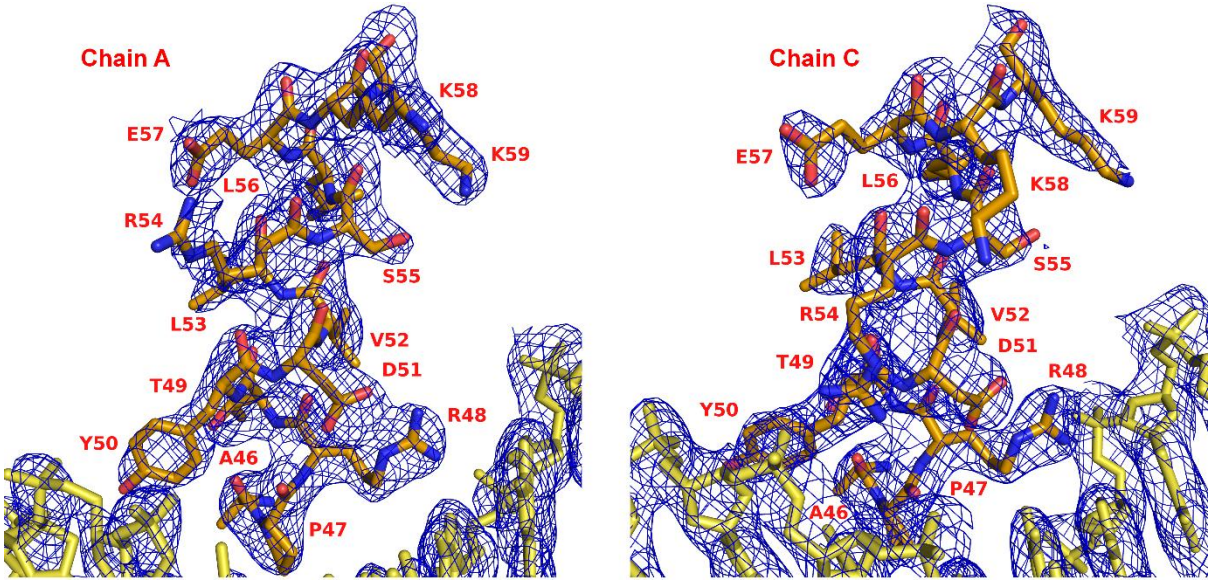


Figure 16. Details of the interaction of the recognition helices with TGM19 in the AC dimer. The $2F_o$-F_c electron density map contoured at 1σ is shown with the amino acids and DNA (yellow) in stick representation. R54 can be seen shifted towards the DNA backbone in case of chain C whereas in case of Chain A it points away from the backbone.

3.6 Structure of DNA-free pfTrmBL2

The DNA-free pfTrmBL2 structure was solved by Molecular Replacement using ecTrmBL2 as the search model and revealed a tetramer in the asymmetric unit. Superposition of both structures was possible with an RMSD of 2.8 Å (calculated by SSM superpose in Coot) and shows only subtle changes upon DNA binding.

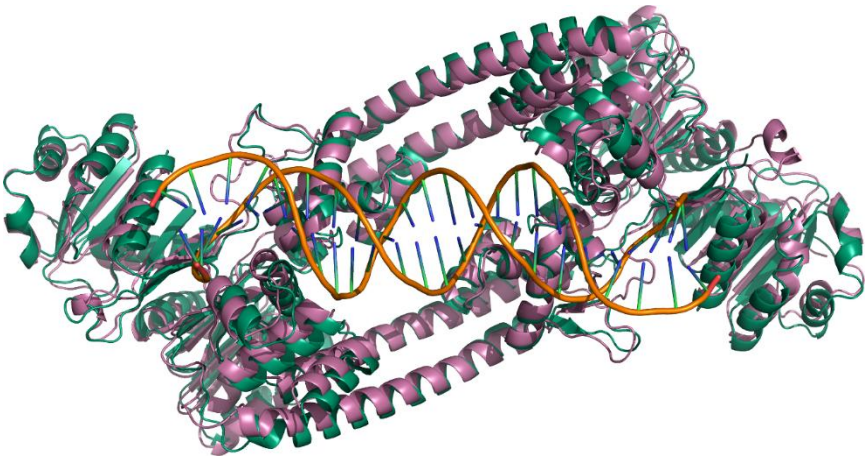


Figure 17. Cartoon representation of the superposition of ecTrmBL2-TGM19 complex (magenta) and DNA-free pfTrmBL2 (green).

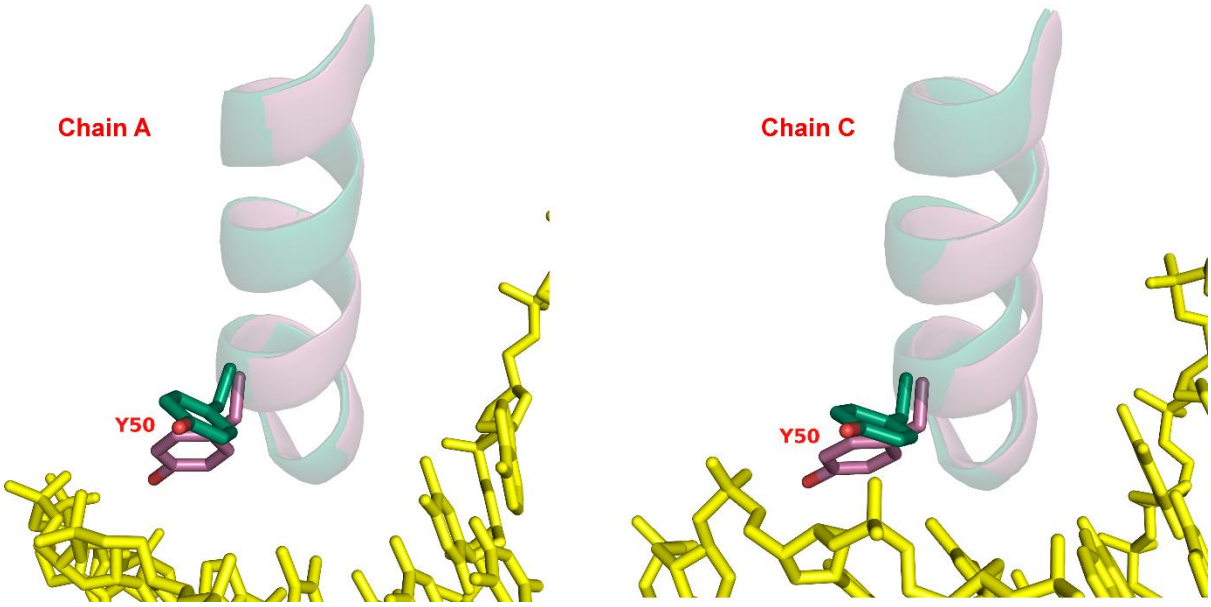


Figure 18. Superposition of H3 helices of DNA-bound (magenta) and DNA-free (blue) structures. The Y50 shifts towards the DNA backbone in case of DNA-bound TrmBL2. DNA is coloured yellow and shown in stick representation.

PDB accession numbers

The structures have been deposited in the PDB under the following accession numbers:

5BOX (ecTrmBL2-TGM19 complex)

5BPD (ecTrmBL2-TGM17 complex, 2.4Å resolution)

5BPI (ecTrmBL2-TGM17 complex, 3.2Å resolution)

5BQT (DNA free pfTrmBL2)

4. DISCUSSION

Challenges in expression, purification and crystallization of TrmBL2

The protein expression levels of ecTrmBL2 from pET24d construct expressed in BL21 (DE3) strain were very high and close to 5mg of pure protein could be obtained from 6g of cell pellet. A major problem however was that a fraction of ecTrmBL2 was cleaved at the N-terminal end during expression in *E.coli* BL21 (DE3) cells. Changing the expression strain to SF120 which lacks multiple proteases or BL21 (DE3) STAR which has a mutation in the RNaseE gene (*rne131*) which improves mRNA stability did not solve the problem of in vivo TrmBL2 cleavage. It was observed that the in vivo cleavage started right after induction and therefore reducing the post induction time did not help. Decreasing the growth temperature to 30°C or 18°C also didn't prevent cleavage. Induction with lower concentration of IPTG like 0.5mM or 0.1mM did not improve the results either. The difference between the cleaved and the full length ecTrmBL2 was only 1-2kDa, which made a separation difficult. For the heat treatment step, the NaCl and the glycerol concentrations in the lysis buffer needed to be optimized and it was found that 20% Glycerol and a minimum of 150mM NaCl was required to prevent precipitation.

Separation was tried over a Heparin column and also over a Q-Sepharose ion exchange column. Purification over a 30ml Q-Sepharose column yielded the best results even though the protein eluted as multiple peaks which needed to be analysed on an SDS-PAGE to select for fractions with minimum amount of impurities. The reason for the protein eluting as multiple peaks was the prior heating of the lysate at 80°C which would in addition to the full length dimeric TrmBL2 result in the formation of higher oligomers and small aggregates. These multiple oligomeric forms of TrmBL2 would carry different surface charges and elute as multiple peaks over the ion-exchange column.

About 85% of the protein eluting from the ion-exchange column could not be used further as it was not possible to separate the cleaved TrmBL2 from the full length protein. This loss of protein could not be overcome by purification with methods like Ni-affinity chromatography using various C-terminal or N-terminal His-tagged versions of TrmBL2. The cleaved product could also not be separated by using 1M NaCl during gel filtration chromatography which indicated a strong hydrophobic interaction between the two. Addition of mild detergents like

Tween-20 to the purification buffers also did not improve the results. However the remaining ~15% of the protein could be purified and concentrated to 10mg/ml.

ecTrmBL2 crystals were obtained under many conditions but only a few conditions yielded diffraction quality crystals. However, the major problem was that the crystals grew as stacks of thin plates which made their dissociation into single crystals difficult. Various approaches to obtain single crystals like streak seeding, microseeding, macroseeding, growth under oil, changing the precipitant and/or protein concentrations, incubating the crystals at various temperatures (4°C, 8°C or 18°C), growth in sitting or hanging drop plates, extensive additive screening and/or transferring the crystals to different Cryo solutions did not work or resulted in growth of low diffraction quality crystals. Diffraction data was therefore collected from these thin plate stacks by focussing the beam at a very narrow region of a carefully oriented crystal targeting areas where the thickness of the stack was least.

The absence of a model that could be used for Molecular Replacement made it necessary to solve the phases by heavy atom derivatization. Soaking of the crystals with heavy atoms like Pt, Cd or Hg resulted in fracturing and dissolution of the crystals. Various heavy atom compounds and concentrations were tried and it was found that crystals could be derivatized by soaking in solutions containing 5-10 mM K_2PtCl_4 without significantly affecting their morphology and diffraction quality. However, the heavy atom data collected from these crystals lacked a sufficiently strong anomalous signal required to solve the phases. It was therefore decided to label the protein with Selenomethionine. Owing to the low expression rates of Selenomethionine labelled proteins, it was necessary to start purification from a cell pellet obtained from a minimum of 10L culture. After purification and concentration, SeMet labelled crystals of the ecTrmBL2-TGM17 complex could be reproduced in the same conditions used for growth of native crystals. After screening hundreds of ecTrmBL2-TGM17 SeMet crystals, phases were finally obtained from a single crystal.

TrmBL2 binds DNA in a non-specific manner

TrmBL2 was discovered together with TrmB and TrmBL1, both of which bind specific DNA sequences. Earlier studies to determine the role of TrmBL2 found that it binds the TGM containing promoters but not exclusively [152, 153]. Attempts to crystallize TrmBL2 with different lengths of TGM sequence were prompted by this finding.

Recently, a novel function of TK0471, the *T. Kodakarensis* homologue of TrmBL2 has come to light. Apart from acting as a global transcriptional regulator, it was observed that TK0471 associates non-specifically with the chromatin forming thick fibrous structures. These thick fibrous structures appear only during the log phase of the archaeal growth suggesting that TK0471 exerts its effects on the chromatin in growth phase dependent manner. Search for a consensus TK0471/TrmBL2 binding site did not yield any results suggesting that it binds DNA in a largely non-specific manner [34].

A more recent study investigating the role of TrmBL2 in archaeal chromosome organization shows that it counteracts the role of histones in DNA packaging and actively competes with them for DNA binding. The study also shows that TrmBL2 exhibits concentration dependent high and low affinity DNA binding modes and the affinity for DNA is affected by the salt concentration of the medium [154].

The ecTrmBL2-TGM complex structure which shows a superposition of three binding modes, a likely consequence of non-specific binding is in agreement quite well with the above findings.

Structural similarities with TrmB

First discovered in Thermococcus litoralis and Pyrococcus furiosus, TrmB is a transcriptional regulator which controls the transcription of genes encoding the ABC transporter for trehalose and maltose, the TM system, with maltose and trehalose as inducers. In addition it also controls the expression of the ABC transporter for maltodextrins, the MD system, with maltotriose and sucrose as inducers [155]. TrmBL2 bears only 23% sequence identity with TrmB but has a strikingly similar three dimensional structure and modular organization of domains [156]. In one of the two TrmBL2 chains (Chain C in AC dimer and Chain D in BD dimer) the ewHTH and the C-terminal domain show a “trans” arrangement with respect to the

dimerization helix, similar to the one found in TrmB (Figure 19). The ewHTH domains of the two proteins superimpose with an RMSD of 1.83Å, the dimerization helices with 0.82Å and the C-terminal domains with 2.07Å (Calculated by SSM Superpose in Coot).

Apart from these similarities, the two proteins differ in many ways. TrmB crystallizes as a monomer in the asymmetric unit whereas TrmBL2 crystallizes as a tetramer. Unlike TrmBL2, a perfect twofold crystallographic symmetry relates the two TrmB monomers in the crystallographic dimer. The dimerization helix and the loop connecting it to the C-terminal domain is longer in case of TrmBL2. A prominent difference among the two proteins is the lack of the sugar binding domain in TrmBL2 found at the C-terminal end of the effector binding domain (EBD) of TrmB (Figure 19). Moreover, the extensive interaction interface which one of the ewHTH domains of the TrmBL2 dimer forms with the C-terminal domain of the oppositely arranged monomer is not seen in case of TrmB (Figure 20).

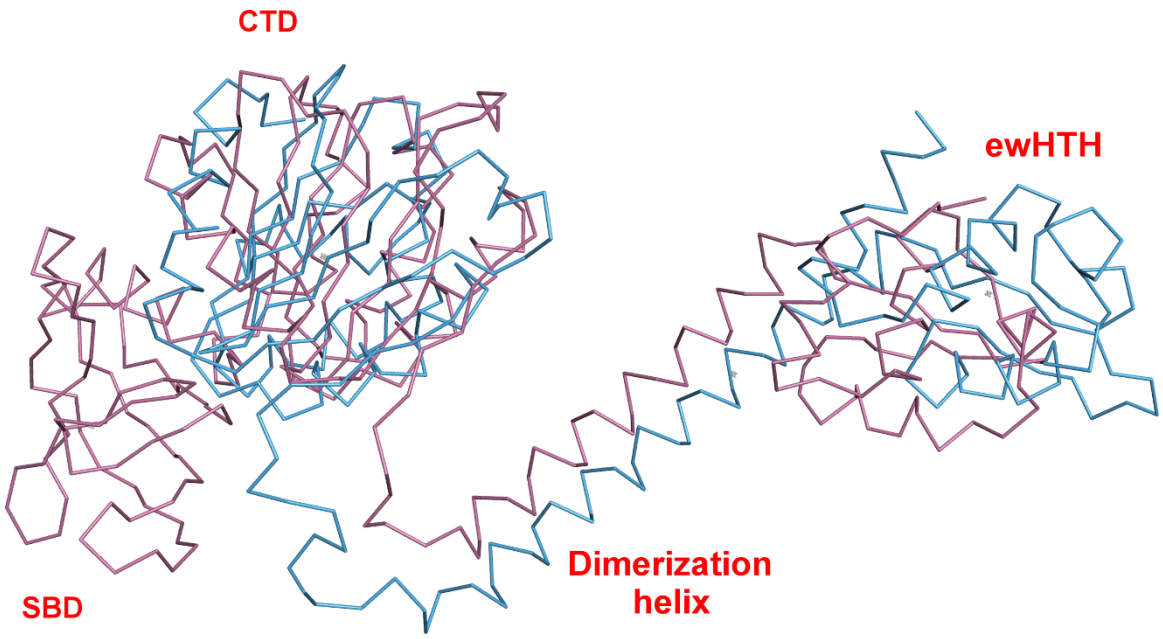


Figure 19. Superposition of C_α backbone of TrmBL2 (blue) on TrmB (purple). The CTD of this chain of TrmBL2 and that of TrmB show a similar “trans” arrangement to the ewHTH domain relative to the dimerization helix. TrmBL2 lacks the sugar binding domain (SBD) present at the C-terminal end of TrmB Effector binding domain (EBD) [156].

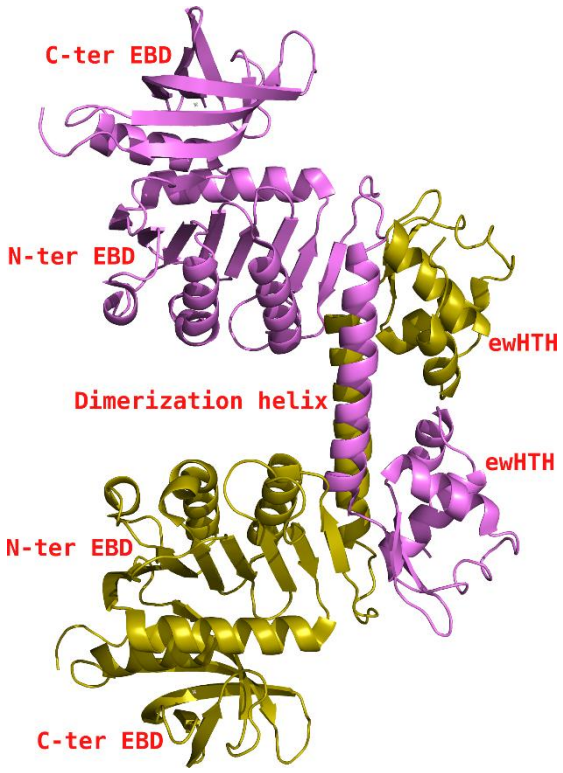


Figure 20. TrmB crystallographic dimer. PDB: 3QPH

Consequences of dimer asymmetry

A unique feature of the TrmBL2 structure is the different orientation of the CTDs within a dimer, an arrangement not seen in any of the repressor proteins reported so far. The 13-residue loop connecting the dimerization helix to the C-terminal domain allows for the flexibility of the C-terminal domains to adopt different conformations. As a result of this asymmetry, the C-terminal domains of the two dimers bury large surfaces which serve to primarily strengthen the tetramer stability.

A second result of this dimer asymmetry is that two of the ewHTH domains in the tetramer, the distal ones, form extensive interactions with the C-terminal domains of the dimeric partners (Figure 10, a) whereas the same is not true for the proximal ewHTH domains. In a biological context, this means that a conformational change in the C-terminal domains which could arise from protein-protein interactions would be transmitted more strongly to the distal ewHTH domains. This in turn could suggest a regulatory role for the distal ewHTH domains. It is important to mention that in the absence of any interacting partners for TrmBL2 such a scenario is essentially theoretical although a role of TrmBL2 in preventing the access of other proteins to DNA has earlier been suggested [34].

Comparison with TtgV structure

Although structures of a large number of dimeric bacterial and archaeal repressors are available, the same is not the case for complete tetrameric assemblies. TtgV represents one example of a bacterial repressor whose complete tetrameric structure, alone and in complex with DNA, has been solved (Figure 21) [157]. Upon DNA binding the N-terminal domains in TtgV undergo significant conformational changes with respect to the C-terminal domains. In contrast, TrmBL2 has very similar DNA-bound and DNA-free structures. The way the four N-terminal DNA binding domains are arranged on the DNA is also different between TtgV and TrmBL2. Whereas in case of TtgV the N-terminal domains bind only to once face of the DNA, in TrmBL2, owing to their staggered arrangement, the N-terminal domains make a larger fraction of the DNA inaccessible to macromolecules. On the other hand the 19bp dsDNA in complex with TrmBL2 is more accessible to small molecules like water than the 42bp dsDNA in complex with TtgV, as in both complexes, calculations show that 20% of DNA surface becomes inaccessible to solvent molecules upon complex formation [158]. This is quite in agreement with the fact that TtgV is a specific binder and therefore contacts the DNA more intimately than TrmBL2 which is a non-specific binder and forms very few contacts with the DNA bases.

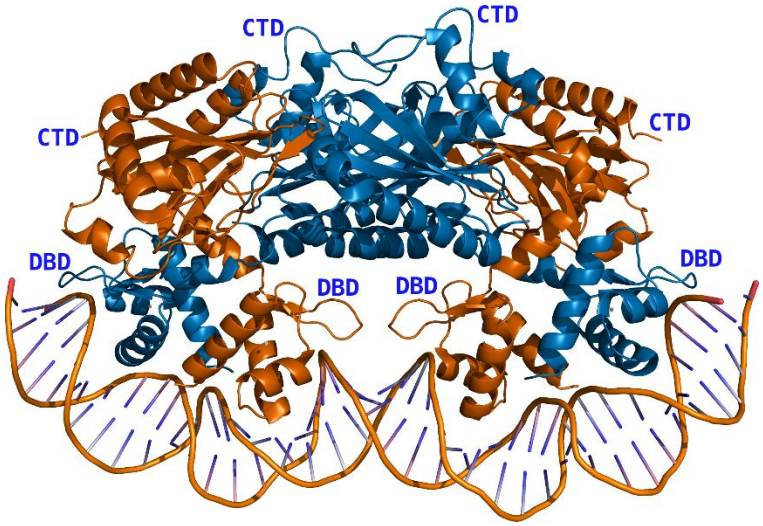


Figure 21: TtgV tetramer bound to DNA (PDB ID: 2XRO). The symmetry related chains are shown in the same colour. The DNA binding domains (DBD) and the C-terminal domains (CTD) are labelled for each monomer.

CTD homologs of TrmBL2

A search for structural homologs using the PDBeFold server [159] showed that the C-terminal domain of TrmBL2 bears structural homology to the phospholipase D domain (PLD) [160]. However the $HxK(x)_4D$ motif (HKD motif) essential for phospholipase activity is absent in TrmBL2. PLD-like domains are also found in cardiolipin synthases, phosphatidylserine synthases, poxvirus envelope proteins, a *Yersinia* murine toxin and a small number of endonucleases [161, 162]. In case of TrmBL2 the C-terminal domain seems to function as a tetramerization domain and could in principle also act as a scaffold for protein-protein interactions [34].

The coiled coil domain

The coiled coil domain as a dimerization motif, is a common feature found in many bacterial and eukaryotic transcription factors. In case of TrmB, TrmBL2, BmrR and proteins of similar dimeric structure it places the DNA binding domains in a manner that allows for their simultaneous interaction with adjacent major grooves. The dimerization helix α4 in TrmBL2 is longer than in TrmB but of similar length as in Sso10a [163] and its hydrophobic interface promotes the dimer formation. The two antiparallel alpha helices in TrmBL2 are almost parallel (178° angle as calculated in Pymol). A similar orientation of the dimerization helices is also found in TrmB (175°) [156]. This contrasts with other dimeric DNA binding proteins like BmrR [164] (161°), Sso10a (157°) and CAP [165] (156°).

The N-terminal ewHTH domain

TrmBL2 belongs to the wHTH family of DNA binding proteins which are widely distributed across all the three domains of life and are involved in almost every aspect of nucleic acid metabolism [80]. Within the wHTH family, TrmBL2 falls under the extended winged helix turn helix (ewHTH) subfamily which is characterized by extra helices α0 and α4 at the N and C-termini respectively, of the canonical wHTH domains [100]. The closest structural homologs based on the PDBeFold Q-score are PKZ Zalpha (PDB 4LB5), a zebrafish Z-DNA dependent protein kinase PKZ, meiotic recombination protein HOP2 (PDB 2MH2) and C-terminal domains of human replication protein, hRPA32 (PDB 1Z1D), all eukaryotic DNA binding proteins.

Role of surface electrostatics in DNA binding

Most of the wHTH proteins carry a strong positive potential on the surface that is involved in DNA binding. In the majority of cases the strongly electropositive recognition helix α3 is involved in base specific contacts. In cases where other structural features are employed for nucleobase recognition in the major groove like wing1 in case of hFRX1, the wing1 is more electropositive than the recognition helix [86, 166]. Electrostatic surface comparison of wHTH domains of TrmBL2 with other members of this family whose DNA bound structures are available revealed that the DNA-facing surface of TrmBL2 is significantly less electropositive and more apolar (Figure 22). This lack of electropositive surface in the DNA binding region of TrmBL2 is reflected in the low number of polar interactions with the DNA bases and less intimate binding to the DNA as reflected by solvent accessible surface area calculations (see Comparison with TtgV structure, Figure 21). Interestingly, the surface charge within the α3 helix of the distal and proximal ewHTH domains also varies with the proximal ewHTH domains displaying a slightly greater positive surface than the distal one. This difference is reflected in the extra number of amino acids which the proximal ewHTH domains present to the DNA (Figure 15, c).

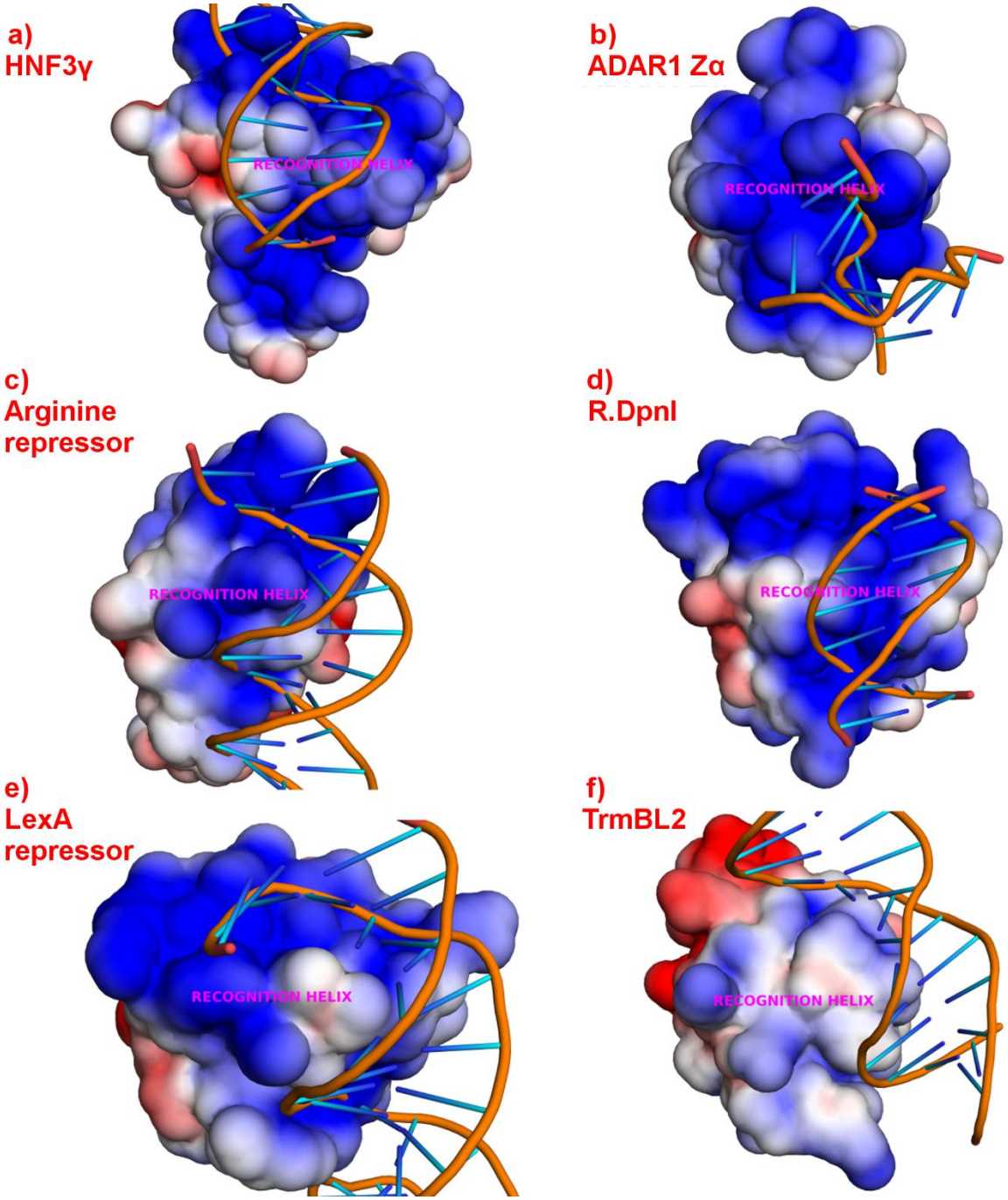


Figure 22. Comparison of the surface potential representation of wHTH domains of bacterial repressors with the distal ewHTH domain of TrmBL2. The domains are shown in the same orientation with the recognition helix labelled and roughly parallel to the plane of the paper. Unlike the other wHTH domains, TrmBL2 presents a largely apolar surface to the DNA reminiscent of archaeal architectural proteins. PDB IDs: a) 1VTN; b) 1QBJ; c) 2P5L; d) 4ESJ; e) 3JSP

Functional and structural similarities with bacterial H-NS

In an attempt to explain the concurrent role of transcriptional repression and chromosome shaping, it has been suggested that TrmBL2 functions in a similar manner as bacterial H-NS which in involved in gene silencing and also forms stiff nucleoprotein filaments [27, 31, 167]. A structural comparison of the two reveals that H-NS also contains a C-terminal wHTH domain [168] and an N-terminal domain [169] which has been proposed to facilitate oligomerization. In the absence of a full length H-NS-DNA complex structure, the relative orientation of the N and the C-terminal domains with respect to the unstructured linker is not known [170]. The order of the domains in bacterial H-NS (N-ter-CTD-linker-wHTH-C-ter) is reversed as compared to TrmBL2 (N-ter-wHTH-coiled coil-CTD-C-ter) indicating that a similar function would be conceivable only by convergent evolution.

TrmBL2 exhibits properties characteristic of archaeal chromatin proteins

Like in eukaryotes, a strict categorization of archaeal DNA binding proteins as either transcriptional factors or architectural proteins is not that straightforward as most of the these proteins exhibit roles that combine these two functions [52]. As an example Alba, a highly conserved chromatin protein, imparts different structural features to chromatin depending on its concentration, with sequence specific DNA binding suggested at low concentrations and sequence independent cooperative binding at high concentration [171]. However, in *Methanococcus maripaludis,* a mesophilic archaea, Alba has been shown to bind a specific 18 bp DNA motif and regulate the expression of several genes involved in Carbon dioxide assimilation [47].

Other examples include proteins Sa-Lrp of *S. acidocaldarius* and the Lrs-14 family in *S. solfataricus,* which act both as global transcriptional regulators and as nucleoid structuring proteins [172-175].

Studies on TrmBL2 function have also concluded that it plays a role in gene regulation as well as chromosome shaping [34]. It has been observed that TrmBL2 binds to both coding and non-coding regions of the genome. Similar to the binding of HMtA2 histones to coding regions [176], binding of TrmBL2 does not halt transcription. On the other hand, like in HMtA2, TrmBL2 binding to promoter regions is suggested to block the access of RNAP and therefore repress transcription of these genes.

The findings that TrmBL2 binds to DNA in a sequence specific manner when present at low concentration and switches to a largely non-specific cooperative binding mode at high concentrations [154] is reminiscent to the concentration dependent binding of Alba [171]. It is important to add that no consensus binding sequence has been found for TrmBL2 during these studies.

The observed non-specific binding to DNA of TrmBL2 is supported by its crystal structure. The low polarity of the DNA binding region and the slight asymmetry of proximal and distal ewHTH domains fit to a protein design that avoids specific polar contacts with DNA. The blurred electron density of the nucleobases could be explained by a superposition of three TGM motifs approximately related to each other by screw axis symmetry operations shifting the dsDNA by three base pairs on either end. TrmBL2 is comprised of two dimers each of which is structurally related to the repressor proteins of TrmB family which bind a specific operator sequence by placing a pair of ewHTH domains in adjacent major grooves. It is therefore not surprising that in TrmBL2 such a dimer of dimers designed for non-specific binding will use mismatching of the four ewHTH binding sites and apolar interaction surfaces to offer several low affinity binding sites to DNA.

The design of the DNA binding surface in TrmBL2 is reminiscent of archaeal chromatin proteins Sac7d and Sso7d of Sul7d family. The crystal structure of Sso7d in complex with DNA shows that this protein presents an apolar surface to the DNA [51]. Hyperthermophilic archaea thrive at extreme temperatures and are susceptible to a host of insults which include thermal denaturation of DNA and radiolysis [177]. To counter the effect of extreme temperatures on DNA structure, archaea have evolved a host of chromatin binding proteins which include in addition to the Sul7d proteins, proteins like HTa, MC1, Cren7 and Alba. All of these proteins show very low nucleotide sequence specificity and stabilize the DNA against denaturation at extreme temperatures [178-183].

TrmBL2 with its ewHTH domains which are infrequent in archaeal architectural proteins obviously has evolutionary roots in the specific repressor protein families of archaea and bacteria. The development of an apolar interaction surface with DNA which TrmBL2 shares with Sso7d could be a case of convergent evolution.

Despite these similarities with the archaeal chromatin proteins, a clear role for TrmBL2 is still not defined. Although the deletion strains of TrmBL2 in *T. kodakarensis* show repression of many genes, when compared to specific transcriptional regulators, the repression is modest (~5%). Moreover a large number of the genes that are repressed by TrmBL2 are annotated as hypothetical proteins [34] . Furthermore the proposed role of TrmBL2 in protecting genomic DNA from melting at high temperatures [154] is not experimentally confirmed. Further experiments are therefore required to fully understand the role of TrmBL2 in hyperthermophilic archaea.

PART II: TrmB

5. INTRODUCTION

TrmB family of transcription factors have been shown to control the uptake and subsequent metabolism of various sugars in several archaeal phyla. TrmB (Transcriptional regulator for the maltose sytem) (PF1743) was first discovered in *Pyrococcus furiosus* as a repressor of TM operon which encodes the high-affinity ABC transporter for Trehalose and Maltose [184]. The same protein was later found to regulate the MD (maltodextrins) operon as well. Biochemical studies showed that maltose and trehalose abrogated repression by TrmB at the TM operon whereas maltodextrins and sucrose had the same effect at the MD operon [155]. Glucose was found to increase repression at both the operons. It is interesting to note that *Pyrococcus furiosus* is not able to transport and utilize glucose as a carbon source although it does encode an ADP-dependent glucokinase [185]. Glucose for cellular energy requirements is derived from the metabolism of other sugars.

TrmB complexed with sucrose crystallizes as a monomer in the asymmetric unit. The structure features an N-terminal ewHTH domain, a coiled coil domain which promotes dimerization as seen in one of the crystallographic dimers and a C-terminal effector binding domain (EBD). The C-terminal part of the EBD is the sugar binding domain (SBD) which is seen bound to sucrose in this structure [156]. An earlier structure of TrmB EBD was solved with bound maltose [186]. Earlier biochemical and molecular biology studies has suggested a critical role for Y50 in DNA binding [187] which is present in the recognition helix α4 of TrmB.

The full length TrmB crystallized with sucrose had an accidental V161A mutation which rendered this protein incapable of binding to the MD promoter as confirmed by in vitro transcription assays (Michael Krug, Dissertation, 2009, Universität Konstanz). Expression of wild type TrmB was therefore necessary for any future crystallization attempts. The major goal of this project was the crystallization of TrmB with TM and MD promoters so as to shed light on the differential regulation of these promoters and also to provide insights into the role of different inducer molecules.

6. MATERIALS AND METHODS

6.1 Expression and purification of SUMO cleaving enzyme Ulp1 protease

C-terminal 6x His tagged Ulp1 protease cloned in Kanamycin and Chloramphenicol resistant pET24 vector was transformed into *E.coli* BL21 STAR pRARE strain[3]. The transformed cells were grown at 37°C and protein induction was done with 1mM IPTG at OD_{600} of 0.5. Post induction, the culture was grown overnight at 20°C. Cells were harvested and stored at -80°C.

For purification, the cells were resuspended in lysis buffer containing 40mM HEPES pH 7.5, 150mM KCl, 20mM β-mercaptoethanol and 10µg/ml DNase and lysed using the cell disrupter. The lysate was centrifuged at 16000 rpm and the supernatant containing the protein was incubated with 1ml Ni-Sepharose beads at 4°C for 60 minutes. The beads were washed with 10ml of lysis buffer and this step was repeated 5 times. Protein elution was done by adding 2ml of 40mM HEPES pH 7.5, 150mM KCl, 20mM β-me and 250mM Imidazole. The eluate was dialyzed overnight against 40mM HEPES pH 7.5, 100mM KCl and 10mM β-me and concentrated in a 10kDa cut-off centricon.

6.2 Expression and purification of His-SUMO-TrmB

TrmB was cloned in Kanamycin resistant pSUMO vector (pSUMO-TrmB) yielding His_6-SUMO tag at its N-terminus. The resulting construct was transformed into *E.Coli* BL21 (DE3) strain[4]. Cells were grown at 30°C and protein expression was induced with 1mM IPTG at OD_{600} of 0.8. Cell culture was grown for 5 hours post induction before harvesting at 4°C and storage at -80°C.

For purification[5], the cells were resuspended in 20mM Tris pH 8.0, 150mM NaCl and 2mM $MgCl_2$ (Buffer 1) and lysed by passing two times through Cell Disrupter (Constant Systems Limited, UK) at 2.6 Bar. The lysate was centrifuged at 100,000g at 4°C and the resulting supernatant containing the protein was incubated with 3ml Protino Ni-IDA beads (Machery-Nagel) at 4°C for 60 minutes. The beads were washed extensively with 100ml Buffer 1 followed by washing with 100ml Buffer 2 (20mM Tris pH 8.0, 500mM NaCl) and finally with

[3] The transformed cells were obtained from Steffan Preißler, AG Deuerling, Universität Konstanz
[4] The construct was made by Jutta Nesper, Universität Konstanz
[5] This purification protocol was designed by Jutta Nesper, Universität Konstanz

100ml Buffer 1. For elution, 30ml of Buffer 3 (20mM Tris pH 8.0, 150mM NaCl, 2mM $MgCl_2$, and 300mM imidazole) was added in 10ml steps and eluate was collected in 1ml fractions. The protein containing fractions were identified by adding 100µl of Bradford solution to 2µl of protein sample. Fractions containing the protein were pooled and an overnight digestion with 6x His-tagged Ulp1 protease was setup (8µg protease/mg TrmB). The digestion was set up in an 8kDa dialysis tubing which was dialyzed overnight against 5L of Dialysis buffer (20mM Tris pH 8.0, 200mM NaCl and 2mM $MgCl_2$) at 4°C and later centrifuged at 100,000g for 60 minutes. The supernatant was incubated with 5ml Protino Ni-IDA beads at room temperature with constant mixing for 60 minutes. The beads were allowed to settle down and the supernatant contained the pure TrmB. The beads were later washed with dialysis buffer and the cleaved His-SUMO and Ulp1-protease were eluted from the beads by elution buffer containing 20mM Tris pH 8.0, 200mM NaCl and 2mM $MgCl_2$ and 500mM Imidazole.

6.3 Expression and purification of TrmB-C-His

TrmB cloned in Ampicillin resistant pCS19 vector with C-terminal 6x His-tag was transformed into *E.Coli* BL21 (DE3) STAR strain[6]. Cells were grown in LB medium at 37°C and protein expression was induced by adding 1mM IPTG at OD_{600} of 0.6. Post induction the culture was grown for another 5 hours before harvesting at 4°C and storage of the pellet at -80°C.

For purification, cells were resuspended in a Buffer A containing 10mM CHES pH 9.0, 200mM NaCl, 50mM imidazole, 3% (v/v) dioxane and 5mM sucrose. Cells were lysed by French Press (16,000 psi) and the lysate was centrifuged at 30,000 rpm for 60 minutes. The supernatant was heat treated at 85°C for 20 minutes followed by another round of centrifugation (30,000 rpm, 60 minutes). The resulting supernatant was filtered through a 0.22µm filter to remove any particulate matter and loaded onto a 1ml Ni His-Trap HP column pre-equilibrated with Buffer A. After extensive washing of the column with 20CV of Buffer A, elution was carried out using a linear imidazole gradient with Buffer B (10mM CHES pH 9.0, 200mM NaCl, 0.5M imidazole, 3% (v/v) dioxane and 5mM sucrose). Fraction containing the protein were determined by running an SDS-PAGE of the peak fractions. Protein containing fractions were pooled and concentrated in a 10kDa cut-off Vivaspin concentrator. To remove the aggregates,

[6] Construct obtained from our collaboration partners at the University of Regensburg

the protein solution was centrifuged at 40,000 rpm for 25 minutes before setting up crystallization trials.

6.4 Expression and purification of tag-free TrmB

Codon optimized, tag-free TrmB cloned in pET24d vector, ordered from Trenzyme GmbH Konstanz, was transformed into *E.Coli* BL21 (DE3) cells. The cells were grown at 37°C and protein induction was done with 1mM IPTG at OD_{600} of 0.6. The culture was grown for further 5 hours before harvesting at 4°C and storage at -80°C.

For purification the cells were resuspended in 50ml of Buffer A containing 10mM CHES pH 9.0, 200mM NaCl, 3% (v/v) dioxane and 5mM glucose. Cells were lysed by French Press (3 times at 16,000 psi), the lysate was heated at 85°C for 25 minutes and centrifuged at 35,000 rpm for 60 minutes. The supernatant was filtered and loaded onto a 5ml Q-XL ion exchange column or a 5ml Heparin FF column pre-equilibrated with Buffer A. Elution was achieved by a linear salt gradient with Buffer B (10mM CHES pH 9.0, 1M NaCl, 3% (v/v) dioxane and 5mM glucose). After analyzing the peak fractions on SDS-gel, the protein containing fractions were pooled, concentrated and loaded onto a Superdex 75 column equilibrated with 10mM CHES, 5mM glucose and 3% (v/v) dioxane.

7. RESULTS

For a detailed description of expression and purification protocols, refer to the Materials and Methods section

7.1 SUMO cleaving enzyme Ulp1 protease: Purification

Ulp1 protease bearing a 6x-His tag at the C-terminus was purified by Ni-affinity chromatography. During concentration, a large fraction of the protein got precipitated but there was enough protein in the supernatant that had a concentration of 1mg/ml, sufficient enough to be used for the cleavage of SUMO tag (Figure 23).

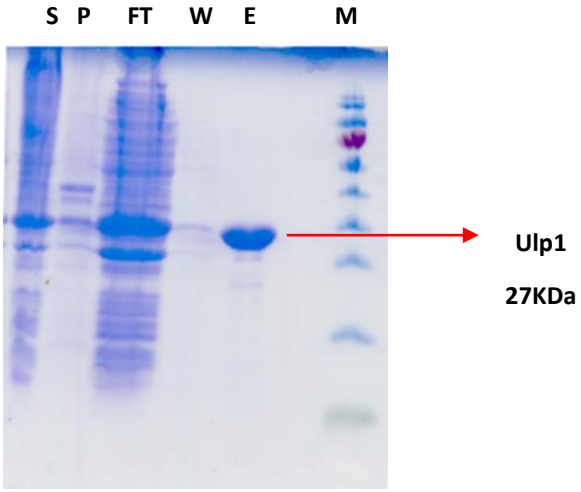


Figure 23. SDS-PAGE analysis of Ulp1 fractions S = Supernatant; P = Pellet; FT = Flow through; W = Wash; E = Eluate; M = Marker

7.2 His-SUMO-TrmB: Purification

The N-terminal His-SUMO tagged TrmB was purified initially by Ni-affinity chromatography. After purification, the SUMO tag was cleaved by incubating the protein with Ulp1 protease. The incubation mixture was passed over a Ni-affinity column which retained the His-tagged proteins (Ulp1, Cleaved His-SUMO and uncleaved His-SUMO-TrmB). The flowthrough contained the cleaved and purified TrmB.

During purification a large fraction of the TrmB got precipitated. Protein purification over prepacked Ni-HisTrap column (Figure 24) did not improve the results. Upon concentration, further protein precipitation occurred and this could not be avoided by adding stabilizing

agents like dioxane [156]. It was observed that after digestion with Ulp1 protease, a large fraction of the protein gets precipitated and also that the cleaved His-SUMO bound non-covalently to TrmB and retained a fraction of it on the Ni-affinity column (Figure 25). This non-covalent interaction could not be disrupted even by using high concentrations of salt (up to 1M NaCl). Overall, the purification procedure couldn't be optimized for large scale production of TrmB, a prerequisite for crystallization.

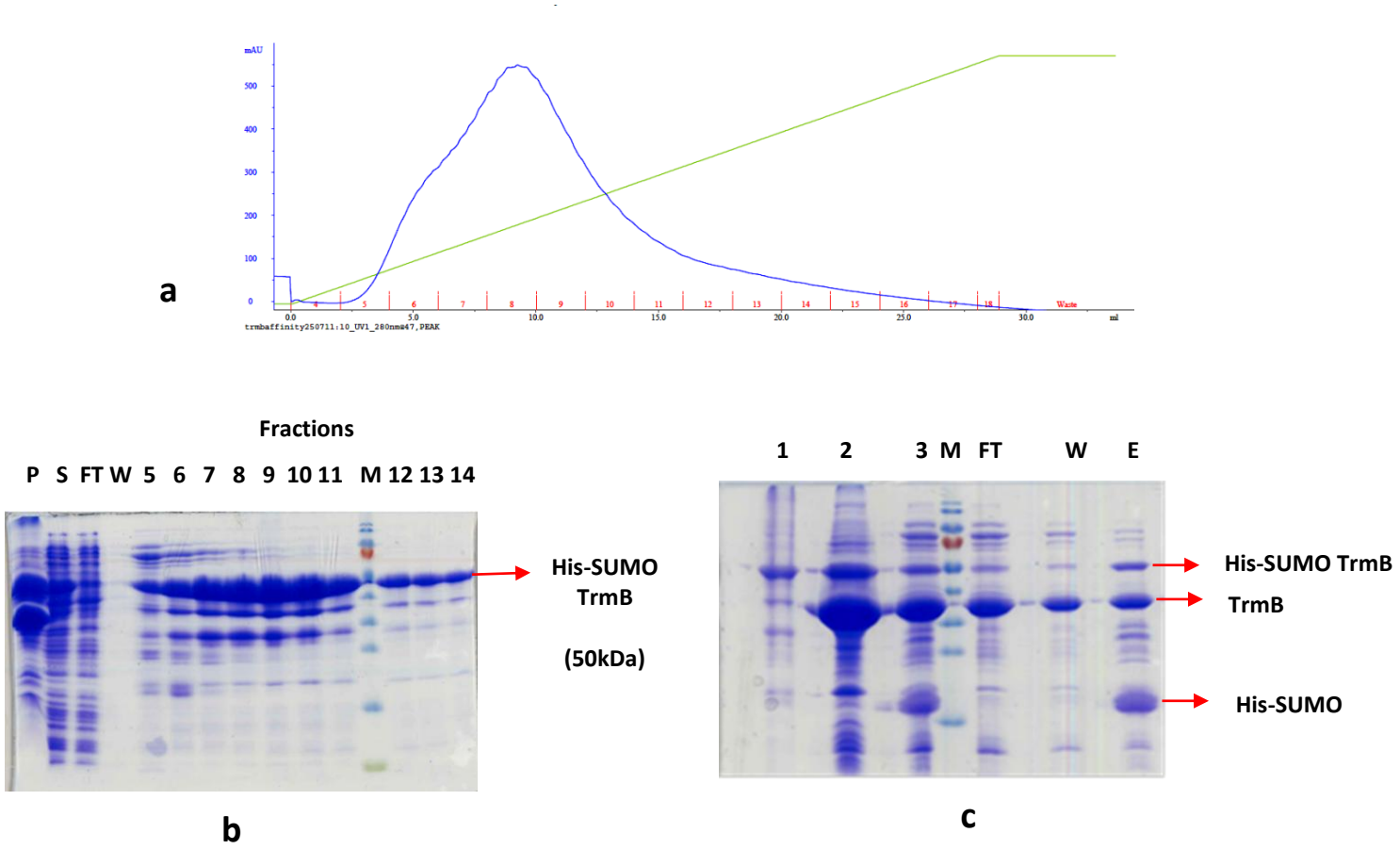


Figure 24.

a: Elution profile of His-SUMO-TrmB from 1ml Ni-HisTrap column.
X-axis: Volume; Y-axis: UV-absorbance

b: SDS-PAGE analysis of purification. P = Pellet; S = Supernatant; FT = Flowthrough; W = Wash; M = Marker

c: SDS-PAGE analysis of His-SUMO cleavage by Ulp1 protease

1 = Uncleaved His-SUMO TrmB; 2 = Cleaved His-SUMO TrmB (precipitate); 3 = Cleaved His-SUMO TrmB (Supernatant); FT = Flowthrough; W = Wash; E = Eluate

7.3 TrmB-C-His: Purification

For purification of C-terminal His-tagged TrmB, the lysate was heat treated to remove most of the *E.Coli* protein impurities and later purified over a Ni-HisTrap column. Protein elution was carried out by applying a linear imidazole gradient.

During Ni-affinity chromatography, the protein elutes at two different salt concentrations. The fractions that elute at higher salt (53% Buffer B) concentration are more pure as seen from the SDS-PAGE. This construct expresses well and could be concentrated up to 8mg/ml. Crystals were set up alone or in complex with 26bp TM promoter (5'GCAAAATATATATA CTTTTAGTATAT3') but no crystals were obtained.

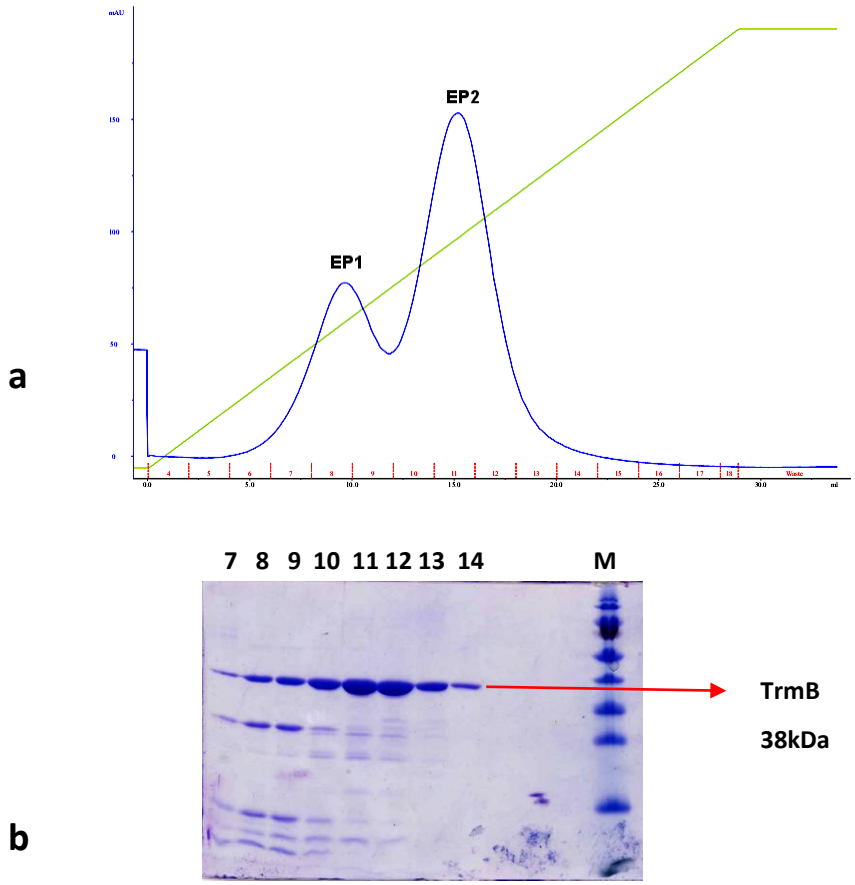


Figure 25.

a: Elution profile of TrmB C-His from a Ni-HisTrap column.
EP1 (Elution Peak 1) = 34% Buffer B, EP2 (Elution Peak 2) = 53% Buffer B.

b: SDS-PAGE analysis of the two elution peaks EP1 (Fractions 7, 8, 9) and EP2 (Fractions10, 11, 12, 13, 14); M = Marker.

7.4 Tag-free TrmB: Purification

The purification procedure involved heat treatment of the lysate to remove the protein impurities from *E.Coli*, purification over Ion exchange column followed by gel filtration chromatography.

During expression of this construct, a fraction of TrmB was cleaved which remained as an impurity and couldn't be separated by either Ion-exchange chromatography or Gel filtration.

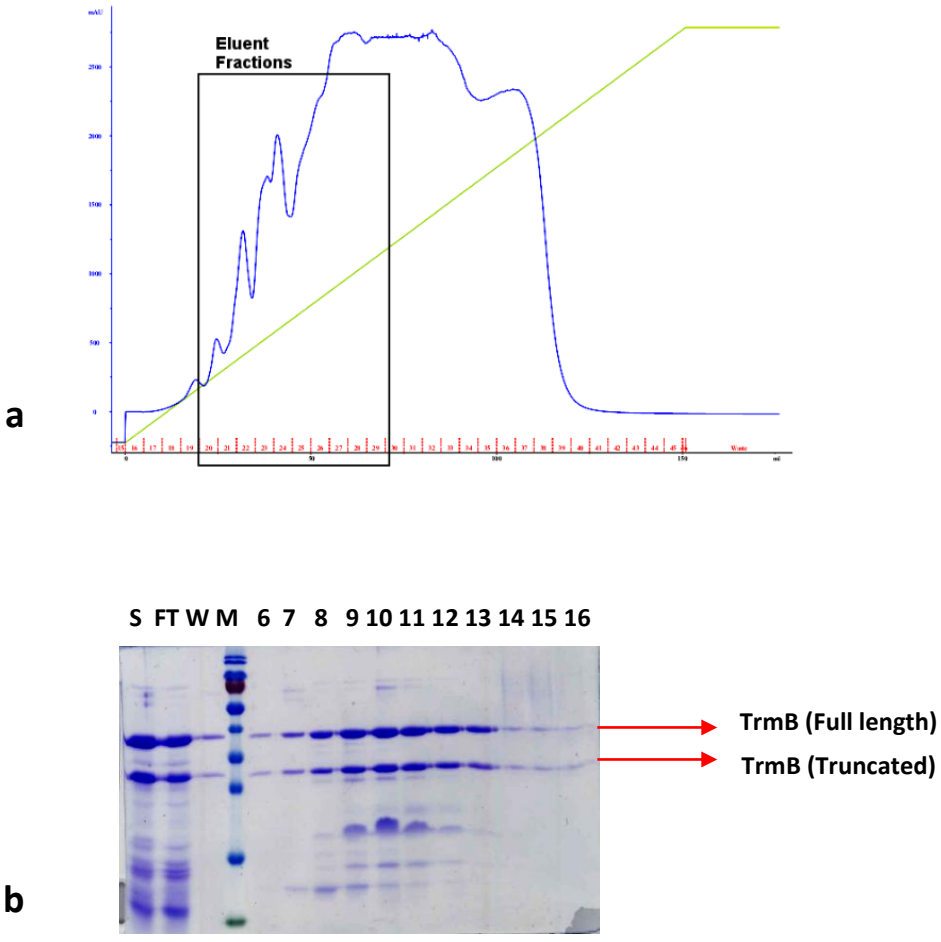
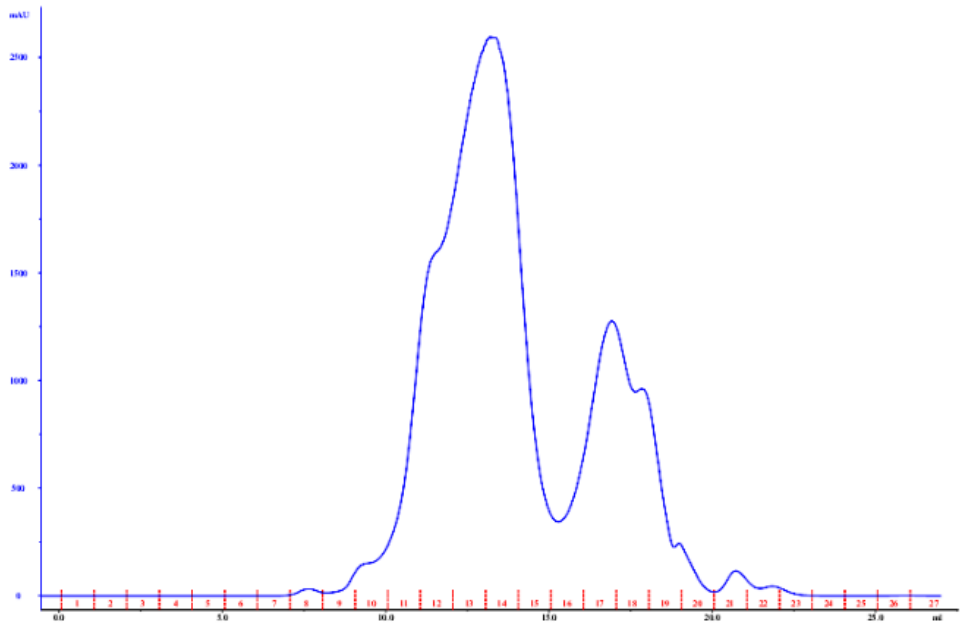


Figure 26.

a: Elution profile of tag-free TrmB from a Q-Sepharose ion exchange column. Inset: Fractions eluting between14-47% Buffer B.

b: SDS-PAGE analysis. S = Supernatant; FT = Flowthrough; W = Wash; M = Marker;
Eluate fractions 6 -16

a



b

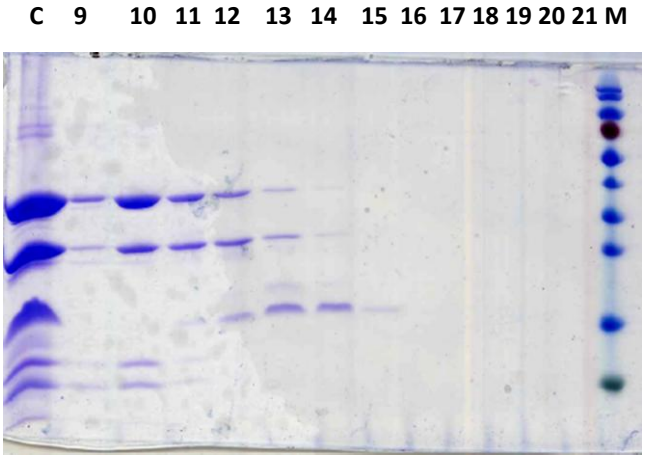


Figure 27.

a: Elution profile of tag-free TrmB from a Superdex75 gel filtration column.

b: SDS-PAGE analysis of peak fractions. C = Concentrated protein before loading on the column, M = Marker.

8. DISCUSSION

A major issue with the purification of TrmB is its tendency to stick to polysaccharide based matrices [155] which makes purification over Agarose based columns difficult. As Gel filtration is usually used as a final purification step to obtain homogenous protein preparations, the same is not possible with TrmB. The heat treatment of TrmB results in the generation of small aggregates and higher oligomers which cannot be completely removed by ultracentrifugation and presence of which hinders crystallization. Purification over Bio-Gel P, a polyacrylamide matrix was also tried but it did not improve the results.

The TrmB-C-His construct could be purified but concentrations beyond 8mg/ml resulted in precipitation of the protein which could not be avoided by adding stabilizing agents like dioxane or L-Arginine. Crystallization trials at this concentration alone and with TM promoter DNA did not yield any crystals. Moreover, the C-terminal 6x-Histag was not cleavable for this construct, presence of which can affect crystallization.

The tag-free TrmB construct gets proteolysed at the N-terminus during expression but unlike TrmBL2, the truncated product could not be separated by various purification methods like Heparin-affinity, Ni-affinity or Gel filtration chromatography. To separate the truncated product of TrmB, other purification methods like hydrophobic interaction chromatography should be tried.

APPENDIX

Protein: TrmBL2

Organism: *Pyrococcus furiosus* DSM 3638

Gene: PF0496

GenBank identifier: AAL80620.1

UniProt ID: Q8U3H1

Gene sequence:

ATGAGCAAAGATAGAATGGTAGAACTATTACAAGAACACTTTGAGTTGAACTTATACGAGGCTAGAGCATATGTTGCCTT
AGTAGCGTTTGGTGTTCTTACTCCAGCAGAATTAGCAAGTGTTTCGGAAGTTCCAGCGCCAAGAACTTACGATGTACTTAG
GAGTTTAGAAAAGAAAGGATTTGCCATGACTCAACCAGGAAAGACAAACAAGTACAGGCCAGTTCACCCAGCAAACGTT
CTTGAAAAGTTCATCCAGGACTGGCAAGAGAGGGTTAAGGAAGAACTTGAGGCTAAGAAGAAAGCTAAGGAGGAACTC
CTTGAATTAATGGCCCCACTAATAGAGACAGAGGTTCCAAAGTACGGAGTTGAGAGAGTTTGGGTGGTTAGAGGAATTA
AGAACTCAACATTAAAGACCAAAGAAATGCTGGAAGAAGCTCAAAATGAAATACTATTGGCAGACGATGGATTCATAGC
AGTTAACCTTGAAGATGACATAATAAAGGCAGTTGACAGAGGAGTTAAGACAAAGATCCTTCTAACAAAGAACTTACTTC
CAAGGCTAAAGGCCTCCAAGATCATTGACTATGCAAAAGAGGGTAAGCTCGAGCTGAGGGCACTTGATAAGTTTGATCTT
CCAATGCTGATCTGTGATGAAGAGGTATTCTTTGCACTGGAGGACTTGGCAGCTAGATACTTCAACTATGAGACCCAGGT
TTGGATTAAGGATCACAGAGTTGTAGCTCTCTTCAAAGAGAAGTTCAACGAATACTGGGAGAAGGCTGAGAAGGTTTGA

Amino acid sequence:

MSKDRMVELLQEHFELNLYEARAYVALVAFGVLTPAELASVSEVPAPRTYDVLRSLEKKGFAMTQPGKTNKYRPVHPANVLEK
FIQDWQERVKEELEAKKKAKEELLELMAPLIETEVPKYGVERVWVVRGIKNSTLKTKEMLEEAQNEILLADDGFIAVNLEDDIIKA
VDRGVKTKILLTKNLLPRLKASKIIDYAKEGKLELRALDKFDLPMLICDEEVFFALEDLAARYFNYETQVWIKDHRVVALFKEKFNE
YWEKAEKV

Number of amino acids: 264

Molecular weight: 30606.5 Da

Theoretical pI: 5.55[7]

[7] Calculated by Expasy ProtParam (http://web.expasy.org/protparam/)

TrmBL2 constructs used during the course of this work:

1. TrmBL2 (hereafter named pfTrmBL2)

Tag: N-terminal 6x-Histidine

Expression system: *Pyrococcus furiosus*

Strain: MURPf10 ([136, 137])

TrmBL2 expressing *Pyrococcus furiosus* cells were provided by our collaboration partners at the University of Regensburg.

2. TrmBL2

Tag: N-terminal 6x-Histidine

Vector: pQE31

Expression system: *Escherichia coli*

Strain: SF120

Antibiotic resistance: Ampicillin

The DMSO stock was provided by Prof. Winfried Boos, Universität Konstanz.

3. TrmBL2 (hereafter named ecTrmBL2)

Tag: None

Vector: pET24d

Expression system: *Escherichia coli*

Strain: BL21 (DE3)

Antibiotic resistance: Kanamycin

This construct was ordered from Trenzyme GmbH, Konstanz.

Protein: TrmB

Organism: *Pyrococcus furiosus* DSM 3638

Gene: PF1743

GenBank identifier: AAL81867.1

UniProt ID: Q9HGZ9

Gene sequence:

ATGGAGATTCCCCCAGAAATTTCACATGCTTTGAGTGAGATTGGATTTACAAAGTATGAAATCCTCACTTACTGGACTCTCT
TGGTTTATGGGCCCAGTACTGCGAAGGAAATATCCACAAAAAGCGGGATTCCATACAACAGGGTCTATGATACAATATCC
TCCTTAAAACTTAGAGGATTTGTGACTGAAATTGAAGGGACACCAAAGGTATACGCTGCCTACTCACCAAGAATAGCATTT
TTCAGGTTCAAAAAAGAACTCGAAGACATAATGAAAAAGCTCGAAATTGAGCTTAATAATGTAAAAAAGGAAGAACAGA
GACCTGCAATATGGAGGAGCAGGAGTTTTGACGAAGCCATTGAGATGTTCAGAGAGTCACTCTATTCAGCTAAAAATGAG
GTTATAGTAGTTACTCCAAGCGAATTCTTTGAAACAATAAGAGAAGATTTGATCAAAACTCTCGAGAGAGGTGTAACGGT
GTCCCTCTATATCGACAAAATCCCGGATCTATCAGAGTTCAAAGGGAAAGGGAATTTCTTTGTTAGGCAGTTCTACAAGCT
GAACCACTTAATAGGCATGACTGATGGAAAAGAGGTAGTCACAATTCAGAATGCCACTTTTGACTCTATTGGACCTCCTTC
ATTTAAGTCCACTTATCCTGAGATAATATTCTCCCAATACAGTCTCATAATAGAAATTTTCAAGGAATCCACACTCGAAAAA
GAGATTATCGGCAATCCAAAAGATATTAGATTCTTCGCTATGTTTCATGCAGTAGACTTTGTTAAAAATCACTTAAAAAACA
GAAACATCTATGCTGAGATAACTGGAAAAAACTTGGAGTCAGGCAGATTAGAAACCCTTACCGGAAGGGTTGTAGGATA
CACACTCTCTCTCAGGGAAGCCGTTAACAATATCCATCTCGAAACTGAAAACGGGGTTGTAAAGGTTGGAGGCATGTTTG
CGGTTATTGAAGATTATGAAAGTACTGAAATAAAATTCATTATGGGGTGA

Amino acid sequence:

MEIPPEISHALSEIGFTKYEILTYWTLLVYGPSTAKEISTKSGIPYNRVYDTISSLKLRGFVTEIEGTPKVYAAYSPRIAFFRFKKELEDI
MKKLEIELNNVKKEEQRPAIWRSRSFDEAIEMFRESLYSAKNEVIVVTPSEFFETIREDLIKTLERGVTVSLYIDKIPDLSEFKGKGNF
FVRQFYKLNHLIGMTDGKEVVTIQNATFDSIGPPSFKSTYPEIIFSQYSLIIEIFKESTLEKEIIGNPKDIRFFAMFHAVDFVKNHLKN
RNIYAEITGKNLESGRLETLTGRVVGYTLSLREAVNNIHLETENGVVKVGGMFAVIEDYESTEIKFIMG

Number of amino acids: 338

Molecular weight: 38838.6 Da

Theoretical pI: 5.76 [8]

[8] Calculated by Expasy ProtParam (http://web.expasy.org/protparam/)

TrmB constructs used during the course of this work

1. TrmB

Tag: N-terminal His-SUMO[9]

Vector: pSUMO

Expression system: *Escherichia coli*

Strain: BL21 (DE3)

Antibiotic resistance: Kanamycin

2. Ulp1 (SUMO protease)[10]

Tag: C-terminal 6x-Histidine

Vector: pET24

Expression system: *Escherichia coli*

Strain. BL21 (DE3) STAR pRARE

Antibiotic resistance: Chloramphenicol + Kanamycin

3. TrmB[11]

Tag: C-terminal 6x-Histidine

Vector: pCS19

Expression system: *Escherichia coli*

Strain: BL21 (DE3) STAR

Antibiotic resistance: Ampicillin

4. TrmB

Tag: None

Vector: pET24d

Expression system. *Escherichia coli*

Strain: BL21 (DE3)

Antibiotic resistance: Kanamycin

[9] Construct designed by Jutta Nesper.
[10] Transformed cells obtained from Steffan Preißler, AG Deuerling, Universität Konstanz
[11] Plasmid obtained from Dr Winfried Hausner, Universität Regensburg

Diffraction data sets used for solving the structures

1. Data set used to calculate the phases by Se-SAD:
/nfs/loop1/synchrotron/SLS-2014/apr11/misbha/SeMet-A2-III/SeMet-A2-III-peak_t_1_0????.cbf

2. For ecTrmBL2-TGM17 complex (PDB ID: 5BPD, 2.4Å resolution):
/nfs/loop1/synchrotron/SLS-2013/sep12/ma/L2-TGM17-H5-MPD_2_0????.cbf

3. For ecTrmBL2-TGM17 complex (PDB ID: 5BPI, 3.2Å resolution):
/nfs/loop1/synchrotron/SLS-2013/oct04/ma/TGM-17-H5-MPD-Pt-Remote_1_0????.cbf

4. For ecTrmBL2-TGM19 complex (PDB ID: 5BOX):
/nfs/loop1/synchrotron/SLS-2013/aug23/ma/L2-TGM-6-5-60MPD_1_0????.cbf

5. For DNA-free TrmBL2 (PDB ID: 5BQT):
/nfs/loop1/synchrotron/SLS-2012/feb29/ma/O5-A2-3_2_0????.cbf

BIBLIOGRAPHY

1. Woese, C.R. and G.E. Fox, *Phylogenetic structure of the prokaryotic domain: the primary kingdoms.* Proc Natl Acad Sci U S A, 1977. **74**(11): p. 5088-90.

2. Albers, S.V. and B.H. Meyer, *The archaeal cell envelope.* Nat Rev Microbiol, 2011. **9**(6): p. 414-26.

3. Mondav, R., et al., *Discovery of a novel methanogen prevalent in thawing permafrost.* Nat Commun, 2014. **5**: p. 3212.

4. Olsen, G.J. and C.R. Woese, *Archaeal genomics: an overview.* Cell, 1997. **89**(7): p. 991-4.

5. Bernander, R., *Chromosome replication, nucleoid segregation and cell division in archaea.* Trends Microbiol, 2000. **8**(6): p. 278-83.

6. Samson, R.Y., et al., *A role for the ESCRT system in cell division in archaea.* Science, 2008. **322**(5908): p. 1710-3.

7. Wang, X. and J. Lutkenhaus, *FtsZ ring: the eubacterial division apparatus conserved in archaebacteria.* Mol Microbiol, 1996. **21**(2): p. 313-9.

8. Tenorio-Salgado, S., A. Huerta-Saquero, and E. Perez-Rueda, *New insights on gene regulation in archaea.* Comput Biol Chem, 2011. **35**(6): p. 341-6.

9. Woese, C.R., O. Kandler, and M.L. Wheelis, *Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya.* Proc Natl Acad Sci U S A, 1990. **87**(12): p. 4576-9.

10. Brochier-Armanet, C., et al., *Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota.* Nat Rev Microbiol, 2008. **6**(3): p. 245-52.

11. Lindas, A.C. and R. Bernander, *The cell cycle of archaea.* Nat Rev Microbiol, 2013. **11**(9): p. 627-38.

12. Gargaud, M., et al., *Encyclopedia of Astrobiology*. 2011, Springer-Verlag Berlin Heidelberg,: Berlin, Heidelberg. p. 1 online resource.

13. Pester, M., C. Schleper, and M. Wagner, *The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology.* Curr Opin Microbiol, 2011. **14**(3): p. 300-6.

14. Luger, K., M.L. Dechassa, and D.J. Tremethick, *New insights into nucleosome and chromatin structure: an ordered state or a disordered affair?* Nat Rev Mol Cell Biol, 2012. **13**(7): p. 436-47.

15. Kornberg, R.D., *Chromatin structure: a repeating unit of histones and DNA.* Science, 1974. **184**(4139): p. 868-71.

16. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 A resolution.* Nature, 1997. **389**(6648): p. 251-60.

17. Becker, P.B. and J.L. Workman, *Nucleosome remodeling and epigenetics.* Cold Spring Harb Perspect Biol, 2013. **5**(9).

18. Smolle, M. and J.L. Workman, *Transcription-associated histone modifications and cryptic transcription.* Biochim Biophys Acta, 2013. **1829**(1): p. 84-97.

19. Higgins, N.P., *The bacterial chromosome*. 2005, Washington, D.C.: ASM Press. xv, 559 p., 8 p. of plates.

20. Dillon, S.C. and C.J. Dorman, *Bacterial nucleoid-associated proteins, nucleoid structure and gene expression.* Nat Rev Microbiol, 2010. **8**(3): p. 185-95.

21. Grainger, D.C., et al., *Association of nucleoid proteins with coding and non-coding segments of the Escherichia coli genome.* Nucleic Acids Res, 2006. **34**(16): p. 4642-52.

22. Oshima, T., et al., *Escherichia coli histone-like protein H-NS preferentially binds to horizontally acquired DNA in association with RNA polymerase.* DNA Res, 2006. **13**(4): p. 141-53.

23. Navarre, W.W., et al., *Selective silencing of foreign DNA with low GC content by the H-NS protein in Salmonella.* Science, 2006. **313**(5784): p. 236-8.

24. Lucchini, S., et al., *H-NS mediates the silencing of laterally acquired genes in bacteria.* PLoS Pathog, 2006. **2**(8): p. e81.

25. Lang, B., et al., *High-affinity DNA binding sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes.* Nucleic Acids Res, 2007. **35**(18): p. 6330-7.

26. Dorman, C.J., *H-NS: a universal regulator for a dynamic genome.* Nat Rev Microbiol, 2004. **2**(5): p. 391-400.

27. Fang, F.C. and S. Rimsky, *New insights into transcriptional regulation by H-NS.* Curr Opin Microbiol, 2008. **11**(2): p. 113-20.

28. Smyth, C.P., et al., *Oligomerization of the chromatin-structuring protein H-NS.* Mol Microbiol, 2000. **36**(4): p. 962-72.

29. Rimsky, S., *Structure of the histone-like protein H-NS and its role in regulation and genome superstructure.* Curr Opin Microbiol, 2004. **7**(2): p. 109-14.

30. Dame, R.T., M.C. Noom, and G.J. Wuite, *Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation.* Nature, 2006. **444**(7117): p. 387-90.

31. Liu, Y., et al., *A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes.* Genes Dev, 2010. **24**(4): p. 339-44.

32. Vora, T., A.K. Hottes, and S. Tavazoie, *Protein occupancy landscape of a bacterial genome.* Mol Cell, 2009. **35**(2): p. 247-53.

33. Wang, W., et al., *Chromosome organization by a nucleoid-associated protein in live bacteria.* Science, 2011. **333**(6048): p. 1445-9.

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

35. Musgrave, D., P. Forterre, and A. Slesarev, *Negative constrained DNA supercoiling in archaeal nucleosomes.* Mol Microbiol, 2000. **35**(2): p. 341-9.

36. Pavlov, N.A., et al., *Identification, cloning and characterization of a new DNA-binding protein from the hyperthermophilic methanogen Methanopyrus kandleri.* Nucleic Acids Res, 2002. **30**(3): p. 685-94.

37. Tomschik, M., et al., *The archaeal histone-fold protein HMf organizes DNA into bona fide chromatin fibers.* Structure, 2001. **9**(12): p. 1201-11.

38. Driessen, R.P. and R.T. Dame, *Nucleoid-associated proteins in Crenarchaea.* Biochem Soc Trans, 2011. **39**(1): p. 116-21.

39. Edmondson, S.P. and J.W. Shriver, *DNA binding proteins Sac7d and Sso7d from Sulfolobus.* Methods Enzymol, 2001. **334**: p. 129-45.

40. Lurz, R., et al., *Electron microscopic study of DNA complexes with proteins from the Archaebacterium Sulfolobus acidocaldarius.* EMBO J, 1986. **5**(13): p. 3715-21.

41. Anderson, I., et al., *Genome sequence of Thermofilum pendens reveals an exceptional loss of biosynthetic pathways without genome reduction.* J Bacteriol, 2008. **190**(8): p. 2957-65.

42. Feng, Y., H. Yao, and J. Wang, *Crystal structure of the crenarchaeal conserved chromatin protein Cren7 and double-stranded DNA complex.* Protein Sci, 2010. **19**(6): p. 1253-7.

43. Hardy, C.D. and P.K. Martin, *Biochemical characterization of DNA-binding proteins from Pyrobaculum aerophilum and Aeropyrum pernix.* Extremophiles, 2008. **12**(2): p. 235-46.

44. Luo, X., et al., *CC1, a novel crenarchaeal DNA binding protein.* J Bacteriol, 2007. **189**(2): p. 403-9.

45. Marsh, V.L., A.T. McGeoch, and S.D. Bell, *Influence of chromatin and single strand binding proteins on the activity of an archaeal MCM.* J Mol Biol, 2006. **357**(5): p. 1345-50.

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

47. Liu, Y., et al., *The Sac10b homolog in Methanococcus maripaludis binds DNA at specific sites.* J Bacteriol, 2009. **191**(7): p. 2315-29.

48. Bell, S.D., et al., *The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation.* Science, 2002. **296**(5565): p. 148-51.

49. Heinicke, I., 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**(1): p. 76-87.

50. Kahsai, M.A., et al., *Solution structure, stability, and flexibility of Sso10a: a hyperthermophile coiled-coil DNA-binding protein.* Biochemistry, 2005. **44**(8): p. 2822-32.

51. Gao, Y.G., et al., *The crystal structure of the hyperthermophile chromosomal protein Sso7d bound to DNA.* Nat Struct Biol, 1998. **5**(9): p. 782-6.

52. Peeters, E., et al., *The interplay between nucleoid organization and transcription in archaeal genomes.* Nat Rev Microbiol, 2015. **13**(6): p. 333-41.

53. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid.* Nature, 1953. **171**(4356): p. 737-8.

54. Leslie, A.G., et al., *Polymorphism of DNA double helices.* J Mol Biol, 1980. **143**(1): p. 49-72.

55. Shakked, Z. and D. Rabinovich, *The effect of the base sequence on the fine structure of the DNA double helix.* Prog Biophys Mol Biol, 1986. **47**(3): p. 159-95.

56. Lu, X.J., Z. Shakked, and W.K. Olson, *A-form conformational motifs in ligand-bound DNA structures.* J Mol Biol, 2000. **300**(4): p. 819-40.

57. Shakked, Z., et al., *The conformation of the DNA double helix in the crystal is dependent on its environment.* Nature, 1989. **342**(6248): p. 456-60.

58. Wang, A.H., et al., *Molecular structure of a left-handed double helical DNA fragment at atomic resolution.* Nature, 1979. **282**(5740): p. 680-6.

59. Arnott, S., et al., *Left-handed DNA helices.* Nature, 1980. **283**(5749): p. 743-5.

60. Rohs, R., et al., *Origins of specificity in protein-DNA recognition.* Annu Rev Biochem, 2010. **79**: p. 233-69.

61. Nelson, H.C., et al., *The structure of an oligo(dA).oligo(dT) tract and its biological implications.* Nature, 1987. **330**(6145): p. 221-6.

62. Hizver, J., et al., *DNA bending by an adenine--thymine tract and its role in gene regulation.* Proc Natl Acad Sci U S A, 2001. **98**(15): p. 8490-5.

63. Haran, T.E. and U. Mohanty, *The unique structure of A-tracts and intrinsic DNA bending.* Q Rev Biophys, 2009. **42**(1): p. 41-81.

64. Mack, D.R., T.K. Chiu, and R.E. Dickerson, *Intrinsic bending and deformability at the T-A step of CCTTTAAAGG: a comparative analysis of T-A and A-T steps within A-tracts.* J Mol Biol, 2001. **312**(5): p. 1037-49.

65. Olson, W.K., et al., *DNA sequence-dependent deformability deduced from protein-DNA crystal complexes.* Proc Natl Acad Sci U S A, 1998. **95**(19): p. 11163-8.

66. Lu, X.J. and W.K. Olson, *3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures.* Nat Protoc, 2008. **3**(7): p. 1213-27.

67. Rohs, R., H. Sklenar, and Z. Shakked, *Structural and energetic origins of sequence-specific DNA bending: Monte Carlo simulations of papillomavirus E2-DNA binding sites.* Structure, 2005. **13**(10): p. 1499-509.

68. Petrey, D. and B. Honig, *GRASP2: visualization, surface properties, and electrostatics of macromolecular structures and sequences.* Methods Enzymol, 2003. **374**: p. 492-509.

69. Honig, B. and A. Nicholls, *Classical electrostatics in biology and chemistry.* Science, 1995. **268**(5214): p. 1144-9.

70. Rocchia, W., et al., *Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects.* J Comput Chem, 2002. **23**(1): p. 128-37.

71. Garvie, C.W. and C. Wolberger, *Recognition of specific DNA sequences.* Mol Cell, 2001. **8**(5): p. 937-46.

72. Schumacher, M.A., et al., *Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices.* Science, 1994. **266**(5186): p. 763-70.

73. Lewis, M., et al., *Crystal structure of the lactose operon repressor and its complexes with DNA and inducer.* Science, 1996. **271**(5253): p. 1247-54.

74. Van Roey, P., et al., *Intertwined structure of the DNA-binding domain of intron endonuclease I-TevI with its substrate.* EMBO J, 2001. **20**(14): p. 3631-7.

75. Edgell, D.R., et al., *Intron-encoded homing endonuclease I-TevI also functions as a transcriptional autorepressor.* Nat Struct Mol Biol, 2004. **11**(10): p. 936-44.

76. Shen, B.W., et al., *DNA binding and cleavage by the HNH homing endonuclease I-HmuI.* J Mol Biol, 2004. **342**(1): p. 43-56.

77. Aravind, L., et al., *The many faces of the helix-turn-helix domain: transcription regulation and beyond.* FEMS Microbiol Rev, 2005. **29**(2): p. 231-62.

78. Brennan, R.G. and B.W. Matthews, *The helix-turn-helix DNA binding motif.* J Biol Chem, 1989. **264**(4): p. 1903-6.

79. Daniels, D.S., et al., *DNA binding and nucleotide flipping by the human DNA repair protein AGT.* Nat Struct Mol Biol, 2004. **11**(8): p. 714-20.

80. Harami, G.M., M. Gyimesi, and M. Kovacs, *From keys to bulldozers: expanding roles for winged helix domains in nucleic-acid-binding proteins.* Trends Biochem Sci, 2013. **38**(7): p. 364-71.

81. Clark, K.L., et al., *Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5.* Nature, 1993. **364**(6436): p. 412-20.

82. Benayoun, B.A., S. Caburet, and R.A. Veitia, *Forkhead transcription factors: key players in health and disease.* Trends Genet, 2011. **27**(6): p. 224-32.

83. Zheng, N., et al., *Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP.* Genes Dev, 1999. **13**(6): p. 666-74.

84. Zhang, A.P., Y.Z. Pigli, and P.A. Rice, *Structure of the LexA-DNA complex and implications for SOS box measurement.* Nature, 2010. **466**(7308): p. 883-6.

85. Sharrocks, A.D., *The ETS-domain transcription factor family.* Nat Rev Mol Cell Biol, 2001. **2**(11): p. 827-37.

86. Gajiwala, K.S., et al., *Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding.* Nature, 2000. **403**(6772): p. 916-21.

87. Siwek, W., et al., *Crystal structure and mechanism of action of the N6-methyladenine-dependent type IIM restriction endonuclease R.DpnI.* Nucleic Acids Res, 2012. **40**(15): p. 7563-72.

88. Schwartz, T., et al., *Crystal structure of the Zalpha domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA.* Science, 1999. **284**(5421): p. 1841-5.

89. Placido, D., et al., *A left-handed RNA double helix bound by the Z alpha domain of the RNA-editing enzyme ADAR1.* Structure, 2007. **15**(4): p. 395-404.

90. Schwartz, T., et al., *Structure of the DLM-1-Z-DNA complex reveals a conserved family of Z-DNA-binding proteins.* Nat Struct Biol, 2001. **8**(9): p. 761-5.

91. Kahmann, J.D., et al., *The solution structure of the N-terminal domain of E3L shows a tyrosine conformation that may explain its reduced affinity to Z-DNA in vitro.* Proc Natl Acad Sci U S A, 2004. **101**(9): p. 2712-7.

92. Rothenburg, S., et al., *A PKR-like eukaryotic initiation factor 2alpha kinase from zebrafish contains Z-DNA binding domains instead of dsRNA binding domains.* Proc Natl Acad Sci U S A, 2005. **102**(5): p. 1602-7.

93. Teplova, M., et al., *Structural basis for recognition and sequestration of UUU(OH) 3' temini of nascent RNA polymerase III transcripts by La, a rheumatic disease autoantigen.* Mol Cell, 2006. **21**(1): p. 75-85.

94. Yoshizawa, S., et al., *Structural basis for mRNA recognition by elongation factor SelB.* Nat Struct Mol Biol, 2005. **12**(2): p. 198-203.

95. Soler, N., D. Fourmy, and S. Yoshizawa, *Structural insight into a molecular switch in tandem winged-helix motifs from elongation factor SelB.* J Mol Biol, 2007. **370**(4): p. 728-41.

96. Wolin, S.L. and T. Cedervall, *The La protein.* Annu Rev Biochem, 2002. **71**: p. 375-403.

97. Alfano, C., et al., *Structural analysis of cooperative RNA binding by the La motif and central RRM domain of human La protein.* Nat Struct Mol Biol, 2004. **11**(4): p. 323-9.

98. Kitano, K., S.Y. Kim, and T. Hakoshima, *Structural basis for DNA strand separation by the unconventional winged-helix domain of RecQ helicase WRN.* Structure, 2010. **18**(2): p. 177-87.

99. Deng, X., et al., *Structure of the full-length human RPA14/32 complex gives insights into the mechanism of DNA binding and complex formation.* J Mol Biol, 2007. **374**(4): p. 865-76.

100. Meinhart, A., J. Blobel, and P. Cramer, *An extended winged helix domain in general transcription factor E/IIE alpha.* J Biol Chem, 2003. **278**(48): p. 48267-74.

101. Ellenberger, T.E., et al., *The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex.* Cell, 1992. **71**(7): p. 1223-37.

102. Weiss, M.A., et al., *Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA.* Nature, 1990. **347**(6293): p. 575-8.

103. Kim, Y., et al., *Crystal structure of a yeast TBP/TATA-box complex.* Nature, 1993. **365**(6446): p. 512-20.

104. Kim, J.L., D.B. Nikolov, and S.K. Burley, *Co-crystal structure of TBP recognizing the minor groove of a TATA element.* Nature, 1993. **365**(6446): p. 520-7.

105. Cho, Y., et al., *Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations.* Science, 1994. **265**(5170): p. 346-55.

106. Ghosh, G., et al., *Structure of NF-kappa B p50 homodimer bound to a kappa B site.* Nature, 1995. **373**(6512): p. 303-10.

107. Muller, C.W., et al., *Structure of the NF-kappa B p50 homodimer bound to DNA.* Nature, 1995. **373**(6512): p. 311-7.

108. Tahirov, T.H., et al., *Structural analyses of DNA recognition by the AML1/Runx-1 Runt domain and its allosteric control by CBFbeta.* Cell, 2001. **104**(5): p. 755-67.

109. Kovall, R.A. and W.A. Hendrickson, *Crystal structure of the nuclear effector of Notch signaling, CSL, bound to DNA.* EMBO J, 2004. **23**(17): p. 3441-51.

110. Sidote, D.J., et al., *Structure of the Staphylococcus aureus AgrA LytTR domain bound to DNA reveals a beta fold with an unusual mode of binding.* Structure, 2008. **16**(5): p. 727-35.

111. Pavletich, N.P. and C.O. Pabo, *Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A.* Science, 1991. **252**(5007): p. 809-17.

112. Schreiter, E.R. and C.L. Drennan, *Ribbon-helix-helix transcription factors: variations on a theme.* Nat Rev Microbiol, 2007. **5**(9): p. 710-20.

113. Somers, W.S. and S.E. Phillips, *Crystal structure of the met repressor-operator complex at 2.8 A resolution reveals DNA recognition by beta-strands.* Nature, 1992. **359**(6394): p. 387-93.

114. Raumann, B.E., et al., *DNA recognition by beta-sheets in the Arc repressor-operator crystal structure.* Nature, 1994. **367**(6465): p. 754-7.

115. Otwinowski, Z., et al., *Crystal structure of trp repressor/operator complex at atomic resolution.* Nature, 1988. **335**(6188): p. 321-9.

116. Harrison, S.C. and A.K. Aggarwal, *DNA recognition by proteins with the helix-turn-helix motif.* Annu Rev Biochem, 1990. **59**: p. 933-69.

117. Rastinejad, F., et al., *Structure of the RXR-RAR DNA-binding complex on the retinoic acid response element DR1.* EMBO J, 2000. **19**(5): p. 1045-54.

118. Aggarwal, A.K., et al., *Recognition of a DNA operator by the repressor of phage 434: a view at high resolution.* Science, 1988. **242**(4880): p. 899-907.

119. Wolberger, C., et al., *Structure of a phage 434 Cro/DNA complex.* Nature, 1988. **335**(6193): p. 789-95.

120. Watkins, D., et al., *P22 c2 repressor-operator complex: mechanisms of direct and indirect readout.* Biochemistry, 2008. **47**(8): p. 2325-38.

121. Max, K.E., et al., *Common mode of DNA binding to cold shock domains. Crystal structure of hexathymidine bound to the domain-swapped form of a major cold shock protein from Bacillus caldolyticus.* FEBS J, 2007. **274**(5): p. 1265-79.

122. Max, K.E., et al., *T-rich DNA single strands bind to a preformed site on the bacterial cold shock protein Bs-CspB.* J Mol Biol, 2006. **360**(3): p. 702-14.

123. Bewley, C.A., A.M. Gronenborn, and G.M. Clore, *Minor groove-binding architectural proteins: structure, function, and DNA recognition.* Annu Rev Biophys Biomol Struct, 1998. **27**: p. 105-31.

124. Joshi, R., et al., *Functional specificity of a Hox protein mediated by the recognition of minor groove structure.* Cell, 2007. **131**(3): p. 530-43.

125. Rohs, R., et al., *Nuance in the double-helix and its role in protein-DNA recognition.* Curr Opin Struct Biol, 2009. **19**(2): p. 171-7.

126. Rohs, R., et al., *The role of DNA shape in protein-DNA recognition.* Nature, 2009. **461**(7268): p. 1248-53.

127. Gajiwala, K.S. and S.K. Burley, *Winged helix proteins.* Curr Opin Struct Biol, 2000. **10**(1): p. 110-6.

128. Fairall, L., et al., *The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition.* Nature, 1993. **366**(6454): p. 483-7.

129. Hegde, R.S., *The papillomavirus E2 proteins: structure, function, and biology.* Annu Rev Biophys Biomol Struct, 2002. **31**: p. 343-60.

130. Kim, S.S., et al., *The structural basis of DNA target discrimination by papillomavirus E2 proteins.* J Biol Chem, 2000. **275**(40): p. 31245-54.

131. Travers, A.A., *DNA conformation and protein binding.* Annu Rev Biochem, 1989. **58**: p. 427-52.

132. Koudelka, G.B. and P. Carlson, *DNA twisting and the effects of non-contacted bases on affinity of 434 operator for 434 repressor.* Nature, 1992. **355**(6355): p. 89-91.

133. Patikoglou, G. and S.K. Burley, *Eukaryotic transcription factor-DNA complexes.* Annu Rev Biophys Biomol Struct, 1997. **26**: p. 289-325.

134. Tolstorukov, M.Y., et al., *A novel roll-and-slide mechanism of DNA folding in chromatin: implications for nucleosome positioning.* J Mol Biol, 2007. **371**(3): p. 725-38.

135. Satchwell, S.C., H.R. Drew, and A.A. Travers, *Sequence periodicities in chicken nucleosome core DNA.* J Mol Biol, 1986. **191**(4): p. 659-75.

136. Waege, I., et al., *Shuttle vector-based transformation system for Pyrococcus furiosus.* Appl Environ Microbiol, 2010. **76**(10): p. 3308-13.

137. Sommer, B., et al., *Activation of a chimeric Rpb5/RpoH subunit using library selection.* PLoS One, 2014. **9**(1): p. e87485.

138. Van Duyne, G.D., et al., *Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin.* J Mol Biol, 1993. **229**(1): p. 105-24.

139. Kabsch, W., *Xds.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 125-32.

140. Adams, P.D., et al., *Recent developments in the PHENIX software for automated crystallographic structure determination.* J Synchrotron Radiat, 2004. **11**(Pt 1): p. 53-5.

141. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.

142. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.

143. Painter, J. and E.A. Merritt, *Optimal description of a protein structure in terms of multiple groups undergoing TLS motion.* Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 4): p. 439-50.

144. McCoy, A.J., et al., *Phaser crystallographic software.* J Appl Crystallogr, 2007. **40**(Pt 4): p. 658-674.

145. DiMaio, F., et al., *Improved low-resolution crystallographic refinement with Phenix and Rosetta.* Nat Methods, 2013. **10**(11): p. 1102-4.

146. Kabsch, W. and C. Sander, *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.* Biopolymers, 1983. **22**(12): p. 2577-637.

147. Robert, X. and P. Gouet, *Deciphering key features in protein structures with the new ENDscript server.* Nucleic Acids Res, 2014. **42**(Web Server issue): p. W320-4.

148. Baker, N.A., et al., *Electrostatics of nanosystems: application to microtubules and the ribosome.* Proc Natl Acad Sci U S A, 2001. **98**(18): p. 10037-41.

149. Dolinsky, T.J., et al., *PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations.* Nucleic Acids Res, 2004. **32**(Web Server issue): p. W665-7.

150. Dolinsky, T.J., et al., *PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W522-5.

151. Vangone, A., et al., *COCOMAPS: a web application to analyze and visualize contacts at the interface of biomolecular complexes.* Bioinformatics, 2011. **27**(20): p. 2915-6.

152. Lee, S.J., et al., *Characterization of the TrmB-like protein, PF0124, a TGM-recognizing global transcriptional regulator of the hyperthermophilic archaeon Pyrococcus furiosus.* Mol Microbiol, 2007. **65**(2): p. 305-18.

153. Lee, S.J., 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): p. 247-56.

154. Efremov, A.K., 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.

155. Lee, S.J., et al., *TrmB, a sugar sensing regulator of ABC transporter genes in Pyrococcus furiosus exhibits dual promoter specificity and is controlled by different inducers.* Mol Microbiol, 2005. **57**(6): p. 1797-807.

156. Krug, M., et al., *The three-dimensional structure of TrmB, a transcriptional regulator of dual function in the hyperthermophilic archaeon Pyrococcus furiosus in complex with sucrose.* Protein Sci, 2013. **22**(6): p. 800-8.

157. Lu, D., et al., *Crystal structure of TtgV in complex with its DNA operator reveals a general model for cooperative DNA binding of tetrameric gene regulators.* Genes Dev, 2010. **24**(22): p. 2556-65.

158. Ribeiro, J., F. Melo, and A. Schuller, *PDIviz: analysis and visualization of protein-DNA binding interfaces.* Bioinformatics, 2015.

159. Krissinel, E. and K. Henrick, *Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2256-68.

160. Stuckey, J.A. and J.E. Dixon, *Crystal structure of a phospholipase D family member.* Nat Struct Biol, 1999. **6**(3): p. 278-84.

161. Koonin, E.V., *A duplicated catalytic motif in a new superfamily of phosphohydrolases and phospholipid synthases that includes poxvirus envelope proteins.* Trends Biochem Sci, 1996. **21**(7): p. 242-3.

162. Ponting, C.P. and I.D. Kerr, *A novel family of phospholipase D homologues that includes phospholipid synthases and putative endonucleases: identification of duplicated repeats and potential active site residues.* Protein Sci, 1996. **5**(5): p. 914-22.

163. Chen, L., et al., *The hyperthermophile protein Sso10a is a dimer of winged helix DNA-binding domains linked by an antiparallel coiled coil rod.* J Mol Biol, 2004. **341**(1): p. 73-91.

164. Heldwein, E.E. and R.G. Brennan, *Crystal structure of the transcription activator BmrR bound to DNA and a drug.* Nature, 2001. **409**(6818): p. 378-82.

165. Schultz, S.C., G.C. Shields, and T.A. Steitz, *Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees.* Science, 1991. **253**(5023): p. 1001-7.

166. Wolberger, C. and R. Campbell, *New perch for the winged helix.* Nat Struct Biol, 2000. **7**(4): p. 261-2.

167. Amit, R., A.B. Oppenheim, and J. Stavans, *Increased bending rigidity of single DNA molecules by H-NS, a temperature and osmolarity sensor.* Biophys J, 2003. **84**(4): p. 2467-73.

168. Shindo, H., et al., *Solution structure of the DNA binding domain of a nucleoid-associated protein, H-NS, from Escherichia coli.* FEBS Lett, 1995. **360**(2): p. 125-31.

169. Arold, S.T., et al., *H-NS forms a superhelical protein scaffold for DNA condensation.* Proc Natl Acad Sci U S A, 2010. **107**(36): p. 15728-32.

170. Esposito, D., et al., *H-NS oligomerization domain structure reveals the mechanism for high order self-association of the intact protein.* J Mol Biol, 2002. **324**(4): p. 841-50.

171. Laurens, N., et al., *Alba shapes the archaeal genome using a delicate balance of bridging and stiffening the DNA.* Nat Commun, 2012. **3**: p. 1328.

172. Vassart, A., et al., *Sa-Lrp from Sulfolobus acidocaldarius is a versatile, glutamine-responsive, and architectural transcriptional regulator.* Microbiologyopen, 2013. **2**(1): p. 75-93.

173. Napoli, A., et al., *A novel member of the bacterial-archaeal regulator family is a nonspecific dna-binding protein and induces positive supercoiling.* J Biol Chem, 2001. **276**(14): p. 10745-52.

174. Fiorentino, G., et al., *Transcriptional regulation of the gene encoding an alcohol dehydrogenase in the archaeon Sulfolobus solfataricus involves multiple factors and control elements.* J Bacteriol, 2003. **185**(13): p. 3926-34.

175. Kessler, A., et al., *A novel archaeal regulatory protein, Sta1, activates transcription from viral promoters.* Nucleic Acids Res, 2006. **34**(17): p. 4837-45.

176. Xie, Y. and J.N. Reeve, *Transcription by an archaeal RNA polymerase is slowed but not blocked by an archaeal nucleosome.* J Bacteriol, 2004. **186**(11): p. 3492-8.

177. Zhang, Z., L. Guo, and L. Huang, *Archaeal chromatin proteins.* Sci China Life Sci, 2012. **55**(5): p. 377-85.

178. Cam, E.L., et al., *DNA bending induced by the archaebacterial histone-like protein MC1.* J Mol Biol, 1999. **285**(3): p. 1011-21.

179. Crnigoj, M., et al., *Interactions of archaeal chromatin proteins Alba1 and Alba2 with nucleic acids.* PLoS One, 2013. **8**(2): p. e58237.

180. DeLange, R.J., G.R. Green, and D.G. Searcy, *A histone-like protein (HTa) from Thermoplasma acidophilum. I. Purification and properties.* J Biol Chem, 1981. **256**(2): p. 900-4.

181. Guo, L., et al., *Biochemical and structural characterization of Cren7, a novel chromatin protein conserved among Crenarchaea.* Nucleic Acids Res, 2008. **36**(4): p. 1129-37.

182. McAfee, J.G., et al., *Equilibrium DNA binding of Sac7d protein from the hyperthermophile Sulfolobus acidocaldarius: fluorescence and circular dichroism studies.* Biochemistry, 1996. **35**(13): p. 4034-45.

183. McCrary, B.S., S.P. Edmondson, and J.W. Shriver, *Hyperthermophile protein folding thermodynamics: differential scanning calorimetry and chemical denaturation of Sac7d.* J Mol Biol, 1996. **264**(4): p. 784-805.

184. Lee, S.J., et al., *TrmB, a sugar-specific transcriptional regulator of the trehalose/maltose ABC transporter from the hyperthermophilic archaeon Thermococcus litoralis.* J Biol Chem, 2003. **278**(2): p. 983-90.

185. Kengen, S.W., et al., *Purification and characterization of a novel ADP-dependent glucokinase from the hyperthermophilic archaeon Pyrococcus furiosus.* J Biol Chem, 1995. **270**(51): p. 30453-7.

186. Krug, M., et al., *Crystal structure of the sugar binding domain of the archaeal transcriptional regulator TrmB.* J Biol Chem, 2006. **281**(16): p. 10976-82.

187. Lee, S.J., et al., *Differential signal transduction via TrmB, a sugar sensing transcriptional repressor of Pyrococcus furiosus.* Mol Microbiol, 2007. **64**(6): p. 1499-505.

ACKNOWLEDGEMENTS

First of all I am deeply thankful to Prof. Wolfram Welte for providing me the opportunity to carry out my doctoral work under his able supervision. He has been enormously supportive during the course of my PhD. His critical inputs at difficult phases of my projects have made this work a success.

I would like to thank Dr. Karsten Schäfer for teaching me the basics of crystallographic computing and sitting together patiently with me for long hours explaining every critical detail in the most lucid manner.

For tougher crystallographic problems, Prof. Kay Diederichs was always there to help with his incisive inputs. I thank him for his guidance during the entire course of my PhD.

I had many useful discussions with Prof. Hans-Jürgen Apell and I also thank him for being generous in providing me the space in his lab for writing my thesis.

I express my gratitude to Prof. Winfried Boos for showing great interest in my projects and providing useful suggestions during lab discussions.

I thank Prof. Andreas Marx for agreeing to review my work and providing important suggestions.

Claudia Hoffmann for helping me out with both administrative and non-administrative issues.

All the members of the Welte and Diederichs group for providing an excellent working environment in the lab.

Beamline staff at the Swiss Light Source especially Dr. Meitian Wang for his guidance regarding data collection.

Prof M K Mathew of NCBS Bangalore, India for providing me the opportunity to work in his lab and supporting me in realizing my academic endeavours.

My friends Muzamil, Rais, Tafheem, Jeelani, Sameer and Mujeeb for the nice time I had with them.

My parents Khursheed Anwar and Gul-i-Raina for investing all their time and resources and making me what I am today.

My wife Sabahat for her tremendous support during my PhD.

And lastly my daughter Aaminah for the lighter moments at home after a hard day's work.