Department of Biochemistry and Metabolism

DNA topoisomerase VI: characterisation and investigation as a target for inhibitors

Adam Allen

January 2023

This thesis is submitted in partial fulfilment of the requirements of the degree of Doctor Philosophy at the University of East Anglia.

Statement

The work submitted within this thesis is entirely my own, except where due reference has been paid, and has not been submitted to this or any other university as part of any degree.

Table of contents

Chapter **Page**

Abstract

Novel antimicrobials that target *Plasmodium falciparum*, the causative agent of severe malaria, are needed to combat widespread drug resistance. Herbicide resistance is also prevalent and poses a substantial threat to global crop production. DNA topoisomerases (topos) are a group of enzymes that regulate the topological state of DNA during vital cell processes such as DNA replication and transcription. Because of their fundamental cellular importance, topos are major molecular targets for chemotherapeutic and antimicrobial drugs. Topo VI is found ubiquitously in archaea and plants, where it plays critical roles in chromosome segregation and endoreduplication, respectively. Topo VI is also found sporadically in bacteria, algae, and other protists, but it is unclear whether the putative topo VI genes present in *Plasmodium* constitute a canonical topo VI complex. The present study features a series of bioinformatic and biochemical approaches to improve our knowledge of topo VI, and to determine its suitability as a target for inhibitors. Phylogenetic analysis has been used to show that topo VI is widely distributed across all three domains of life but is not present in *Plasmodium*. Furthermore, a coupled ATPase assay has shown that the rate of ATP hydrolysis by topo VI from the archaeon *Methanosarcina mazei* (MmTopo VI) is enhanced ~2-fold in the presence of positively-supercoiled DNA over that in the presence of negatively-supercoiled DNA. A novel high-throughput relaxation assay was also developed that has identified two potent inhibitors of MmTopo VI and yeast topo II. MmTopo VI is a useful model for the study of eukaryotic topo VI, and the plant enzyme should be pursued as a target for herbicides.

Acknowledgments

A PhD project cannot be undertaken alone, and it would be wildly inappropriate to not mention those who have helped me along the way. First and foremost, I must of course thank my project supervisor, Tony Maxwell, who deserves more than the few sentences of appreciation I am about to give him. Tony gave me a second chance to join his lab by offering me a new project that I was much more suitable for, and since then, has shared his near-infinite wisdom with me to guide me on an incredible journey of scientific research. However, what I am most grateful for is Tony's ability to support his students through difficult times, and to mitigate their emotional wellbeing. I could not have asked for a better mentor, and I hope to remain his mentee for the foreseeable future.

I am also very lucky to have been placed in a lab group that has the most perfect blend of scientific professionalism and immature shenanigans. I would like to specifically thank Lesley Mitchenall, Lipeng Feng, and Monica Agarwal, for sharing with me their remarkable scientific expertise, and always being there for me when I required a helping hand. I also could not have imagined having no one to share my PhD journey with, and so I must thank Harriet Gooch and Vicky Baskerville, who I will forever hold close to my heart: we did it guys. And to the remainder of the lab group, thank you, for your ever-present support and kindness.

I will never take for granted the amazing support network that I have been surrounded by for the past four years in Norwich, and I will always be indebted to my friends and loved ones. I would also be a fool to not express my utmost gratitude to my three besties: Emma, Socks, and Soup, who helped me through a difficult time and gave me joy when not much else could. I have not been alone, and so I must dedicate this thesis to everyone at the John Innes Centre and beyond for pushing me over the line.

Abbreviations

% (v/v)	Percentage volume per volume
% (w/v)	Percentage weight per volume
-	Negatively
+	Positively
2YT	2-yeast-tryptone
ABS	Asexual blood stage
ACT	Artemisinin-based combination therapy
ADP	Adenosine diphosphate
ADPNP	Adenosine 5′-(β,γ-imido)triphosphate
AIM	Autoinduction media
ATP	Adenosine triphosphate
At	*Arabidopsis thaliana*
BLAST	Basic Local Alignment Search Tool
bp	Base pair
BR	Brassinosteroid
C	Total number of chromosomes
CASP	Critical Assessment of Structure Prediction
cc	Closed circular
CCP4MG	Collaborative Computational Project Number 4 Molecular Graphics
Co-IP	Coimmunoprecipitation
CTD	C-terminal domain
DB	Dilution buffer
DHJ	Double Holliday junction
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSB	Double-strand break
DTT	Dithiothreitol
DxD	Aspartate–any residue–aspartate
EDTA	Ethylenediaminetetraacetic acid
EtBr	Ethidium bromide
Exo1	Exonuclease 1
FPLC	Fast protein liquid chromatography

FRET	Fluorescence resonance energy transfer
FT	Flow through
G	Gap
G	Gate
GESAMT	General Efficient Structural Alignment of Macromolecular Targets
GHKL	Gyrase, Hsp90, histidine kinase, MutL
GHL	Gyrase, Hsp90, MutL
Glu	Glutamate
GRC	glyphosate-resistant crop
H2TH	Helix-2-turn-helix
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HGT	Horizontal gene transfer
HPLC	High-performance liquid chromatography
Hr	Hour
Hsp90	Heat shock protein 90
HTH	Helix-turn-helix
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kb	kilobases
kD	Kilodaltons
kDNA	Kinetoplast DNA
L	Load
LB	Lysogeny broth
LDH	Lactose dehydrogenase
Lk	Linking number
M	Marker
M	Mitotic
MfpA	*Mycobacterium* fluoroquinolone resistance protein A
Min	Minutes
Mm	*Methanosarcina mazei*
MRN	Mre11, Rad50, Nbs1
MRX	Mre11, Rad50, Xrs2
MW	Molecular weight
MWCO	Molecular weight cut-off

N	Number of homologous pairs of chromosomes
NADH	Nicotinamide adenine dinucleotide
NEB	New England Biolabs
NCBI	National Center for Biotechnology Information
OD	Optical density
O/N	Overnight
p	Putative
P_i	Inorganic phosphate
PCR	Polymerase chain reaction
PEP	Phosphoenolpyruvate
Pf	*Plasmodium falciparum*
PK	Pyruvate kinase
PRP	Pentapeptide repeat protein
Qnr	Quinolone resistance protein
RBC	Red blood cell
RGRP	Arginine-glycine-arginine-proline
RMM	Rec114, Mer2, Mei4
RNA	Ribonucleic acid
RPM	Revolutions per minute
S	Seconds
S	Synthesis
SAR	Stramenopiles, alveolates, Rhizaria
SDSA	Synthesis-dependent strand annealing
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
scDNA	Supercoiled DNA
SmD	Small domain
STEB	Sucrose, Tris, EDTA, bromophenol blue
SD	Synthetic dropout
T	Transport
TEV	Tobacco etch virus
Topo	Topoisomerase
Topo VI-A	DNA topoisomerase VI A subunit
Topo VI-B	DNA topoisomerase VI B subunit
Tw	Twist

UFBoot	Ultrafast bootstrap approximation approach
UV	Ultraviolet
W	Wash
WG	Wheat germ
WHD	Winged-helix domain
Wr	Writhe
WT	Wild-type
YPD	Yeast extract peptone dextrose

List of tables

Page

List of Equations

Page

List of figures

Page

Chapter 1

General introduction

1.1 DNA topoisomerases and linking number

The helical structure of duplex DNA has a fundamental consequence: for DNA metabolism to occur, the two intertwined strands of DNA must separate to expose their bases to the enzymatic machinery. However, the processes that induce local helix unwinding generate a mechanical strain on the DNA that must be relieved by the formation of various topological states. During transcription, a relative rotation between RNA polymerase and the DNA causes overwinding of the DNA ahead of the transcription complex and underwinding behind it, resulting in positive and negative DNA supercoiling, respectively[1,2]. Positive superhelical strain is also produced in front of the replication fork during DNA replication[3], and swivelling of the fork causes the daughter duplexes behind it to intertwine and mature into DNA catenanes[3,4]. Furthermore, during the recombination of inverted DNA repeats, a single supercoiled DNA molecule can also entangle with itself and form a knot[5,6]. To solve these problems, a broad group of enzymes called DNA topoisomerases (topos) have evolved to regulate changes in the topological state of DNA[7].

Topos modify DNA topology by cutting one or both strands of DNA, and either passing the DNA strands through the resulting break or rotating around the uncleaved strand, before resealing the phosphodiester backbone[7]. Transient DNA cleavage by topos involves the formation of a phosphotyrosyl linkage between an active-site tyrosine and a phosphodiester bond[7]. Topos are classified according to whether they introduce transient single-strand breaks (type I) or double-strand breaks (DSBs) (type II) in DNA during the strand-passage reaction[8]. All type I topos have an odd number (I, III, V) and all type II enzymes have an even number (II, IV, VI, VIII), except for DNA gyrase (type II) and reverse gyrase (type I) that are named alternatively due to their unique supercoiling activities[8]. The enzymes are further subdivided into five distinct families (IA, IB, IC, IIA, IIB) based on mechanistic and sequence similarities[8].

Topo enzymes are unique in their ability to change the linking number (Lk) of DNA. The Lk is a topological property of closed-circular DNA (ccDNA) that is defined by the number of times

the two strands of the duplex wind around each other when the helix axis is constrained to lie in a plane[9]. Type I and II topos change the Lk by 1 and 2, respectively, and DNA molecules that are identical except for their Lk are known as topoisomers. The Lk of a ccDNA molecule is always an integer value that is a sum of the twist (Tw) and the writhe (Wr), where Tw is the number of helical turns around the helix axis and Wr is the number of times the helix axis coils in space (**Equation 1.1**)[10]. A fundamental property of ccDNA is that its Lk is fixed in the absence of DNA breaks, and so any change to either Tw or Wr must be compensated by an equal and opposite change in the other[10].

$$Lk = Tw + Wr$$

Equation 1.1| The relationship between twist (Tw), writhe (Wr), and linking number (Lk)

1.2 Classification of type I topoisomerases

The first topo was discovered in 1971 in *Escherichia coli* and is now called topo I[11]. This enzyme is ubiquitous in all three domains of life[12], where it plays a major role in the relaxation of negative supercoils and the decatenation of nicked DNA[13]. Prokaryotic topo I (type IA) functions via an enzyme-bridging strand-passage mechanism, whereby the enzyme is covalently attached to both the 5'-phosphate of the broken DNA strand, known as the gate (G) segment, and to the second DNA strand, known as the transport (T) segment[14]. Topo IA therefore creates a bridge between the two DNA strands allowing the T-segment to be passed through the G-segment before the break is resealed[14]. Unlike the prokaryotic enzyme, eukaryotic topo I (type IB) is able to relax positive supercoils in addition to negative supercoils[15] and functions via a controlled-rotation mechanism[16]. Instead of performing strand-passage, type IB topos covalently attach to the 3'-phosphate of the broken DNA strand and rotates around the 5' non-covalent end before resealing[16]. Type IB topos are essential in mammals[17], and the human enzyme has become an important target in anticancer therapy[18].

Another member of the type IA topo family, topo III, is found in all three domains of life[12] and is homologous to prokaryotic topo I[19]. Like eukaryotic topo I, topo III can relax both negative and positive supercoils[20], however, it differs in its additional ability to decatenate RNA molecules[21] and to resolve precatenanes during chromosome segregation[22]. Topo III acts via a similar mechanism to prokaryotic topo I but can perform strand-passage using either a

single-stranded or a double-stranded T-segment[23]. In higher eukaryotes, topo III exists as two isoforms, IIIα and IIIβ, and topo IIIα is further divided into nuclear and mitochondrial isoforms[24]. Topo IIIβ is the only isoform that has RNA topoisomerase activity[25], and nuclear topo IIIα is essential for chromosome segregation[24]. The final member of the type IA topo family is reverse gyrase, which is unique among type I topos in that it has the remarkable ability to utilise ATP to introduce positive supercoils into DNA, as well as possessing ATP-dependent negative supercoil relaxation activity[26]. Reverse gyrase is present in all hyperthermophilic prokaryotes as well as in some thermophilic bacteria and is likely to be an adaptation for maintaining genome integrity in high temperatures[27]. This enzyme is also unusual in that it is formed by the fusion of a type IA topo to a helicase-like domain[28], suggesting that the mechanism of positive supercoiling has evolved extraordinarily by combining helicase and topoisomerase activities.

The sole member of the type IC topo family is topo V, which has only been discovered in the archaeal genus *Methanopyrus*[27]. Like the type IB enzymes, topo V can relax both negative and positive supercoils via a controlled-rotation mechanism and forms a covalent linkage to the 3' end of the broken DNA strand[29]. However, topo V shares no sequence homology to other type IB enzymes[30] and its core domain exhibits a unique fold not exhibited by any other protein, including the topos[31]. Topo V also possess an atypical active site whereby the spatial arrangement of the catalytic tyrosine is different to other topos[31], and this suggests that the enzyme functions via an alternative cleavage-religation mechanism.

1.3 Classification of type II topoisomerases

Type II topos are ATP-dependent enzymes that are essential for chromosome segregation due to their ability to decatenate newly replicated intertwined DNA molecules, and can resolve knotted and supercoiled DNA topoisomers (**Figure 1.1**)[7]. Type II topos share many common structural modules (**Figure 1.2**), including a GHKL ATPase domain, a transducer domain, a metal-binding toprim domain, and a DNA-cleaving winged-helix domain (WHD)[7]. The type IIA family is typified by eukaryotic topo II, and bacterial topo IV and DNA gyrase, the latter of which is also found ubiquitously in plants and sporadically in archaea and protists[12]. DNA gyrase is a unique member of the type II family in that it can wrap a segment of covalently closed double-stranded DNA around itself and introduce negative supercoils[32]. Most

eukaryotes possess a single topo II enzyme, however, two genetically distinct isoforms of topo II (α and β) are present in vertebrates[33,34], which exhibit differential expression patterns and tissue specificity[35]. Topo II is a homodimeric complex and DNA gyrase and topo IV form heterotetramers comprising two A subunits (GyrA/ParC) and two B subunits (GyrB/ParE). Unlike the type IIB enzymes, type IIA topos possess a tower domain, which forms reinforcing DNA-binding contacts[36], and a coiled-coil domain that forms an additional domain interface[7].

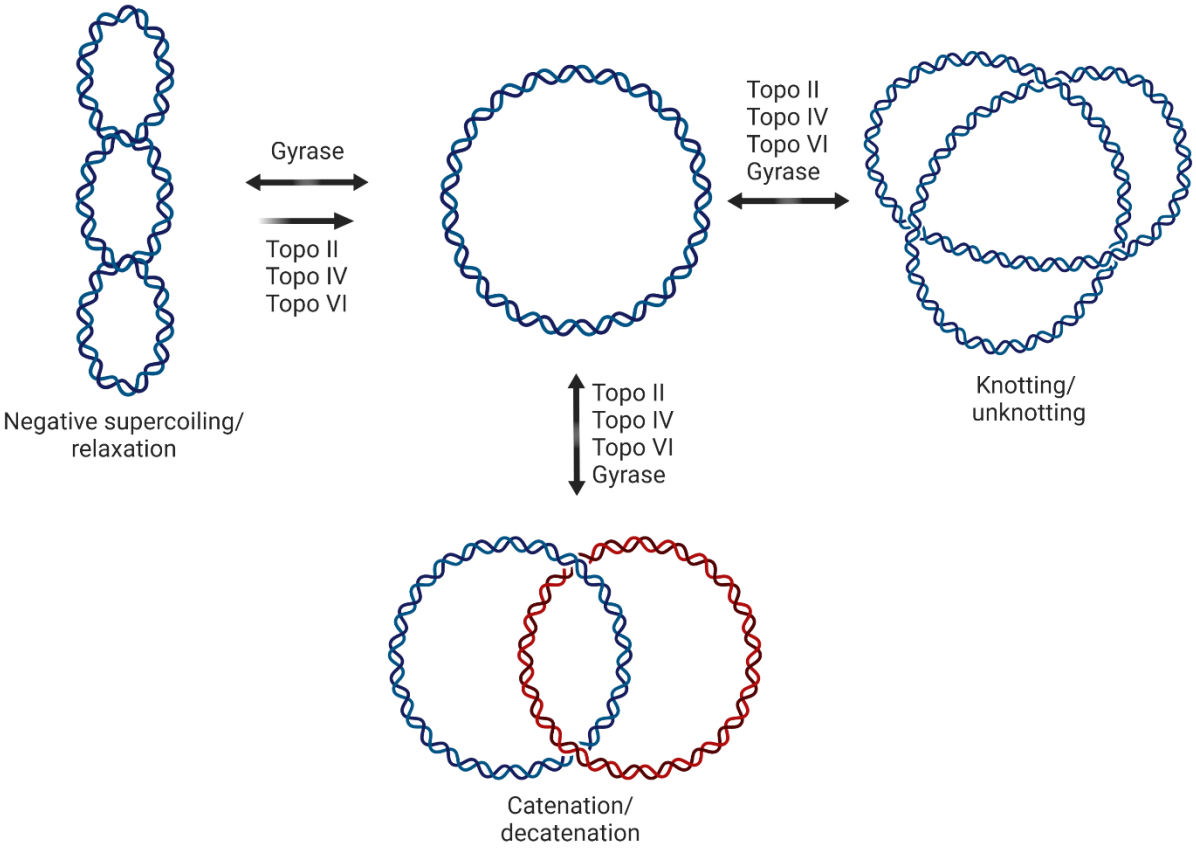

Figure 1.1| DNA topological transformation by type II topoisomerases: *Type II topoisomerases cut both strands of DNA transiently to resolve supercoils, knots, and catenanes. Arrows indicate the specific reactions catalysed by individual type II topoisomerases. Created with BioRender.com*

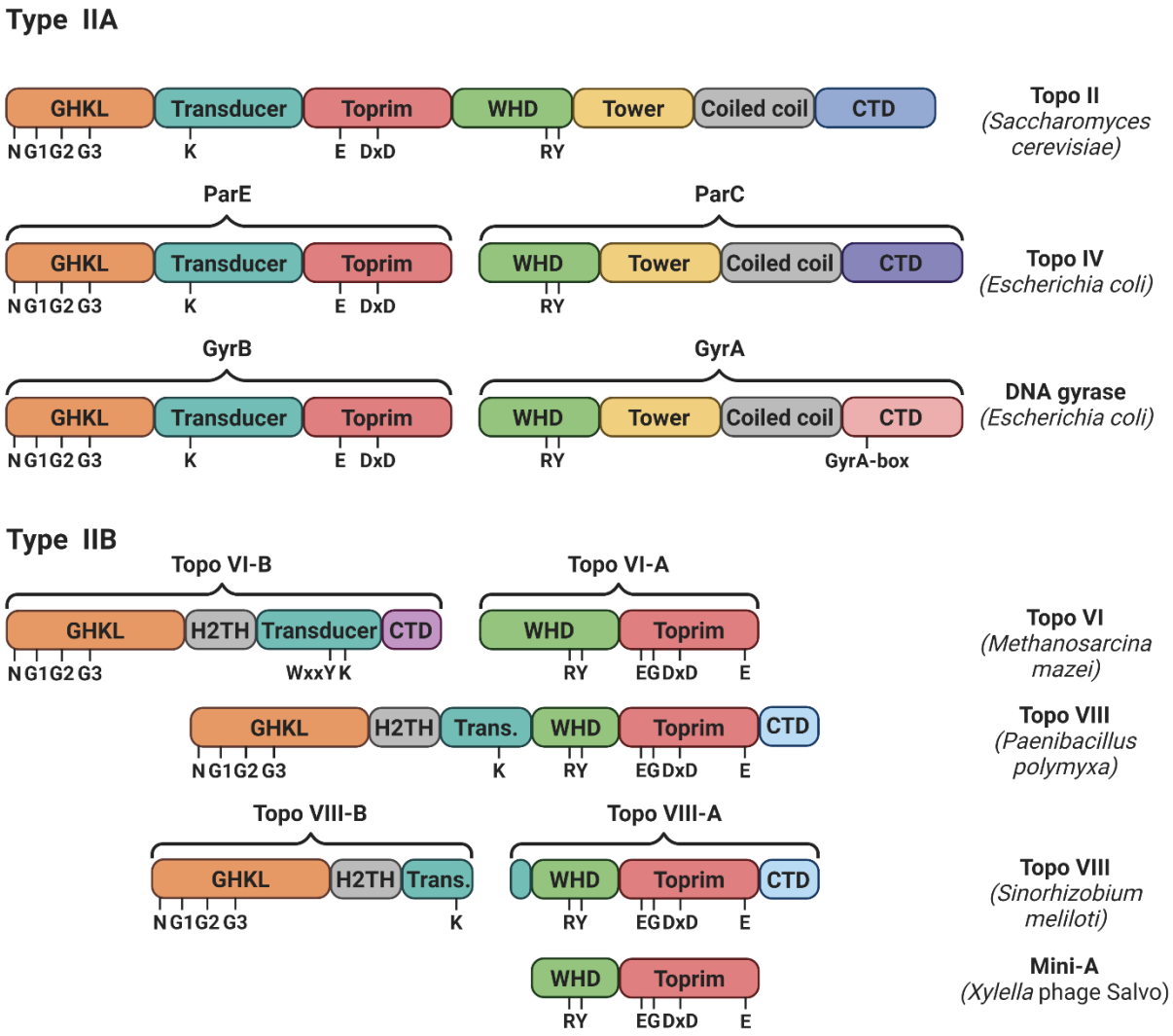

Figure 1.2| Comparative primary domain structures of type II topoisomerases: *The key structural modules of type II topoisomerases include a GHKL ATPase domain, a transducer domain, a metal-binding toprim domain, and a DNA-cleaving winged-helix domain (WHD). Type IIA enzymes possess an additional tower domain and coiled-coil domain, and type IIB enzymes possess an additional helix-2-turn-helix (H2TH) domain. Domains are not drawn to scale, and linker sequences are not included. Black bars indicate key motifs and invariant residues. Created with BioRender.com.*

Although type IIA topos exhibit similar domain architecture, key differences appear in their C-terminal domains (CTDs). In topo II, the CTD is thought to have a regulatory role, and seems to influence dimerisation and nuclear localisation[37]. In DNA gyrase and topo IV, however, the CTD plays an important mechanistic role and possesses key structural features that determine substrate specificity[38,39]. The CTD of DNA gyrase adopts a six-bladed β-pinwheel fold and contains a highly conserved GyrA-box motif, found on a loop connecting blades 1 and 6, which is essential for wrapping and bending the DNA for supercoiling[40-42]. The CTD of topo IV forms

a broken form of the β-pinwheel fold seen in the DNA gyrase CTD, in which the first and last blades are further apart due to the absent of a GyrA-box[43]. Furthermore, the CTD of DNA gyrase is connected to the N-terminal region by a flexible linker, which allows the CTD to rotate and orientate the DNA for wrapping[44,45]. The CTD of topo IV, however, is anchored to the N-terminal domains by a well-ordered linker, which permits only minimal rotation[39,43]. These structural distinctions prevent topo IV from fully wrapping DNA and thus prevent the enzyme from supercoiling DNA.

The discovery of topo VI in the archaeon S*accharolobus shibatae* led to the establishment of the type IIB family due to mechanistic and evolutionary divergence[46]. While the type IIA topos generate DSBs with four-nucleotide overhangs[7], archaeal topo VI creates a two-nucleotide stagger[47]. Topo VI is almost ubiquitous in the archaea domain of life[8,48], and due to the greater availability of sequencing data, has now also been identified in plants[49], bacteria[27], and protists[50]. Topo VI forms a heterotetrameric complex comprising two A subunits (topo VI-A) two B subunits (topo VI-B) (**Figure 1.3**)[46] and is responsible for chromosome segregation[46] and maintaining relaxed intracellular DNA in archaea[51]. Various archaeal species have also acquired DNA gyrase from bacteria via horizontal gene transfer, which allows these organisms to maintain negatively-supercoiled plasmids[51-53]. Interestingly, topo VI is reported to be absent in the archaea order Thermoplasmatales, where supercoil relaxation and decatenation are instead performed by DNA gyrase[8].

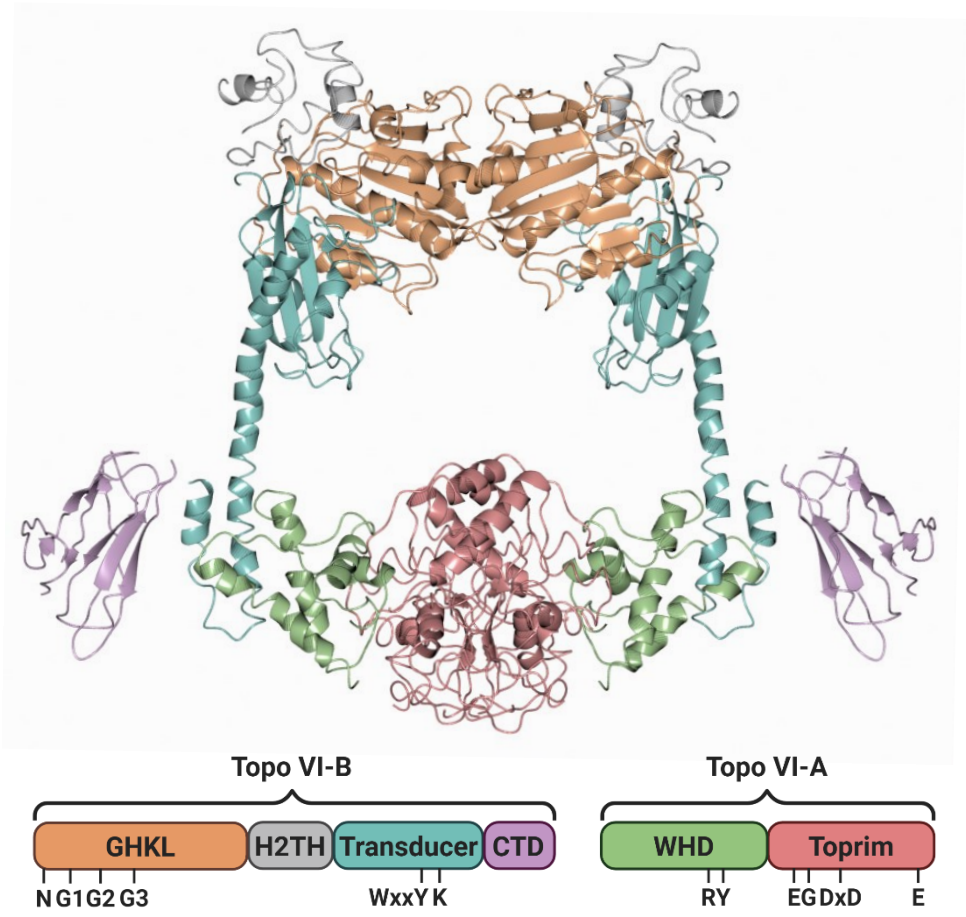

Figure 1.3| Crystal structure of *Methanosarcina mazei* topo VI: *The crystal structure of **M. mazei** topo VI (PDB code: 2Q2E) determined at 4 Å resolution exhibits two A subunits and two B subunits assembled into a heterotetramer. Topo VI-B comprises a GHKL ATPase domain, a helix-2-tun-helix domain, and a transducer domain. Topo VI-A consists of a DNA-cleaving winged-helix domain and a metal-binding toprim domain. The domain architecture diagram is drawn to scale, does not include linker sequences, and was created with BioRender.com*

Another member of the type IIB topo family, topo VIII, is found in free and integrated plasmids in archaea and bacteria[54]. Topo VIII usually takes the form of a single polypeptide homologous to a fusion of topo VI-B and topo VI-A but can also exist as two separate proteins[54]. Notably, the A subunit of topo VIII (topo VIII-A) corresponds to a variant homologous to the C-terminus of topo VI-B fused to topo VI-A (**Figure 1.2**)[54]. Topo VIII enzymes exhibit shorter WHDs and transducer domains than topo VI but possess a unique C-terminal region[54]. Topo VIII from *Paenibacillus polymyxa* has been shown to possess ATP-dependent negative and positive supercoil relaxation and decatenation activities[54], as is typical for the type IIB topos. Remarkably, however, topo VIII from *Microscilla marina* was reported to possess ATP-independent relaxation activity[54], which is notably exhibited by bacterial gyrase[55,56]. The topo

VIII genes are therefore likely to be encode functional enzymes that are required for the maintenance of the host plasmid. Furthermore, a new family of proteins homologous to the A subunit of type IIB topos have recently been discovered in viruses and integrated proviral sequences in archaea and bacteria[57]. These proteins have been called 'Mini-A' since they are characteristically shorter than other topo VI-A homologues[57]. Type IIB topos possess a structural module absent in the type IIA enzymes called the helix-2-turn-helix (H2TH) domain, which is implicated in supercoiled DNA specificity and in coupling ATP hydrolysis with strand-passage[58].

1.4 The GHKL ATPase superfamily

Type II topos possess an evolutionary conserved region common to a superfamily of structurally related ATPases called the GHKL (gyrase, Hsp90, histidine kinase, MutL) domain[59]. GHKL-type ATPases adopt an ATP-binding module called the Bergerat fold, which is an α-β sandwich characterised by the presence of four conserved motifs: N-box, G1-box, G2-box, and G3-box[59,60]. The N-box (ExxxNxxD) contains a conserved asparagine that coordinates an Mg^{2+} ion[61-63], and the G1-box (DNGxG) and G2-box (GxxG) contain conserved glycines that form two hinge segments conferring flexibility to a disordered loop called the ATP-lid[59,62]. The G3-box glycine confers flexibility to another disordered loop that possesses neutral polar residues that form hydrogen bonds with a hydrophobic residue in the G1-box[59]. ATP makes specific contacts with a conserved glutamate in the N-box[62], a conserved aspartate in the G1-box[64], and conserved glycines in the G2-box[59,62]. Mutating the G1-box aspartate drastically decreases nucleotide affinity[65], while mutation of the N-box glutamate or the first glycine in the G2-box abolishes ATPase activity but permits ATP binding[63,66]. The N-box glutamate also serves as a general base by polarising a water molecule for nucleophilic attack on the γ-phosphate of ATP[59,63].

Binding of ATP to the Bergerat motifs induces a conformational change within the GHKL domain, whereby the ATP-lid is repositioned to enclose the ATP-binding pocket and orientate the γ-phosphate for hydrolysis[59,67]. The GHL (gyrase, Hsp90, MutL) subclass of GHKL enzymes are distinct from the histidine kinases and possess a transducer domain and additional conserved motifs. GHL enzymes exist as homodimers in solution and dimerise upon binding of ATP and can therefore serve as molecular clamps. A region at the N-terminus of the GHKL

monomers in GHL proteins, called the strap, interacts with the ATP-lid of the opposite monomer, and stabilises the dimer[61,62].

The transducer domain of GHL enzymes forms an unusual left-handed β-α-β crossover that contains a conserved basic residue located on a region called the switch loop[62,67-69]. In topo VI, the transducer domain features a long α-helix (~68 Å in *Methanosarcina mazei* topo VI[70]) that spans the internal cavity[61,71] and an invariant WxxY motif, located on a loop, that is implicated in G-segment binding[58,72]. ATP-binding induces a rotation of the topo VI transducer domain which shifts the switch-loop into the ATP-binding pocket and positions a conserved lysine for interaction with the γ-phosphate[61,67]. Mutating the switch-lysine to alanine has been shown to inhibit ATP binding in *E. coli* gyrase[73], and mutations to the WxxY motif in *Methanosarcina mazei* topo VI (MmTopo VI) impaired DNA binding and abolished DNA relaxation[58]. The transducer domain thus functions in intersubunit communication, by coupling the structural signals from ATP-binding to a downstream catalytic domain.

1.5 Divalent metal ion-dependent cleavage in type II topoisomerases

The active-site tyrosine of type II topos is found within the WHD, which forms a compact α/β DNA-binding fold similar to the helix-turn-helix (HTH) motif[74]. Mutation of an invariant arginine (WHD-R), that lies next to the catalytic tyrosine, to alanine has been shown to significantly reduce the DNA relaxation and cleavage activity of *Saccharomyces cerevisiae* topo II[75]. To generate DSBs, type II enzymes must form a dimeric interface between their WHDs to position their active-site tyrosines for dual nucleophilic attacks on opposite strands of a DNA duplex.

Type II topos require divalent metal ions to generate DSBs[7] and their WHDs are adjacent to a toprim domain, a structurally conserved domain found in topos, primases, and nucleases[76]. Toprim domains are characterised by a conserved glutamate (toprim-E1) and two conserved aspartates (DxD) which coordinate the essential divalent metal ion[76]. Mutation of the toprim-E1 or either of the DxD aspartates has been shown to inhibit DNA binding in *Escherichia coli* topo I[77,78]. Topo VI homologues also contain a conserved glutamate (toprim-E2) in the active site that hydrogen bonds with the WHD-R and the catalytic tyrosine and contain a conserved glycine (toprim-G) that sits in the DNA-binding pocket[60,79]. Substitution of the toprim-E2 with

alanine has been shown to reduce the activity of *S. cerevisiae* Spo11, a homologue of topo VI-A[60,80], while mutation of the toprim-G inactivates *Schizosaccharomyces pombe* Spo11[81].

1.6 Type II topoisomerase mechanism

Type IIA topos employ a three-gate mechanism for strand-passage, whereby a section of one DNA duplex called the T-segment is trapped at the GHKL dimer interface (N-gate), passed through a DSB in another DNA duplex called the G segment which lies at the WHD interface (DNA gate), and released at the coiled-coil dimer interface (C-gate)[82,83]. The C-gate can only open once the T-segment has passed through the DNA-gate and the G-segment has been rejoined, thereby coupling transient DNA cleavage with strand-passage[84,85]. Type II topos require ATP to perform DNA relaxation and decatenation, even though these reactions can be thermodynamically favourable[7]. It has thus been postulated that type II enzymes manipulate the free energy of ATP hydrolysis to stringently control the opening of the DNA-gate during strand-passage to prevent the formation of permanent DSBs[86].

Topo VI lacks a C-gate and must therefore compensate by tightly controlling G-segment cleavage to ensure that DSBs only occur in the context of strand-passage. The strand-passage reaction in topo VI occurs via a simplified two-gate mechanism, whereby a T-segment is captured at the ATP-gate and passed through a DSB in the G segment at the DNA-gate (**Figure 1.4**)[61,70,79]. The topo VI-A dimer serves as the DNA-gate[79], and the ATP-gate is formed by the GHKL clamp[61]. ATP binding triggers topo VI-B dimerisation which captures the T-segment in the internal cavity and induces G-segment cleavage by the topo VI-A dimer[61,70,79]. The T-segment is then passed through the DNA gate, and the G-segment is resealed. Phosphate release from the hydrolysis of one ATP stimulates strand-passage, before hydrolysis of the second ATP and a final conformational change cause the release of two ADP and one phosphate to reset the enzyme for another cycle[67,87].

Topo VI has been shown to bind preferentially to the DNA crossings present in supercoiled DNA, and to perform DNA relaxation in a distributive manner[58]. Unlike the extremely processive topo II, which relaxes the majority of DNA supercoils while remaining bound to a single DNA duplex[88], topo VI dissociates after resolving a single crossing[58]. When the ATP-gate is open, the topo VI-A dimer active-site tyrosines sit in an inappropriate orientation for phosphotyrosyl linkage[79]. Binding a G-segment accelerates the ATP-dependent dimerisation

of topo VI-B, which is then stabilised by trapping a T-segment[58]. Topo VI-B dimerisation then induces G-segment bending and causes the WHDs to rotate and reposition the catalytic tyrosines closer to the Mg^{2+} binding site and DNA backbone, which primes both the DNA and the topo VI-A dimer for cleavage[58,70,89]. Topo VI therefore strictly coordinates T-segment binding, ATP binding, and G-segment cleavage to couple transient DSB formation with productive strand-passage.

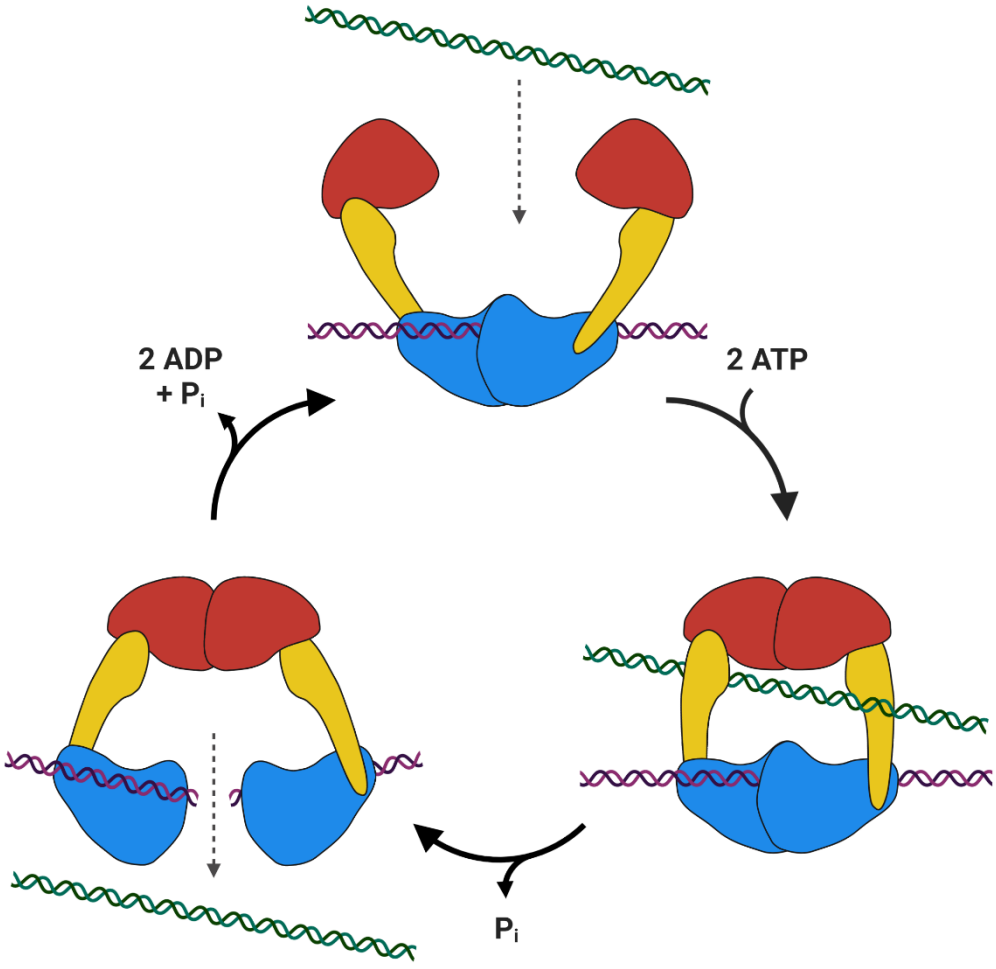

Figure 1.4| Mechanism of strand-passage by topo VI: *Topo VI employs a two-gate mechanism for strand-passage, whereby the T-segment (green) is captured at the ATP-gate (red), passed through the G segment (purple), and released at the DNA-gate (blue). The transducer domain (yellow) forms an internal cavity and couples ATP-binding with G-segment cleavage. The ATP-gate is formed by the GHKL ATPase domains, and the DNA-gate comprises the winged-helix and toprim domains. Created with BioRender.com.*

1.7 Spo11 and meiotic recombination

It has been shown that topo VI-A is highly homologous to Spo11, a eukaryotic protein that introduces meiotic DSBs[60]. Meiosis is the specialised cell division responsible for generating haploid gametes in sexually-reproducing organisms. A key event during meiosis is chromosome segregation, where paired chromosome homologues are separated into daughter cells. Faithful chromosome segregation in the first nuclear division is dependent on meiotic recombination, a process which induces the formation of physical links between homologous chromosomes called chiasmata and mediates the exchange of DNA. Meiotic recombination is initiated during prophase I by the formation of programmed DNA DSBs[90,91]. Meiotic DSB formation in *S. cerevisiae* requires the products of ten genes, of which only Spo11 is functionally conserved[92]. Spo11 was shown to be bound 5′ to DSB intermediates in *S. cerevisiae*[93-95] and to be orthologous to topo VI-A[60]. Spo11 has since been discovered in almost all sequenced eukaryotic genomes and is therefore acknowledged to be an evolutionarily-conserved catalytic subunit of meiotic DSB formation[96].

Spo11-induced DSBs are not resealed, and instead, the covalent Spo11-DNA intermediates are subject to bidirectional resection by exonuclease 1 (Exo1), and the Mre11/Rad50/Xrs2 (MRX) complex in yeast and the Mre11/Rad50/Nbs1 (MRN) complex in mammals[92]. The resulting 3′ single-stranded overhangs can then invade a homologous template and give rise to either crossover or non-crossover products via different recombinant intermediates (**Figure 1.5**)[97]. Most crossovers arise via the double Holliday junction (DHJ) pathway, whereby both 3′ ends anneal to the adjacent template for gap-filling DNA synthesis[98,99]. Non-crossovers are mostly formed by synthesis-dependent strand annealing (SDSA), during which only one 3′ overhang anneals to the template[99,100].

Spo11 features a toprim domain and a WHD, and possesses all the key invariant residues found in topo VI-A[60]. Crystallized *Methanocaldococcus jannaschii* topo VI-A (PDB code: 1D3Y) forms a U-shaped dimer[79], and Spo11 has been shown to self-interact in vivo in *S. cerevisiae*[101]. Only one tyrosine is conserved in Spo11 and type IIB topos, which must therefore be the catalytic residue responsible for phosphotyrosyl linkage[54,57,60]. Indeed, mutation of the corresponding residue in *S. cerevisiae* Spo11 to phenylalanine generates mutants deficient in meiotic recombination[60].

Although most organisms encode a single *spo11* gene, the genomes of plants and protists can encode multiple copies[49,50]. Interestingly, Spo11 appears to be absent in dictyostelids of the genera *Dictyostelium*, *Polysphondylium*, and *Acytostelium*, even though these organisms are thought to be sexual[102-105]. Conversely, several eukaryotes that encode Spo11 in their genomes, such as the excavate *Giardia* and the amoeba *Entamoeba*, do not seem to undergo meiosis[105]. As well as its conserved function in meiotic DSB formation, Spo11 seems to play a noncatalytic role in homologous chromosome pairing prior to entry into prophase I, although this mechanism is poorly understood[106].

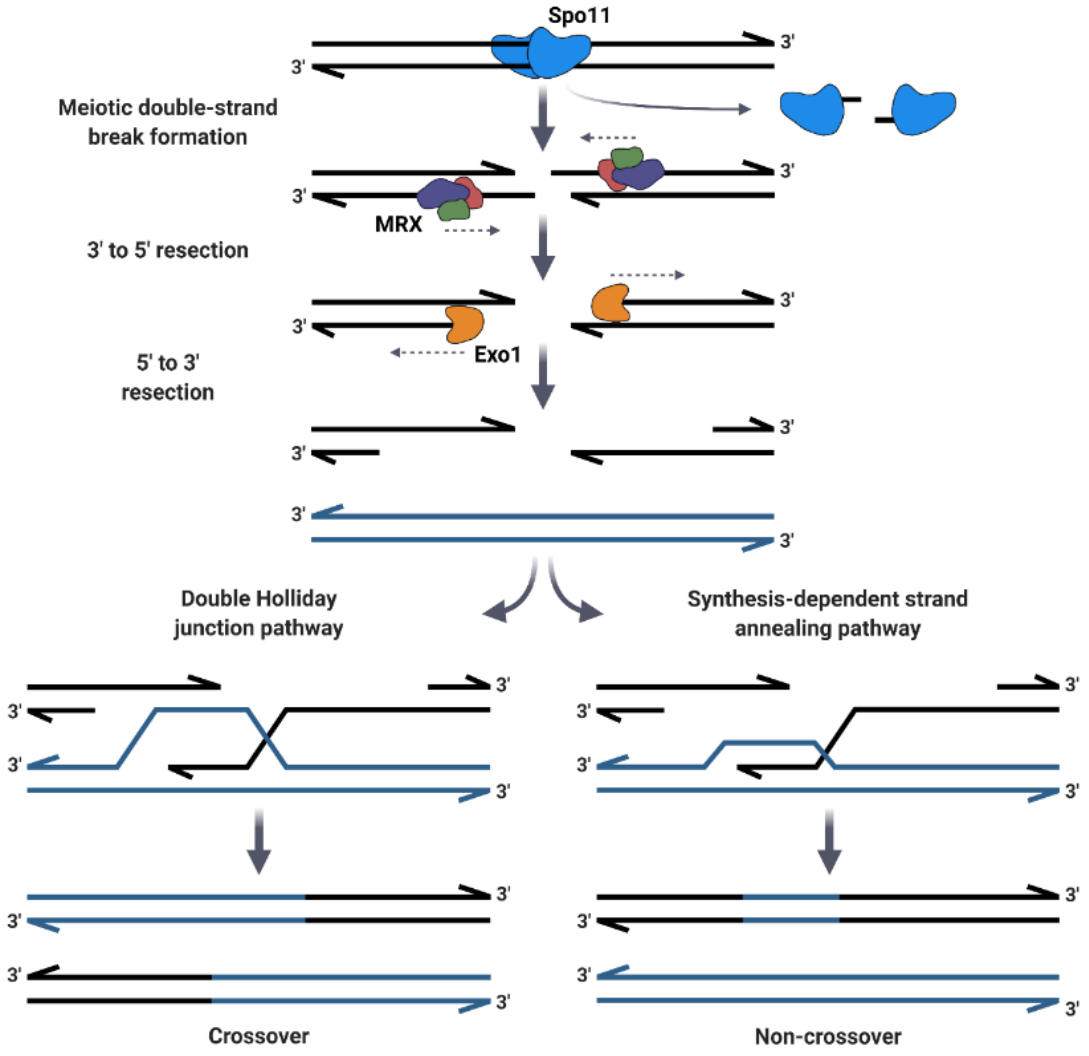

Figure 1.5| Model of DNA double-strand break formation in meiotic recombination: *Spo11 initiates meiotic recombination by inducing DNA double-strand breaks (DSBs), generating covalent protein-DNA intermediates which, in* ***Saccharomyces cerevisiae****, are resected by the Mre11/Rad50/Xrs2 (MRX) complex and exonuclease 1 (Exo1). The DSB can be repaired via the double Holliday junction pathway or the synthesis-dependent strand annealing pathway to yield crossover or non-crossover products, respectively. Created with BioRender.com.*

1.8 *Arabidopsis thaliana* topo VI

The discovery of topo VI in plants paves the way for this enzyme to be pursued as a herbicide target[49]. Weed control significantly improves crop yield and is essential for agricultural productivity. However, due to the constant use of herbicides, resistant weed strains have emerged that pose a substantial threat to global crop production. The International Herbicide-Resistant Weed Database has recorded 513 unique cases of herbicide resistance in 267 species, for 21 of the 31 known molecular sites of herbicide action[107]. Glyphosate is the world's most widely used herbicide and is primarily used in non-selective treatment for broad-spectrum weed control before crop planting[108]. Glyphosate can also be utilised as a selective herbicide in transgenic glyphosate-resistant crops (GRCs), such as soybean and maize, to remove emerging weeds after planting without reducing crop yield[108]. Glyphosate resistance is well-documented, stressing the need for identifying herbicides with novel modes of action[109].

Arabidopsis thaliana has three non-redundant Spo11 paralogues (SPO11-1, -2, -3) and a topo VI-B homologue (TOP6B) and these proteins are widely distributed among plant and algae lineages[49]. Disruption of the *A. thaliana SPO11-1* (*AtSPO11-1*)[110] and *SPO11-2* (*AtSPO11-2*)[111] genes gives rise to a phenotype indicative of severely reduced meiotic recombination. It has also been shown that mutation of the active site tyrosines in At*SPO11-1* and At*SPO11-2* to phenylalanine could not complement their respective T-DNA insertion lines, but that wild-type genomic DNA could[112]. These results suggest that a SPO11-1/SPO11-2 heterodimer is required for meiotic DSB formation in plants.

Endoreduplication is the replication of chromosomal DNA in the absence of mitosis, resulting in increased nuclear ploidy[113]. In plants, endopolyploidy is a mechanism for sustaining growth in response to abiotic stress[114] and can contribute to increasing cell size[113]. *A. thaliana* SPO11-3 (AtSPO11-3) and TOP6B (AtTOP6B) are highly expressed in somatic tissues and have been shown to interact with each other in a yeast two-hybrid assay[49]. Hartung et al also showed that T-DNA insertion mutants of At*SPO11-3* and At*TOP6B* have an identical dwarf phenotype, attributed to reduced cell proliferation and the failure to progress past the 8C ploidy level[115]. Coincidently, Sugimoto-Shirasu et al had identified *spo11-3* and *top6b* at a similar time in a screen of growth-retarded mutants deficient in cell elongation in *A. thaliana*[116]. The mutant plants were fertile, suggesting that neither AtSPO11-3 nor AtTOP6B play a major role in

meiosis[116]. Interestingly, an earlier study had also cloned *AtSPO11-3* and *AtTOP6B* from a screen for plants defective in brassinosteroid (BR) biosynthesis[117]. BR-insensitive mutants exhibit a phenotype characterised by cell elongation defects and dwarfism which is indistinguishable from that of endoreduplication mutants[118]. Taken together, these results indicate that plants contain a functional topo VI formed by SPO11-3 and TOP6B that is essential for endoreduplication. However, there is no evidence of special DNA replication intermediates associated with high endopolyploidy and a plant topo VI has yet to be isolated for biochemical characterisation, and so it remains unclear why topo VI specifically is required for endoreplication in plants.

There is evidence that topo VI plays a key cellular role in plants independent of the resolution of DNA topoisomers. A microarray analysis has identified 321 *A. thaliana* genes that are downregulated by at least two-fold in *spo11-3* and *top6b* mutant plants[117]. Furthermore, virus-induced gene silencing of *TOP6B* in *Gossypium arboreum* has been shown to increase plant sensitivity to drought stress, and overexpression of *G. arboreum* topo VI in transgenic *Arabidopsis* plants can confer tolerance to drought treatment[119]. Transgenic *Arabidopsis* lines overexpressing *Oryza sativa* topo VI exhibit a more than two-fold change in the expression of 322 genes and appear to gain tolerance to high salinity and dehydration[120]. These studies implicate plant topo VI in the transcriptional regulation of stress-response genes and suggest a novel mechanism for the contribution of this enzyme to plant endoreduplication.

Topos often function as part of multi-enzyme complexes and are regulated via physical interactions with other DNA-associated proteins[121]. For example, topo IV has been shown to interact with XerCD recombinases and the DNA-binding protein MatP to regulate chromosome segregation in *E. coli*[122]. The loss of function of two evolutionarily conserved plant proteins in *A. thaliana*, RHL1[123] and BIN4[124], causes precocious endocycle arrest and a dwarf phenotype comparable to *spo11-3* and *top6b* mutants. *RHL1* was identified in a screen for mutants deficient in root hair formation[125], and was shown to encode a small protein with a functional nuclear localisation signal[126]. BIN4 was identified alongside SPO11-3 and TOP6B in a screen for BR-insensitive mutants[117] and was shown to possess an AT-hook motif and a putative nuclear localisation signal[124]. AT-hooks are short DNA-binding motifs that use a conserved arginine-glycine-arginine-proline (RGRP) peptide core to interact with the minor groove of AT-rich DNA sequences[127]. RHL1 and BIN4 have both been shown to interact with

AtSPO11-3, as well as with each other, in vivo, and both exhibit in vitro DNA binding activity[123,124]. Interestingly, RHL1 and BIN4 have been reported to share partial sequence homology with the CTD of mammalian topo IIα[123,124], a poorly characterised region implicated in nuclear localisation and in the regulation of catalytic activity[128]. These discoveries suggest that RHL1 and BIN4 function as accessory components of the plant topo VI complex. Given the success of topos as drug targets, plant topo VI is a potential target for herbicides.

1.9 *Plasmodium* and its putative topo VI

It has been shown that the genome of the highly virulent apicomplexan *Plasmodium falciparum*, which causes the most severe form of malaria, possesses putative topo VI genes[129]. Malaria remains a substantial global health problem, causing an estimated 241 million clinical cases and 627,000 deaths worldwide in 2020[130], and fast-acting artemisinin-based combination therapies (ACTs) are being used as first-line treatment[131]. Resistance to ACT treatment was first seen in western Cambodia, and a prevalence of multi-drug resistant *P. falciparum* is now spreading across southeast Asia[132]. To combat the threat of widespread ACT failure, novel molecular candidates must be placed into the antimalarial drug development pipeline.

Malaria is a vector-borne disease, and the intracellular *Plasmodium* parasites undergo cyclical infections of humans and female *Anopheles* mosquitoes. *Plasmodium* can serially differentiate into morphologically distinct forms (**Figure 1.7**), alternating between invasive stages (sporozoite, merozoite, ookinete), sexual stages (microgametocyte, macrogametocyte), and asexual stages (oocyst, schizont)[133]. An infected *Anopheles* mosquito injects sporozoites into the blood circulation of the human host, which infect hepatocytes. During a process called schizogony, the sporozoites undergo multiple rounds of DNA replication and mature into hepatic schizonts, which rupture and release thousands of merozoites into the bloodstream where they invade red blood cells (RBCs). The merozoites then mature into trophozoites which undergo further schizogony and develop into erythrocytic schizonts, which burst to release merozoites that invade other RBCs. The asexual blood stages (ABSs) of *Plasmodium* are responsible for the clinical manifestations of malaria and are therefore the targets of most antimalarial drugs[134].

A subpopulation of erythrocytic parasites will differentiate into male microgametocytes and female macrogametocytes to initiate the sporogonic cycle[135]. Once ingested by an *Anopheles* blood meal, the gametocytes become activated and undergo a process called gametogenesis. During female gametogenesis the macrogametocytes develop into macrogametes, and during male gametogenesis the microgametocytes undergo rapid DNA replication to give rise to multiple flagellate microgametes. Fertilisation of the macrogamete results in a zygote which differentiates into an ookinete and invades the *Anopheles* midgut lumen. The ookinete then undergoes multiple rounds of DNA replication in a process called sporogony, to give rise to an oocyst, which then ruptures to release sporozoites which migrate to the *Anopheles* salivary gland. During the mosquito's next blood meal, the sporozoites are injected into another human host, perpetuating the *Plasmodium* life cycle. Schizogony, sporogony, and male gametogenesis are processes of endopolyploidy which give rise to polyploid nuclei[136,137].

A phylogenetic analysis has described the broad distribution of the SPO11-1 and SPO11-2 subfamilies in eukaryotes and the presence of multiple SPO111 paralogues in the protozoan phylums Amoebozoa, Euglenozoa, Percolozoa, and Apicomplexa[50]. *Plasmodium* encode two SPO11 paralogues[129,138] (herein referred to as SPO11-A and SPO11-B), that have been likened to the SPO11-1 and SPO11-2 subfamilies[50], and a putative TOP6B (pTOP6B)[129]. The pTOP6B in *Plasmodium* and other Apicomplexa exhibit significant sequence divergence from the topo VI-B homologues in plants and archaea except at the N-terminal region[50]. Given that endopolyploidy is abundant in plants and *Plasmodium*, it is plausible that a canonical topo VI in *Plasmodium* could function analogously to plant topo VI.

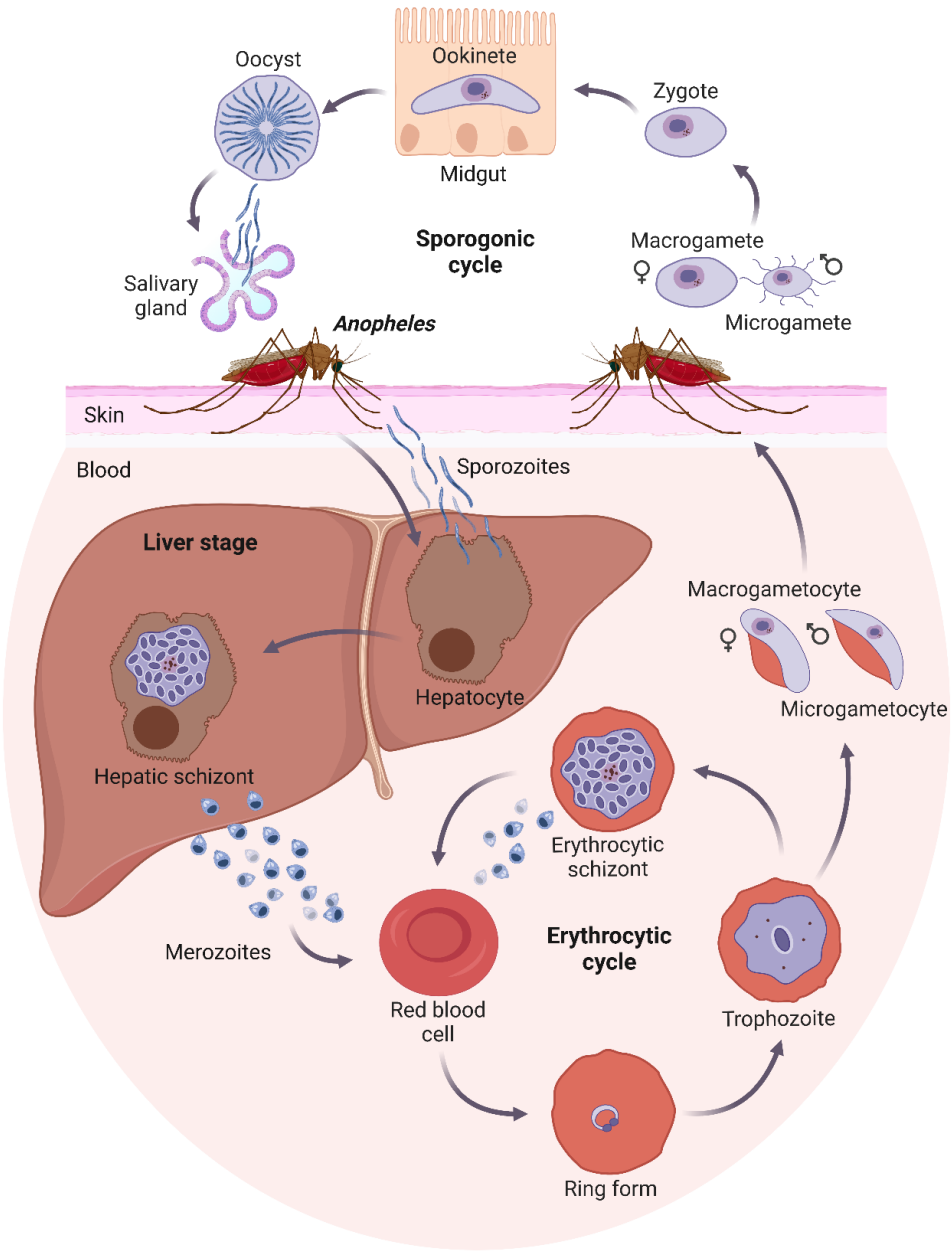

Figure 1.6| The *Plasmodium* life cycle: *Plasmodium* *has a complex developmental cycle within two hosts, involving blood and liver stages in humans and gut and salivary stages in mosquitos. Created with BioRender.com.*

An expression vector harbouring both SPO11-A and the pTOP6B from *P. falciparum* has been shown to functionally complement an *S. cerevisiae* topo II null mutant, suggesting that the putative topo VI (pTopo VI) in *Plasmodium* possesses type IIB topo activity[139]. Furthermore, the *Plasmodium* pTopo VI genes have also been characterised in a series of genome-wide knockout and transcriptomics studies. Zhang et al used piggyBac transposon insertional mutagenesis to measure the relative fitness cost of 5,399 knockout genes in *P. falciparum*[140].

In this study, piggyBac insertions were deleterious in the *SPO11-B* gene during ABS growth but were viable in the *SPO11-A* and *pTOP6B* genes[140]. A similar study by Bushell et al used barcode-tagged mutagenesis to measure the growth rate phenotypes for 2,578 *Plasmodium berghei* knockout genes[141]. Here, *SPO11-B* was deemed essential for normal ABS growth, while *SPO11-A* and *pTOP6B* were dispensable[141]. The same group also performed transcriptome analysis using RNA-sequencing and discovered that *SPO11-A* and *pTOP6B* were upregulated during the sexual developmental stages[141]. Another study by Hall et al looked at the transcriptome profiling of *P. berghei* by microarray analysis and found that *pTOP6B* is upregulated in gametocytes compared to the asexual stages[142]. Together, these studies highlight the similar functional and expression profiles exhibited by SPO11-A and pTOP6B and implicate SPO11-A and not SPO11-B as the functional partner of pTOP6B in *Plasmodium*. These studies also suggest that the *Plasmodium* pTopo VI functions primarily during the sexual developmental stages of the apicomplexan's life cycle. Given the success of topos as drug targets, the *Plasmodium* pTopo VI complex is a potential target for antimalarial drugs.

1.10 The role of eukaryotic topo VI in endopolyploidy

The eukaryotic cell cycle is a process whereby somatic cells divide into two identical daughter cells, and involves two gap (G1, G2) phases, a synthesis (S) phase, and a mitotic (M) phase. The G1 phase is a period of cell growth and protein synthesis to prepare the cell for DNA replication, the S phase is where chromosomes are duplicated into two sister chromosomes, and in the G2 phase cells undergo further metabolic changes to prepare for mitosis. The M phase consists of mitosis, which is a process of nuclear division where the two sister chromosomes segregate to opposite poles of the cell; and cytokinesis, where the cytoplasm divides into two daughter cells. The number of homologous pairs of chromosomes (N) and the total number of chromosomes (C) in the cell remains unchanged after the mitotic cell cycle.

Endopolyploidy is prevalent in certain eukaryotic cells that participate in variations of the cell cycle where cytokinesis is aborted. During an endoreduplicative cell cycle, the G2 phase and the entire M phase are avoided resulting in a doubling of the C ploidy level but maintainence of the N ploidy level after each successive endocycle (**Figure 1.8**)[113]. In plants, the

endopolyploid level can be as high as 16C in leaf epidermal cells, and 64C in leaf trichomes[143]. Topo VI in plants functions specifically in endoreduplicating cells by resolving the chromosome entanglements that occur above the 8C ploidy level[115] which seemingly cannot be processed by topo II nor gyrase. However, high endopolyploidy is also well documented throughout the animal kingdom, including mammals, worms, and arthropods, where topo VI is absent[144]. Follicular cells in *Drosophila*, for example, can increase their ploidy level to 16C following three successive endocycles[145], and endoreduplicating trophoblast cells in rodents can elevate their ploidy level up to 512C[146]. Topo VI therefore does not possess a unique capability of processing endoreduplication intermediates, and so its requirement in plants remains a mystery.

Eukaryotic cells can also increase their endopolyploidy by undergoing another atypical variation of the cell cycle called endomitosis. The endomitotic cell cycle is devoid of cytokinesis, like endoreduplication, but involves a partial M phase in which mitosis is aborted after anaphase A[144]. The sister chromatids in endomitosing cells consequently become encapsulated into separate membranes following chromosome replication forming a single polylobulated nucleus[144]. Unlike endoreduplication, each successive cycle of endomitosis doubles both the N and C ploidy levels of the cell (**Figure 1.8**). Endomitosis is common in fungi[147], but occurs sporadically in animals and plants[144,148].

The *Plasmodium* life cycle involves both endomitotic events in the human host (schizogony and sporogony) and an endoreduplicative event in the *Anopheles* vector (male gametogenesis)[136,137]. The erythrocytic schizonts can reach a ploidy level of up to 16N 16C[149], while hepatic schizonts and oocysts can reach ploidy levels in the hundreds and thousands, respectively[150,151]. The presence of pTopo VI genes in *Plasmodium* has led to speculation that the enzyme could act during schizogony to sustain high endopolyploidy in a function analogous to plant topo VI[129,152]. However, the endopolyploid state of schizonts and other endomitosing cells differs greatly from that of endoreduplicating plant cells, due to the separation of sister chromatids (**Figure 1.8**). Furthermore, whole-genome mutagenesis studies in *Plasmodium* suggest the pTopo VI subunits are not required for erythrocytic schizogony but are upregulated instead during gametogenesis[140-142]. The three successive endocycles which occur during the endoreduplication of haploid male microgametocytes in *Plasmodium,* are followed imminently by chromosome segregation and binary fission[136].

Therefore, while endoreduplicating haploid plant cells can maintain a ploidy level of up to 2N 64C[114,143], endoreduplicating *Plasmodium* cells can only reach a transient 1N 8C ploidy level[136]. Given that endopolyploidy in eukaryotes does not correlate with the presence of topo VI, and that only limited endoreduplication occurs in *Plasmodium*, it is likely that the *Plasmodium* pTopo VI does not play a role in resolving endoreduplicative DNA intermediates.

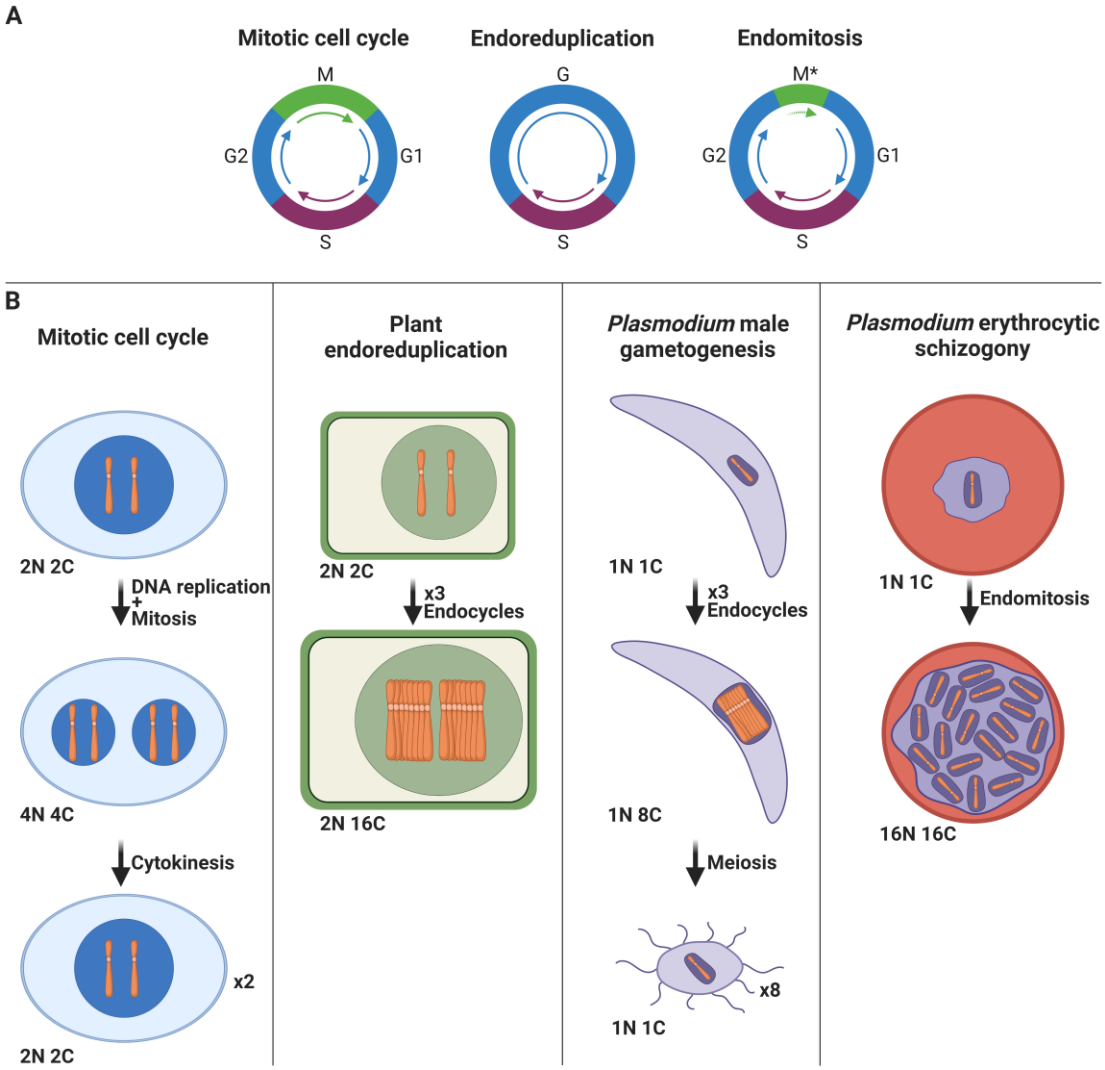

Figure 1.7| Eukaryotic endopolyploidy: ***(A)*** *The mitotic cell cycle involves two gap (G1, G2) phases, a synthesis (S) phase, and a mitotic (M) phase. Endoreduplicating cells skip the G2 and M phase, and endomitosing cells undergo a partial M phase (M*) that is aborted at anaphase A.* ***(B)*** *During the mitotic cell cycle, eukaryotic cells undergo DNA replication, mitosis, and cytokinesis to maintain their original ploidy level. Endoreduplicating plant cells successively replicate their DNA without dividing, increasing their C ploidy level. Male gametogenesis in* ***Plasmodium*** *consists of an endoreduplicative stage that raises the C ploidy level, and a meiotic stage that restores the original ploidy level. During schizogony, Plasmodium cells undergo an endomitotic cell cycle which increases both the C and N ploidy levels. Created with BioRender.com.*

1.11 Aims and objectives

There are many gaps in our understanding of the topo VI enzyme and I will attempt to fill some of them in this thesis. The very first problem I was presented with as I embarked upon my PhD was whether the malaria parasite *Plasmodium* possessed a functional topo VI. Although this genus possesses homologues of both topo VI-A and topo VI-B, there are discrepancies in the literature as to their classification. The second problem I faced was elucidating the broader function of topo VI in eukaryotes, given the special phenotypes of topo VI-deficient plants and the unusual requirement of accessory proteins for topo VI in plants. Furthermore, the presence of topo VI in bacteria is mentioned only in passing in the literature and the significance of this detail is yet to be explored. To gain insights into these uncertainties, a series of bioinformatic analyses were performed in Chapter 3 to mine the key pieces of information present in the amino acid sequences of topo VI across the three domains of life.

Another vital issue that I wished to resolve was the difficulty in isolating an active preparation of eukaryotic topo VI. Further understanding of the cellular roles of topo VI can be gained from in vitro biochemical characterisation of the physical complexes, and so this was a fundamental hurdle in the study of type IIB topos. So far, only limited success has been made in the isolation of the *Arabidopsis* enzyme and Chapter 4 will lay out novel methods to express and purify topo VI from different eukaryotic and bacterial species. Additionally, a series of mutants and fragments of an archaeal topo VI were generated for structural and mechanistic analyses in Chapter 5. The eukaryotic topo VI enzymes warrant the greatest interest of all the type IIB topos given their significance to the pharmaceutical and agricultural industries. However, alternative homologues in archaea can be useful models for studying the plant and *Plasmodium* enzymes given the ease at which they can be isolated. Several mysteries still surround the mechanism of topo VI, such as the role of ATP hydrolysis, the significance the enzyme's supercoil chiral discrimination, and the direction of its strand-passage mechanism. As a prelude to investigating the eukaryotic enzymes, Chapter 5 describes a series of spectrophotometric and gel-based assays that attempt to address these mysteries using topo VI from the archaeon *M. mazei*.

The final piece of my PhD project aims to home in on a fundamental concern regarding topo VI: is it suitable for targeting with inhibitors? Studying the archaeal enzyme identified

interesting mechanistic insights and investigating the eukaryotic enzymes introduced us to fascinating new features of cellular biology. However, for topo VI to be granted widespread attention, it must be druggable. Plant and *Plasmodium* topo VI could represent novel targets for herbicides and antimalarials, respectively, and Chapter 6 aims to kickstart the lead compound discovery pipeline by identifying novel inhibitors of the archaeal enzyme.

In this thesis, I will seek to gain a greater understanding of a neglected enzyme called DNA topoisomerase VI. I will employ a combination of bioinformatic and biochemical approaches to expand our knowledge of the enzyme and will develop a novel tool to investigate its potential as a druggable target. I hope to convince the reader that topo VI deserves greater recognition and that it could one day sit alongside its type IIA cousins on the topo wall of fame.

Chapter 2

Materials and methods

2.1 General materials and methods

2.1.1 Biochemicals, reagents, and equipment

Laboratory chemicals were supplied by Sigma-Aldrich, unless stated otherwise. Restriction enzymes were from New England BioLabs, codon optimised genes were from GenScript, and oligonucleotides were purchased from Sigma-Genosys. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was from Formedium, agarose from Melford Laboratories, and cOmplete™ Mini EDTA-free Protease Inhibitor Cocktail tablets were from Roche. Inspiralis supplied all substrate DNA (pBR322) for biochemical assays, unless stated otherwise.

2.1.2 Gel electrophoresis

Proteins were analysed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) on 10% (w/v) TruPAGE Precast polyacrylamide gels with TruPAGE TEA-Tricine SDS Running Buffer and TruPAGE LDS Sample Buffer. Protein samples were run alongside New England BioLabs Blue Prestained Protein Standard Broad marker and visualised by Expedeon InstantBlue Protein Stain. DNA was analysed by agarose gel electrophoresis on 1% (w/v) agarose gels with Formedium TAE Buffer. DNA samples were run with Purple Gel Loading Dye (New England Biolabs) alongside 1 kb Plus DNA Ladder (New England Biolabs), and gels were stained with ethidium bromide (EtBr). If stated, agarose gels were run in the presence of 500 $ng.mL^{-1}$ EtBr.

2.2 Phylogenetic analysis of the Spo11 and topo VI-B superfamilies

2.2.1 Collection of Spo11 and type IIB topoisomerase sequences

Protein sequences were obtained using the online National Center for Biotechnology Information (NCBI) protein database tool (https://www.ncbi.nlm.nih.gov/protein), by searching for sequences annotated with DNA topoisomerase (topo) VI, type II topo, Spo11, or MTOPVIB identifiers. The protein-protein Basic Local Alignment Search Tool (BLAST)[153] was used to identify protein sequences that lacked annotation by using known topo VI/Spo11 sequences as queries. Sequences from the Mini-A, topo VIII-A, and topo VIII-B families were taken from the supplementary data of Takahashi et al. (2020)[57]. Pooled Spo11/topo VI

homologue sequences were subject to a preliminary alignment, and only sequences that contained the conserved catalytic tyrosine residue (Y106 in *Methanosarcina mazei* topo VI) were retained. Enzyme host organisms were subjected to taxonomic classification using the NCBI Taxonomy Browser[154].

2.2.2 Phylogenetic analysis and structural modelling

Sequences were aligned using the EMBL-EBI Multiple Sequence Alignment tool[155] and the T-Coffee algorithm[156]. Maximum-likelihood phylogenetic trees were constructed using the W-IQ-TREE server[157] and IQ-TREE algorithm[158], and the ultrafast bootstrap approximation approach[159] was used to estimate the phylogenetic support of clades. Phylogenetic trees were visualised using Geneious Prime desktop software[160], and the ESPript 3.0 programme[161] was used to render sequence similarities from the aligned amino acid sequences[161]. Protein structures were modelled using the AlphaFold2 algorithm[162] and ColabFold software[163] and visualised using the CCP4 Molecular Graphics (CCP4MG) program suite[164]. Superimposition of protein structures was performed on CCP4MG using the General Efficient Structural Alignment of Macromolecular Targets (GESAMT) algorithm[165].

2.3 Topo VI purification and expression trials

2.3.1 Molecular cloning

DNA for cloning was isolated using a Qiagen QIAprep Spin Miniprep Kit and purified using a Takara NucleoSpin Gel and PCR clean-up kit. The *M. mazei* topo VI A and B subunits have previously been cloned into the polycistronic and polyhistidine-tagged pST39 plasmid[166], and this was gifted to the lab by James Berger (Johns Hopkins University, USA). *M*. *mazei* topo VI (MmTopo VI) mutants were generated using the pST39 plasmid as a template and a Q5 Site-Directed Mutagenesis Kit (New England BioLabs) using their corresponding mutagenic primers (**Table 2.1**). The topo VI subunits from *Nitrospira moscoviensis, Sedimentisphaera cyanobacteriorum, Pajaroellobacter abortibovis*, and *Arabidopsis thaliana* were cloned into the pST39 vector, and the putative topo VI subunits from *Plasmodium falciparum* were cloned into the 12-URA-B, 12-TRP-B, and 12-ADE-B vectors (gifted by Scott Gradia, QB3 Berkeley, USA) (**Table 2.2**). The top6B-GHKL and top6B-trans truncate constructs from *M. mazei (*residues 1-233 and 1-488), *A*. *thaliana (*residues 1-306 and 1-576*)*, and *P*. *falciparum* (residues 1-230 and 1-450) were cloned into the pET28-MHL vector (Addgene plasmid number 26096). The Takara In-Fusion HD cloning kit and In-Fusion cloning primers (Table 2.1)

were used to clone genes and gene truncates into their corresponding expression vectors. The list of strains used for recombinant protein expression is shown in **Table 2.3** and all cloning was verified by Sanger sequencing. Kanamycin and chloramphenicol were used at working concentrations of 50 µg/mL and 30 µg/mL, respectively.

Table 2.1| Primer oligonucleotide sequences

Primer identifier	Oligonucleotide sequence
MmT6_G137A_Forward	CCGCCCGCATGCGACGCAGATTG
MmT6_G137A_Reverse	AACCAGTCCCTGACCTCGTC
MmT6_D294E_Forward	TCAGCCTTCTGAAATCGTGGAATAC
MmT6_D294E_Reverse	AGCCCCAGGAATTTGGCT
MmT6_R137N_Forward	CAGCCTCCAGAACGAGTATTTCCATATG
MmT6_R137N_Reverse	GTCAGGATTTCCAGGTCC
MmT6_K172N_Forward	CCACTGCCAGAACGATGTGGGAG
MmT6_K172N_Reverse	ATATTCCGCTCTCCGCGC
MmT6_R235N_Forward	ACGGTCAACCAACAGGATAATCAAGC
MmT6_R235N_Reverse	GCAGGCTGACCTTTCAGG
MmT6_K239N_Forward	CAGGATAATCAACCGCATGAACGAAGAACTCGG
MmT6_K239N_Reverse	CGGGTTGACCGTGCAGGC
MmT6_R240N_Forward	GATAATCAAGAACATGAACGAAGAACTCGGGATTCC
MmT6_R240N_Reverse	CTGCGGGTTGACCGTGCA
MmT6_K286N_Forward	ACCGGCAGCCAACTTCCTGGGGC
MmT6_K286N_Reverse	GTTGCCATGAATTCCGAAAGGTGAGC
MmT6_GHKL_Forward	TTGTATTTCCAGGGCATGGAAACCCCCATTGCAG
MmT6_GHKL_Reverse	CAAGCTTCGTCATCATTCTTCCGCAGGCTCAGGC
MmT6_Trans_Forward	TTGTATTTCCAGGGCATGGAAACCCCCATTGCAGAAGA
MmT6_Trans_Reverse	CAAGCTTCGTCATCATGCTGCTAATTTCGGAAGAACCT
AtT6_GHKL_Forward	TTGTATTTCCAGGGCATGTCAACTGGTTCTTCTCACCA
AtT6_GHKL_Reverse	CAAGCTTCGTCATCAGTATGGAGTAATAACAGCCATTTGT
AtT6_Trans_Forward	TTGTATTTCCAGGGCATGTCAACTGGTTCTTCTCACC
AtT6_Trans_Reverse	CAAGCTTCGTCATCATTGTTCTCTTGCTTGCAATCTC
PfT6_GHKL_Forward	TTGTATTTCCAGGGCATGTCAACTGGTTCTTCTCA
PfT6_GHKL_Reverse	CAAGCTTCGTCATCAAACCAATTTAATGTAAGTATAGATT
PfT6_Trans_Forward	TTGTATTTCCAGGGCATGTCAACTGGTTCTTCTCACCA
PfT6_Trans_Reverse	CAAGCTTCGTCATCAGTAATCTGAAATAGATTCAAATTCA
PfT6B_12-URA-B_Forward	TACTTCCAATCCAATGAAACTTTGAACGATAAGA
PfT6B_12-URA-B_Reverse	TTATCCACTTCCAATTTACATAATATCATTAATTTCATT
PfSPO11A_12-TRP-B_Forward	TACTTCCAATCCAATCCAAGATTGGATATTATTTGTTC
PfSPO11A_12-TRP-B_Reverse	TTATCCACTTCCAATTTATAACAATTCTTTAATTC
PfSPO11B_12-ADE-B_Forward	TACTTCCAATCCAATATTAATTGTAATAGAGTTAATAAGA
PfSPO11B_12-ADE-B_Reverse	TTATCCACTTCCAATTTATGTCAACCATTCTTTTCT
AtT6B_pST39_Forward	ACGGTTTCCCTCTAGGCTGGTGACGATTTGGTTGA
AtT6B_pST39_Reverse	TACAGGGCCCGGATCTTACAACATTAATCTAAAAACAAAT
AtT6BA_pST39_Forward	GACGGCCAGTGAATTGCAGATAAAAAGAAAAGAAAGAGAT
AtT6BA_pST39_Reverse	AACCCTGGCGAAGCTTTACAACCAATCTTGTTGTTGCAA
AtT6AR_pST39_Forward	GGGTTTTCCCAGTCGGTTAGAGCAAGTAGTTCCAAGAAGG
AtT6BAR_pST39_Reverse	AATTGTTATCCGCTGTTATGCCTTGGAGGAGGACTTAGC
AtT6BARB_pST39_Forward	CAATTTCACATCCGGTCCTCCAGTTCCAGAGAAGG
AtT6BARB_pST39_Reverse	ATCTCTCGAGACGCGTCACTTTTTGGCTTTTGGCTTC
NmT6A_pST39_Forward	GACGGCCAGTGAATTATGCATCACCACCACCACC
NmT6A_pST39_Reverse	AACCCTGGCGAAGCTGTCCAGCCAATCACGCTCG
NmT6BA_pST39_Forward	ACGGTTTCCCTCTAGATGCATCACCACCACCACC
NmT6BA_pST39_Reverse	TACAGGGCCCGGATCTTTCTTACCTCTTGGCCTGTGC
PaT6A_pST39_Forward	GACGGCCAGTGAATTATGCATCACCACCACCACC
PaT6A_pST39_Reverse	AACCCTGGCGAAGCTCGTCAGGTAATCTTTGTCGCGC
PaT6BA_pST39_Forward	ACGGTTTCCCTCTAGATGCATCACCACCACCAC
PaT6BA_pST39_Reverse	TACAGGGCCCGGATCCGCAAAACCATGATTGTGG
ScT6B_pST39_Forward	ACGGTTTCCCTCTAGATGCATCACCACCACCACC
ScT6B_pST39_Reverse	TACAGGGCCCGGATCGAATTTCCGGGAGCGCTC
ScT6BA_pST39_Forward	GACGGCCAGTGAATTATGCATCACCACCACCACC
ScT6BA_pST39_Reverse	AACCCTGGCGAAGCTTGGCAGAAACTTATTAGTATTAGCG

Table 2.2| Expression vectors for recombinant protein expression

Gene origin	Gene orientation	Vector	Selection marker	Expression organism
Methanosarcina mazei	*top6B_top6A*	pST39	Kanamycin	*Escherichia coli*
Methanosarcina mazei	*top6B* -GHKL	pET28-MHL	Kanamycin	*Escherichia coli*
Methanosarcina mazei	*top6B* -trans	pET28-MHL	Kanamycin	*Escherichia coli*
Nitrospira moscoviensis	*top6B_top6A*	pST39	Kanamycin	*Escherichia coli*
Pajaroellobacter abortibovis	*top6B_top6A*	pST39	Kanamycin	*Escherichia coli*
Sedimentisphaera cyanobacteriorum	*top6B_top6A*	pST39	Kanamycin	*Escherichia coli*
Arabidopsis thaliana	*TOP6B _SPO11-3 _RHL1 _BIN4*	pST39	Kanamycin	*Escherichia coli*
Plasmodium falciparum	*pTOP6B*	12-URA-B	Uracil	*Saccharomyces cerevisiae*
Plasmodium falciparum	*SPO11-A*	12-TRP-B	Tryptophan	*Saccharomyces cerevisiae*
Plasmodium falciparum	*SPO11-B*	12-ADE-B	Adenine	*Saccharomyces cerevisiae*

Table 2.3| Expression strains for recombinant protein expression

Expression strain	Organism	Resistance	Source
Rosetta 2(DE3)pLysS	*Escherichia coli*	Chloramphenicol	Novagen
BL21(DE3)pLysS	*Escherichia coli*	Chloramphenicol	Novagen
Lemo21(DE3)	*Escherichia coli*	Chloramphenicol	New England Biolabs
Tuner(DE3)pLacI	*Escherichia coli*	None	Novagen
SY991	*Saccharomyces cerevisiae*	None	Tomlin et al, 2001
BMA64-1A	*Saccharomyces cerevisiae*	None	Entian et al, 2007
MYA-3332	*Saccharomyces cerevisiae*	None	ATCC

2.3.2 Transformations

For bacterial transformations, tubes of competent cells were removed from -80°C and thawed on ice for 20 min. 1 µL of DNA (~1-100 ng) was added to each tube, mixed by gentle flicking, and placed on ice for 30 min. Rosetta 2(DE3)pLysS cells were then heat-shocked at 42°C for 30 s, BL21(DE3)pLysS cells for 18 s, Lemo21(DE3) cells for 10 s, and tuner(DE3)pLacI cells for 30 s before being placed back on ice for 2 min. 250 µL of Lysogeny Broth (LB) (**Table 2.4**) was added to each tube, and the cells were incubated at 37°C for 1 hr. 5 µL, 50 µL, and 150 µL of cells were then spread on separate LB selection plates and incubated overnight at 37°C. Yeast cells were transformed as per the protocol in Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual[167]. Growth media plates possessed 2% agar.

Table 2.4| Growth media components: **denotes amino acids removed for auxotrophic selection.*

Growth medium	Component	Concentration ($g.L^{-1}$)
2-Yeast-tryptone	Sodium chloride	5
	Tryptone	16
	Yeast extract	10
Lysogeny broth	Sodium chloride	10
	Tryptone	10
	Yeast extract	5
Auto induction medium	Tryptone	10
	Yeast extract	5
	Ammonium sulphate	3.3
	Glucose	0.5
	Lactose	2
	Magnesium sulphate	0.15
	Potassium dihydrogen phosphate	6.8
Yeast extract peptone dextrose	Bacteriological peptone	20
	Glucose	20
	Yeast extract	10
Synthetic dropout	Yeast nitrogen base	6.7
	Glucose	20
	Adenine*	0.01
	L-Arginine	0.05
	L-Aspartic acid	0.08
	L-Histidine	0.02
	L-Isoleucine	0.05
	L-Leucine	0.1
	L-Lysine	0.05
	L-Methionine	0.02
	L-Phenylalanine	0.05
	L-Threonine	0.1
	L-Tryptophan*	0.05
	L-Tyrosine	0.05
	L-Valine	0.14
	Uracil*	0.02

2.3.3 Protein expression in *Escherichia coli*

The pET series of expression vectors, which includes pET28-MHL, are derived from the pBR322 plasmid and are widely used to support high levels of transcription. This series of vectors utilise the T7 RNA polymerase expression system[166], whereby the target gene is cloned downstream of the T7 promoter, and expression of the T7 gene is prevented by the Lac repressor. This system can be induced using AIM[168], which contains glucose and lactose, with the former being the preferred carbon source. Once the glucose is depleted, lactose is converted into allolactose, releasing the Lac repressor to initiate transcription. The Lac

repressor can also be released by IPTG, a structural homologue of allolactose. The pST39 expression vector was engineered from the pET3a plasmid and contains multiple cloning sites to permit the coexpression of multiple genes under the control of the same promoter[166]. Proteins were expressed in derivatives of BL21(DE3) *Escherichia coli* cells, which contain the lambda DE3 phage construct carrying the T7 RNA polymerase gene. BL21 cells are void of Lon protease and OmpT protease, which degrade heterologous proteins in the cytoplasm and outer membrane, respectively. Rosetta 2(DE3)pLysS is a BL21(DE3)-derived cell line that carries tRNA genes that encode for seven codons that are rarely used in *E*. *coli* (AGA, AGG, AUA, CUA, GGA, CCC, and CGG) on an additional pLysS plasmid. The pLysS plasmid also encodes T7 lysozyme, which is a natural inhibitor of T7 RNA polymerase that can reduce the background expression of genes downstream of the T7 promoter.

For large-scale protein expression in *E*. *coli*, fresh transformants were inoculated into 4 mL of LB and incubated at 37°C and 200 RPM for 5 hr. 500 µL of culture was then added to 10 mL of LB and incubated at 37°C and 200 RPM overnight. To express MmTopo VI and its mutants and truncates, 1 mL of the overnight culture was added to 6 x 1 L of auto-induction media (AIM)[168] (**Table 2.4**) and incubated at 37°C and 200 RPM for 24 hr. For the bacterial and plant topo VI expression trials, induction in AIM was performed for 24 hr and 48 hr at 20°C, 30°C, and 37°C. Expression trials also consisted of adding 1 mL of the overnight culture to 1 L of 2-yeast-tryptone (2YT) media (**Table 2.4**) and incubating at 37°C and 200 RPM until cells reached mid-log phase (OD_{600} = 0.8). Protein expression was then induced by incubating cells with 1 mM IPTG for 4 hr at 37°C or 300 µM IPTG for 16 hr at 20°C. Cells were harvested by centrifugation at 4000 x *g* for 20 min, resuspended in 10 mL of purification buffer A (**Table 2.5**), flash frozen in liquid nitrogen, and stored in -80°C. Cells were lysed under high pressure (12 kpsi) using an Avestin Emulsiflex high pressure homogeniser, and the soluble and insoluble lysates were separated by centrifugation at 40,000 x *g* at 4°C for 1 hr. Pre- and post-induction samples and the soluble and insoluble fractions were analysed by SDS-PAGE.

Table 2.5| Buffer compositions

Buffer	Composition
Purification buffer A	20 mM HEPES (pH 7.5), 10% (v/v) glycerol, 800 mM NaCl, 20 mM Imidazole, 2 mM β-mercaptoethanol, ETDA-free protease inhibitors
Purification buffer B1	20 mM HEPES (pH 7.5), 10% (v/v) glycerol, 150 mM NaCl, 20 mM Imidazole, 2 mM β-mercaptoethanol, and ETDA- free protease inhibitors
Purification buffer B2	20 mM HEPES (pH 7.5), 10% (v/v) glycerol, 150 mM NaCl, 500 mM Imidazole, 2 mM β-mercaptoethanol
Purification buffer C1	20 mM HEPES (pH 7.5), 10% (v/v) glycerol, 50 mM NaCl, 2 mM β-mercaptoethanol,
Purification buffer C2	20 mM HEPES (pH 7.5), 10% (v/v) glycerol, 2 mM β-mercaptoethanol, 800 mM NaCl,
Purification buffer D	20 mM HEPES (pH 7.5), 10% (v/v) glycerol, 300 mM NaCl, 2 mM β-mercaptoethanol, ETDA-free protease inhibitors
Dilution buffer	20 mM HEPES (pH 7.5), 10% (v/v) glycerol, 100 mM potassium glutamate, 2mM β-mercaptoethanol
Minimal Buffer	20 mM bis-tris propane (pH 7.0), 10 mM $MgCl_2$ 1 mM DTT, 100 mM potassium glutamate
ATPase buffer	400 µM NADH, 800 µM phosphoenolpyruvate, 1X minimal buffer, 6-10/9-1.4 units.mL^{-1} pyruvate kinase/lactate dehydrogenase
Yeast lysis buffer	50 mM Tris (pH 7.5), 1 mM EDTA, 1 mM EGTA, 10% glycerol (v/v), 5mM β-mercaptoethanol, ETDA-free protease inhibitors
STEB	40% (w/v) sucrose, 100 mM Tris-HCl (pH 8.0), 100 mM EDTA, 500 µg.mL^{-1} bromophenol blue

2.3.4 Protein expression in S*accharomyces cerevisiae*

For protein expression in yeast, glycerol stocks of *Saccharomyces cerevisiae* transformed with topo VI expression vectors were streaked out on yeast extract peptone dextrose (YPD) (**Table 2.4**) plates and incubated at 30°C for 72 hr. Single colonies were then inoculated into 10 mL of synthetic dropout (SD) media (**Table 2.4**) and incubated at 30°C and 215 RPM overnight. 200 µL of the overnight cultures were added to 20 mL SD media and 2% (w/v) lactic acid and incubated at 30°C and 215 RPM until reaching an OD_{600} of 0.8. Protein expression was then induced by adding galactose to a final concentration of 2% (w/v) and incubating for a further 7 hr. Cells were harvested by centrifugation at 4000 x *g* for 10 min, resuspended in 1 mL of yeast lysis buffer (**Table 2.5**), and lysed by vortexing for 40 x 20 s with 0.5 mm glass beads (Sigma-Aldrich). Pre- and post-induction samples were analysed by SDS-PAGE. *S. cerevisiae* (unknown strain gifted by Nick Burton, Inspiralis, UK) possessing the yeast topo II expression vector YEpTOP2PGAL1[169] was grown in SD media lacking uracil. The *P. falciparum* putative topo VI expression vectors, 12-URA-B_pTOP6B, 12-TRP-B_SPO11-A, and 12-ADE-B_SPO11-B

were each transformed into SY991[170] *S. cerevisiae* cells and were grown in SD media lacking uracil, tryptophan, and adenine, respectively. To test the effect of coexpressing the *P. falciparum* putative topo VI subunits, combinations of the expression vectors were cotransformed into SY991 cells and were grown in SD media lacking the corresponding amino acids (**Table 2.6**). 12-URA-B_pTOP6B and 12-TRP-B_SPO11-A coexpression was also tested in BMA64-1A[171] and MYA3332 *S. cerevisiae* cells. The 12-X-B series of expression vectors and YEpTOP2PGAL1 possess the GAL1 promoter region of *S. cerevisiae*, which is induced when *S. cerevisiae* cells are grown in the presence of galactose and the absence of glucose.

Table 2.6| Combinations of *Plasmodium falciparum* putative topo VI expression vectors

Plasmodium falciparum **putative topo VI expression vector(s)**	**Dropout amino acid/nucleobase(s)**
12-URA-B_pTOP6B	Uracil
12-TRP-B_SPO11-A	Tryptophan
12-ADE-B_SPO11-B	Adenine
12-URA-B_pTOP6B, 12-TRP-B_SPO11-A	Uracil, tryptophan
12-URA-B_pTOP6B, 12-ADE-B_SPO11-B	Uracil, Adenine
12-URA-B_pTOP6B, 12-TRP-B_SPO11-A, 12-ADE-B_SPO11-B	Uracil, tryptophan, adenine

2.3.5 Purification of *Methanosarcina mazei* topo VI and its mutants and truncates

The procedure for purifying MmTopo VI (**Figure 2.1**) was adapted from a protocol developed by Shannon McKie (John Innes Centre, UK)[172]. Purification was performed by fast protein liquid chromatography (FPLC) at 4˚C on an ÄKTA pure system with a flow rate of 1.0 mL.min^{-1} using chromatography columns from GE Healthcare. Soluble lysate was passed over a HisTrap FF 5 mL column and washed with 10 column volumes of purification buffer B1 (**Table 2.4**). Proteins were eluted in 5 ml fractions by applying a linear gradient of purification buffer B2 (**Table 2.4**) in 100 mL. Fractions were analysed by SDS-PAGE, and the theoretical isoelectric point was determined using the online ExPASy ProtParam tool (https://web.expasy.org/protparam/)[173]. Pooled fractions from the HisTrap FF column possessing MmTopo VI were then passed over a 5 mL HiTrap SP HP column followed in tandem by a 5 mL HiTrap Q HP column and washed with purification buffer C1 (**Table 2.4**). The SP column was removed and proteins were eluted from the Q column in a gradient of purification buffer C2 (**Table 2.4**). Tobacco etch virus (TEV) protease was added at a 1:50 concentration ratio (TEV:MmTopo VI) to pooled Q column fractions and incubated overnight

at 4˚C to remove the polyhistidine tags. The cleaved protein was passed back over the FF column, washed with purification buffer B1, and collected in the flow-through. The protein sample was concentrated to 500 µL using a Merck Millipore Amicon Ultra-15 Centrifugal Filter Unit (10,000 MWCO) and passed through a Superdex 200 10/300 GL column. Proteins were eluted in purification buffer D (**Table 2.4**) in 2 mL fractions and individual fractions were concentrated to ~300 µL and analysed by SDS-PAGE. Protein concentration was determined by UV absorbance (280 nm) adjusted by molar extinction coefficient. Concentrated protein was flash frozen in liquid nitrogen and stored in -80°C.

Mutants of MmTopo VI were purified using the same procedure as the wild-type enzyme. The topo VIB-GHKL and topo VIB-trans truncates, however, were purified using a slightly modified procedure that replaces the HiTrap Q HP column with a 5 mL HiTrap Heparin HP and the Superdex 200 10/300 GL column with a Superdex 75 10/300 GL column (**Figure 2.1**). Furthermore, the MmTopo VI truncates were concentrated with an Amicon Ultra-15 Centrifugal Filter Unit with a 3 kDa cut off. The isolated *M*. *mazei* topo VIB-GHKL truncate was validated by positive electrospray ionisation time-of-flight mass spectrometry.

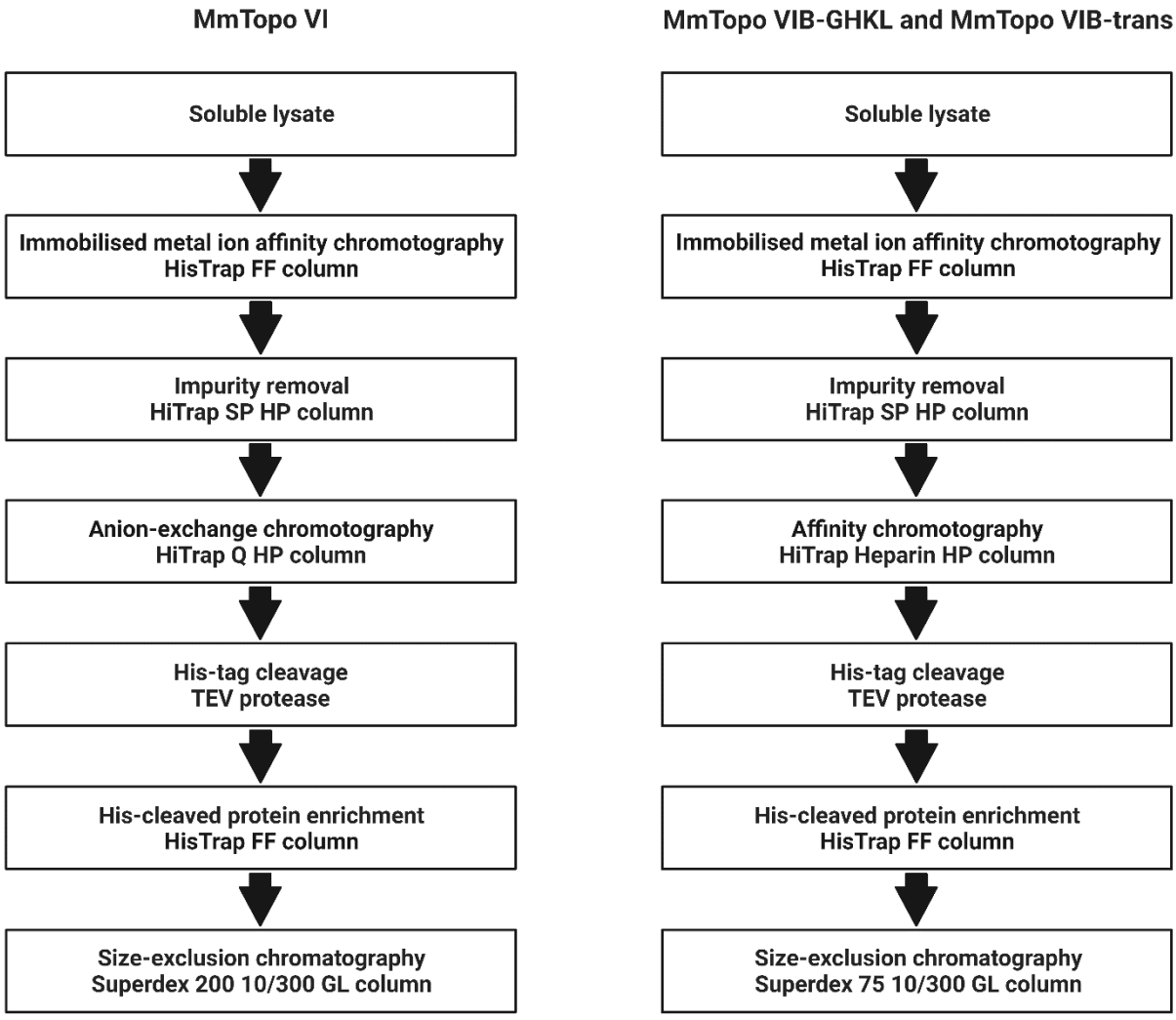

Figure 2.1| *Methanosarcina mazei* topo VI and truncates purification procedures. *Polyhistidine-tagged full-length **Methanosarcina mazei** topo VI and its GHKL domain fragment were purified by immobilised metal ion affinity chromatography, size-exclusion chromatography, and either anion-exchange chromatography or affinity chromatography, respectively.*

2.4 Biochemical characterisation of *Methanosarcina mazei* topo VI

2.4.1 DNA relaxation and decatenation assay

Relaxation of supercoiled (sc) pBR322 was carried out in reaction buffer containing 2.5 nM scDNA, 1X minimal buffer (**Table 2.4**) and 1 mM ATP. MmTopo VI serially-diluted in MmTopo VI dilution buffer (**Table 2.4**) was added to each reaction and incubated at 37˚C for 30 min. The reactions were quenched by adding 30 µL of STEB (**Table 2.4**) and 30 µL of chloroform:isoamyl alcohol (24:1), before samples were vortexed for 10 s, centrifuged at 20,000 x *g* for 5 min, and analysed by agarose gel electrophoresis. Decatenation activity was

analysed in the same reaction conditions, substituting scDNA with 6.7 ng μL^{-1} kinetoplast (k) DNA or 1.3 ng μL^{-1} singly-linked catenated (bis-cat) DNA^{174} (gifted by Nick Burton, Inspiralis, UK). DNA relaxation timecourse assays were performed at either 20˚C or 37˚C and individual reactions were quenched at different time points.

2.4.2 DNA cleavage assay

Cleavage of negatively (-) sc pBR322 was carried out in a reaction buffer containing 5 nM -scDNA, 1X minimal buffer, and 1 mM ADPNP. MmTopo VI serially diluted in MmTopo VI dilution buffer was added and incubated at 37˚C for 1 hr. To release the cleaved DNA products, the samples were treated with 3 µL of SDS (2% w/v) and 1 µL of Proteinase K (20 $mg.mL^{-1}$) and incubated for 1 hr at 37˚C. The reactions were quenched by adding 3 µL of 500 nM EDTA, 30 µL of STEB, and 30 mL of chloroform:isoamyl alcohol (24:1), before samples were vortexed for 10 s, centrifuged at 20,000 x *g* for 5 min, and analysed by agarose gel electrophoresis.

2.4.3 ATPase assay

The ATPase activity of MmTopo VI and its mutants and truncates was tested by combining 50 µL of reaction mixture possessing 1X ATPase buffer (**Table 2.4**), 2 mM ATP, 177 nM pBR322, and enzyme in a Corning Costar 96-well flat-bottom microplate. MmTopo VI and the G137A mutant was tested at a concentration of 1.2 µM, and the MmTopo VI-GHKL and MmTopo VI-trans truncates were tested at 12 µM. pBR322 DNA was tested in linear, relaxed, -sc, and positively (+) sc isoforms. 46.7 µL of reaction mixture omitting ATP was added to the wells and incubated at 37°C. After 10 min, ATP was added to each well and the solution absorbance (340 nm) was measured every 60 s for 90 min at 37°C using a CLARIOstar plate reader (BMG LabTech). Reactions devoid of enzyme, devoid of both enzyme and NADH, and devoid of enzyme in the presence of ADP instead of ATP were used as controls. *Mycobacterium smegmatis* gyrase (gifted by Lipeng Feng, John Innes Centre, UK) was used as a control under the same conditions, but in the presence of 14 nM linear pBR322 and replacing minimal buffer with *Mycobacterium tuberculosis* assay buffer (Inspiralis). Graphs were generated using GraphPad Prism. The ATPase assay conditions were replicated in a gel-based timecourse assay, and agarose gels were run in the presence and absence of EtBr.

2.5 Identification of *Methanosarcina mazei* topo VI inhibitors

2.5.1 High-throughput DNA relaxation assay

A fluorescence-based high-throughput DNA relaxation assay for topos was developed based on the differential fluorescence spectra exhibited by the commercial dye H19 (Profoldin) in response to its binding to -sc or relaxed DNA. Test compounds, dissolved in 1 µL of 100% DMSO, were provided by Syngenta in Greiner Bio-One 384-well microplates. In each well, 15 µL of reaction mixture containing 2 nM MmTopo VI, 6.7 nM -scDNA (ProFoldin), 1 mM ATP, 10 mM $MgCl_2$, 20 mM bis-tris propane (pH 7.5), 100 mM potassium glutamate, 1 mM DTT, 6.7 % (v/v) DMSO, and 10 µg.mL^{-1} test compound was incubated for 45 min at 37°C. After incubation, 10 µL of freshly prepared H19 dye, diluted to 2.2X in H19 Dilution Buffer (ProFoldin), was added to the reaction mixture and incubated again for 5 min at room temperature. The fluorescence intensity (excitation: 485 nm, emission: 535 nm) was then measured using a Tecan Infinite 200 PRO plate reader. Hits with >70% inhibition in the preliminary screen were evaluated in a confirmation dose-response screen by serially-diluting compounds in 100% DMSO with a maximum concentration of 10 mg.mL^{-1}. IC_{50} values were calculated using the four-parameter logistic model in GraphPad Prism.

2.5.2 Biochemical assays

DNA relaxation, decatenation and cleavage assays were performed as stated in section 2.4, but the -sc pBR322 substrate was replaced with -scDNA from ProFoldin at a final concentration of 6.7 nM and reactions were incubated for 45 min. For dose-response analyses, serially-diluted compounds in 100% DMSO were added to each reaction, DNA band intensities were measured using the ImageJ programme, and IC_{50} values were calculated using the four-parameter logistic model in GraphPad Prism. *E*. *coli* gyrase, *E*. *coli* topo IV, yeast topo II, human topo IIα, and wheat germ (WG) topo I were supplied by Inspiralis, and reactions possessing these enzymes were performed by replacing the MmTopo VI dilution buffer and minimal buffer with the corresponding enzyme buffers from Inspiralis. All reactions were performed in a total DMSO concentration of 6.7% (v/v).

2.5.3 Topoisomerase poisoning assay

Test compounds at concentrations of 10-500 µM, dissolved in 100% DMSO, were assessed for their ability to poison MmTopo VI in a cleavage assay using an enzyme concentration of 64 nM. The band intensities of the ADPNP-induced cleavage products were measured using the

ImageJ programme and were plotted on a bar chart using GraphPad Prism. As a control, the cleavage assay was repeated in the presence of ATP instead of ADPNP.

2.5.4 DNA intercalation assay

Test compounds at a concentration of 50 µM, dissolved in 100% DMSO, were assessed for their ability to intercalate into DNA by combining with 1 x WG topo I assay buffer, 2 µL of WG topo I diluted 1/15 in WG dilution buffer, and 6.7 nM relaxed pBR322. The total reaction volume was 30 µL and each reaction was incubated for 45 min at 37˚C. The reactions were quenched by adding 20 µL of water and 50 µL of butanol and the samples were vortexed and centrifuged at 20,000 x g for 1 min. The aqueous layer was transferred to a fresh tube and combined with 50 µL of STEB and 50 µL of chloroform:isoamyl alcohol (24:1). The samples were then vortexed for 10 s, centrifuged at 20,000 x g for 5 min, and analysed by agarose gel electrophoresis. 5 $\mu g.mL^{-1}$ EtBr, dissolved in water, and 300 µM amsacrine, dissolved in 100% DMSO, were used as controls.

2.5.5 DNA relaxation assay with *Arabidopsis thaliana* topo VI

DNA relaxation assays using a preparation of *A*. *thaliana* topo VI (AtTopo VI) (gifted by Monica Agarwal, John Innes Centre, UK) were performed as stated in section 2.5.2 using the MmTopo VI dilution buffer and minimal buffer. AtTopo VI was titrated into the relaxation assay by adding 0.125-8 µL of the enzyme preparation to each reaction, and the final DMSO concentration ranged from 0-6.7% (v/v). Test compounds at concentrations of 10-500 µM were assessed for their ability to inhibit AtTopo VI by adding 4 µL of enzyme to each reaction mixture with a final DMSO concentration of 1.6% (v/v). Control reactions were performed by omitting compound, enzyme, ATP, or $MgCl_2$. Agarose gels were run in the presence and absence of EtBr.

Chapter 3

Phylogenetic analysis of the Spo11 and topo VI-B superfamilies

3.1 Introduction

3.1.1 General introduction

As a first year PhD student, I was fascinated by the homology between the type IIB DNA topoisomerases (topos) and the meiotic double-strand break (DSB) machinery. I was particularly intrigued by the evolutionary histories of these two seemingly unrelated complexes, and how it became so that eukaryotes had hijacked a ubiquitous archaeal enzyme. A key question that will reoccur throughout this thesis is whether topo VI is an appropriate target for herbicides and antiprotozoal drugs, however, the similarity of this enzyme's mechanism to that of an essential eukaryotic process should clearly be of concern. It is therefore necessary to extensively characterise both systems and determine whether they are distinguishable at the sequence level. By doing so, we can quickly identify topo VI-possessing species in eukaryotes that would otherwise remain undetermined. In this chapter I will detail the unusual relationship between the type IIB topo and meiotic DSB complex families and will delve into the mystery that is the putative topo VI (pTopo VI) in the malaria parasite *Plasmodium*. I will then describe my own work elucidating the evolutionary relationships between these systems as a prelude to the biochemistry that will come in later chapters.

3.1.2 Spo11 and topo VI in eukaryotes

Type II DNA topoisomerases (topos) are essential for cellular life and so it came as no surprise when one was eventually discovered in archaea in 1994[46]. What was surprising, however, was that this enzyme formed a heterotetrameric complex like the bacterial type II topos but was sensitive to several inhibitors of eukaryotic type II topos[46]. The same group then went on to publish a landmark paper in 1997 that made two significant discoveries. The first was that this archaeal enzyme, now termed topo VI, shared limited sequence homology to both bacterial topo IV and eukaryotic topo II and was thus the first member of a new subfamily of topos: type IIB[60]. The second discovery was that the A subunit of topo VI (topo VI-A) was homologous to Spo11[60], a conserved meiotic protein ubiquitous in eukaryotes[96]. Spo11 was therefore

confirmed as the enzyme responsible for generating the double-strand breaks (DSBs) that initiate meiotic recombination in eukaryotes and was seemingly doing so via a topo VI-like activity.

Animals and fungi possess a single Spo11 gene and lack a gene product with sequence homology to topo VI-B[96]. This was also thought to be case with all eukaryotes, until the complete genome sequence of the model plant *Arabidopsis thaliana*[175] was published in 2000. Using this resource, *A. thaliana* was shown to possess three non-redundant Spo11 paralogues (SPO11-1, -2, -3)[49,176] that are widely distributed among plant and algae lineages[49,50]. However, only SPO11-1 and SPO11-2 were shown to be necessary for meiotic DSB formation in plants[110-112] and they unusually seem to form a heterodimer[72]. Remarkably, a homologue of archaeal topo VI-B (TOP6B) was identified in plants that could interact with SPO11-3 in a yeast two-hybrid assay[49]. A series of studies then went on to demonstrate that TOP6B and SPO11-3 form a functional topo VI complex in plants that participates in endoreduplication[115-117], and classifying SPO11-3 as a canonical topo VI A subunit and not an archetypal Spo11. The topo VI complex in plants was also shown to be unique in its requirement of two accessory proteins, RHL1[123] and BIN4[124].

The presence of topo VI in plants was surprising, however, the enzyme would also be discovered elsewhere in the eukaryotic domain. The complete genome sequence of the malaria parasite *P. falciparum* was published in 2002[138], and was later found to possess genes encoding two Spo11 paralogues (herein referred to as SPO11-A and SPO11-B) and a putative homologue of topo VI-B (pTOP6B)[50,129]. SPO11-A and pTOP6B have been reported to complement a yeast topo II null mutation[139] and a series of studies have demonstrated that these subunits have similar functional and expression profiles in vivo[140-142]. These pieces of evidence implicate SPO11-A and pTOP6B as canonical topo VI subunits in *Plasmodium*, however, as will be discussed later in the chapter, this may not necessarily be the case.

3.1.3 Meiotic double-strand break machinery

Many proteins have been identified in different eukaryotic lineages that are essential for meiotic DSB formation, but none, other than Spo11, are functionally conserved[96]. Seven gene products are essential for meiotic DSB formation in *S. cerevisiae*: Spo11, Rec102, Rec104, Rec114, Ski8, Mer2, and Mei4[177]. To determine how these proteins cooperate during meiotic

recombination initiation, a combination of genetic and biochemical approaches have been employed for characterising their physical interactions in vivo. Spo11 was shown to interact with Ski8 in a yeast two-hybrid assay[178], and coimmunoprecipitation (Co-IP) experiments have revealed interactions between Spo11 and Rec102[179] and between Rec102 and Rec104[180]. The three remaining gene products, Rec114, Mer2, and Mei4 (RMM), have all been shown to coimmunoprecipitate with each other[181] and yeast two-hybrid assays have also identified interactions between Mei4 and both Rec102 and Rec104[182]. Taken together, these studies suggest that the meiotic DSB machinery in *S. cerevisiae* comprises a core complex of Ski8, Spo11, Rec102, and Rec104, and an accessory RMM complex.

Other experiments have attempted to elucidate the role of the essential meiotic gene products in regulating the activity of Spo11. Mutant *S. cerevisiae* strains lacking REC102, REC104, or REC114 abolish the detection of Spo11 at meiotic recombination hotspots[183], and the interaction of Ski8 with Spo11 was shown to recruit Rec102 and Rec104 to the chromosomes[178]. Furthermore, the RMM complex has been shown to associate with chromosome axis sites[184] and Co-IP experiments have demonstrated that Spo11 can only dimerise in the presence of Rec102, Rec104, and Rec114[101], unlike topo VI-A which can self-associate independently[185]. Overall, this plethora of approaches have indicated that the Spo11 core complex in *S. cerevisiae* becomes localised to the nucleus at hotspots before being tethered to the chromosome axis by the accessory RMM complex. Only then does Spo11 dimerise and induce the formation of meiotic DSBs.

3.1.4 The topo VIB-like family

Topo VI requires its B subunit to regulate DSB formation by its A subunit[46,70], but no canonical topo VI-B homologue has been identified as a Spo11 partner. Thus, it was long thought that the meiotic DSB machinery had only adopted the catalytic A subunit of topo VI, and not the regulatory B subunit. However, in 2016, a momentous breakthrough in our understanding of the mechanisms involved in the initiation of meiotic recombination was made. In a screen for *A. thaliana* mutants with meiotic defects, two allelic lines were identified that could rescue a mutant defective in DSB repair[72]. The group also performed yeast two-hybrid experiments to show that the product of one of these alleles could interact with SPO11-1 and SPO11-2, and performed a bimolecular fluorescence complementation assay to demonstrate that this product also mediates the formation of the SPO11-1/SPO11-2 heterodimer[72]. An astonishing

discovery was then made following structural modelling of the allele product: the protein assembled into a structural homologue of topo VI-B[72]. This homologue, which was termed MTOPVIB, was shown to be ubiquitous in flowering plants, suggesting for the first time that the meiotic DSB complex may adopt a type IIB topo scaffold[72]. MTOPVIB, was shown to exhibit the same domain architecture as topo VI-B; to have homology to the G1-box, G3-box, and WxxY motifs; but to lack most of the key residues for ATP hydrolysis[72].

Using *A*. *thaliana* MTOPVIB as a query for position-specific iterative BLAST searches, a topo VI-B structural homologue was also identified in animals[186]. This subfamily of proteins was called TOPOVIBL and expanded what is now called the topo VIB-like family[186]. In *Mus musculus*, TOPOVIBL was shown to interact with SPO11 in a yeast two-hybrid assay, and *Top6bl−/−* mutants were deficient in meiotic DSB formation[186]. This suggests that TOPOVIBL, like MTOPVIB, forms a topo VI-like complex with Spo11. Despite their high degree of sequence divergence, model structures suggest that MTOPVIB and TOPOVIBL share similar structural elements with the transducer and GHKL domains of topo VI-B but that they are not conserved in the H2TH domain[72,186]. A predicted structure of *A. thaliana* MTOPVIB was shown to possess a H2TH-like fold, termed the small domain (SmD)[186], while models of other members of the topo VIB-like family lacked a H2TH domain completely[186]. As with MTOPVIB, TOPOVIBL lacks many of the invariant ATP-binding residues found in topo VI-B, which seems to be a defining feature of the topo VIB-like family[186].

Structural modelling of essential meiotic DSB gene products in other eukaryotic species has suggested that they can adopt individual transducer-like or GHKL-like structural modules. In these organisms, the topo VIB-like structure seems to be formed from several different proteins. Model structures of Rec6 and mei-P22, which are subunits of the meiotic DSB complex in *S. pombe* and *Drosophila melanogaster*, respectively, as well as Rec102 from *S. cerevisiae*, adopt transducer domain-like folds[186]. Furthermore, the predicted structure of *S. cerevisiae* Rec104 has been shown to resemble a GHKL-like module[187], and it is likely that yet unidentified proteins in *S. pombe*, *D. melanogaster*, and other fungal and insect species, adopt analogous GHKL-like folds. Conservation of the topo VIB-like family in distant eukaryotic lineages suggests that meiotic DSBs are catalysed by a topo VI-like tetramer or by a type IIB topo-like scaffold (**Figure 3.1**). It is worth considering that a meiotic topo VIB-like subunit has not been identified for any protist species, although this is likely to be due to the

difficulty in searching for structural homologues in a genome, and the fact that the essential meiotic genes in these organisms are less characterised.

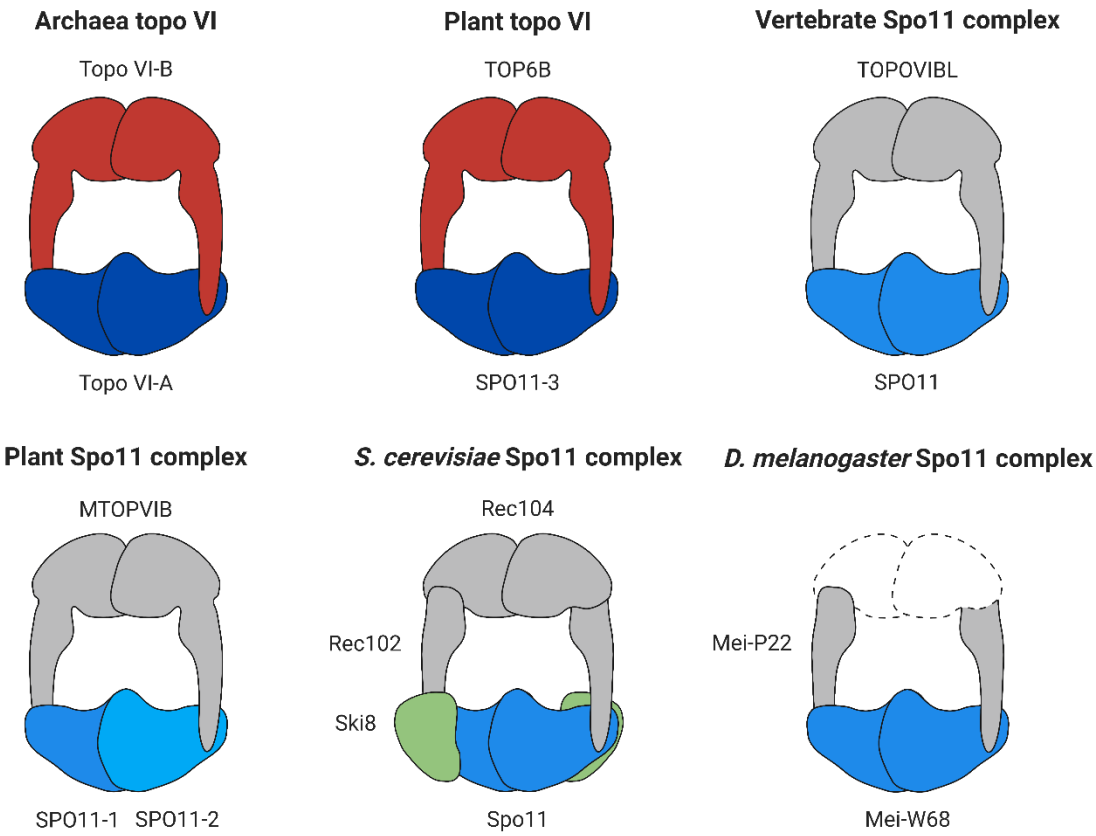

Figure 3.1| The type IIB topoisomerase scaffold: *The three-dimensional structure of topo VI has been adopted by the meiotic double-strand break (DSB) complex. Topo VI comprises two canonical topo VI-B subunits (red) and two canonical topo VI-A subunits (dark blue). The core Spo11 complex comprises a topo VIB-like fold (grey) and two Spo11 subunits (light blue). In plants, meiotic DSBs are catalysed by a heterotetramer of SPO11-1 (light blue) and SPO11-2 (baby blue). In* ***Saccharomyces cerevisiae****, the core Spo11 complex possesses an additional subunit, Ski8 (green), and the* ***Drosophila melanogaster*** *Spo11 complex currently lacks an identified GHKL-like module. Created with BioRender.com.*

The presence of a topo VIB-like Spo11 partner raises questions about the regulation and mechanism of meiotic DSB formation. Topo VI utilises its ATPase B subunit for capturing a T-segment for strand-passage and to regulate G-segment cleavage by the A subunit[46,70], however, the meiotic topo VI-like complex is unlikely to perform strand-passage and seems

to have lost its capacity for ATP hydrolysis[72,186]. This B-like subunit of the Spo11 complex is therefore unlikely to be acting as a molecular clamp and the type IIB topo scaffold must be regulating Spo11 cleavage via an unknown and ATP-independent mechanism. It may be worth speculating that the B subunit of topo VI is imparting another yet-to-be elucidated function that is also necessary for meiotic DSB formation.

3.15 Spo11 evolution and the *Plasmodium* putative topo VI

A consequence of the discovery of the topo VIB-like family is that the presence of a eukaryotic gene product with homology to a canonical topo VI B subunit is no longer sufficient to signify the existence of a topo VI enzyme in that species. Furthermore, although it is usually the case that a single Spo11 gene denotes the absence of topo VI and multiple Spo11 paralogues denotes the presence of topo VI, there are examples of protozoan phyla that possess multiple Spo11 paralogues but no identifiable topo VI-B[50]. Unlike plant topo VI-B, the pTopo VI-B of *Plasmodium* has been shown to possess limited sequence homology with topo VI-B from archaea. This raises the possibility that the subunit may instead be a structural homologue of the canonical topo VI-B and a module of the meiotic topo VI-like complex.

Further doubt can be placed on the occurrence of a canonical topo VI in *Plasmodium,* and other members of the Apicomplexa phylum, given an earlier phylogenetic study in 2007, which sought to characterise the Spo11 and topo VI-B homologues in eukaryotes and archaea. The group produced a phylogenetic tree of eukaryotic Spo11 sequences that placed all proteins into three distinct groups: Spo11-1, Spo11-2, and Spo11-3; and revealed the evolutionary narrative of these genes[50]. Topo VI-A and topo VI-B were seemingly present in the common ancestor of eukaryotes and archaea, and successive gene duplication events in eukaryotes separated the meiotic Spo11-2 from Spo11-3, and Spo11-1 from Spo11-2[50]. Some eukaryotes then suffered lineage-specific gene losses of topo VI-B and some of their Spo11 paralogues[50]. The analysis placed animal and fungal Spo11s into the Spo11-1 group, and protist Spo11s into a combination of the Spo11-1 and Spo11-2 groups[50]. Spo11-3, the canonical topo VI A subunit, was only present in plants and algae, and was always associated with a canonical topo VI B subunit[50]. An exception to this rule, however, was found in the Apicomplexa, whose Spo11 sequences were placed into the meiotic Spo11-1 and Spo11-2 groups despite this phylum also possessing a pTopo VI-B[50].

Aside from their phylogenetic distinctions, there seems to be a key mechanical difference between canonical topo VI-A and meiotic Spo11 homologues. *S. cerevisiae* Spo11[101] and *A. thaliana* SPO11-1/SPO11-2[72] can only dimerise in the presence of their topo VIB-like partners, whereas *Methanocaldococcus jannaschii* topo VI-A[79] and *A. thaliana* SPO11-3[72] can self-associate without a topo VI-B subunit. A yeast-two hybrid experiment has shown that *P. falciparum* SPO11-A is unable to dimerise in the absence of pTOP6B, a result which follows the pattern of the meiotic Spo11 homologues[139]. Overall, given the lack of a canonical topo VI-A and considering the poor homology of its pTopo VI-B with canonical topo VI-B sequences, it remains unclear whether Apicomplexa possess topo VI. Therefore, in this thesis, the Apicomplexa Spo11 paralogues have been termed SPO11-A and SPO11-B and their topo VI-B homologues as putative to reflect the uncertainty surrounding these gene products.

Topo VIB-like homologues have only been identified in plants, animals, and fungi[72,186], although it seems likely that this family is also present in Apicomplexa and other protists. Apicomplexa are members of the superphylum Alveolates, which belong to the Stramenopile-Alveolate-Rhizaria (SAR) supergroup[188]. SAR is a major phylogenetic clade whose members share a common ancestor that has diverged from the Archaeplastida supergroup, which contains red algae, green algae, and plants, and from the protist superphylum Hacrobia[188]. Canonical topo VI-A and topo VI-B homologues are common in Archaeplastida, including in algae, but have also been identified in diatomic members of stramenopiles[49,50]. For these protists, their topo VI sequences clearly resemble those of archaea and there is no uncertainty as to their identity. A discrepancy in the identity of their pTOP6B homologue is therefore unique to Apicomplexa, thus it remains plausible that this protein is the first member of the protist topo VIB-like family.

3.1.6 Aims and objectives

Here, an extensive phylogenetic analysis was performed to fully characterise the topo VI and Spo11 complex subunits from archaea, plants, protists, and bacteria. This work has three aims: to determine whether *Plasmodium* and other Apicomplexa possess topo VI; to describe the distribution of topo VI in bacteria and protists; and to investigate the role of the plant topo VI accessory proteins.

3.2 Results

3.2.1 Collection of Spo11 and type IIB topoisomerase sequences

I performed a phylogenetic analysis to determine the differences between the homologous subunits of the Spo11 and type IIB topo complexes at the sequence level. To do this, I collected annotated sequences using keyword searches on the online NCBI protein database tool and I performed protein-protein BLAST searches to identify unannotated sequences. A total of 174 Spo11/topo VI-A sequences were collected that were represented by the three primary eukaryotic kingdoms, six protist supergroups, four archaea supergroups, and eight bacteria supergroups (**Table 3.1**). A total of 132 sequences identifying as MTOPVIB, pTOP6B, or topo VI-B were also collected that were represented by Embryophyta, commonly known as land plants; five protist supergroups; four archaea supergroups; and eight bacteria supergroups. Furthermore, sequences from other members of the type IIB topo family were collected, including eight topo VIII-A and six topo VIII-B sequences from bacteria, and ten Mini-A sequences from bacteria and viruses. Mini-A sequences were obtained from the supplementary data of Takahashi et al. (2020)[57]. MTOPVIB sequences show weak homology to canonical topo VI-B sequences[186] and so were retained in this study. Other members of the topo VIB-like family, however, were deemed to have insufficient homology for appropriate phylogenetic analysis with topo VI-B.

Every sequence analysed was a member of a unique taxonomic family, expect for Plasmodiidae which appears twice. Due to the uncertain identity of the pTOP6B subunit in *Plasmodium*, I sought to identify other pTOP6B species in Apicomplexa to aid its analysis. Using *P. falciparum* pTOP6B (NCBI Reference Sequence: XP_001350366.1) as a query, the BLASTP programme identified homologous pTOP6B sequences in two genera in the Aconoidasida class (*Babesia*, *Hepatocystis*) and in five genera in the Conoidasida class (*Cryptosporidium*, *Eimeria*, *Besnoitia*, *Hammondia*, *Toxoplasma*) (**Table 3.2**). Of these, only the pTOP6B from *Babesia* and *Hammondia* were annotated as pTopo VI-B subunits. Therefore, a total of ten pTOP6B sequences, including three *Plasmodium* species (*P. berghei*, *P. falciparum*, *P. vivax*), were included in the phylogenetic analysis. Orthologues of *Plasmodium* SPO11-A and SPO11-B were also identified in these Apicomplexa genera (**Table 3.2**).

Table 3.1| Distribution of collected Spo11 and topo VI sequences

Kingdom/supergroup	Spo11/topo VI-A	pTOP6B/topo VI-B
Eukaryotes		
Metazoa	+	-
Fungi	+	-
Archaeplastida	+	+
Cryptista	+	+
Discoba	+	-
Haptista	+	+
SAR	+	+
Sarcomastigota	+	+
Archaea		
Asgard	+	+
DPANN	+	+
Euryarchaeota	+	+
TACK	+	+
Bacteria		
Acidobacteria	+	+
Elusimicrobia	+	+
FCB	+	+
Nitrospirae	+	+
Proteobacteria	+	+
PVC	+	+
Terrabacteria	+	+
Thermotogae	+	+

Table 3.2| Distribution of collected Apicomplexa SPO11 and pTOP6B sequences

Phylum	Class	Order	Family	Genus	SPO11-A	SPO11-B	pTOP6B
Apicomplexa	Aconoidasida	Haemosporida	Plasmodiidae	Hepatocystis	+	+	+
			Plasmodiidae	Plasmodium	+	+	+
		Piroplasmida	Babesiidae	Babesia	+	+	+
	Conoidasida	Eucoccidiorida	Cryptosporidiidae	Cryptosporidium	+	+	+
			Eimeriidae	Eimeria	+	+	+
			Sarcocystidae	Besnoitia	+	+	+
				Hammondia	+	+	+
				Toxoplasma	+	+	+

3.2.2 Apicomplexa possess two canonical Spo11 paralogues

To determine the phylogenetic relationships within the Spo11/topo VI-A superfamily in the three domains of life, I subjected a total of 192 homologues to phylogenetic tree analysis. The Spo11/topo VI-A superfamily splits into two major clades on a phylogenetic tree: type IIB topo A subunits (topo VI-A, SPO11-3, topo VIII-A, Mini-A) and canonical Spo11 subfamilies (SPO11-

1, SPO11-2) (**Figures 3.2** and **3.3**). SPO11-3 from plants and protists form two distinct clades, and an unrooted phylogenetic tree consigns topo VIII-A and Mini-A as close relatives of archaeal and bacterial topo VI-A (**Figure 3.3**). Archaeal topo VI-A forms two distinct clades, made up of the TACK/Asgard and Euryarchaeota/DPANN supergroups, that are separated by bacterial topo VI-B (**Figure 3.2**). The rooted phylogenetic tree suggests that at least one horizontal gene transfer (HGT) event has occurred from Euryarchaeota archaea to bacteria. Given the unusual placement of '*Verrucomicrobiales bacterium*', which sits away from other bacterial sequences, I investigated the anomalous sequence further. The topo VI-A from *V. bacterium* shares ~99% sequence similarity with an Euryarchaeota archaeon topo VI-A sequence, and so this may represent a recent HGT event or an annotation error. *Sedimentisphaera cyanobacteriorum* topo VI-A, for example, shares only 92% sequence identity with topo VI-A from the archaeon *Conexivisphaera calidus*.

The SPO11-1 subfamily includes plant, fungi, vertebrate, and invertebrate Spo11 sequences, and the SPO11-2 subfamily includes plant, algae, and other protist Spo11 sequences, that all form distinct clades. Anomalies in the SPO11-1 group include Spo11 sequences from the Sarcomastigota protist *Capsaspora owczarzaki* and the Alveolate protist *Stentor coeruleus*, which would be expected to place in the SPO11-2 group. Anomalies in the SPO11-2 group are metazoan Spo11 sequences from the sponge *Amphimedon queenslandica*, and from the bilaterians *Caenorhabditis elegans*, *Schmidtea mediterranea*, *Mesocestoides corti*, and *Stichopus japonicus*, which would be expected to place in the SPO11-1 group. In contradiction with an earlier phylogenetic analysis[50], Spo11 homologues from protists do not occupy the SPO11-1 group. The two Spo11 paralogues in *Plasmodium*, termed here SPO11-A and SPO11-B, are conserved in Apicomplexa. Every genus analysed here that possesses multiple Spo11 paralogues places each of them into separate Spo11 subfamilies. However, Apicomplexa are the one exception, as both SPO11-A and SPO11-B cluster together in a single clade within the SPO11-2 subfamily. This suggests that Apicomplexa possess two canonical Spo11 paralogues and do not possess a canonical topo VI-A. Given that SPO11-A and SPO11-B also cluster with the anomalous metazoan Spo11 sequences, the placement of these paralogues in SPO11-2 remains uncertain.

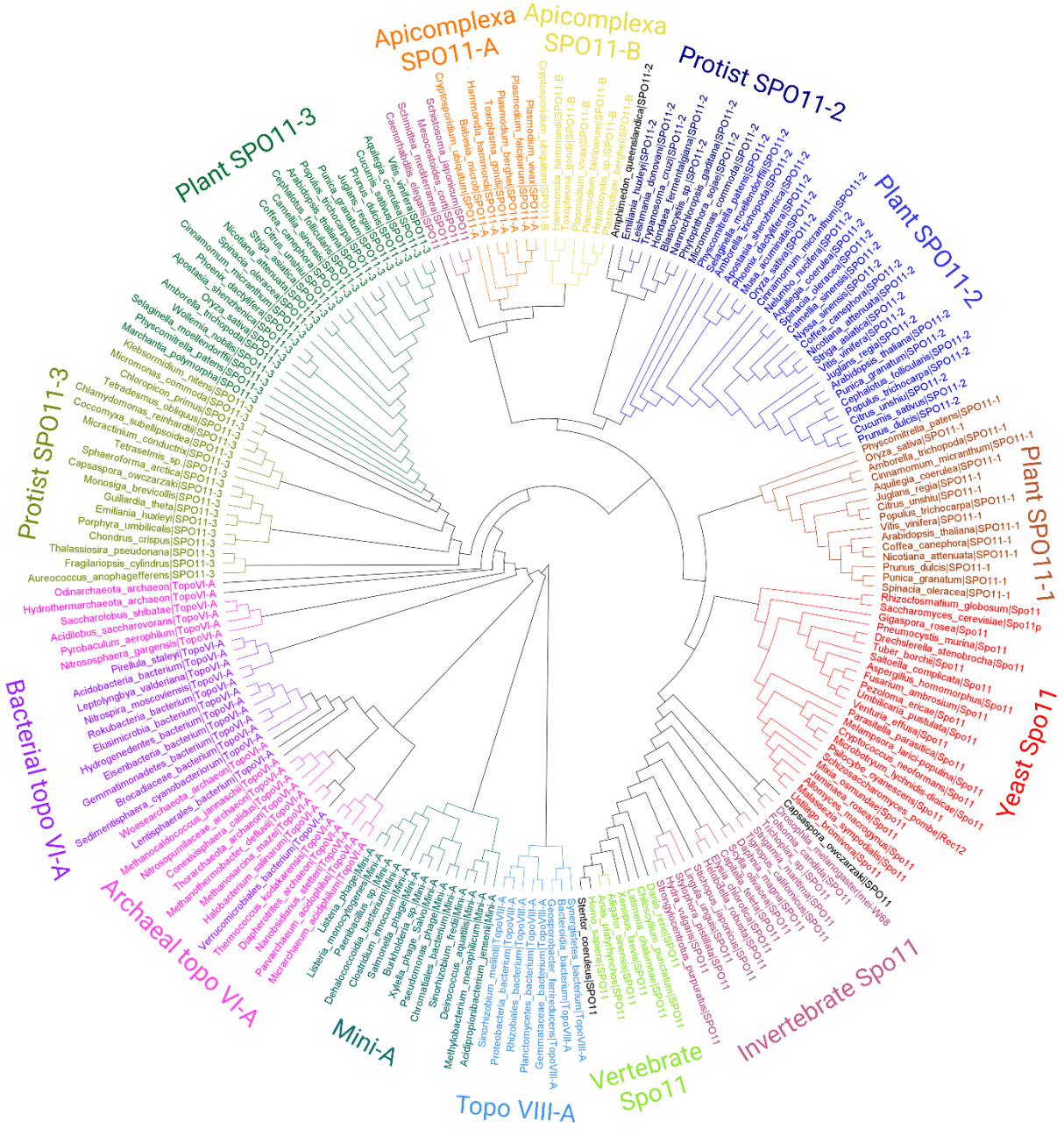

Figure 3.2| Rooted maximum-likelihood phylogenetic tree of the Spo11/topo VI-A superfamily: *The Spo11/topo VI-A superfamily split into two major clades when rooted to topo VIII-A. The canonical type IIB topoisomerase B subunit clade comprises bacteria topo VI-A (pink), archaea topo VI-A (purple), plant SPO11-3 (dark-green), protist SPO11-3 (olive-green), topo VIII-A (light-blue), and Mini-A (teal). The canonical Spo11 clade comprises plant SPO11-1 (brown), plant SPO11-2 (dark-blue), protist SPO11-2 (navy-blue), fungi Spo11 (red), vertebrate Spo11 (lime-green), invertebrate Spo11 (maroon), Apicomplexa SPO11-A (orange), and Apicomplexa SPO11-B (yellow). Anomalously placed organisms are coloured black.*

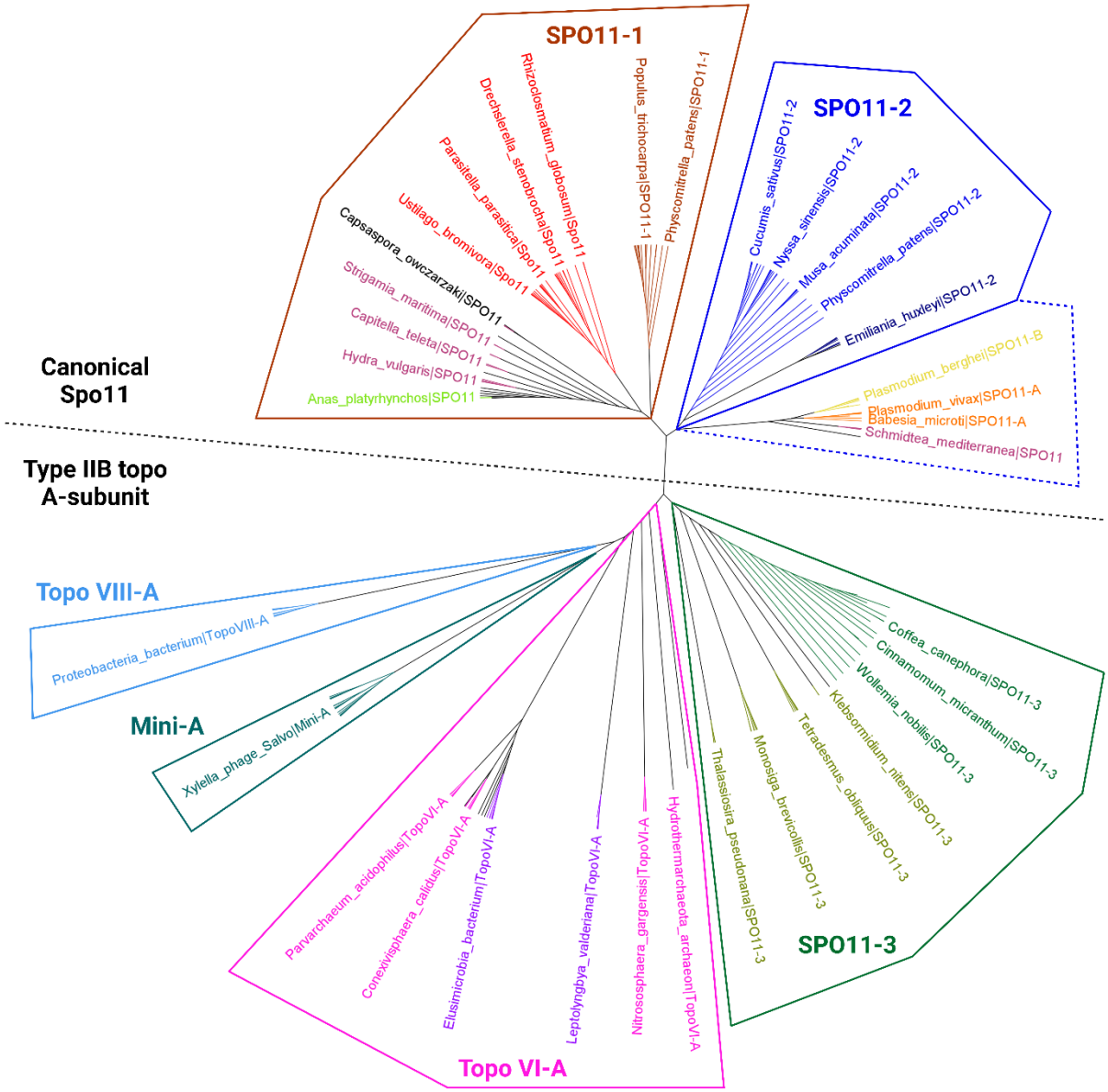

Figure 3.3| Unrooted maximum-likelihood phylogenetic tree of the Spo11 superfamily: *The Spo11/topo VI-A superfamily split into two major clades. The canonical type IIB topoisomerase B subunit clade comprises bacteria topo VI-A (pink), archaea topo VI-A (purple), plant SPO11-3 (dark-green), protist SPO11-3 (olive-green), topo VIII-A (light-blue), and Mini-A (teal). This major clade possesses four groups: topo VIII-A (light-blue), Mini-A (teal), topo VI-A (pink), and SPO11-3 (dark-green). The canonical Spo11 clade comprises plant SPO11-1 (brown), plant SPO11-2 (dark-blue), protist SPO11-2 (navy-blue), fungi Spo11 (red), vertebrate Spo11 (lime-green), invertebrate Spo11 (maroon), Apicomplexa SPO11-A (orange), and Apicomplexa SPO11-B (yellow). This major clade possesses two groups: SPO11-1 (brown) and SPO11-2 (dark-blue). The dotted blue line represents species that are tentatively placed in the SPO11-2 group, and anomalously placed organisms are coloured black.*

To determine whether canonical Spo11 and topo VI-A sequences can be distinguished by individual sequence motifs, I performed a sequence alignment using representatives from each subfamily. All members of the Spo11/topo VI-A superfamily contain key conserved residues present in the WHD, namely the catalytic tyrosine and the adjacent arginine (WHD-R). All members of this superfamily also contain key conserved elements in the toprim domain, these are the metal-coordinating glutamate (Toprim-E1) and DxD motif, a second glutamate (Toprim-E2) that interacts with the catalytic tyrosine and the WHD-R, and an active site glycine (Toprim-G) (**Figure 3.4**). The only exception to the conservation of key residues in the Spo11/topo VI-A superfamily is found in the SPO11-2 subfamily, whose members have all substituted the second aspartate residue in the DxD motif for asparagine (**Figures 3.4** and **3.5**). Plant and protist SPO11-2 sequences all possess this alternative DxN motif, as does *Amphimedon queenslandica* suggesting that this sponge does indeed possess SPO11-2 and that its position in this group is not anomalous (**Figure 3.5**). However, the metazoan Spo11 sequences along with the Apicomplexa Spo11 paralogues that clustered together within the SPO11-2 group do not possess this substitution. Given the presence of other metazoan Spo11 sequences in the SPO11-1 subfamily and that these sequences all possess a DxD motif, it is likely that this mini cluster of metazoan sequences in the SPO11-2 group are in fact SPO11-1s. The identity of the Apicomplexa Spo11 paralogues is more difficult to establish, however, given that the SPO11-1 subfamily contains no protist representatives and that Apicomplexa Spo11s possess a DxD motif. Apicomplexa SPO11-A and SPO11-B are therefore uncharacterised members of the canonical Spo11 major clade.

I was unable to find any conserved sequence motifs that were present exclusively to either of the canonical topo VI-A or canonical Spo11 families that could be used as a marker for distinguishing them. However, this search led me to identifying an aspartate residue (D294 in *M. mazei* topo VI-A) that was conserved in all topo VI-A sequences and most Spo11 sequences (**Figure 3.6**). This residue was particularly interesting as it was present in all Apicomplexa SPO11-B sequences, but was absent in all SPO11-A sequences, and so was the only motif identified that could distinguish between these subfamilies (**Figure 3.6**). The two Spo11 paralogues present in Apicomplexa can therefore be distinguished both phylogenetically and by the presence of this aspartate residue, and the significance of this residue in *M. mazei* topo VI will be assessed in Chapter 5.

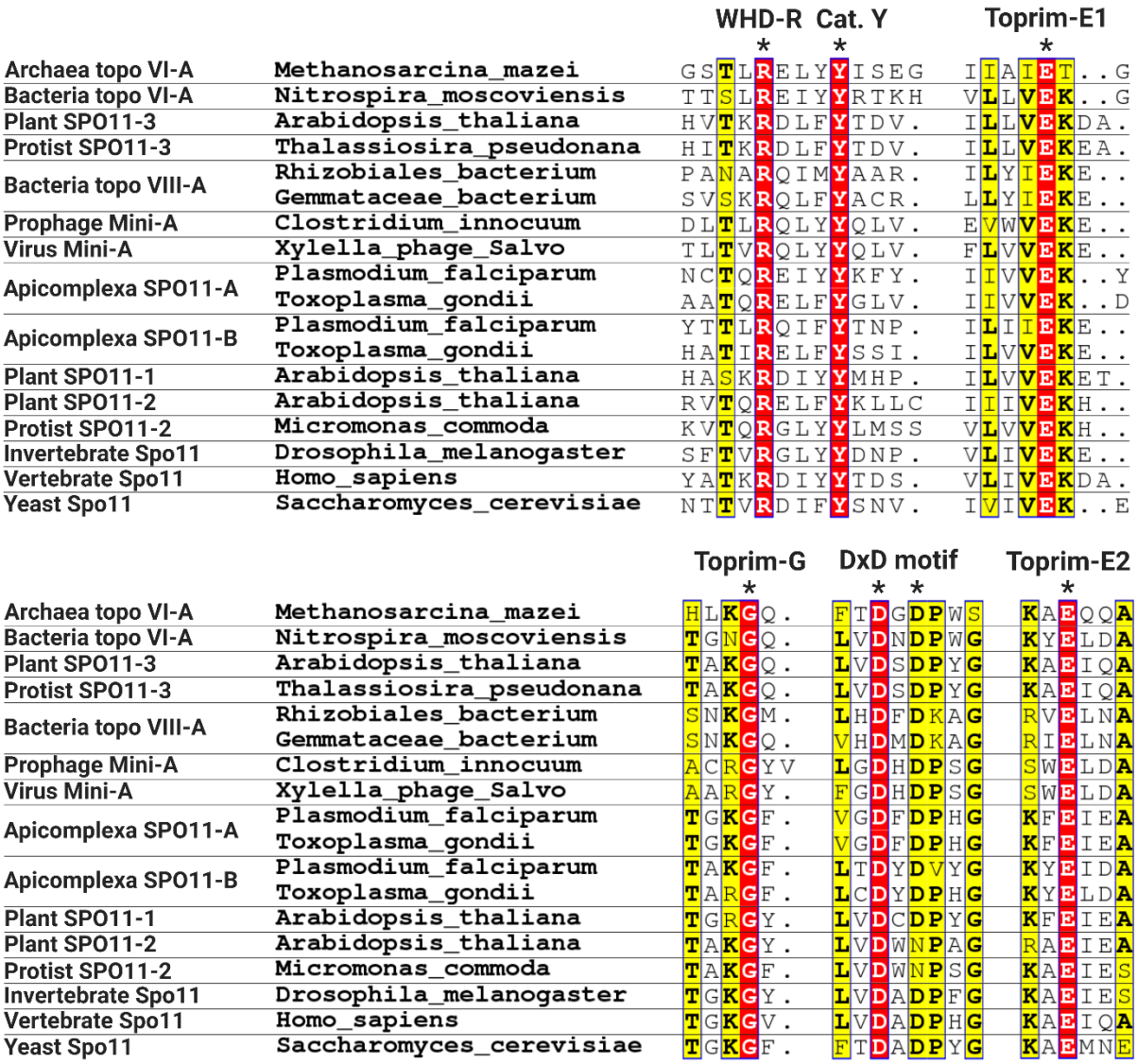

Figure 3.4| Sequence alignment of the Spo11/topo VI-A superfamily: *All members of the Spo11/topo VI-A superfamily contain key invariant resides in the winged-helix domain and toprim domain. In the SPO11-2 subfamily, however, the second aspartate residue in the DxD motif is substituted for asparagine.*

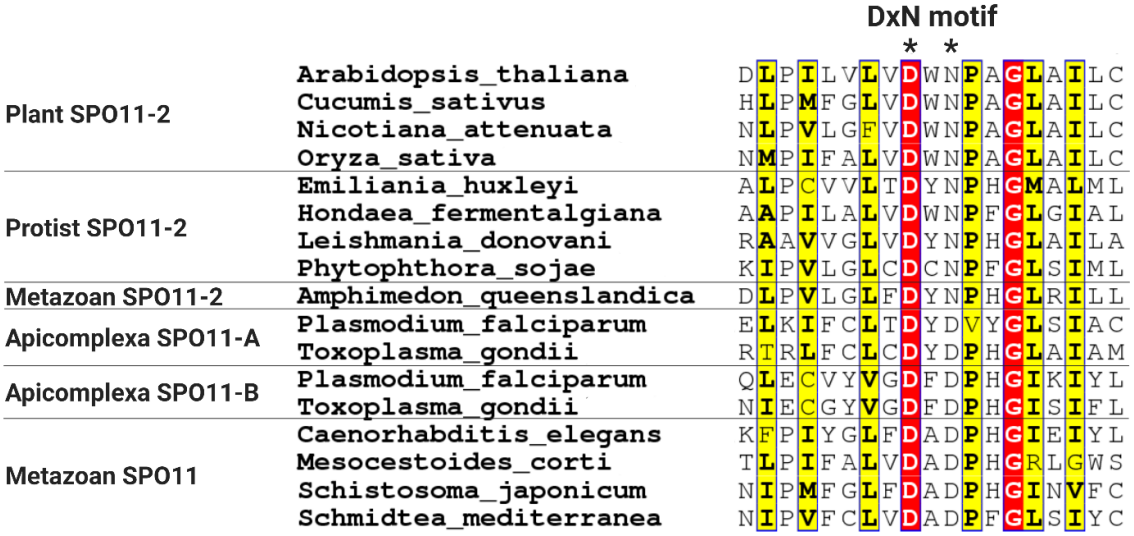

Figure 3.5| Sequence alignment of the SPO11-2 DxN motif: *All members of the SPO11-2 subfamily possess a conserved DxN motif. This is true even of Spo11 from the metazoan* ***Amphimedon queenslandica****. Other metazoans (****Caenorhabditis elegans****,* ***Mesocestoides corti****,* ***Schistosoma japonicum****,* ***Schmidtea mediterranea****) and the apicomplexans, whose SPO11 sequences were tentatively placed in the SPO11-2 subfamily, do not possess the characteristic DxN motif.*

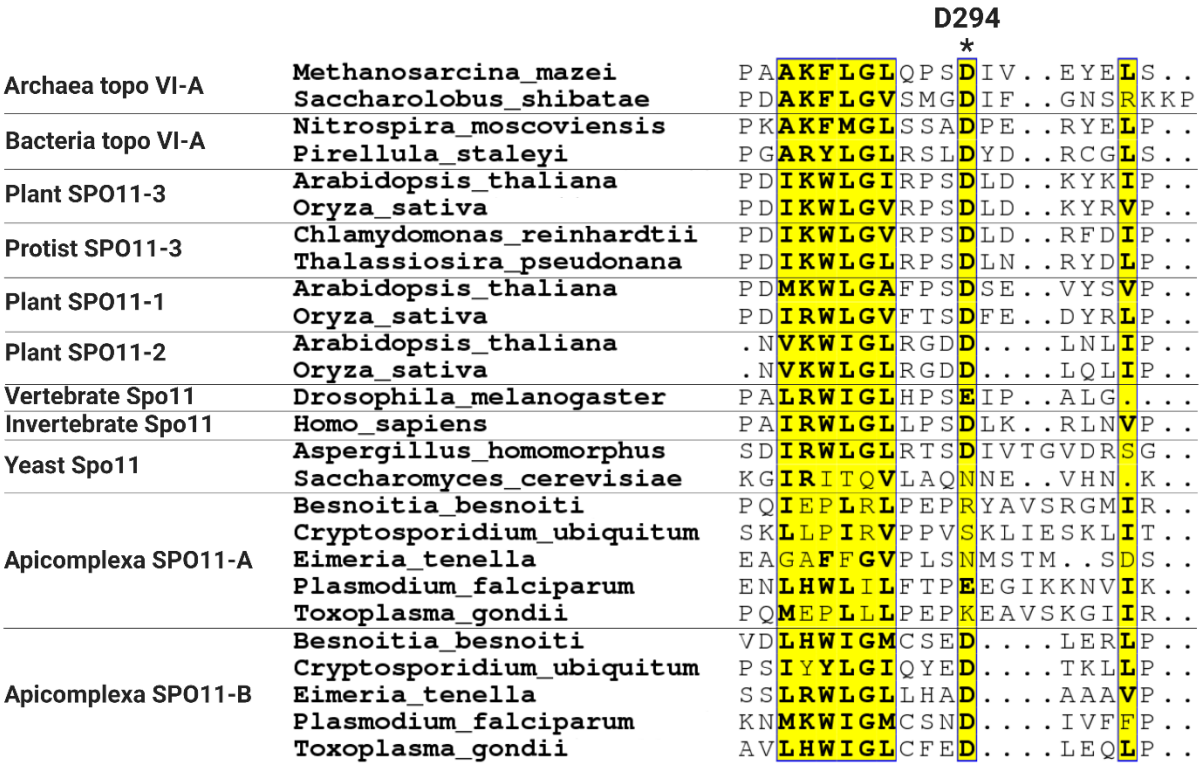

Figure 3.6| Sequence alignment of the *Methanosarcina mazei* topo VI-A D294 residue: *All canonical topo VI-A sequences possess a conserved aspartate residue (D294 in* ***M. mazei*** *topo VI-A). This residue is present in most canonical Spo11 sequences, is present in all SPO11-B sequences, but is absent in all SPO11-A sequences.*

3.2.3 Apicomplexa pTOP6B is distinct from SAR topo VI-B

Phylogenetic analysis suggests that Apicomplexa do not possess a canonical topo VI-A, therefore I sought to elucidate the identity of the pTOP6B subunit. To determine the phylogenetic relationships within the topo VI-B/topo VIB-like superfamily, a total of 132 canonical topo VI-B sequences, 6 topo VIII-B sequences, 13 MTOPVIB sequences, and 8 pTOP6B sequences were subject to phylogenetic tree analysis. A rooted phylogenetic tree of the topo VI-B/topo VIB-like superfamily splits into three major clades: archaeal and bacterial topo VI-B, and topo VIII-B; plant and protist TOP6B; and land plant MTOPVIB (**Figure 3.7**). Like their SPO11-3 partners, plant and protist TOP6B are separated into two distinct clades.

Bacterial topo VI-B forms a single clade that is closely related to the Euryarchaeota archaeal supergroup, as was the case for their topo VI-A partners (**Figure 3.7**). However, unlike with their topo VI-A partners, the two archaeal topo VI-B clades (Euryarchaeota/DPANN, TACK/Asgard) are separated not by bacterial topo VI, but by bacterial topo VIII-B, which forms a single clade. Topo VI-B from the putatively named '*Thermotogae bacterium*', '*Marinimicrobia bacterium*', and '*Opitutaceae bacterium*' are placed outside of the bacterial topo VI clade, as was the case for *V. bacterium* topo VI-A. These subunits share ≥97% sequence similarity with a Euryarchaeota archaeon topo VI-B sequence, and so may represent recent HGT events or incorrect annotations. Two independent HGT events for topo VI-B from Euryarchaeota archaea to bacteria have previously been reported in a smaller study[27], and the rooted phylogenetic tree from my analysis suggests that up to four events may have occurred from the same supergroup (**Figure 3.7**).

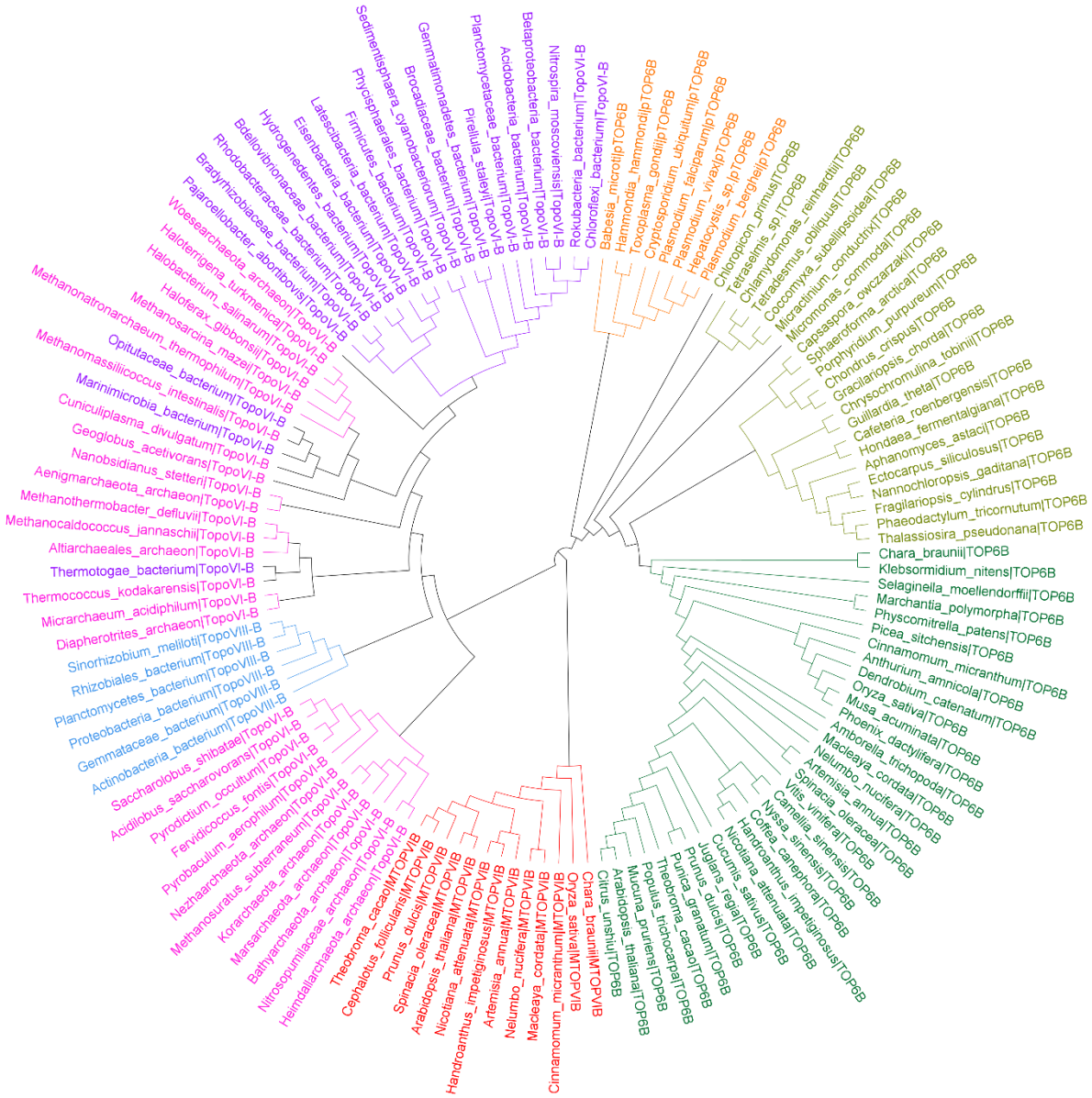

Figure 3.7| Rooted maximum-likelihood phylogenetic tree of the topo VI-B/topo VIB-like superfamily: *The topo VI-B/topo VIB-like superfamily splits into three major clades. The first clade comprises bacteria topo VI-B (pink), archaea topo VI-B (purple), and topo VIII-B (light-blue), the second clade comprises plant TOP6B (dark-green), protist TOP6B (olive green), and Apicomplexa pTOP6B (orange), and the third clade comprises land plant MTOPVIB (red). The tree is rooted at the MTOPVIB clade.*

On the rooted phylogenetic tree, the Apicomplexa pTOP6B sequences form a single clade that lies in the plant and protist TOP6B major clade and sits closer to the protist TOP6B clade (**Figure 3.7**). However, on an unrooted tree Apicomplexa pTOP6B lies between the MTOPVIB clade and the plant and protist TOP6B major clade (**Figure 3.8**). It is therefore unclear from the phylogenetic analysis whether pTOP6B is a member of the topo VIB-like major clade, or the protist TOP6B clade. Nonetheless, both the MTOPVIB and pTOP6B clades are more closely related to the plant and protist TOP6B major clade than to the archaeal and bacterial topo VI-B and topo VIII-B major clade. TOP6B sequences from members of the Ochrophyta phylum in the SAR taxonomic supergroup, including the diatomic *Thalassiosira pseudonana*, seem distantly related to pTOP6B from Apicomplexa on the phylogenetic tree even though they too are member of the SAR supergroup (**Figure 3.8)**.

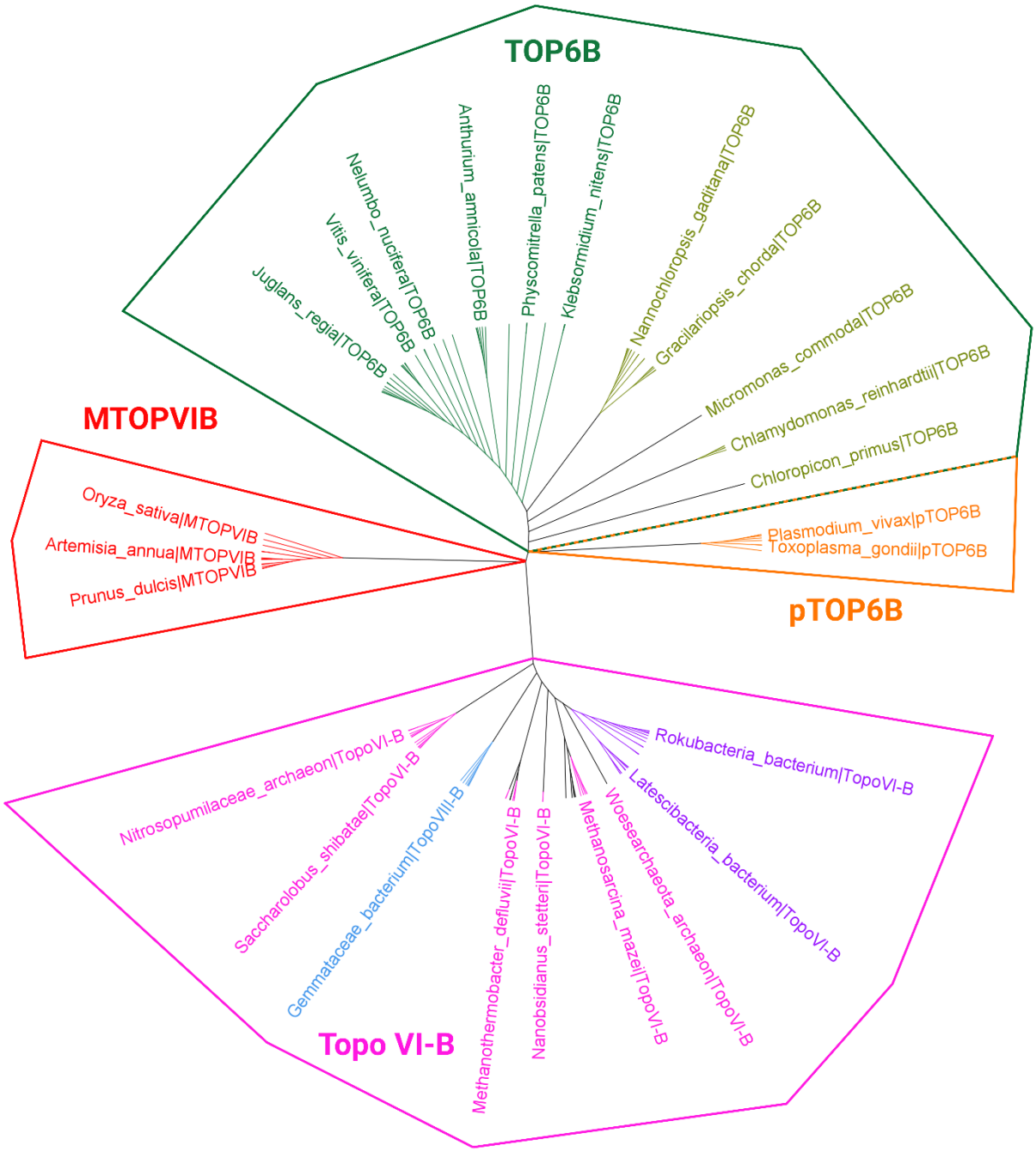

Figure 3.8| Unrooted maximum-likelihood phylogenetic tree of the topo VI-B/topo VIB-like superfamily: *The topo VI-B/topo VIB-like superfamily splits into two major clades. The topo VI-B clade (pink) comprises bacteria topo VI-B (pink), archaea topo VI-B (purple), and topo VIII-B (light-blue), and the TOP6B clade (dark-green) comprises plant TOP6B (dark-green) and protist TOP6B (olive green). Land plant MTOPVIB (red) and Apicomplexa pTOP6B (orange) form separate minor clades between the two major clades. It is unclear whether pTOP6B forms a major topo VIB-like clade with MTOPVIB or whether it is a member of the TOP6B major clade. The dotted orange/dark-green line represents the unclear position of the pTOP6B minor clade.*

No specific sequence motifs have been identified that can be used to distinguish canonical Spo11 and topo VI-A, expect for SPO11-2. However, the topo VI-B and topo VIB-like families are known to exhibit differential sequence homology in their ATP-binding folds[72,186]. To determine whether the identity of Apicomplexa pTOP6B can be elucidated by analysing the presence of individual conserved elements, I performed a sequence alignment using representatives from each subfamily of the topo VI-B/topo VIB-like superfamily. All the canonical topo VI-B and topo VIII-B sequences analysed in this study contain the key ATP-binding residues in the N-box, G1-box, G2-box, and G3-box, and possess the conserved switch-lysine (**Figure 3.9**). An exception is topo VIII-B from *Rhizobiales bacterium* for which the first conserved glycine in the G1-box is substituted for alanine. MTOPVIB sequences all possess the conserved aspartate and glycines in the G1-box, and the conserved G3-box glycine, however, they lack homology in the N-box, G2-box, and switch-lysine regions. Although several MTOPVIB sequences display a lysine residue in the same position as the topo VI-B and topo VIII-B switch-lysine, this region in MTOPVIB is poorly conserved and the location of the lysine is likely to be an artefact of the alignment. MTOPVIB also contains the DNA-binding WxxY motif, which is conserved in topo VI-B but not in topo VIII-B.

All Apicomplexa pTOP6B sequences contain the conserved ATP-contacting residues in the N-box and G2-box, and possess the key switch-lysine, but these sequences lack homology in the G3-box and WxxY motif. The significance of the G3-box glycine in *M. mazei* topo VI will therefore be assessed in Chapter 5. The pTOP6B sequences from the Aconoidasida class of Apicomplexa possess a fully intact G1-box, but pTOP6B from Conoidasida organisms is degenerate. In the G1-box of *Cryptosporidium ubiquitum* pTOP6B the conserved asparagine residue is substituted for threonine, and the G1-box in *Hammondia hammondi* and *Toxoplasma gondii* pTOP6B is missing entirely. In total, of the eleven invariant ATP-binding residues present in all canonical topo VI-B sequences, topo VIII-B members possess either ten or eleven, pTOP6B members possess between six and ten, and MTOPVIB possess four. The absence of key invariant ATP-binding residues is a useful way of distinguishing between topo VI-B and topo VIB-like subunits, however, it remains unclear how many of these residues must be missing before classifying a pTopo VI-B as a topo VIB-like. The absence of a G3-box does not help elucidate the identity of the Apicomplexa pTOP6B, as this motif is present in both topo VI-B and MTOPVIB. All other protist topo VI-B sequences do not lack any of the

conserved ATP-binding residues, and this is true even for close relatives of the Apicomplexa, such as *T. pseudonana*. Taken together, the Apicomplexa pTOP6B, regardless of its identity, is unique among the SAR supergroup, and other protists.

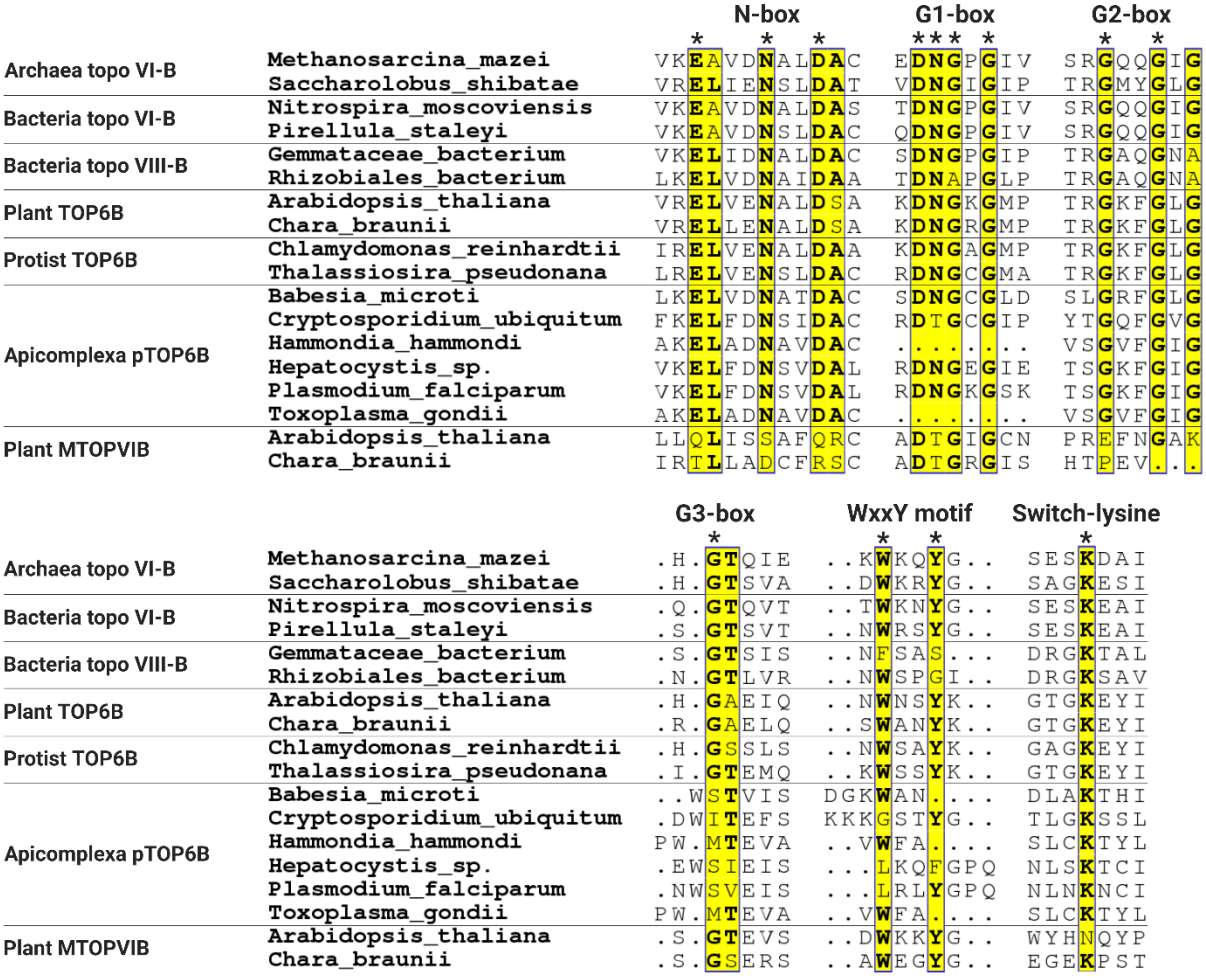

Figure 3.9| Sequence alignment of the topo VI-B/topo VIB-like superfamily: *All canonical topo VI-B members contain key invariant ATP-binding residues in the GHKL domain (N-, G1-, G2, G3-box) and transducer domain (switch-lysine), and a DNA-binding WxxY motif. In MTOPVIB, only the G1-box, G3-box and WxxY motif are conserved, and the presence of the switch lysine in some members, such as* ***Chara braunii****, is likely an artefact of the alignment. Topo VIII-B subunits lack the WxxY motif, and some members, such as* ***Rhizobiales bacterium****, lack the first glycine residue in the G1-box. In Apicomplexa pTOP6B, the G3-box and WxxY motif are missing, and some members, such as* ***Toxoplasma gondii****, also lack a G1-box.*

3.2.4 Apicomplexa possess a topo VIB-like subunit

It is unclear from the phylogenetic analysis whether Apicomplexa pTOP6B is a member of the topo VI-B or the topo VIB-like family. Therefore, I performed 3D modelling using AlphaFold[162] to investigate whether the identity of pTOP6B can be elucidated by analysing the predicted structures of these families. The crystal structure of the topo VI heterotetramer has been solved for both *Methanosarcina mazei*[70] (PDB code: 2Q2E) and *Saccharolobus shibatae*[71] (PDB code: 2ZBK), and their topo VI-Bs feature a H2TH domain implicated in supercoiled DNA specificity[58], an ATPase GHKL domain, and a transducer domain that interacts with topo VI-A. Structural modelling suggests that these three folds are also conserved in topo VI-B from bacteria and eukaryotes (**Figure 3.10**).

M. *mazei* topo VI-B possesses a C-terminal domain (CTD) adjacent to its transducer domain that adopts an immunoglobulin-like fold[70]. A recent study has shown, using AlphaFold, that this C-terminal extension is unique to archaeal species that also possess a gyrase, hence its absence in *S. shibatae* topo VI[48]. My own AlphaFold predictions show that topo VI-B from all bacterial and eukaryotic species tested do not possess an analogous immunoglobulin-like fold (**Figure 3.9**). Furthermore, the transducer domain models of topo VI-B from archaea and bacteria feature a long α-helix (~68 Å in *M. mazei*, ~74 Å in *Nitrospira moscoviensis*) that spans the internal cavity of the enzyme and a short/medium-sized α-helix (~11 Å in *M. mazei*, ~31 Å in *N. moscoviensis*) that kinks back parallel with the long helix (**Figure 3.10**). The predicted structures of topo VI-B from eukaryotes, however, feature an additional medium-sized α-helix (~38 Å in *A. thaliana* TOP6B) adjacent to the short/medium helix.

The predicted structures of TOP6B from the eukaryotic supergroup Archaeplastida, which includes land plants, green algae, and red algae, possess an additional α-helix motif in the GHKL domain which is unique to this taxonomic group (**Figure 3.10**). This α-helix is ~34 Å long in *A. thaliana* TOP6B and may play a role in regulating the ATPase activity of topo VI in Archaeplastida. A sequence alignment of this region with sequence representatives from the Archaeplastida phyla Chlorophyta, Rhodophyta, and Streptophyta, reveals that this α-helix is particularly conserved in the N-terminal portion and possesses between seven and twelve positively charged lysine/arginine residues (**Figure 3.11**). The α-helix may therefore regulate the enzyme's activity by interacting with DNA above or below the ATP-gate.

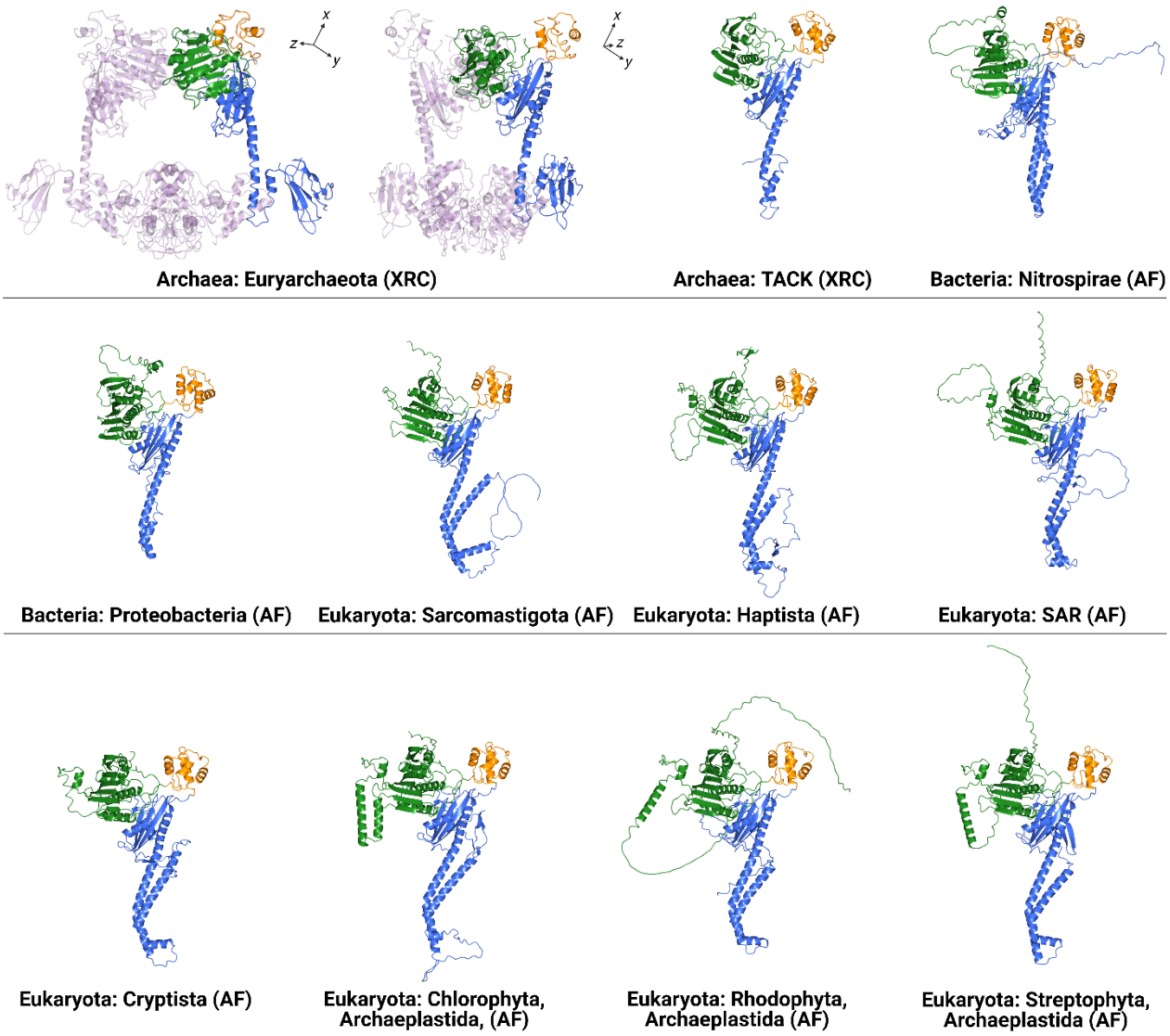

Figure 3.10| Topo VI-B structural prediction: *The X-ray crystal (XRC) structure of* ***Methanosarcina mazei*** *(Euryarchaeota) topo VI (PDB code: 2Q2E) is a heterotetramer formed of two topo VI-As and two topo VI-Bs. The* ***M. mazei*** *topo VI-B structure is comparable to the* ***Saccharolobus shibatae*** *(TACK) topo VI-B XRC structure (PDB code: 2ZBK), which both feature a transducer domain (blue), a H2TH domain (orange), and a GHKL domain (green). AlphaFold (AF) structural modelling predicts that the topo VI-B structure is also conserved in bacteria and eukaryotes. The bacterial topo VI-B AF structures are represented by* ***Nitrospira moscoviensis*** *(Nitrospirae) and* ***Pajaroellobacter abortibovis*** *(Proteobacteria). The eukaryotic topo VI-B AF structures are represented by* ***Chlamydomonas reinhardtii*** *(Chlorophyta),* ***Chondrus crispus*** *(Rhodophyta),* ***Arabidopsis thaliana*** *(Streptophyta),* ***Capsaspora owczarzaki*** *(Sarcomastigota),* ***Chrysochromulina tobinii*** *(Haptista),* ***Thalassiosira pseudonana*** *(SAR), and* ***Guillardia theta*** *(Cryptista).*

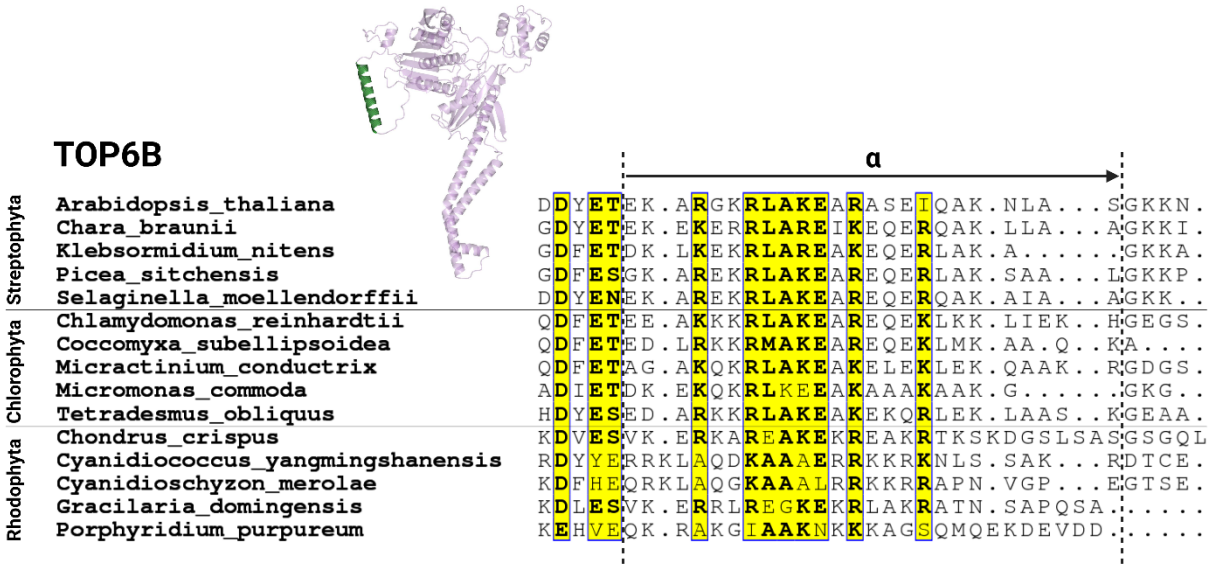

Figure 3.11| Sequence alignment of the Archaeplastida TOP6B GHKL α-helix: *The AlphaFold structures of TOP6B from members of the Archaeplastida eukaryotic superphylum possess a unique α-helix (green) in their GHKL domains. This region is conserved in the Streptophyta, Chlorophyta, and Rhodophyta phyla, and possesses a series of residues with positive sidechains.*

Pre-AlphaFold structural modelling of topo VIB-like subunits has demonstrated the absence of a H2TH-like domain in TOPOVIBL from vertebrates and invertebrates[186], but the presence of a H2TH-like SmD in MTOPVIB[72]. Structural modelling of the topo VIB-like family performed here has shown that both TOPOVIBL and MTOPVIB possess GHKL-like and transducer-like folds, but do not possess a H2TH-like domain (**Figure 3.12**). Therefore, the presence of a H2TH domain seems to be characteristic of canonical topo VI-B and is a useful marker for distinguishing from topo VIB-like. The transducer-like fold in the models of TOPOVIBL resemble the transducer domains of eukaryotic topo VI, featuring a long α-helix adjacent to a small helix and a medium-sized helix, with lengths of 66 Å, 16 Å, and 29 Å, respectively, in *Homo sapiens* (**Figure 3.12**).

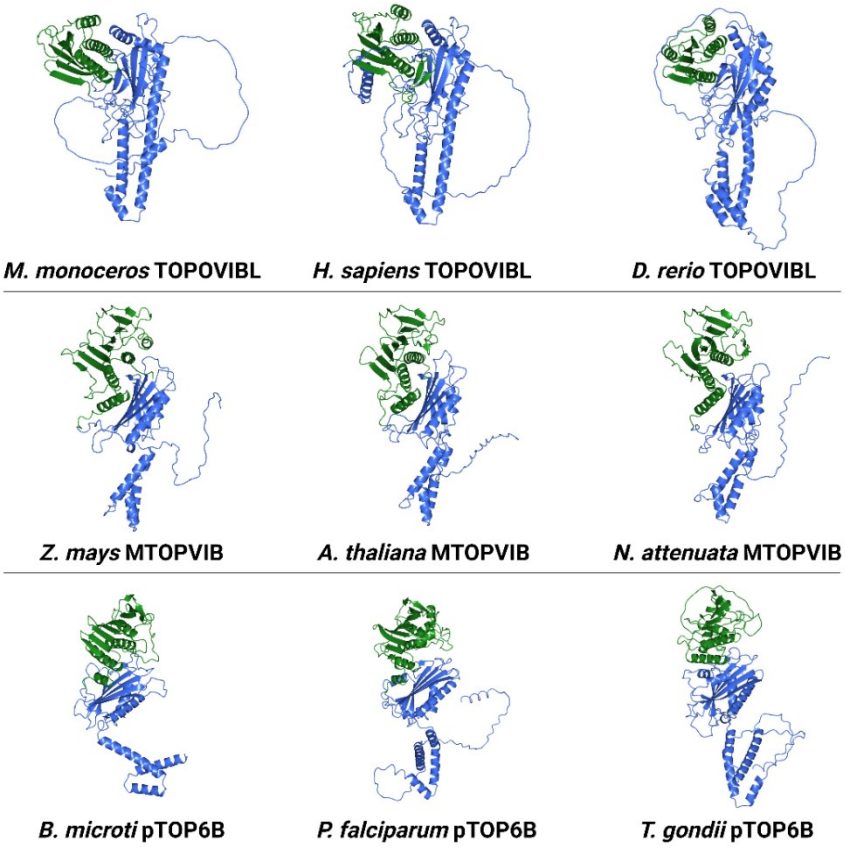

Figure 3.12| Topo VIB-like structural prediction: *AlphaFold structural modelling predicts that the topo VIB-like family possess a transducer-like domain (blue) and a GHKL-like domain (green), but do not possess a H2TH-like domain. This is true of both the TOPOVIBL and the MTOPVIB subfamilies. The Apicomplexa pTOP6B predicted structures are analogous to those of MTOPVIB. The TOPOVIBL structures are represented by* ***Monodon monoceros****,* ***Homo sapiens****, and* ***Danio rerio****. The MTOPVIB structures are represented by* ***Zea mays****,* ***Arabidopsis thaliana****, and* ***Nicotiana attenuata****. The pTOP6B structures are represented by* ***Babesia microti****,* ***Plasmodium falciparum****, and* ***Toxoplasma gondii****.*

The predicted structures of Apicomplexa pTOP6B strongly resemble those of MTOPVIB, with both subfamilies lacking a H2TH-like fold and a long transducer helix, and instead possessing a transducer-like fold formed of a bundle of α-helices (**Figure 3.12**). MTOPVIB from *Zea mays*, *A. thaliana*, and *Nicotiana benthamiana* possess five α-helices in this bundle with a maximum length of 27-28 Å, whereas pTOP6B from *P. falciparum*, *T. gondii*, and *B. microti* possess four helices in this bundle with a maximum length of 36-41 Å. The structural similarity is demonstrated by the superposition of the AlphaFold structures of *A. thaliana* MTOPVIB and *P. falciparum* pTOP6B. These two structures possess comparable GHKL-like and transducer-

like domains that occupy similar positions in 3D space (**Figure 3.13**). The model of pTOP6B is clearly unrelated to the models of protist TOP6B, particularly that of the SAR species *T. pseudonana*. Taken together, these results suggest that Apicomplexa pTOP6B is a member of the topo VIB-like family and the first protist member of the MTOPVIB subfamily. However, using the sequence of *P. falciparum* pTOP6B as a query for position-specific iterative and protein-protein BLAST search, no other protist MTOPVIB sequences were identified outside the Apicomplexa. Given that Apicomplexa appear to possess two canonical Spo11 subunits, topo VI does not seem to be present in this phylum.

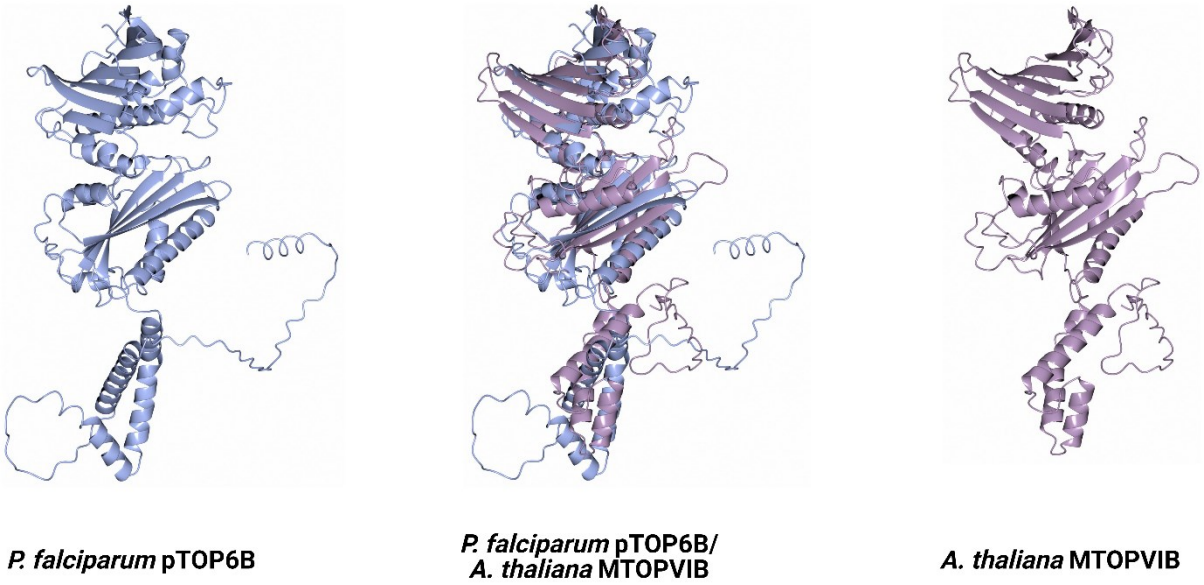

Figure 3.13| Superimposition of pTOP6B and MTOPVIB: *The AlphaFold structures of* ***Plasmodium falciparum*** *pTOP6B (ice blue) and* ***Arabidopsis thaliana*** *MTOPVIB (lilac) are similar and superimpose effectively. The two subunits are thus likely members of the same subfamily.*

3.2.5 Distribution of topo VI in the three domains of life

Using a combination of phylogenetic analyses and structural modelling I have shown that it is possible to distinguish between the subunits of the topo VI and topo VI-like complexes. Using these tools, I can now describe which organisms tested here in the three domains of life possess both canonical subunits of topo VI. All sequenced archaea genomes in the four taxonomic supergroups have been reported to possess topo VI, except the taxonomic order

Thermoplasmatales[8,48]. In this study, topo VI was found in all archaeal orders tested in the four supergroups, including the Thermoplasmatales (**Table 3.3**). The organism in question, the extreme acidophile *Cuniculiplasma divulgatum*, is the first and only species of the Thermoplasmatales to be shown to possess topo VI. Using the sequences of topo VI-A and topo VI-B from *C. divulgatum* as query protein-protein BLAST searches, no other topo VI sequences were found in the Thermoplasmatales.

Topo VI has previously only been reported in the Deltaproteobacteria, Planctomycetia, and Oligoflexia taxonomic classes of bacteria, which lie within the Proteobacteria and PVC supergroups[27]. In this study, I have identified both canonical subunits of topo VI in many more classes within five additional bacterial supergroups, and within the two newly proposed phyla Candidatus Eisenbacteria and Candidatus Rokubacteria (**Table 3.4**). Many of the bacterial species identified here that possess topo VI are poorly characterised organisms that lack complete taxonomic categorisation, and only eleven of these bacterial families possess a given genus and species name. Therefore, topo VI is widely distributed in the bacterial domain, but unlike in the archaeal domain, is sparsely distributed within individual families. Furthermore, all topo VI-possessing bacteria identified in this study have been previously characterised as gram-negative.

Eukaryotic topo VI was first discovered in *A. thaliana* and was shown to be widespread in plants[49]. Topo VI was later shown to also be present in protists within the same phylum as plants, the Streptophyta, and to also be present in protists in the Chlorophyta and Rhodophyta phyla, as well as in the protist supergroup Haptista[50]. In this study, I have identified both canonical subunits of topo VI in the protist supergroups Cryptista and Haptista, as well as in many classes of the SAR Stramenopile phylum Ochrophyta (**Table 3.5**). However, using the Ochrophyta sequences of TOP6B as query protein-protein BLAST searches, no other TOP6B sequences were identified in other stramenopiles or in the other SAR major clades Alveolata and Rhizaria. A recent formation of the eukaryote 'tree of life' separates all eukaryotic supergroups into two major clades[189]. Interestingly, one of these clades contains all the plant and protist supergroups that have been shown to possess topo VI, while the other clade contains the rest of the eukaryotes, such as animals, fungi, and other protist supergroups, that have been shown to lack topo VI. Like in archaea, topo VI is ubiquitous in the Archaeplastida supergroup but it appears only sporadically in related protists.

Table 3.3| Distribution of topo VI in archaea

Archaea Supergroup	Phylum	Class	Order	Representative genus
Asgard	Candidatus Heimdallarchaeota			
DPANN	Candidatus Aenigmarchaeota			
	Candidatus Altiarchaeota			
	Candidatus Micrarchaeota			
	Candidatus Woesearchaeota			
	Diapherotrites			
	Nanoarchaeota		Nanoarchaeales	Nanobsidianus
Euryarchaeota	Euryarchaeota	Archaeoglobi	Archaeoglobales	Geoglobus
		Halobacteria	Halobacteriales	Halobacterium
			Haloferacales	Haloferax
			Natrialbales	Haloterrigena
		Methanobacteria	Methanobacteriales	Methanothermobacter
		Methanococci	Methanococcales	Methanocaldococcus
		Methanomicrobia	Methanosarcinales	Methanosarcina
		Methanonatronarchaeia	Methanonatronarchaeales	Methanonatronarchaeum
		Thermococci	Thermococcales	Thermococcus
		Thermoplasmata	Methanomassiliicoccales	Methanomassiliicoccus
			Thermoplasmatales	Cuniculiplasma
TACK	Candidatus Bathyarchaeota			
	Candidatus Korarchaeota			
	Candidatus Marsarchaeota			
	Candidatus Nezhaarchaeota			
	Candidatus Verstraetearchaeota			
	Crenarchaeota	Thermoprotei	Acidilobales	Acidilobus
			Desulfurococcales	Pyrodictium
			Fervidicoccales	Fervidicoccus
			Sulfolobales	Saccharolobus
			Thermoproteales	Pyrobaculum
	Thaumarchaeota	Nitrososphaeria	Nitrosopumilales	

Table 3.4| Distribution of topo VI in bacteria

Bacteria Supergroup	Phylum	Class	Order	Family	Genus
Candidatus Eisenbacteria	Candidatus Eisenbacteria				
Candidatus Rokubacteria	Candidatus Rokubacteria				
Elusimicrobia	Elusimicrobia				
FCB	Acidobacteria				
	Candidatus Hydrogenedentes				
	Candidatus Latescibacteria				
	Candidatus Marinimicrobia				
	Gemmatimonadetes				
Nitrospirae	Nitrospirae	Nitrospira	Nitrospirales	Nitrospiraceae	Nitrospira
Proteobacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	Bradyrhizobium
				Hyphomicrobiaceae	
			Rhodobacterales	Rhodobacteraceae	Marimonas
		Betaproteobacteria	Burkholderiales	Burkholderiaceae	
		Deltaproteobacteria	Myxococcales	Anaeromyxobacteraceae	Anaeromyxobacter
				Polyangiaceae	Pajaroellobacter
		Gammaproteobacteria	Chromatiales		
		Oligoflexia	Bdellovibrionales	Bdellovibrionaceae	Bdellovibrio
PVC	Lentisphaerae	Lentisphaeria	Lentisphaerales		
	Planctomycetes	Candidatus Brocadiae	Candidatus Brocadiales		
		Phycisphaerae	Phycisphaerales	Phycisphaeraceae	
			Sedimentisphaerales	Sedimentisphaeraceae	Sedimentisphaera
		Planctomycetia	Pirellulales	Lacipirellulaceae	Lacipirellula
				Pirellulaceae	Pirellula
				Thermoguttaceae	Bythopirellula
			Planctomycetales	Planctomycetaceae	
	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales		
Terrabacteria	Armatimonadetes				
	Chloroflexi				
	Cyanobacteria	Cyanophyceae	Synechococcales	Leptolyngbyaceae	
	Firmicutes				
Thermotogae	Thermotogae				

Table 3.5| Distribution of topo VI in eukaryotes

Eukaryotes Supergroup	Phylum	Class	Order	Family	Genus
Archaeplastida	Chlorophyta	Chlorodendrophyceae	Chlorodendrales	Chlorodendraceae	Tetraselmis
		Chlorophyceae	Chlamydomonadales	Chlamydomonadaceae	Chlamydomonas
			Sphaeropleales	Scenedesmaceae	Tetradesmus
		Chloropicophyceae	Chloropicales	Chloropicaceae	Chloropicon
		Mamiellophyceae	Mamiellales	Mamiellaceae	Micromonas
		Trebouxiophyceae	Chlorellales	Chlorellaceae	Micractinium
			Chlorococcales	Coccomyxaceae	Coccomyxa
	Rhodophyta	Bangiophyceae	Cyanidiales	Cyanidiaceae	Cyanidioschyzon
			Porphyridiales	Porphyridiaceae	Porphyridium
		Florideophyceae	Gigartinales	Gigartinaceae	Chondrus
			Gracilariales	Gracilariaceae	Gracilariopsis
	Streptophyta	Bryopsida	Funariales	Funariaceae	Physcomitrella
		Charophyceae	Charales	Characeae	Chara
		Klebsormidiophyceae	Klebsormidiales	Klebsormidiaceae	Klebsormidium
		Lycopodiopsida	Selaginellales	Selaginellaceae	Selaginella
		Magnoliopsida	Alismatales	Araceae	Anthurium
			Amborellales	Amborellaceae	Amborella
			Arecales	Arecaceae	Phoenix
			Asparagales	Orchidaceae	Dendrobium
			Asterales	Asteraceae	Artemisia
			Brassicales	Brassicaceae	Arabidopsis
			Caryophyllales	Chenopodiaceae	Spinacia
			Cornales	Nyssaceae	Nyssa
			Cucurbitales	Cucurbitaceae	Cucumis
			Ericales	Theaceae	Camellia
			Fabales	Fabaceae	Mucuna
			Fagales	Juglandaceae	Juglans
			Gentianales	Rubiaceae	Coffea
			Lamiales	Bignoniaceae	Handroanthus
			Laurales	Lauraceae	Cinnamomum
			Malpighiales	Salicaceae	Populus
			Malvales	Malvaceae	Theobroma
			Myrtales	Lythraceae	Punica
			Poales	Poaceae	Oryza
			Proteales	Nelumbonaceae	Nelumbo
			Ranunculales	Papaveraceae	Macleaya
			Rosales	Rosaceae	Prunus
			Sapindales	Rutaceae	Citrus
			Solanales	Solanaceae	Nicotiana
			Vitales	Vitaceae	Vitis
			Zingiberales	Musaceae	Musa
		Marchantiopsida	Marchantiales	Marchantiaceae	Marchantia
		Pinopsida	Pinales	Pinaceae	Picea
Cryptista	Cryptophyta	Cryptophyceae	Pyrenomonadales	Geminigeraceae	Guillardia
Haptista	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina
SAR	Ochrophyta	Bacillariophyceae	Bacillariales	Bacillariaceae	Fragilariopsis
			Naviculales	Phaeodactylaceae	Phaeodactylum
		Bigyra	Bicosoecida	Cafeteriaceae	Cafeteria
			Thraustochytrida	Thraustochytriaceae	Hondaea
		Coscinodiscophyceae	Thalassiosirales	Thalassiosiraceae	Thalassiosira
		Eustigmatophyceae	Eustigmatales	Monodopsidaceae	Nannochloropsis
		Oomycota	Saprolegniales	Saprolegniaceae	Aphanomyces
		Phaeophyceae	Ectocarpales	Ectocarpaceae	Ectocarpus
Sarcomastigota	Choanozoa	Filasterea	Ministeriida	Capsasporidae	Capsaspora
		Ichthyosporea	Ichthyophonida	Ichthyophonidae	Sphaeroforma

3.2.6 Bacterial topo VI is usually indictive of the absence of topo IV

All archaea, except certain members of the Thermoplasmatales taxonomic order, possess topo VI, some also possess gyrase, but no archaeal topo IV has been identified[12]. Bacteria seem to have acquired topo VI via HGT from the archaeal supergroup Euryarchaeota, and this process is analogous to the acquisition of bacterial gyrase by some species of archaea[8]. The two gyrase subunits, GyrA and GyrB, have been found in all sequenced bacterial genomes, whereas the two homologous topo IV subunits, ParC and ParE, respectively, are present in most bacteria[27]. An exception is the bacteria *Aquifex aeolicus*, which possess a chimeric type IIA topo formed of GyrB and ParC[190]. To investigate the role of topo VI in bacteria, I sought to determine the composition of type IIA topos in bacterial species that have been shown to possess topo VI.

I collected putative GyrA/ParC sequences from topo VI-possessing bacteria using the sequence of *E. coli* ParC as a query protein-protein BLAST search. Only the proteomes of the eleven topo VI-possessing bacteria that have a given genus and species name were analysed to ensure that the results were relevant to a single organism. The proteomes of nine of the topo VI-possessing bacterial species possessed just a single putative GyrA/ParC subunit, and two of the proteomes possessed two putative GyrA/ParC subunits. These sequences were subject to phylogenetic analysis alongside known GyrA and ParC sequences to elucidate their identity. GyrA and ParC separate into two distinct clades, with all eleven topo VI-possessing bacteria containing putative GyrA/ParC sequences in the GyrA clade (**Figure 3.14**). Two of these topo VI-possessing bacteria, *Marimonas lutisalis* and *Bradyrhizobium sediminis* contain putative GyrA/ParC sequences in both clades.

GyrA and ParC sequences can also be distinguished by the presence of a GyrA-box, which is unique to GyrA and essential for the supercoiling activity of gyrase[40,41]. Therefore, I performed a sequence alignment of the gyrA-box region using the same sequences to confirm the result of the phylogenetic analysis. As expected, all the known bacterial GyrA sequences possessed the five invariant gyrA-box residues, whereas the same region in known ParC sequences was missing (**Figure 3.15**). This alignment validated the results of the phylogenetic analysis, as all topo VI-possessing bacteria contained a putative GyrA/ParC sequence that possessed the conserved GyrA-box. Furthermore, *Marimonas lutisalis* and *Bradyrhizobium sediminis* possessed an addition putative GyrA/ParC sequence that lacked a GyrA-box.

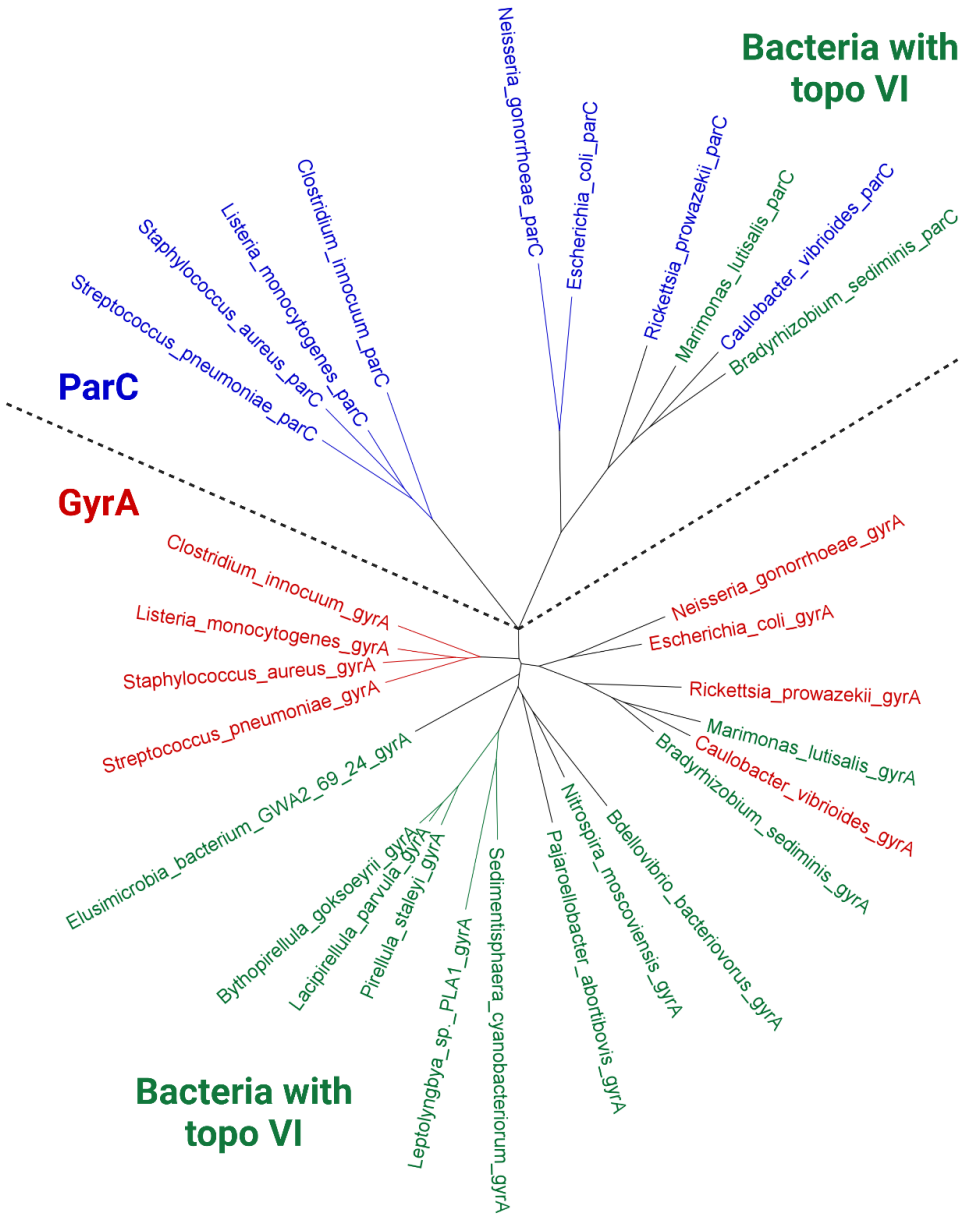

Figure 3.14| Unrooted maximum-likelihood phylogenetic tree of bacterial GyrA and ParC: *The GyrA (red) and ParC (blue) subunits of gyrase and topo IV, respectively, split into two distinct clades. Most of the bacteria that have been shown to possess topo VI (green) all contain a GyrA but not a ParC. The exceptions,* ***Marimonas lutisalis*** *and* ***Bradyrhizobium sediminis****, contain both GyrA and ParC.*

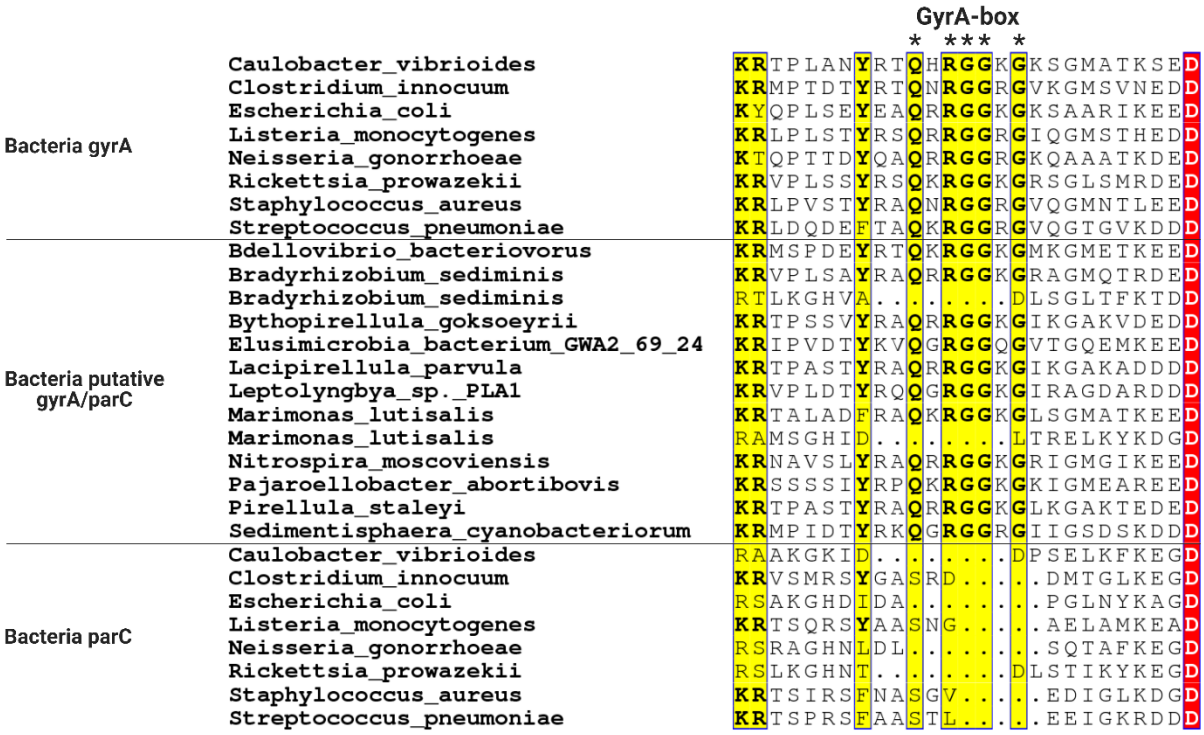

Figure 3.15| Sequence alignment of the bacterial GyrA-box: *Bacterial GyrA and ParC sequences were aligned at the GyrA-box region alongside putative GyrA/ParC sequences from bacterial species that possess topo VI. The gyrA-box is unique to GyrA and can be used to distinguish sequences from ParC. All bacteria that possess topo VI also possess GyrA, and the bacteria* ***Bradyrhizobium sediminis*** *and* ***Marimonas lutisalis*** *possess topo IV in addition to gyrase.*

Given that the proteome of *A. aeolicus* possesses just a single subunit from each of the GyrA/GyrB and ParC/ParE pairs, it is necessary to confirm that an organism contains a full complement of gyrase/topo IV genes. Therefore, I performed a protein-protein BLAST search using *E. coli* ParE as a query to identify GyrB/ParE sequences in the topo VI-possessing bacteria. These sequences were subject to phylogenetic analysis alongside GyrB and ParE sequences to elucidate their identities. GyrB and ParE each separate into two distinct clades, with all eleven topo VI-possessing bacteria containing putative GyrB/ParE sequences in one of the GyrB clades (**Figure 3.16**). *Marimonas lutisalis* and *Bradyrhizobium sediminis* also contain putative GyrB/ParE sequences in one of the ParE clades, confirming that all species possess the full complement of gyrase/topo IV genes.

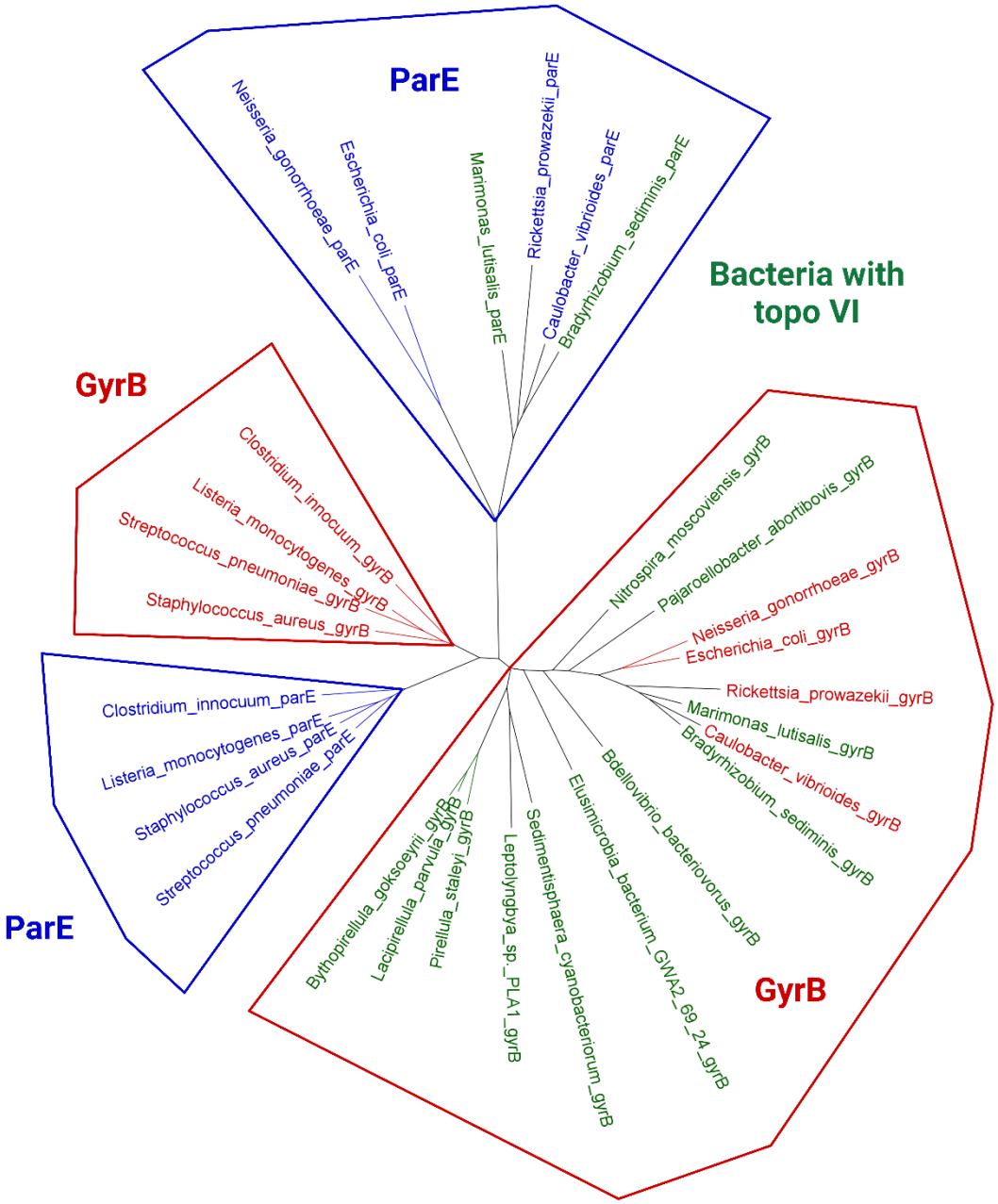

Figure 3.16| Unrooted maximum-likelihood phylogenetic tree of bacterial GyrB and ParE: *The GyrA (red) and ParC (blue) subunits of gyrase and topo IV, respectively, split into two distinct clades each. Most of the bacteria that have been shown to possess topo VI (green) all contain a GyrB but not a ParE. The exceptions,* ***Marimonas lutisalis*** *and* ***Bradyrhizobium sediminis****, contain both GyrB and ParE.*

Taken together, these results suggest that the presence of topo VI in bacteria is usually indicative of the absence of topo IV (**Table 3.6**). Interestingly, two of the topo VI-possessing bacterial species identified here, *Marimonas lutisalis* and *Bradyrhizobium sediminis*, possess both gyrase and topo IV meaning that they are the only known species of bacteria with three type II topos, and the only known organisms in the three domains of life that possess gyrase, topo IV and topo VI. Topo VI-possessing eukaryotes also possess three type II topos but possess topo II instead of topo IV.

Table 3.6| Distribution of type II topoisomerase subunits in topo VI-possessing bacteria

	Topo VI		Gyrase		Topo IV	
Organism	**Topo VI-A**	**Topo VI-B**	**GyrA**	**GyrB**	**ParC**	**ParE**
Bdellovibrio bacteriovorus	+	+	+	+	-	-
Bradyrhizobium sediminis	+	+	+	+	+	+
Bythopirellula goksoeyrii	+	+	+	+	-	-
Elusimicrobia bacterium GWA2_69_24	+	+	+	+	-	-
Lacipirellula parvula	+	+	+	+	-	-
Leptolyngbya sp. PLA1	+	+	+	+	-	-
Marimonas lutisalis	+	+	+	+	+	+
Nitrospira moscoviensis	+	+	+	+	-	-
Pajaroellobacter abortibovis	+	+	+	+	-	-
Pirellula staleyi	+	+	+	+	-	-
Sedimentisphaera cyanobacteriorum	+	+	+	+	-	-

3.2.7 RHL1 and BIN4 have a conserved structural core

In plants there is an unusual requirement for two accessory proteins for the topo VI complex: RHL1[123] and BIN4[124]. These accessory proteins have been reported to possess partial sequence homology with the CTD of mammalian topo IIα[123,124], however this homology was very poor in my own alignments and so was not examined further. To determine whether these accessory proteins are distributed across the three domains of life, I used RHL1 and BIN4 sequences from *A. thaliana* as query protein-protein BLAST searches. Homologues of RHL1 and BIN4 were present in all topo VI-possessing eukaryotic phyla but were not found in archaea or bacteria. The accessory proteins were also not found in Apicomplexa, compounding the evidence against this phylum possessing topo VI.

I then performed AlphaFold structural modelling to investigate whether the predicted structures of RHL1 and BIN4 can shed light on their functional contributions to the eukaryotic topo VI complex. Models of both RHL1 and BIN4 are largely disordered and lack secondary structure definition, however, they possess a conserved core consisting of a four-stranded antiparallel β-sheet adjacent to a β-hairpin (**Figures 3.17** and **3.18**). In some structures, such as *Chrysochromulina tobinii* RHL1 and *A. thaliana* BIN4, the β-sheet and β-hairpin of this module possess an additional parallel β-strand. The conserved structural cores of RHL1 and BIN4 are found at their C-terminal and N-terminal portions, respectively, and these domains occupy similar positions in 3D space when superimposed (**Figure 3.19**). Therefore, these accessory proteins are likely to be performing analogous roles within the eukaryotic topo VI complex.

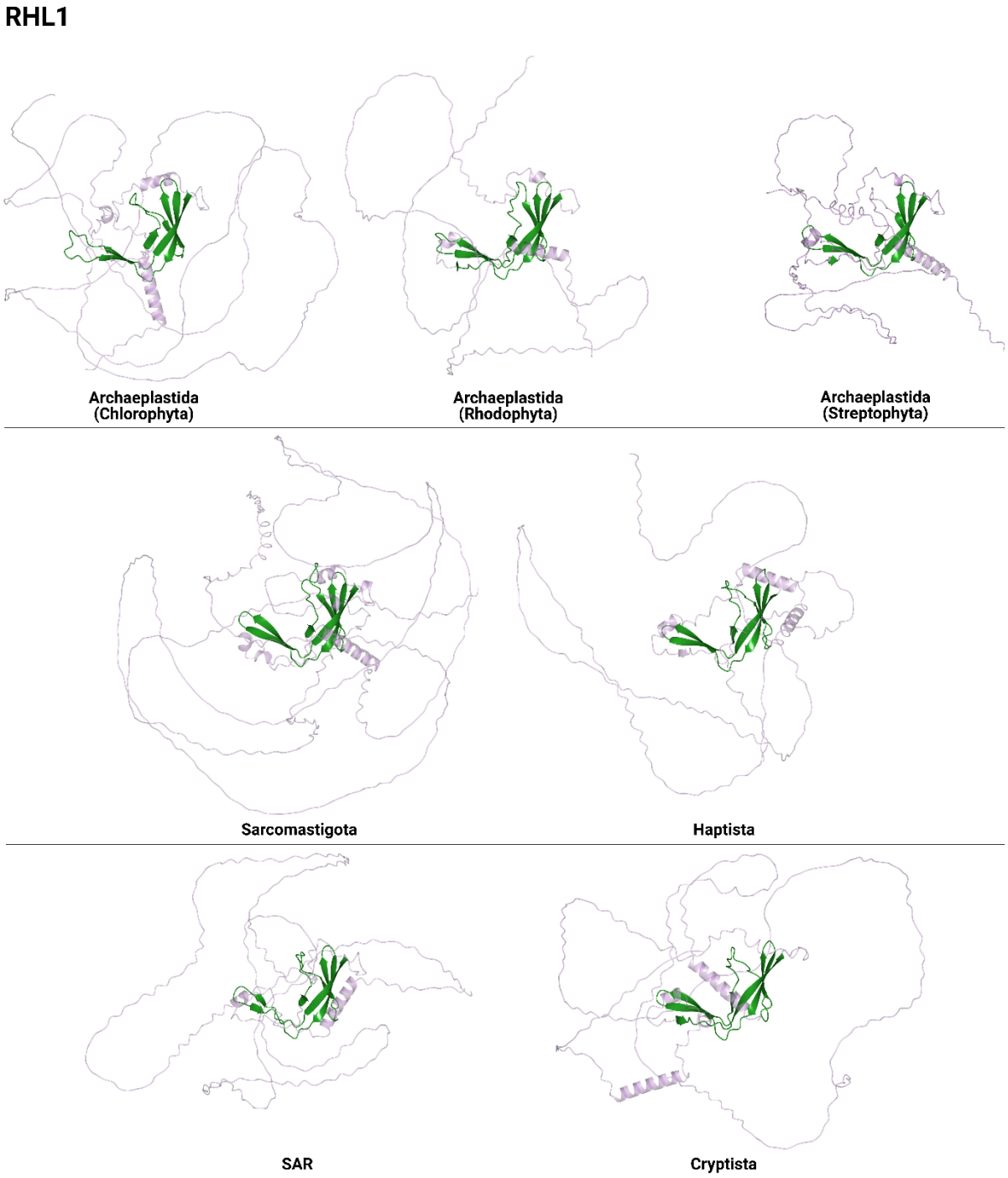

Figure 3.17| RHL1 structural prediction: *AlphaFold structural modelling predicts that RHL1 is largely disordered but has a conserved core (green) consisting of a four-stranded antiparallel β-sheet adjacent to a β-hairpin. The RHL1 structures are represented by* ***Raphidocelis subcapitata*** *(Chlorophyta),* ***Gracilariopsis chorda*** *(Rhodophyta),* ***Arabidopsis thaliana*** *(Streptophyta),* ***Capsaspora owczarzaki*** *(Sarcomastigota),* ***Chrysochromulina tobinii*** *(Haptista),* ***Nannochloropsis salina*** *(SAR), and* ***Guillardia theta*** *(Cryptista).*

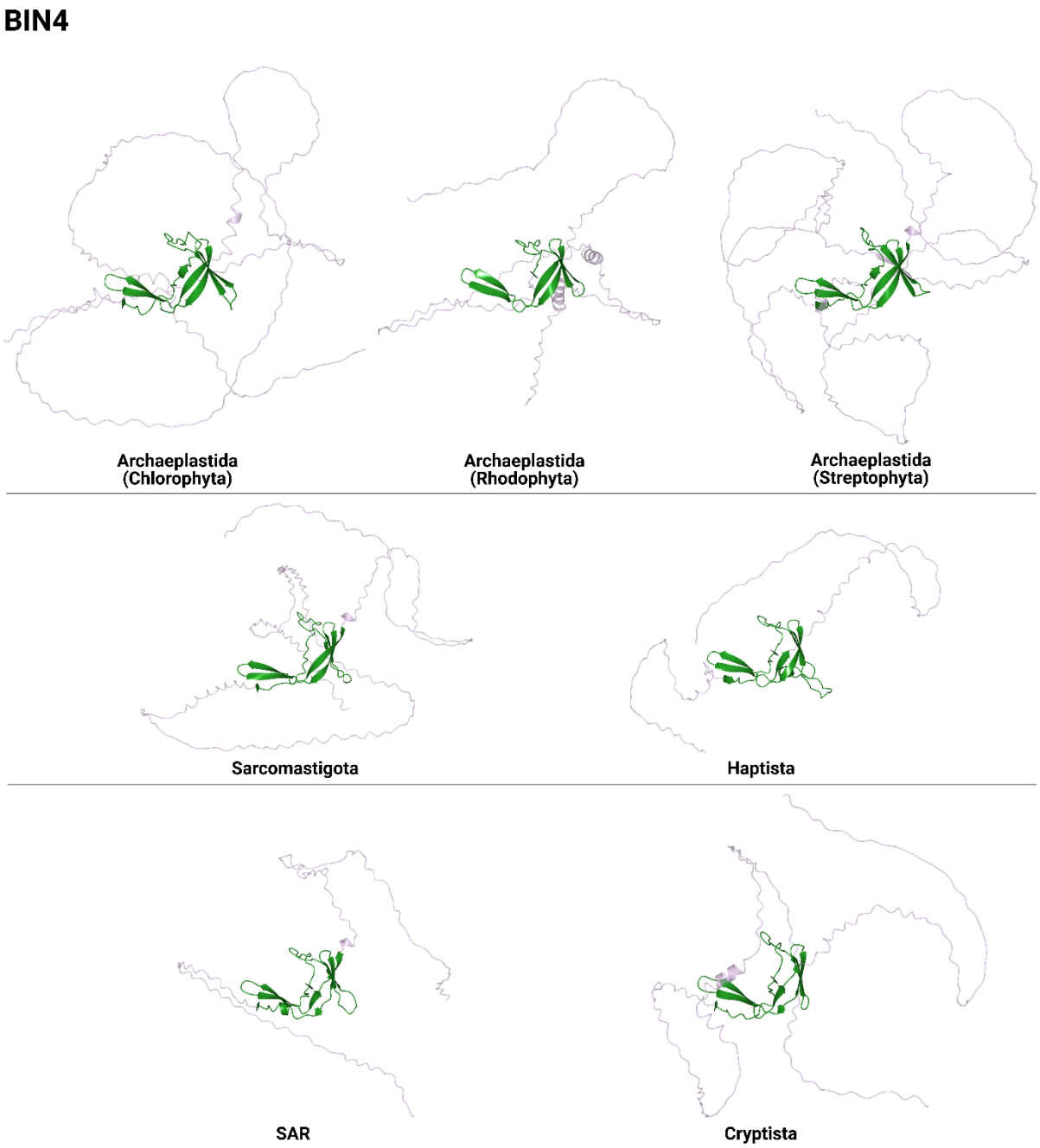

Figure 3.18| BIN4 structural prediction: *AlphaFold structural modelling predicts that BIN4 is largely disordered but has a conserved core (green) consisting of a four-stranded antiparallel β-sheet adjacent to a β-hairpin. The BIN4 structures are represented by* ***Chlorella sorokiniana*** *(Chlorophyta),* ***Gracilariopsis chorda*** *(Rhodophyta),* ***Arabidopsis thaliana*** *(Streptophyta),* ***Sphaeroforma arctica*** *(Sarcomastigota),* ***Chrysochromulina tobinii*** *(Haptista),* ***Ectocarpus siliculosus*** *(SAR), and* ***Guillardia theta*** *(Cryptista).*

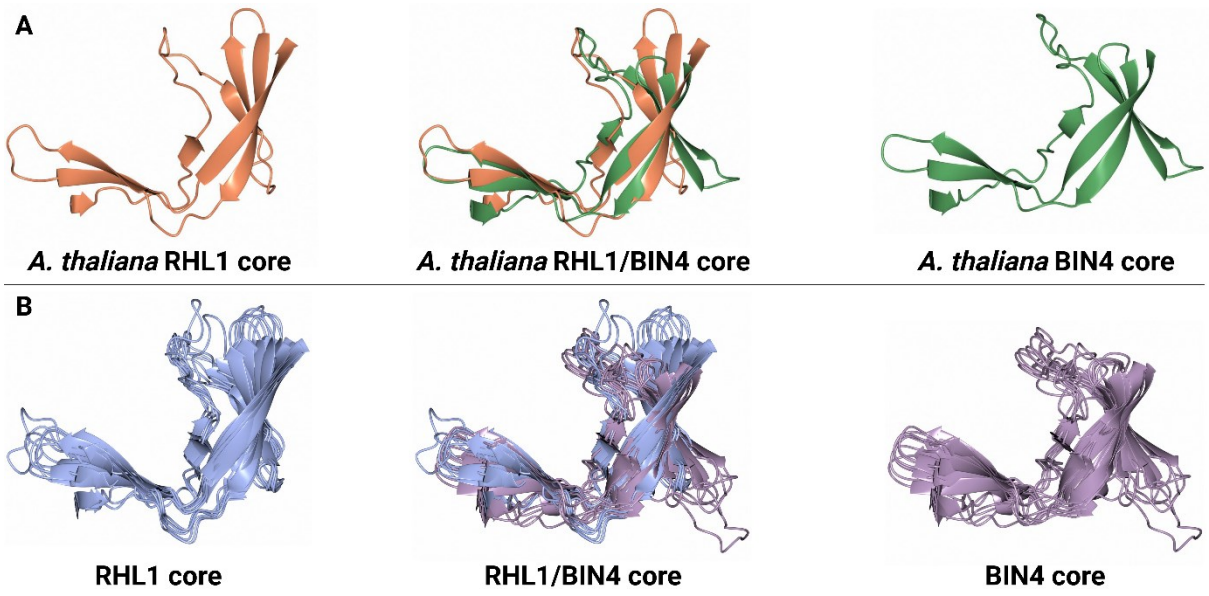

Figure 3.19| Superimposition of the RHL1 and BIN4 structural core: (**A**) *The AlphaFold-modelled structural cores of RHL1 and BIN4 from Arabidopsis thaliana are similar and superimpose effectively. (**B**) The AlphaFold-modelled structural cores of RHL1 and BIN4 from all tested species also super impose effectively. The RHL1 structures are represented by **Raphidocelis subcapitata**, **Gracilariopsis chorda**, **A. thaliana**, **Capsaspora owczarzaki**, **Chrysochromulina tobinii**, **Nannochloropsis salina**, and **Guillardia theta**. The BIN4 structures are represented by **Chlorella sorokiniana**, **G. chorda**, **A. thaliana**, **Sphaeroforma arctica**, **C. tobinii**, **Ectocarpus siliculosus**, and **G. theta**.*

To determine whether the β-sheet/hairpin module in RHL1 and BIN4 is conserved at the sequence level, I performed a sequence alignment of this region. The structural core was highly conserved within RHL1 and BIN4 homologues but was not conserved between them (**Figures 3.20** and **3.21**). The sequence homology was stronger in regions corresponding to individual β-strands and was conserved across all topo VI-possessing eukaryotic taxonomic supergroups. Four invariant residues are present in the RHL1 homologues: a glycine at the start of strand 3, a proline at the start of strand 4, and two glycines in strand 5 (**Figure 3.20**). BIN4 possesses seven invariant residues that are located within two conserved sequence motifs: a GDxGxxG motif at the start of strand 3 and a DxKG motif at the end of strand 4 **(Figure 3.21)**. Taken together, RHL1 and BIN4 are distinct proteins that possess a structurally conserved β-sheet/hairpin core.

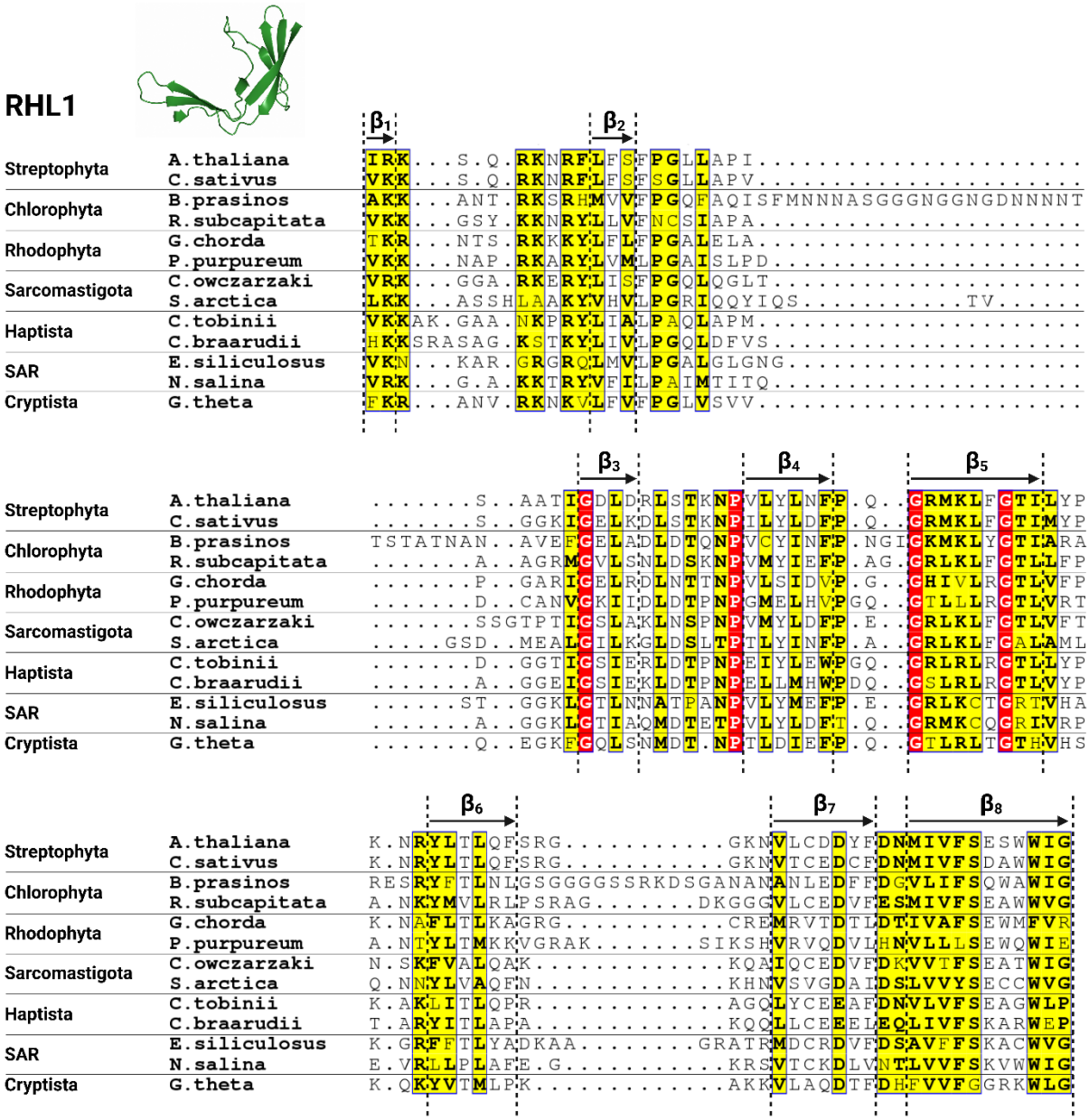

Figure 3.20| Sequence alignment of the conserved RHL1 structural core: *The AlphaFold structure of the conserved* ***Arabidopsis thaliana*** *RHL1 core (green) consists of a four-stranded antiparallel β-sheet (β2-5, β8) adjacent to a β-hairpin (β1, β6-7). The RHL1 structural core possesses four invariant residues (red).*

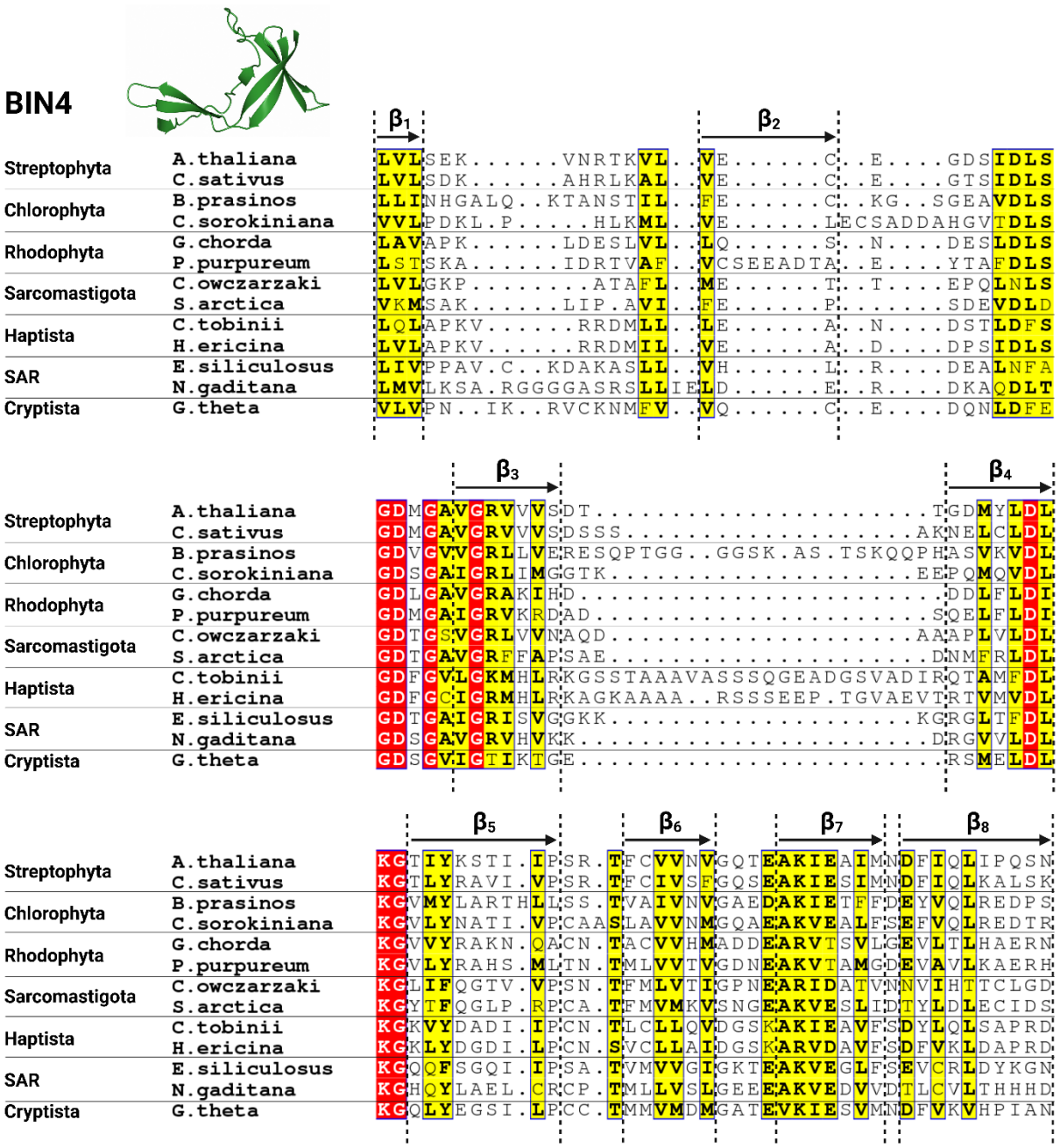

Figure 3.21| Sequence alignment of the conserved BIN4 structural core: *The AlphaFold structure of the conserved* ***Arabidopsis thaliana*** *BIN4 core (green) consists of a four-stranded antiparallel β-sheet (β2-5, β8) adjacent to a β-hairpin (β1, β6-7). The BIN4 structural core possesses seven invariant residues (red).*

3.3 Discussion

3.3.1 General discussion

For many years topo VI was thought to be an enzyme unique to archaea and was largely neglected as an "archaeal curiosity"[191]. Excitement started to surround topo VI following the publications of the complete genomic sequences of the model plant *A. thaliana*[175] and the malaria parasite *Plasmodium falciparum*[138] in 2000 and 2002, respectively. Using these resources, it took less than a year to discover that gene products homologous to topo VI-A and topo VI-B were present in both organisms[49,129]. Suddenly, what was just an archaeal curiosity was now an enzyme of agricultural and pharmacological interest. Further excitement arose following a monumental discovery in 2016: the topo VI scaffold had been adopted by eukaryotes to initiate homologous recombination[72,186]. The topo VI DSB machinery seemingly possesses unique properties that are absent in other type II topos, and it is necessary to establish its evolution in the tree of life for these properties to be revealed.

3.3.2 Topo VI and topo VI-like are distinguishable

The A subunit of topo VI is a useful nuclease module and has been adapted into the essential meiotic enzyme Spo11. Given that topo VI-A and canonical Spo11 participate in different cellular activities, and that they each interact with a different B subunit partner, it is reasonable to predict that their specialisation would be observable at the sequence level. Phylogenetic analysis of topo VI-A and Spo11 has previously provided evidence that this is indeed the case[50], and the extensive phylogenetic analyses presented in this chapter provides further evidence that the A subunits of the topo VI and topo VI-like complexes can be distinguished at the sequence level. However, given that the DSB mechanism of topo VI-A and canonical Spo11 is conserved, both subunits possess the same key invariant catalytic residues, and no specific motifs were found to be present in one subunit and not the other. Therefore, the sequence alignment and phylogenetic tree software seem to be identifying subtle changes at the sequence level that are not noticeable by manual inspection.

The B subunit of topo VI plays a key role in regulating DSB formation by its A subunit partner, and a similar function must be true of the B subunit of the meiotic topo VI-like complex. However, unlike topo VI-B, topo VIB-like has a degenerate version of the GHKL domain that lacks invariant ATP-binding residues, suggesting that this subunit either does not require ATP to regulate DSB formation or that ATP plays a less significant role than it does in topo VI-B. If

ATP is not required by the topo VI-like complex to generate meiotic DSBs, then we must consider whether topo VI-B is imparting an ATP-independent regulation of DSB formation that has been retained by topo VIB-like, or whether topo VIB-like has evolved a unique mechanism of DSB regulation that is absent in topo VI-B.

AlphaFold is a powerful tool that claims to have solved the protein folding problem by accurately predicting a protein's 3D structure from its amino acid sequence to within the width of one atom[162]. The Critical Assessment of Structure Prediction (CASP) provides independent assessments of protein structural modelling and has scored AlphaFold with an average backbone prediction accuracy of 80% making it the world leader in structural prediction[192]. Pre-AlphaFold structural modelling of topo VIB-like from animals and fungi[186] has suggested that these subunits do not possess the H2TH domain found in the crystal structures of topo VI-B from *M*. *mazei*[70] and *S*. *shibatae*[61]. The H2TH domain has been shown to contribute to the specificity of topo VI for supercoiled DNA and to the coupling of ATP hydrolysis with strand-passage[58]. Using AlphaFold, I have shown that the HT2H is indeed absent in TOP6BL, as well as in MTOPVIB (**Figure 3.12**). Although MTOPVIB had previously been reported to possess a H2TH-like domain[72], the AlphaFold prediction is far more likely to be accurate. I have also used AlphaFold to show that the H2TH domain is conserved in topo VI-B from all analysed sequences across the three domains of life (**Figure 3.8**), meaning that the presence or absence of this domain can distinguish between topo VI-B and topo VIB-like. The absence of a H2TH domain is not surprising given that the meiotic topo VI-like complex is unlikely to be performing strand-passage, and suggests that the regulatory function of topo VIB-like is less stringent than that of topo VI-B.

3.3.3 The topo VI-like complex in Apicomplexa

Topo VI-A and canonical Spo11 can be distinguished phylogenetically, while a combination of structural modelling and phylogenetic analysis is required to distinguish between topo VI-B and topo VIB-like. With this knowledge, it was possible to determine that the putative topo VI gene products in the malarial parasite *Plasmodium*, as well as in other members of the Apicomplexa phylum, assemble into a topo VI-like complex, and that this phylum does not possess topo VI. The Apicomplexa pTOP6B subunit is the first topo VIB-like subunit to be identified in protists, and structural modelling suggests that it is a member of the MTOPVIB subfamily. This discovery expands the topo VIB-like family further across the eukaryotic

domain and provides another hint that the topo VI-like scaffold is conserved in eukaryotes. The Apicomplexa MTOPVIB shows a stronger homology to the N-terminus of topo VI-B than all other topo VIB-like subunits which explains why it is the only subfamily that can be pulled from protein-protein BLAST searches using topo VI-B as a query. Although the TOPVIBL subfamily could be identified using *A*. *thaliana* MTOPVIB as a query position-specific iterative BLAST search[186], my own searches using *P. falciparum* pTOP6B could not identify MTOPVIB homologues outside of the Apicomplexa phylum. Although it seems highly likely that the topo VIB-like family is conserved in all eukaryotes, finding this subunit in certain protist proteomes may prove to be elusive given the difficulty in searching for structural homologues.

A series of genome-wide knockout studies have demonstrated that SPO11-A and pTOP6B from *Plasmodium* have similar expression profiles in vivo[140-142], and various transcriptomics studies have shown that these genes are upregulated in the sexual developmental stages of the parasite's lifecycle[141,142]. *Plasmodium* SPO11-B, on the other hand, has been shown to be essential in the asexual developmental stages[140,141]. These studies, taken together with the results of this chapter, suggest that SPO11-A and pTOP6B form a meiotic topo VI-like complex in Apicomplexa. Interestingly, SPO11-B seems to perform a meiosis-independent function that does not require pTOP6B, and it remains unknown whether this activity requires a different topo VIB-like partner. The unusual asexual behaviour of Apicomplexa SPO11-B, together with its phylogenetic distinction from other eukaryotic Spo11 subfamilies, raises the possibility that this nuclease has evolved a novel function that is unique to this subclass.

In this chapter, I have concluded that *Plasmodium*, and other Apicomplexa, do not possess topo VI. This result contradicts the work of Chalapareddy et al, who have claimed that SPO11-A and pTOP6B could complement a yeast topo II null mutation[139], and that a crude yeast cell extract possessing overexpressed SPO11-A and pTOP6B possessed decatenation activity[139]. This work strongly suggests that SPO11-A and pTOP6B form a canonical topo VI complex and not a meiotic topo VI-like complex. The genetic and biochemical data presented in this study should be repeated by an independent group before being accepted as a valid discrepancy. If topo VI is not present in *Plasmodium*, then it can no longer be considered as a valid target for antimalarial drugs.

3.3.4 Evolution of the topo VI and topo VI-like complexes

Topo VI is ubiquitous in archaea, but only occurs sporadically in bacteria, and bacterial topo VI sequences emerge within phylogenetic groups of archaeal topo VI sequences (**Figures 3.2** and **3.6**). This suggests, as previously predicted by Forterre et al[27], that the topo VI genes were introduced to bacteria from archaea via HGT. Archaeal topo VI-A and eukaryotic SPO11-3 group close together on an unrooted phylogenetic tree (**Figure 3.3**), suggesting that the topo VI subunits were present in a common ancestor of eukaryotes and archaea, as suggested by Malik et al[50]. The same study also proposed that a gene duplication event separated SPO11-2 from SPO11-3, and that a second gene duplication event separated SPO11-1 from SPO11-2. However, given the positions of the SPO11-1 and SPO11-2 clades on a phylogenetic tree in this chapter (**Figures 3.2** and **3.3**), it is unclear whether the first duplication event was the separation of SPO11-3 into SPO11-2 or of SPO11-3 into SPO11-1. The absence of SPO11-1 or SPO11-2 in some organisms is likely to have been caused by lineage-specific gene losses[50].

In eukaryotes, the canonical Spo11 homologues were retained to perform their essential role in meiosis, however lineage-specific gene losses of TOP6B and SPO11-3 have occurred. A recent formation of the eukaryote 'tree of life' by Burki et al in 2020 separates all eukaryotic supergroups into two major clades[189]. The first clade includes the taxonomic supergroups SAR, Telonemia, Haptista, Cryptista, Archaeplastida, Ancoracysta, and Picozoa, and the second clade includes Amorphea, CRuMs, Excavates, Ancyromonadida, and Hemimastigophora. Topo VI seems to have been lost in the second clade and retained in the first clade, except for Telonemia, Ancoracysta, and Picozoa, as well as the SAR groups Rhizaria and Alveolates, which do not possess topo VI. Another exception is the presence of topo VI in the two unicellular holozoans *C. owczarzaki* and *Sphaeroforma arctica*, which are close animal relatives that sit in the Amorphea group in the second eukaryotic clade. This presence of topo VI in these two organisms suggests that the enzyme was lost in the second clade after its divergence from the first clade and that topo VI was lost independently in animals and fungi after the kingdoms diverged.

The MTOPVIB subunit from plants and Apicomplexa possesses similar domain architecture with topo VI-B, shares partial sequence homology with the GHKL domain, and sits close to the eukaryotic TOP6B clade on an unrooted phylogenetic tree (**Figure 3.8**). MTOPVIB therefore seems to have been formed from topo VI-B via a gene duplication event, which preceded

extensive sequence divergence, including in the ATPase fold, and the loss of the H2TH domain. Given that MTOPVIB, TOPVIBL, and other members of the topo VIB-like family possess very limited sequence homology with each other, it is difficult to track the evolutionary history of this subunit. It seems likely that TOPVIBL was also formed via a gene duplication event of a eukaryotic TOP6B subunit, however, it is unclear how the modular members of this family, such as *S*. *cerevisiae* Rec102 and Rec104, were formed.

When Hartung and Puchta identified the SPO11-1 and SPO11-2 homologues in *A*. *thaliana*, they noticed that SPO11-2 possessed a variation of the DxD metal binding motif in which the second aspartate residue was substituted for asparagine[176]. This is unusual, because the DxD motif is highly conserved in all toprim-domain possessing enzymes, including topoisomerases, nucleases, and primases[76], and mutating the second DxD aspartate in *E. coli* topo I has been shown to drastically reduce DNA-binding[78]. The sequence alignments performed in this chapter have confirmed that all members of the SPO11-2 phylogenetic clade possess this alternative DxN motif (**Figures 3.4** and **3.5**), and that the motif is not found outside of this clade. In plants, the meiotic topo VI-like complex is a dissymmetric heterotetramer composed of SPO11-1, SPO11-2, and MTOPVIB[110-112], however, all protists analysed in this chapter, as well as the sponge *A*. *queenslandica*, lack SPO11-1 and so are likely to form a SPO11-2 dimer. Unlike in topoisomerases, a DxD to DxN substitution in topo VI-like complexes is seemingly tolerated and does not impair meiotic DSB formation. It is unclear what the effect of the substitution is on metal binding and on DNA cleavage or whether the substitution is neutral or a specialisation, and it would be useful to test its effect on meiotic DSB formation in organisms that lack SPO11-2, such as yeast or animals.

It was previously reported by Malik et al that Apicomplexa possess SPO11-1 and SPO11-2[50], however, the phylogenetic data presented in this chapter suggests that both paralogues are neither SPO11-1 nor SPO11-2. The Spo11 paralogues in Apicomplexa, unlike those from other eukaryotes, cluster together on a phylogenetic tree (**Figures 3.2** and **3.3**) which suggests that they are separated by a gene duplication event unique to this phylum. Given their lack of a DxN motif and their phylogenetic clustering with metazoan SPO11-1 sequences, the Apicomplexa Spo11 paralogues seem to have been formed via a gene duplication of SPO11-1, and this event is likely to have occurred after the divergence of Apicomplexa from the stramenopiles.

The topo VI complex in plants requires two accessory proteins, RHL1[123] and BIN4[124], which seem to be present in all topo VI-possessing eukaryotes but absent in bacteria, archaea and eukaryotes that lack topo VI. It is therefore likely that RHL1 and BIN4 were not present in the common ancestor of eukaryotes and archaea and emerged early following their divergence. It is also worth proposing that these accessory proteins are required for all eukaryotic topo VI enzymes, although this has not been tested outside of plants. Both RHL1 and BIN4 have been shown to possess functional nuclear localisation signals[124,126] and to exhibit in vitro DNA binding activity[123,124], but it is unclear what essential role they are contributing to the topo VI complex. The loss of topo VI in some eukaryotic lineages seems to have been accompanied by the loss of RHL1 and BIN4, suggesting that these accessory proteins do not have a topo VI-independent role in eukaryotes.

3.3.5 Topo VI is a highly specialised type II topoisomerase

The type II topos are essential for cellular life and are typified by gyrase and topo IV in bacteria, by topo VI in archaea, and by topo II in eukaryotes. At least one of these enzymes is required to remove the positive supercoils that form ahead of the replication and transcription machinery and to decatenate sister chromatids during chromosome segregation. Topo II, topo IV, and topo VI are highly specialised at performing these reactions, however gyrase is primarily involved in the introduction of negative supercoils into the genome. Therefore, although gyrase is essential in bacteria, most bacteria also possess topo IV to help alleviate the topological burden of nucleic acid metabolism. Topo IV is only found in bacteria and topo II is only found in eukaryotes, whereas gyrase and topo VI are widespread across all three domains of life. Gyrase-possessing archaea have been shown to maintain a negatively-supercoiled genome[51,53] and eukaryotic gyrase has been shown to complement *E. coli* strains carrying temperature-sensitive *gyrA* and *gyrB* mutations[193], suggesting that the unique negative supercoiling activity of bacterial gyrase is retained in these domains.

Unlike gyrase, the widespread distribution of topo VI in the three domains is puzzling. Some members of Alphaproteobacteria possess topo VI in addition to topo IV and some eukaryotes have retained topo VI despite already possessing topo II. Furthermore, although topo VI has been transferred from archaea to bacteria, bacterial topo IV has not been transferred to archaea. Therefore, topo VI is a highly specialised type II topo that may be a more efficient enzyme than topo II and topo IV and may possess a unique and yet-be-identified property

that significantly improves the fitness of these organisms. This property may explain the adoption of topo VI for meiotic DSB formation and its role in endoreduplication in plants. Endopolyploidy occurs in eukaryotes that do not possess topo VI[144], such as animals, and so it is not certain that topo VI in protists performs an analogous role to that of plants. The presence of topo VI in bacteria is usually indictive of the absence of topo IV, and so topo VI in this domain is likely to be required for its canonical type II function in DNA replication and transcription and chromosome segregation. However, eukaryotic topo VI represents an apparent type II topo redundancy given the ever-present topo II. It is therefore worth speculating that topo VI in eukaryotes has evolved a unique and highly specialised role, in coordination with RHL1 and BIN4, that does not significantly overlap with the activities of topo II.

Given their near-identical domain architecture, topo II seems to have been formed from a gene fusion event of the topo IV subunits ParC and ParE, which suggests that topo IV was present in the common ancestor of eukaryotes and bacteria, and therefore of archaea. Furthermore, type IIA and type IIB topos have an unrelated domain architecture and are likely to have originated by recruiting the various topo-associated domains independently. Therefore, it seems plausible that topo IV originated before topo VI, was retained in bacteria, lost in archaea, and developed into topo II in eukaryotes. Having diverged from bacteria, the common ancestor of archaea and eukaryotes then formed topo VI, which seemingly led to the loss of topo IV in archaea owing to its superior activity or unique specialisation. However, topo IV is a degenerate version of gyrase and is likely to have arisen via a gene duplication event of gyrase followed by the loss of the gyrA-box[194]. This suggests that gyrase too was present in the common ancestor of bacteria, archaea, and eukaryotes, which is problematic as it indicates that gyrase was lost and then reacquired by archaea and eukaryotes. It thus remains difficult to fully establish the evolutionary history of the type II topos, however the distribution of these enzymes hints that topo VI is a highly specialised member of this family.

3.3.6 Conclusion

Topo VI is a highly specialised type II topo that is widely distributed across the three domains of life. Contrary to previous reports, *Plasmodium*, and other members of the Apicomplexa phylum, do not seem to possess canonical topo VI subunits and instead possess the first protist member of the topo VIB-like family. The unique properties of topo VI have deemed it

a useful enzyme for adoption as the meiotic DSB machinery, and these topo VI-like subunits can be distinguished from their canonical topo VI homologues using structural modelling and phylogenetic analysis. Future work should focus on elucidating the unique properties of topo VI that distinguish the enzyme from topo II and topo IV, determining the effect of the eukaryotic topo VI accessory proteins, and identifying the missing eukaryotic topo VIB-like subunits.

Chapter 4

Topo VI purification and expression trials

4.1 Introduction

4.1.1 General introduction

Acquiring soluble preparations of an enzyme of interest is vital for its complete biochemical study in vitro. It is necessary to produce high quantities of a protein before it can be purified, and this can be achieved by growing and harvesting the cells or tissues of the host organism or by expressing the protein recombinantly. The production of recombinant proteins is often performed in *Escherichia coli* cells, due to their fast growth rate, cost-effectiveness, ease of genetic manipulation, and the extensive collection of molecular tools that have been developed for them. Many eukaryotic proteins require post-translational modifications, such as glycosylation, myristoylation, and palmitoylation, which are often major contributors to a protein's stability, solubility, and activity. These modifications cannot be introduced by *E. coli*, and so many eukaryotic proteins are expressed in alternative organisms such as yeast or insect cells. Although eukaryotic expression systems can improve the stability and solubility of a protein of interest, they are often expensive and produce relatively low yields. Once a protein of interest has been overexpressed, the soluble cell lysate can be subject to fast protein liquid chromatography (FPLC) for an automated, reproducible, and high-resolution purification of proteins. In this chapter, I will briefly discuss how type II topoisomerases (topos) have been successfully expressed in the literature and describe my own attempts to express novel topo VI species for purification and biochemical characterisation.

4.1.2 Expression of type IIA topoisomerases

Active bacterial gyrase[54,195] and topo IV[196] can be isolated by expressing their subunits individually on separate plasmids in *E. coli*. To reconstitute the active heterotetramers of these enzymes, equimolar amounts of their isolated subunits are added together at high concentrations. Eukaryotic topo IIA topos, unlike those from bacteria, form a homodimeric complex and so the expression and purification of just a single subunit is necessary to constitute the active enzyme. Active *Saccharomyces cerevisiae* topo II can be overexpressed in yeast cell culture[197] and active human topo IIα can be expressed recombinantly in *Spodoptera frugiperda* insect cells[198]. Active type IIA topos have also been successfully

isolated from plants and *Plasmodium*, with plant topo II being extracted from *Arabidopsis thaliana* seedlings[199], *A. thaliana* gyrA and gyrB being co-expressed recombinantly in *S. frugiperda*[200], and *Plasmodium berghei* topo II being extracted from *P. berghei*-infected red blood cells in mice[201].

4.1.3 Expression of type IIB topoisomerases

Topo VI from *Saccharolobus shibatae* was first isolated as a heterotetramer directly from *S. shibatae* cell culture[46] but it was later shown that the enzyme could be isolated by expressing its subunits individually on separate plasmids in *E. coli*, as is done for bacterial gyrase and topo IV[185]. However, although a soluble preparation of the *S. shibatae* topo VI B subunit (topo VI-B) could be expressed and purified, the A subunit (topo VI-A) was expressed in inclusion bodies and could only be purified after denaturation and renaturation with a ~15-fold lower yield than topo VI-B[185]. This is unusual, given that a soluble preparation of topo VI-A from the archaeon *Methanocaldococcus jannaschii* could be expressed and purified outside of inclusion bodies[79]. Nonetheless, the soluble preparations of the two *S. shibatae* topo VI subunits were mixed together to reconstitute an active heterotetramer[185]. Given the difficulty in isolating its A subunit, the coexpression of both topo VI-A and topo VI-B from *S. shibatae* was attempted by cloning the respective genes into a dual-expression vector[70]. This greatly enhanced the yields of *S. shibatae* topo VI and simplified its purification procedure[70].

S. shibatae is a hyperthermophile that possesses a topo VI with an optimum activity of ~80°C and this is challenging to maintain in experiments. Therefore, the expression of topo VI from a mesophilic archaeon was attempted to permit its vitro study at more suitable temperatures. Having gained important insights from the *S. shibatae* enzyme, topo VI-A and topo VI-B from the archaeon *Methanosarcina mazei* was expressed in an adapted dual-expression vector that possessed an N-terminal polyhistidine-tag[70]. This enzyme could be purified in high yields and to near homogeneity, and performed optimally at a much more appropriate 37°C[202]. Soluble preparations of dimeric topo VIII from three bacterial species have also been produced by recombinant expression in *E. coli*[54].

4.1.4 Aims and objectives

Here, I will isolate *M. mazei* topo VI (MmTopo VI) and fragments of its ATPase B subunit for biochemical characterisation in Chapter 5. I will also isolate a series of mutants of MmTopo VI to investigate the key questions that linger over its mechanism. Given that the mechanistic

study of topo VI is largely based on the mesophilic archaeal enzyme, I also sought to express topo VI from bacteria and eukaryotes using a variety of expression systems and methods. The biochemical characterisation of novel topo VI species will help us to gain a deeper understanding of the enzyme itself, while also providing useful insights into the meiotic topo VI-like complex and the evolution of type II topoisomerases across the three domains of life.

4.2 Results

4.2.1 Isolation of *Methanosarcina mazei* topo VI

The structural and biochemical characterisation of topo VI has mostly involved the enzyme from *M*. *mazei*, owing to its ease of isolation and practical optimum temperature. In Chapter 5, I will be addressing some of the gaps in our mechanistic understanding of topo VI, and so to do this I sought to isolate the *M*. *mazei* enzyme myself using a protocol developed by Shannon McKie (John Innes Centre, UK)[202]. A previous study has coexpressed the A and B subunits of MmTopo VI (MmTopo VI-A, MmTopo VI-B) in a polycistronic expression vector[58] derived from pST39[166], and I was gifted this vector by James Berger (Johns Hopkins University, USA). To express MmTopo VI, I transformed the vector into a Rosetta 2(DE3)pLysS *E*. *coli* cell line, which has the ability to make additional tRNAs for codons that are rarely used in *E*. *coli*. These cells are useful for the recombinant expression of non-bacterial enzymes, such as those from archaea. After a 24-hr incubation in auto-induction media (AIM)[168] at 37°C, the expected expression bands of ~42 kDa and ~69 kDa for MmTopo VI-A and MmTopo VI-B, respectively, were observed by SDS-PAGE (**Figure 4.1**) confirming that the enzyme was successfully expressed.

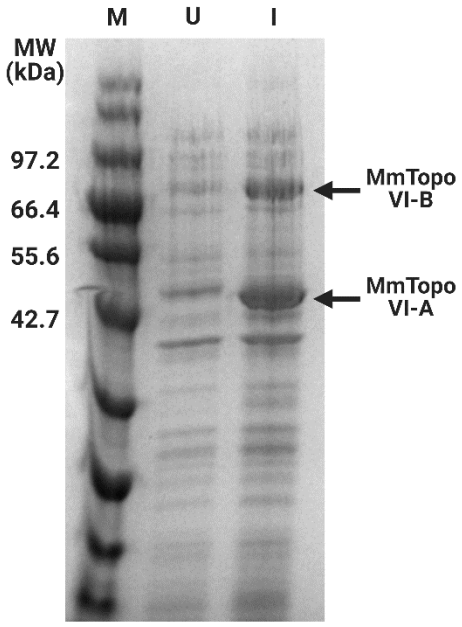

Figure 4.1| *Methanosarcina mazei* topo VI expression: ***M. mazei*** *topo VI-A and topo VI-B were coexpressed in a polycistronic polyhistidine-tagged vector in* ***E. coli****. Samples were taken from uninduced (**U**) cells after 2 hr and from induced (**I**) cells after 24 hr in auto-induction media and were analysed by SDS-PAGE alongside a protein marker (**M**). The expected bands of ~42 kDa and ~69 kDa for MmTopo VI-A and MmTopo VI-B, respectively, were visible in the **I** sample.*

To isolate the polyhistidine-tagged MmTopo VI enzyme, I first obtained the soluble cell lysate by centrifugation. I then performed immobilized metal ion affinity chromatography using a HisTrap FF column, which uses packed nickel ions to bind the polyhistidine residues. Next, I pooled MmTopo VI-possessing fractions eluted from the FF column (**Figure 4.2**) and performed a second purification step by anion-exchange chromatography at pH 7.5 using a HiTrap Q HP column. MmTopo VI has a theoretical isoelectric point of 5.55 that implies a net negative surface charge at pH 7.5, therefore the enzyme will bind to the positively-charged quaternary ammonium anion exchange resin. I then pooled MmTopo VI-possessing fractions from the Q column (**Figure 4.3)** and incubated them with TEV protease to remove the polyhistidine tags. This sample was then passed back through the FF column once more to enrich for polyhistidine-cleaved MmTopo VI, before being subjected to a final purification step by size-exclusion chromatography using a Superdex 200 Increase 10/300 GL. Lastly, I concentrated individual MmTopo VI-possessing fractions from the Superdex column (**Figure 4.4**) to ~500 µg mL^{-1} and stored them at -80°C. The isolated MmTopo VI enzyme was purified to near homogeneity and will be suitable for background-sensitive biochemical assays.

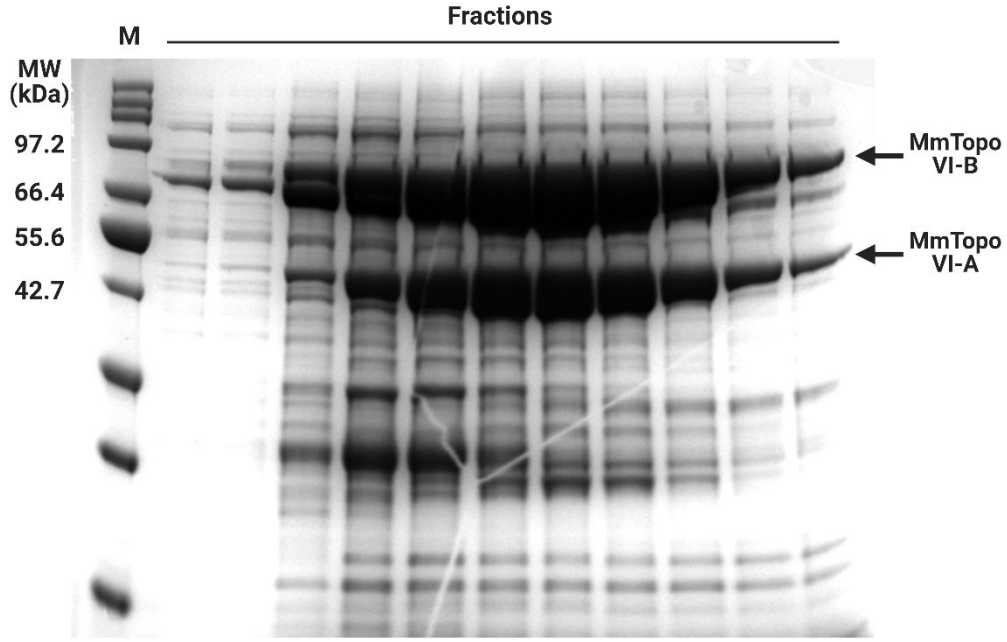

Figure 4.2| Purification of *Methanosarcina mazei* topo VI by immobilized metal ion affinity chromatography: *Soluble cell lysate possessing polyhistidine-tagged **M. mazei** topo VI was passed through a HisTrap FF column at pH 7.5 Fractions were analysed by SDS-PAGE alongside a protein marker (**M**).*

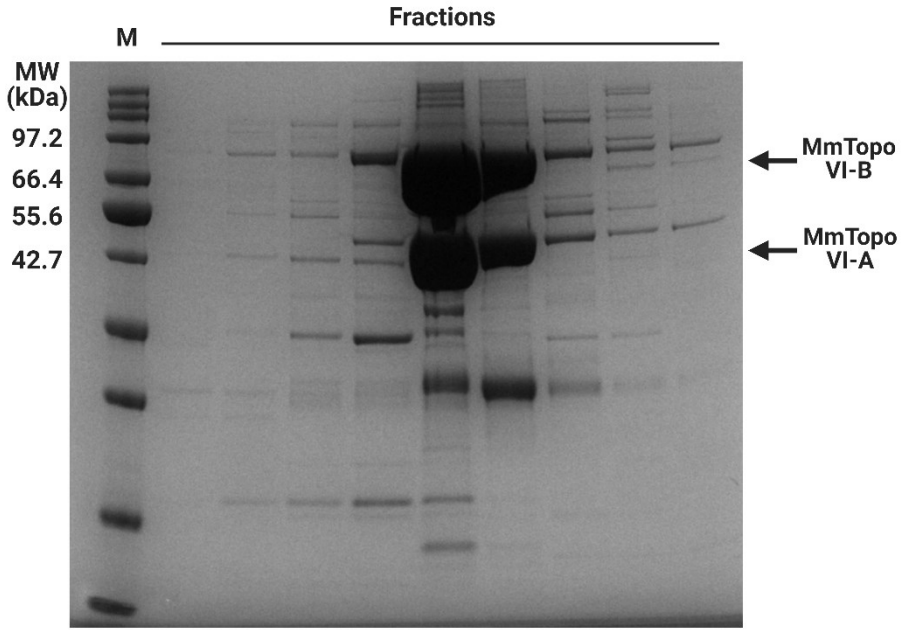

Figure 4.3| Purification of *Methanosarcina mazei* topo VI by anion-exchange chromatography: *Pooled fractions from the HisTrap FF column possessing **M. mazei** topo VI were passed through a HiTrap Q HP column at pH 7.5. Fractions were analysed by SDS-PAGE alongside a protein marker (**M**).*

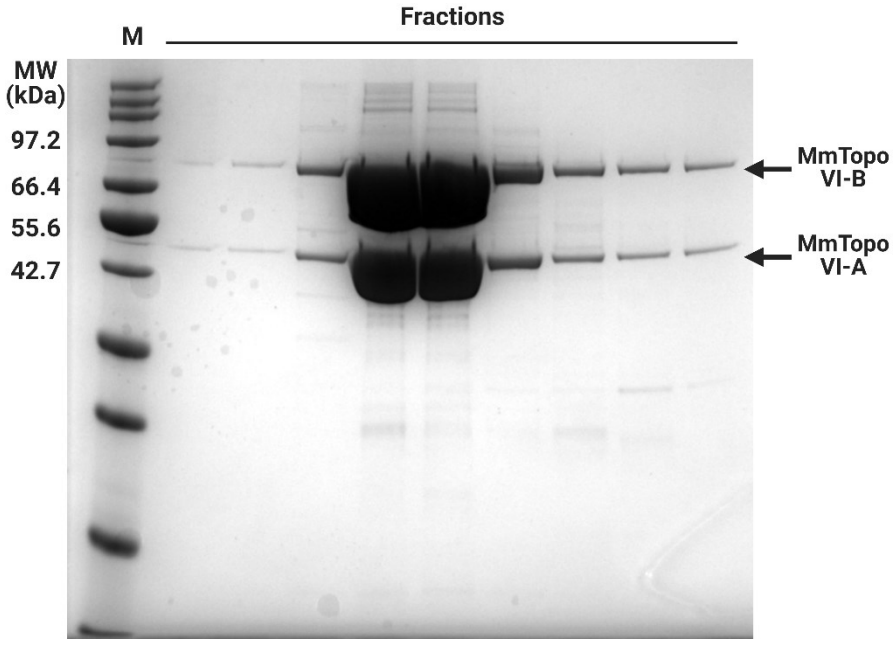

Figure 4.4| Purification of *Methanosarcina mazei* topo VI by size-exclusion chromatography: *Polyhistidine-cleaved **M. mazei** topo VI was passed through a Superdex 200 10/300 GL column at pH 7.5. Fractions were analysed by SDS-PAGE alongside a protein marker (**M**).*

4.2.2 Isolation of *Methanosarcina mazei* topo VI mutants

Site-directed mutagenesis is a powerful tool that enables us to probe the key residues that influence the structural and functional properties of an enzyme of interest. In Chapter 5, I will discuss the design of a series of *M. mazei* topo VI mutants that will permit me to investigate the enzyme's mechanistic uncertainties. Here, I will detail the isolation of these mutants as a prelude to their biochemical characterisation in Chapter 5. The first series of mutants were identified from the phylogenetic analyses that were conducted for Chapter 3, namely a glycine to alanine mutation in the G3-box region (G173A) in MmTopo VI-B, and an aspartate to glutamate mutation at residue 294 (D294E) in MmTopo VI-A (**Figure 4.5**). The G173 residue confers flexibility to a disordered loop that possesses key residues that contribute to the structural integrity of the ATP-binding site[59]. This residue is the only conserved GHKL residue characterised in Chapter 3 that is absent in the *Plasmodium falciparum* putative topo VI-B and will be investigated for its significance in ATP utility. The D294 residue is the only conserved residue that is present in all Spo11 and topo VI-A homologues except for Apicomplexa SPO11-A, and the significance of this residue will be investigated using *M. mazei* topo VI. I isolated both the G173A and D294E mutants to near homogeneity after a final gel-filtration step (**Figures 4.6** and **4.7**) with the same method used for the isolation of the wild-type (WT) MmTopo VI enzyme. However, the yields of both mutants were lower than those obtained for the WT enzyme following expression in the same volume of cell culture, with the D294E mutant in particular being obtained in a significantly lower yield. This may reflect the key contributions these residues make to maintaining protein stability.

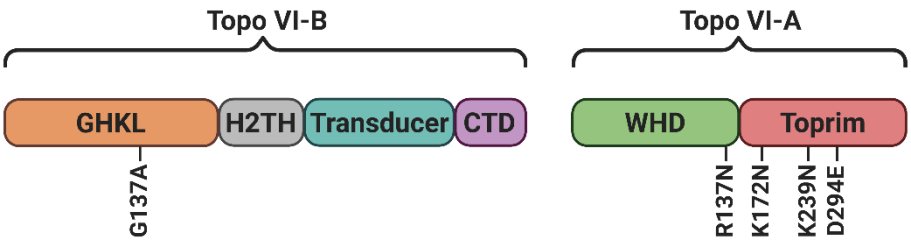

Figure 4.5| *Methanosarcina mazei* topo VI mutation positions: *A series of **M. mazei** topo VI mutants have been isolated in this chapter, namely the topo VI-B mutant G137, and the topo VI-A mutants R137N, K172N, K239N, and D294E.*

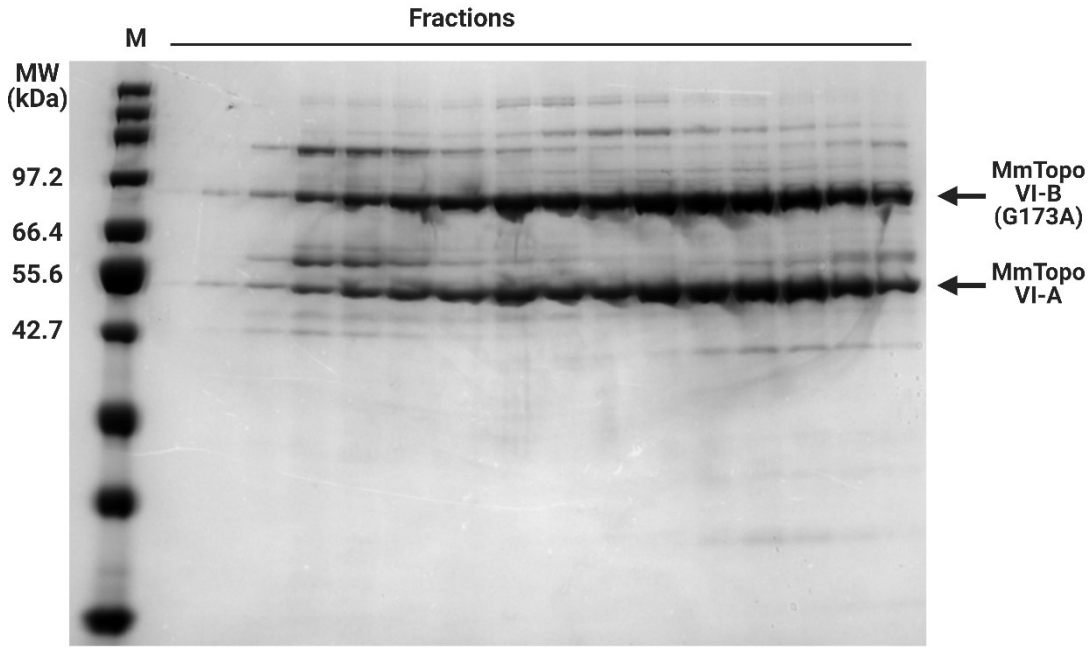

Figure 4.6| Purified *Methanosarcina mazei* topo VI G137A mutant: *The final purification step of the **M. mazei** topo VI G137A mutant. Fractions were obtained from a Superdex 200 10/300 GL column at pH 7.5 and analysed by SDS-PAGE alongside a protein marker (**M**).*

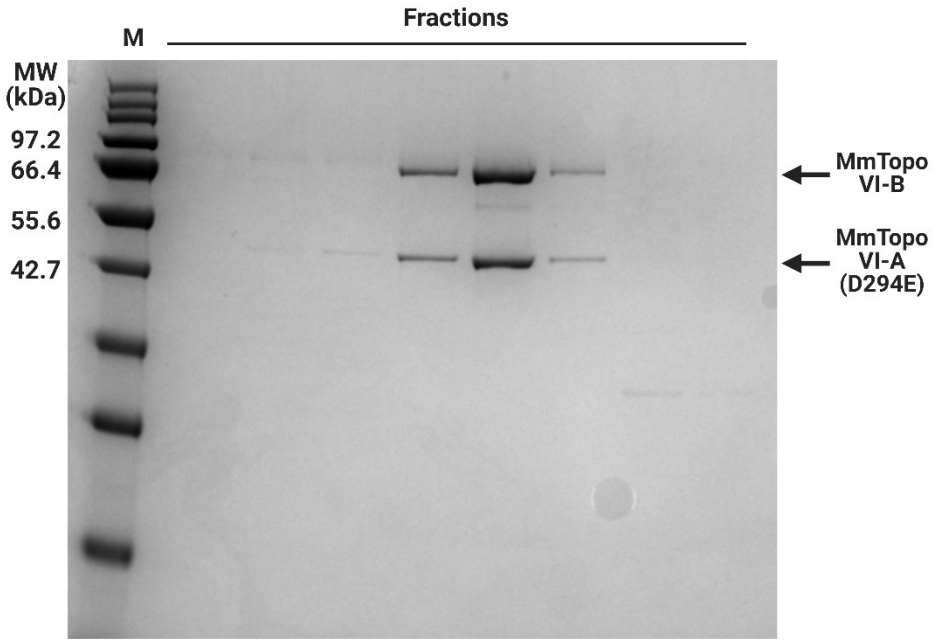

Figure 4.7| Purified *Methanosarcina mazei* topo VI D294E mutant: *The final purification step of the **M. mazei** topo VI D294E mutant. Fractions were obtained from a Superdex 200 10/300 GL column at pH 7.5 and analysed by SDS-PAGE alongside a protein marker (**M**).*

The second series of mutants I designed were substitutions of arginine and lysine residues in MmTopo VI-A to asparagine, namely R137N, K172N, R235N, K239N, R240N, and K286N. These positively-charged residues are located at the entrance to the DNA-gate and will be investigated for their role in influencing the direction of strand-passage in *M*. *mazei* topo VI in Chapter 5. The isolation of these mutants, including mutagenesis, expression, and purification, was carried out under my supervision by Ella Taylor-Cross (University of Bristol, UK), a summer student in our lab. Although all six mutants expressed well under the same conditions used for the WT enzyme, R235N, R240N, and K286N were insoluble. However, the remaining mutants, R137N, K172N, and K239N (**Figure 4.5**), were purified in high yields to near homogeneity after a final gel-filtration step (**Figures 4.8**-**4.10**).

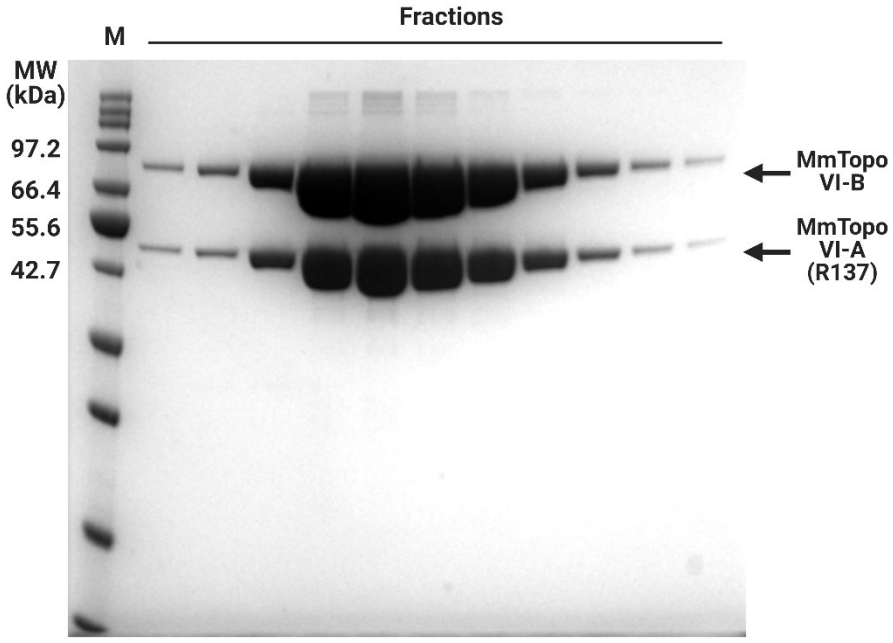

Figure 4.8| Purified *Methanosarcina mazei* topo VI R137N mutant: *The final purification step of the **M. mazei** topo VI R137N mutant. Fractions were obtained from a Superdex 200 10/300 GL column at pH 7.5 and analysed by SDS-PAGE alongside a protein marker (**M**).*

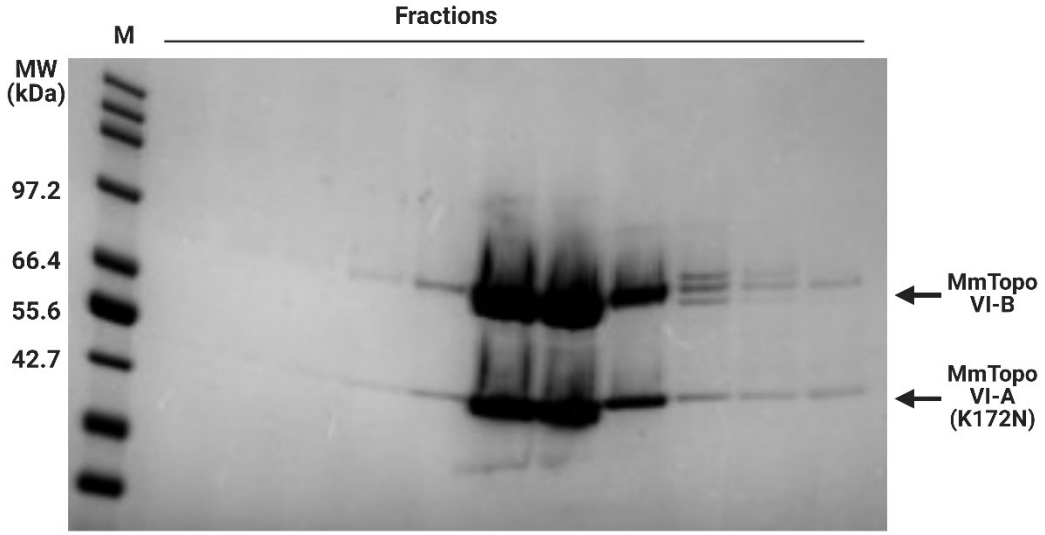

Figure 4.9| Purified *Methanosarcina mazei* topo VI K172N mutant: *The final purification step of the **M. mazei** topo VI K172N mutant. Fractions were obtained from a Superdex 200 10/300 GL column at pH 7.5 and analysed by SDS-PAGE alongside a protein marker (**M**).*

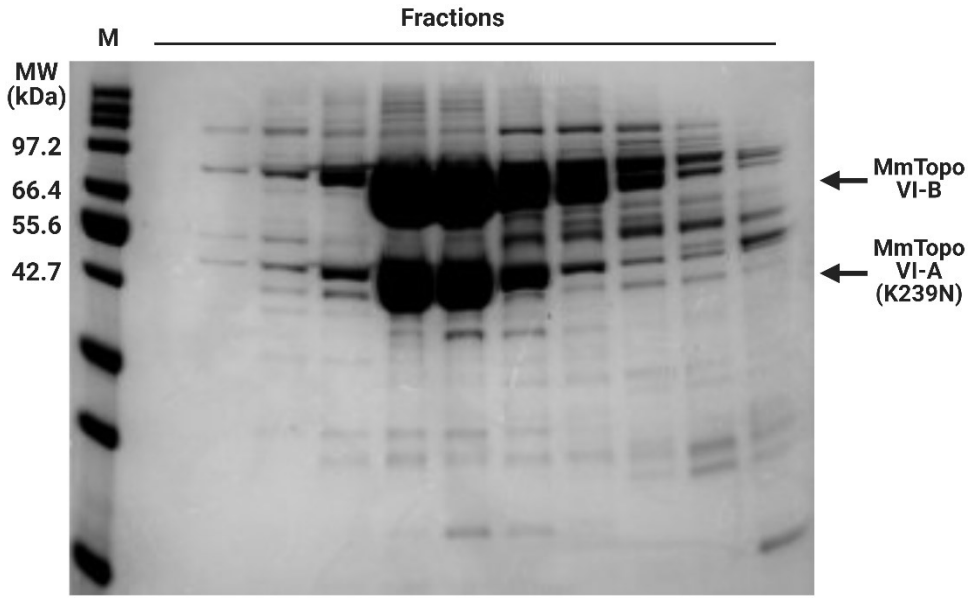

Figure 4.10| Purified *Methanosarcina mazei* topo VI K239N mutant: *The final purification step of the **M. mazei** topo VI K239N mutant. Fractions were obtained from a Superdex 200 10/300 GL column at pH 7.5 and analysed by SDS-PAGE alongside a protein marker (**M**).*

4.2.3 Isolation of *Methanosarcina mazei* topo VI-B truncates

A previous study has isolated a fragment of *S. shibatae* topo VI-B to investigate the mechanism of ATP turnover in this enzyme by X-ray crystallography[67]. In chapter 5, I will discuss how I conducted a similar approach with fragments of MmTopo VI-B to investigate the mechanism of the *M. mazei* enzyme and I will now detail the isolation of these fragments as a prelude. I cloned two constructs containing residues 1-233 and 1-488 of MmTopo VI-B into the pET28-MHL *E. coli* expression vector and transformed the vectors into Rosetta 2(DE3)pLysS cells. Construct 1-233 corresponds to the GHKL domain in isolation (MmTopo VIB-GHKL) and construct 1-488 corresponds to a truncated form of MmTopo VI-B that ends at the terminus of its long transducing α-helix (MmTopo VIB-trans) (**Figure 4.11**). These constructs were designed using the crystal structure of MmTopo VI (PDB code: 2Q2E)[70]. After a 24-hr incubation in AIM at 37°C, an expression band corresponding to MmTopo VIB-GHKL ran between 32 kDa and 46 kDa on an SDS-PAGE gel (**Figure 4.12**), even though the fragment has a molecular weight (MW) of ~26 kDa. The expression band of MmTopo VIB-trans, however, ran at ~54 kDa as expected. Therefore, both fragments were successfully expressed, but the MW of MmTopo VIB-GHKL should be validated.

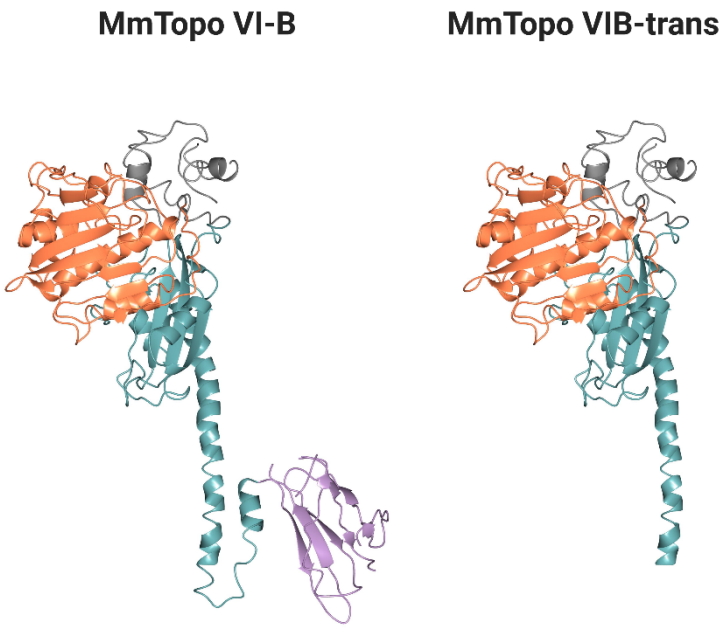

Figure 4.11| *Methanosarcina mazei* topo VI-B trans truncate: *A truncate of **M. mazei** topo VI-B was constructed comprising residues 1-488 (MmTopo VIB-trans). MmTopo VIB-trans is missing the C-terminal domain (lilac) and comprises the complete GHKL domain (orange) and a partial transducer domain (blue) that ends at the terminus of the long transducing α-helix.*

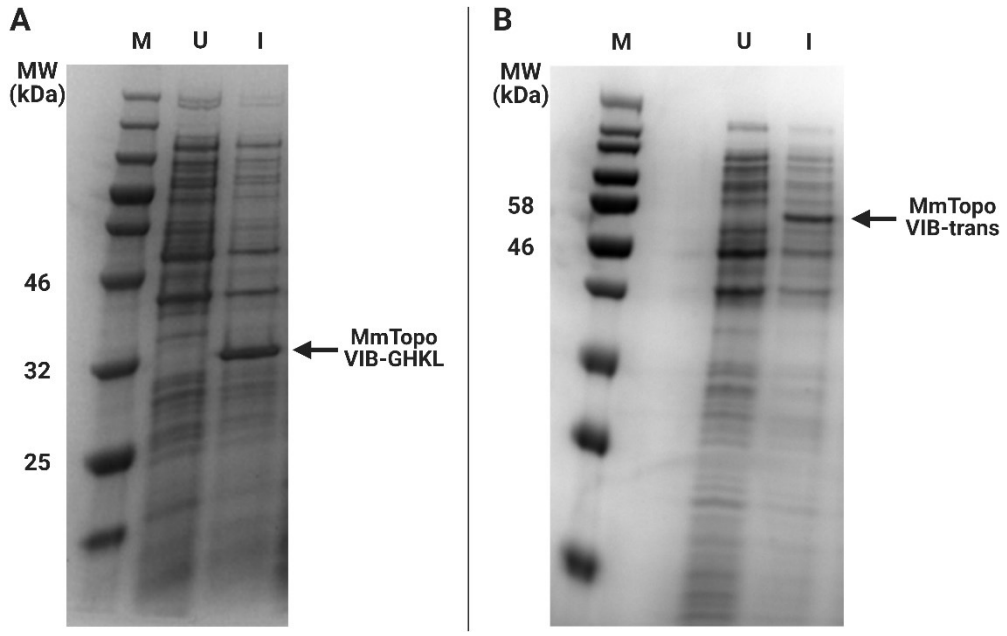

Figure 4.12| *Methanosarcina mazei* topo VI truncate expression: *Truncates of **M. mazei** topo VI-B were expressed in **E. coli**. Samples were taken from uninduced (**U**) cells after 2 hr and from induced (**I**) cells after 24 hr in auto-induction media and were analysed by SDS-PAGE alongside a protein marker (**M**). The expected bands of ~26 kDa and ~54 kDa for MmTopo VIB-GHKL (**A**) and MmTopo VIB-trans (**B**), respectively, were visible in the **I** samples.*

To isolate the polyhistidine-tagged MmTopo VIB-GHKL fragment, I began to perform the same procedure for the purification of the full-length enzyme and subjected the soluble cell lysate to immobilized metal ion affinity chromatography. However, although MmTopo VIB-GHKL has a theoretical isoelectric point of 5.26, the fragment did not bind to a HiTrap Q HP column at pH 7.5. Therefore, I pooled MmTopo VIB-GHKL-possessing fractions from the FF column (**Figure 4.13**) and performed an alternative purification step by affinity chromatography at pH 7.5 using a HiTrap Heparin HP column, which is designed for the purification of DNA-binding proteins. Next, I pooled MmTopo VIB-GHKL-possessing fractions from the heparin column (**Figure 4.14**) and incubated them with TEV protease to remove the polyhistidine tags. This sample was then passed back through the FF column once more to enrich for polyhistidine-cleaved MmTopo VIB-GHKL, before being subjected to a final purification step by size-exclusion chromatography using a Superdex 75 Increase 10/300 GL column. This gel-filtration column is more suitable for the separation of small proteins than the column used for MmTopo VI. Lastly, I concentrated the MmTopo VIB-GHKL-possessing fraction from the Superdex column (**Figure 4.15**) to ~500 µg mL^{-1} and stored it at -80°C. The isolated MmTopo

VIB-GHKL enzyme was purified to near homogeneity and mass spectrometry confirmed that the isolated fragment had the expected MW of ~26 kDa (**Figure 4.16**). The same purification procedure used for MmTopo VI-GHKL was used to isolate MmTopo VIB-trans, yielding highly concentrated fractions after a final gel-filtration step (**Figure 4.17**). The preparation of MmTopo VIB-trans appeared to possess multiple protein contamination, although this is likely to be due to overloading the gel.

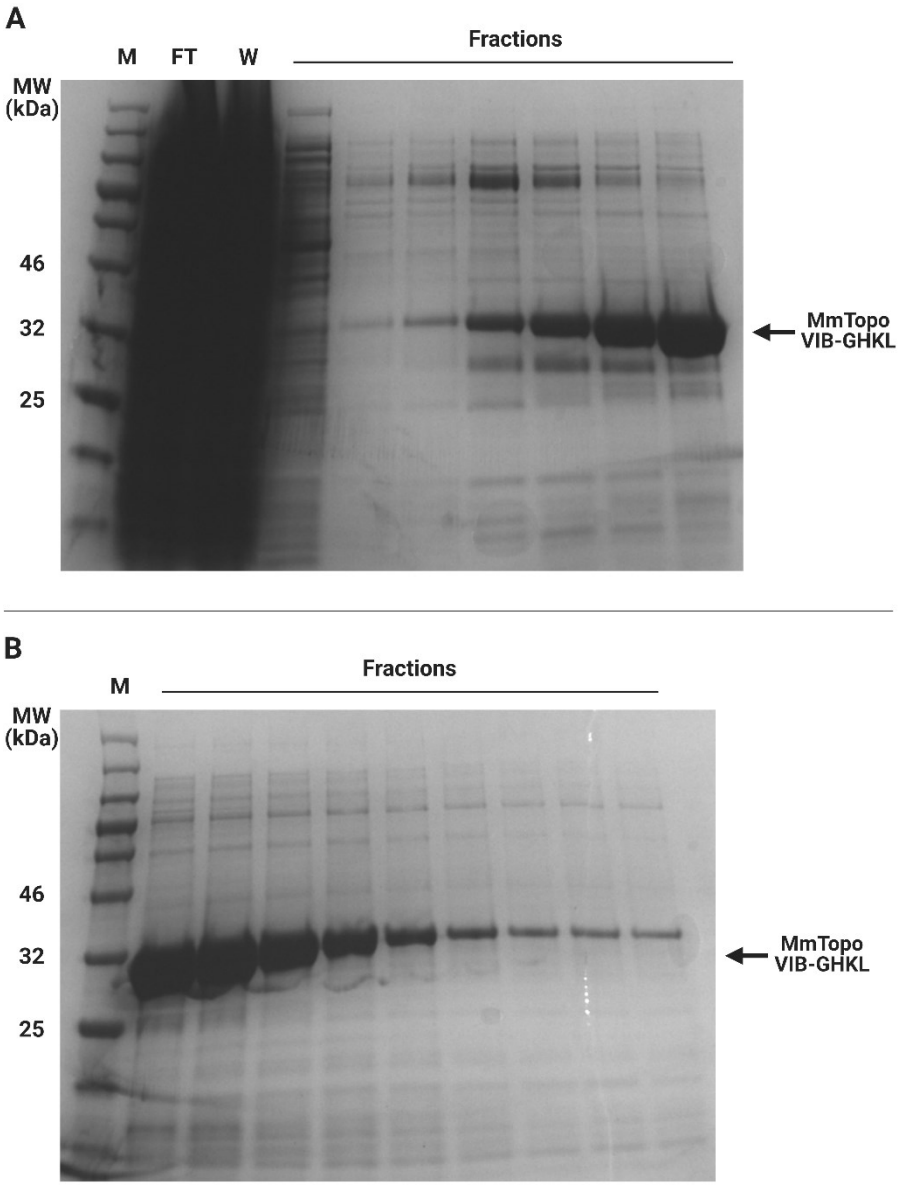

Figure 4.13| Purification of *Methanosarcina mazei* topo VIB-GHKL by immobilized metal ion affinity chromatography: *Soluble cell lysate possessing polyhistidine-tagged **M. mazei** topo VIB-GHKL was passed through a HisTrap FF column at pH 7.5 The column flow-through (**FT**), wash (**W**) and fractions were analysed by SDS-PAGE alongside a protein marker (**M**). Fractions 1-7 (**A**) and fractions 8-16 (**B**) were run on separate gels.*

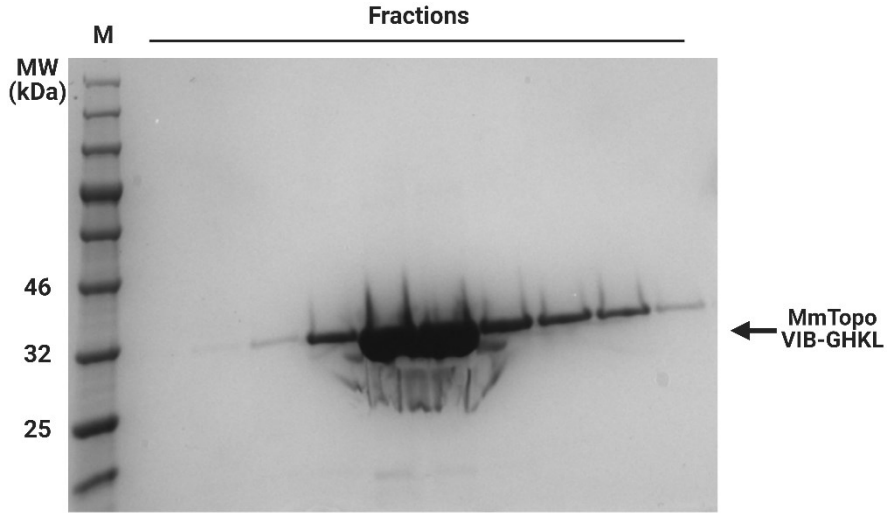

Figure 4.14| Purification of *Methanosarcina mazei* topo VIB-GHKL by affinity chromatography: *Pooled fractions from the HisTrap FF column possessing **M. mazei** topo VIB-GHKL were passed through a HiTrap Heparin HP column at pH 7.5. Fractions were analysed by SDS-PAGE alongside a protein marker (**M**).*

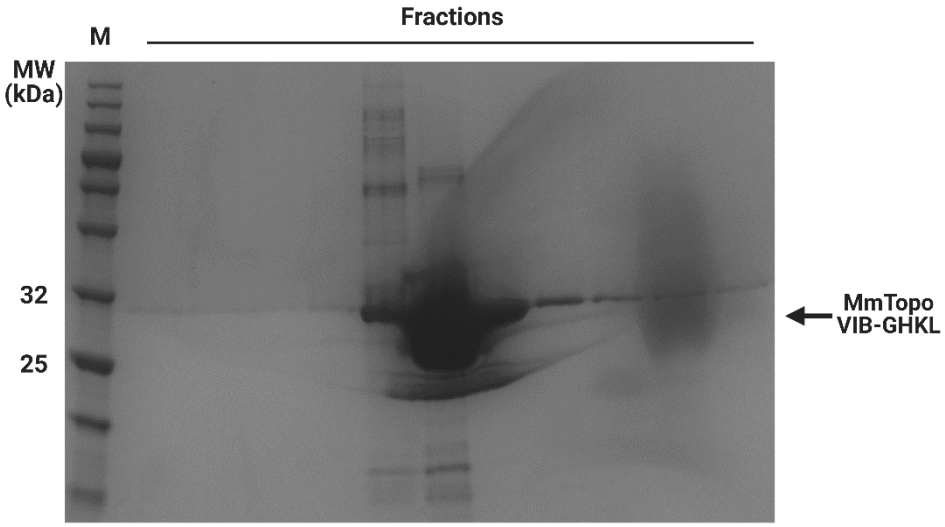

Figure 4.15| Purification of *Methanosarcina mazei* topo VIB-GHKL by size-exclusion chromatography: *Polyhistidine-cleaved **M. mazei** topo VIB-GHKL was passed through a Superdex 75 Increase 10/300 GL column at pH 7.5. Fractions were analysed by SDS-PAGE alongside a protein marker (**M**).*

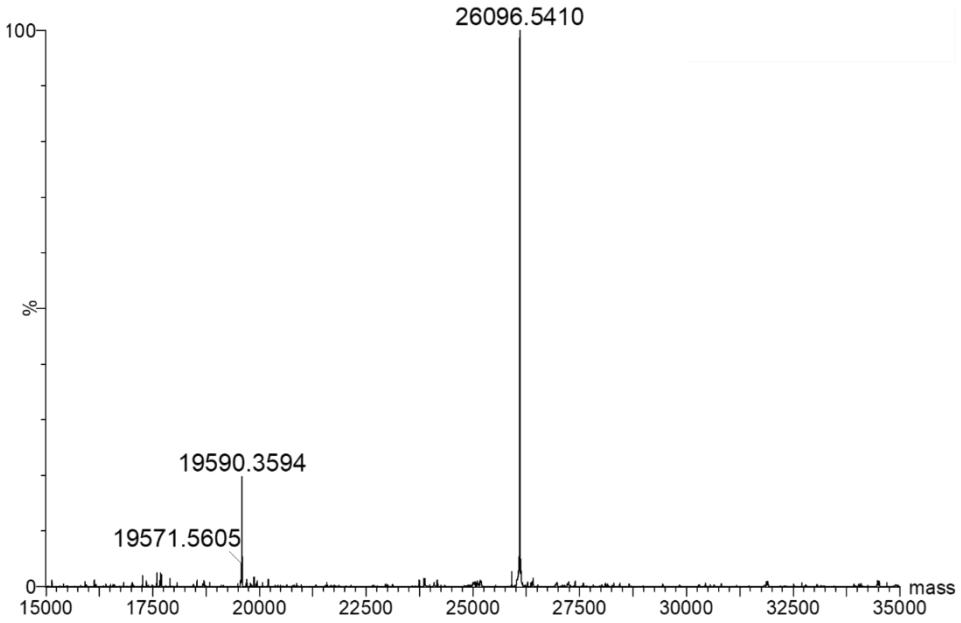

Figure 4.16| Validation of the *Methanosarcina mazei* topo VIB-GHKL molecular weight by mass spectrometry: *The molecular weight of the isolated **M. mazei** topo VIB-GHKL fragment was calculated by positive electrospray ionisation time-of-flight mass spectrometry, yielding the expected value of 26 kDa.*

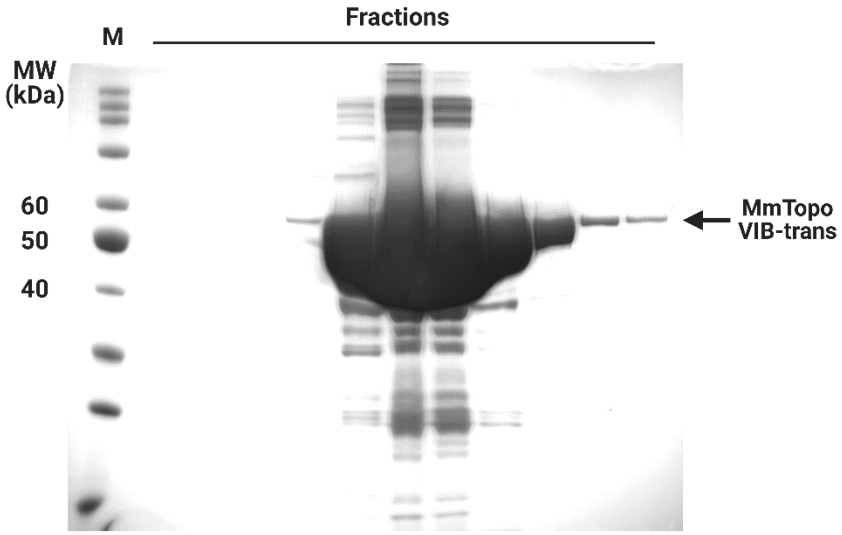

Figure 4.17| Purification of *Methanosarcina mazei* topo VIB-trans by size-exclusion chromatography: *Polyhistidine-cleaved **M. mazei** topo VIB-trans was passed through a Superdex 75 Increase 10/300 GL column at pH 7.5. Fractions were analysed by SDS-PAGE alongside a protein marker (**M**).*

4.2.4 Bacterial topo VI expression trials

I have shown in Chapter 3 that topo VI is widespread in the bacterial domain, and this seems to have occurred via horizontal gene transfer from archaea[27]. I have also shown that all the key invariant residues of functional topo VI complexes in archaea and plants are conserved in bacterial topo VI, and that the predicted structures of bacterial topo VI strongly resemble the archaeal enzymes. If topo VI was not actively expressed in these bacteria, then you would expect to see natural mutations at these conserved sites that would offer no change to the fitness of the bacteria. Given the absolute conservation of these key functional residues, it is reasonable to suggest that these bacteria do indeed possess an active topo VI. To elucidate the role of bacterial topo VI, it is necessary to isolate the enzyme from a species within this domain for in vitro characterisation. It would also be useful to investigate topo VI outside the context of archaea, to identify any key differences in the biochemical properties of topo VI between the two domains. Therefore, I sought to identify three suitable bacterial species from which to pull novel topo VI genes for recombinant expression in *E. coli*.

In Chapter 3, I showed that many of the topo VI-possessing bacteria lacked complete taxonomic categorisation and only eleven of these bacterial species that I identified possessed a given genus and species name. I therefore searched within this group for bacterial species with comparable natural environments to *E. coli* that would favour the expression of the enzyme in this system and possess suitable optimum temperatures for easy biochemical handling. I chose three gram-negative bacterial species from three different taxonomic supergroups that were deemed to have suitable growth conditions for enzyme isolation and characterisation in the lab (**Table 4.1**). Two of these bacteria, namely *Nitrospira moscoviensis*[203] and *Sedimentisphaera cyanobacteriorum*[204] have been cultured in vitro to assess their optimum growth conditions, however, in vitro cultivation of the third species, *Pajaroellobacter abortibovis*, has yet to be achieved[205]. *P. abortibovis* can be found in the tissues of infected bovine foetuses and so is likely to have optimum growth at ~38°C and a pH of ~6.9-7.4 to reflect the conditions of its host[206].

Table 4.1| Optimum growth conditions of three bacterial species suitable for topo VI expression

Bacterial species	Taxonomic supergroup	Optimum temperature (°C)	Optimum Ph
Nitrospira moscoviensis	Nitrospirae	39	7.6 - 8.0
Pajaroellobacter abortibovis	Proteobacteria	38	6.9 - 7.4
Sedimentisphaera cyanobacteriorum	PVC	35	7.5

The expression of archaeal topo VI has been most successful when coexpressing the two subunits in a dual-expression vector with the top6b gene preceding the top6A gene downstream of the promoter region[70]. Therefore, I cloned the topo VI genes of *N*. *moscoviensis*, *S*. *cyanobacteriorum*, and *P*. *abortibovis* in the B-A orientation in the first two multiple cloning sites of the polycistronic expression vector pST39[166]. I then replicated the same expression conditions that have been successful for the expression of *M*. *mazei* topo VI, by transforming the bacterial topo VI expression vectors into Rosetta 2(DE3)pLysS *E*. *coli* cells and incubating them for 24 hr in AIM at 37°C. Decreasing the expression temperature often enhances the solubility of recombinant proteins expressed in *E*. *coli*, however higher temperatures are not always detrimental[207], and so I also tested expression at 30°C and 20°C. After a 24-hr and a 48-hr incubation in AIM at these three temperatures, no expression of topo VI-A or topo VI-B was observed for the three bacterial species (**Figure 4.18**). Given that the bacterial topo VI genes were already codon optimised for *E*. *coli*, I also transformed the expression vectors into BL21(DE3)pLysS *E*. *coli* cells to remove the metabolic burden of the redundant tRNAs. After replicating the expression conditions used for the Rosetta cell line with the BL21 cells, no expression of topo VI-A or topo VI-B was observed for the three bacterial species (**Figure 4.19**).

Given the lack of expression in AIM, I tested an alternative method of inducing protein expression by using Isopropyl β-D-1-thiogalactopyranoside (IPTG) in the nutrient-rich 2× yeast extract-tryptone (2YT) growth medium. However, no expression of topo VI-A or topo VI-B was observed for the three bacterial species following IPTG induction at 20°C for 16 hr and 37°C for 4 hr for the Rosetta and BL21 cells (**Figure 4.20**). Overall, there was no success in expressing topo VI from *N*. *moscoviensis*, *S*. *cyanobacteriorum*, or *P*. *abortibovis*.

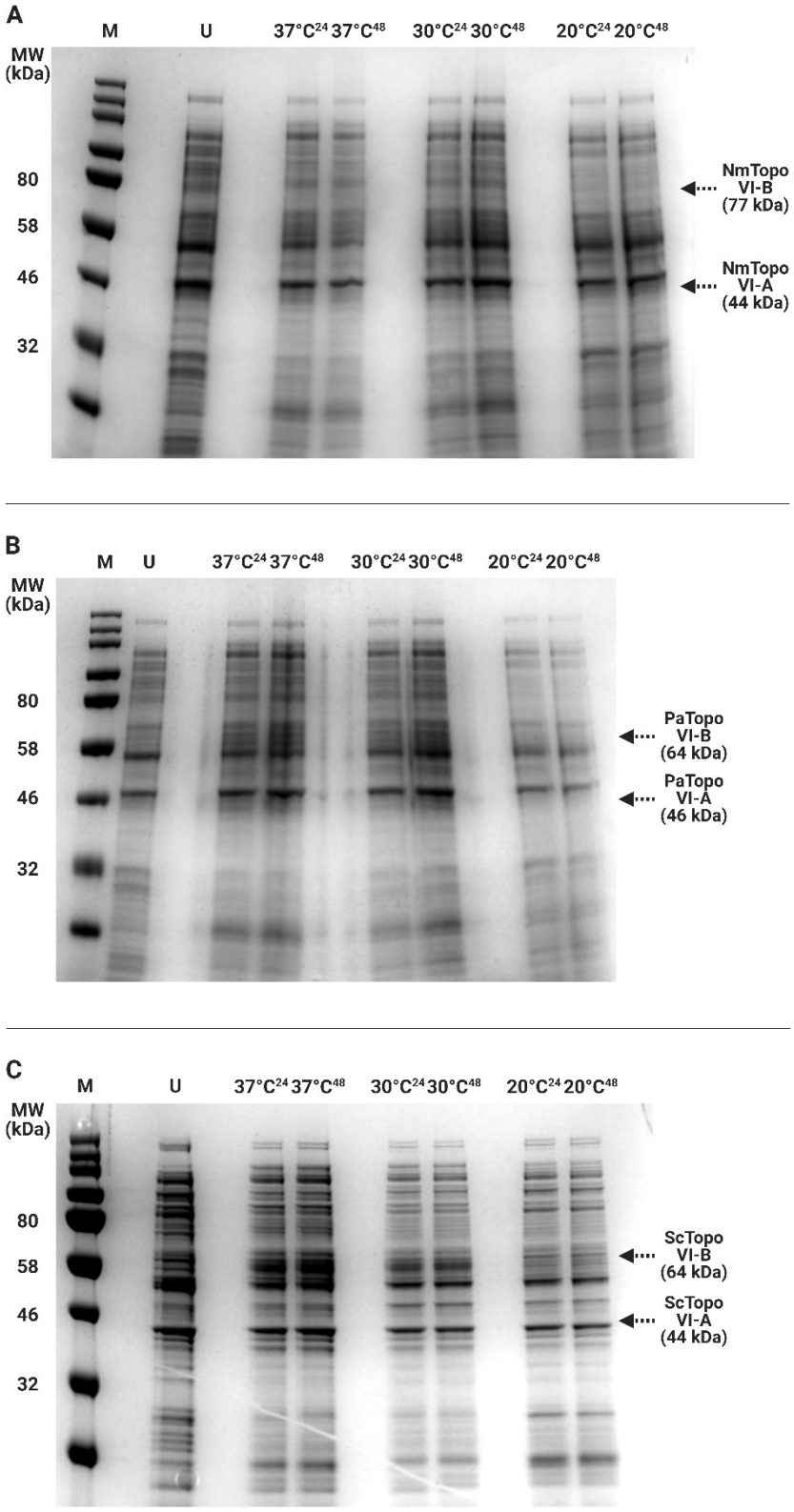

Figure 4.18| Bacterial topo VI expression trials in AIM using Rosetta cells: *The expression of topo VI from **N. moscoviensis** (**A**), **P. abortibovis** (**B**), and **S. cyanobacteriorum** (**C**) was tested in AIM in Rosetta 2(DE3)pLysS cells. Samples taken from uninduced cells (**U**) and cells induced for 24 hr and 48 hr at 37°C, 30°C, and 20°C were analysed by SDS-PAGE alongside a protein marker (**M**). Dotted arrows indicated the predicted molecular weights of topo VI-A and topo VI-B.*

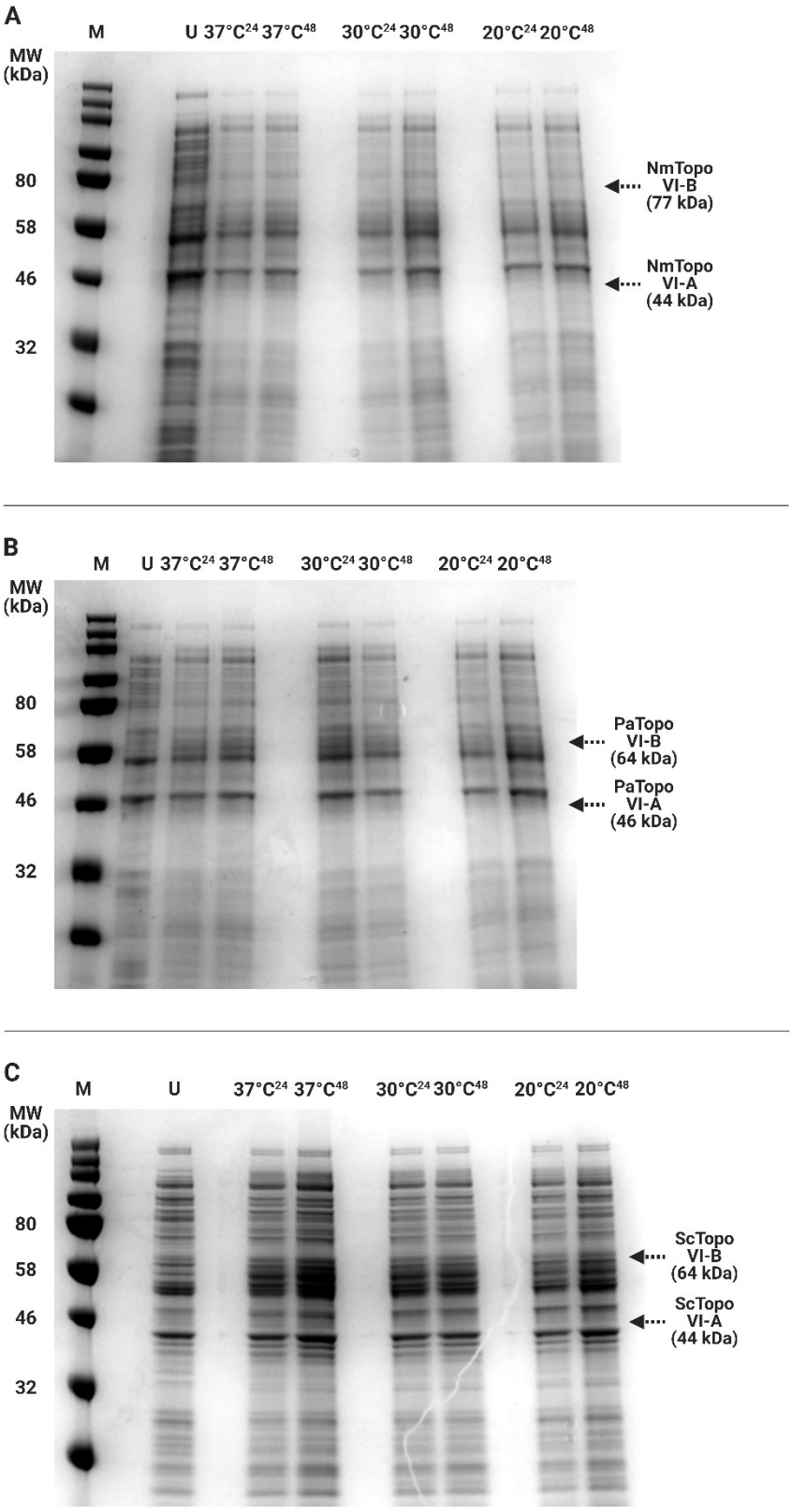

Figure 4.19| Bacterial topo VI expression trials in AIM using BL21 cells: *The expression of topo VI from **N. moscoviensis** (**A**), **P. abortibovis** (**B**), and **S. cyanobacteriorum** (**C**) was tested in AIM in BL21(DE3)pLysS cells. Samples taken from uninduced cells (**U**) and cells induced for 24 hr and 48 hr at 37°C, 30°C, and 20°C were analysed by SDS-PAGE alongside a protein marker (**M**). Dotted arrows indicated the predicted molecular weights of topo VI-A and topo VI-B.*

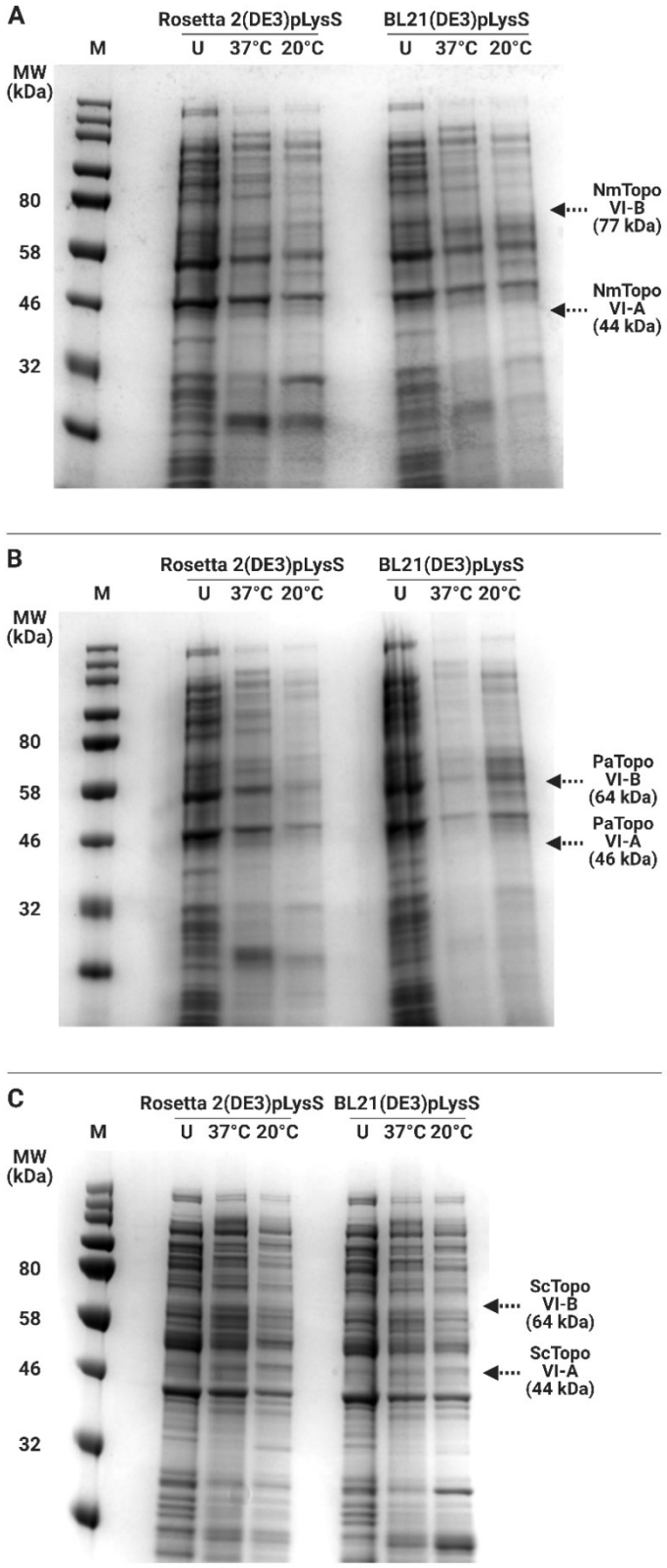

Figure 4.20| Bacterial topo VI expression trials using IPTG induction: *The expression of topo VI from **N**. **moscoviensis** (**A**), **P**. **abortibovis** (**B**), and **S**. **cyanobacteriorum** (**C**) was tested by inducing Rosetta 2(DE3)pLysS cells and BL21(DE3)pLysS cells with IPTG. Samples taken from uninduced cells (**U**) and from cells induced at 37°C for 4 hr and 20°C for 16 hr were analysed by SDS-PAGE alongside a protein marker (**M**). Dotted arrows indicated the predicted molecular weights of topo VI-A and topo VI-B.*

4.2.5 Eukaryotic topo VI expression trials

It is uncertain whether topo VI is actively expressed in bacteria and so it is not concerning that my attempts to express topo VI from this domain have been in vain. Eukaryotic topo VI, however, has been extensively studied and it is highly likely that organisms in this domain do express an active topo VI. Topo VI in plants in particular has been heavily investigated, and here the enzyme seems to function during the process of endoreduplicaiton[115-117] and is implicated in the transcriptional regulation of stress-response genes[117,119,120]. Plant topo VI has been shown to require two accessory proteins, RHL1[123] and BIN4[124], in vivo, and in Chapter 3 I discovered that these proteins are conserved in topo VI-possessing eukaryotes. Given its unusual function and properties, the in vitro characterisation of an isolated plant topo VI complex is an attractive prospect.

Acquiring soluble preparations of a eukaryotic topo VI has proven difficult to achieve, however, Monica Agarwal (John Innes Centre, UK) in the Maxwell research group (personal communication) has attained modest yields of *A*. *thaliana* topo VI (AtTopo VI). Monica succeeded in expressing soluble AtTopo VI in baculovirus-infected *S*. *frugiperda* cells and this expression could only be achieved by coexpressing the two topo VI subunits, AtTOP6B and AtSPO11-3, alongside their two accessory proteins, AtRHL1 and AtBIN4. However, the yield of AtTopo VI was very low and the subunits could only be observed by Western blot analysis. Therefore, only a single-step purification using a HisTrap FF column was suitable and so the enzyme preparation was heavily contaminated. Interestingly, although they seem to be required for the expression of AtTOP6B and AtSPO11-3, AtRHL1 and AtBIN4 could not be seen on a Western Blot following their coexpression, suggesting that they may not be constituents of the catalytic topo VI complex in eukaryotes. To improve the isolation of AtTopo VI, I sought to express the enzyme in *E*. *coli*. This system may vastly improve the yield of AtTopo VI but may also hinder the stability of the enzyme due to the absence of eukaryotic post-translational modifications.

Given that expression of AtTopo VI in the baculovirus-insect cell system required the coexpression of the topo VI subunits and its two accessory proteins, I cloned the *A*. *thaliana* genes in a *TOP6B-TOP6A-RHL1-BIN4* orientation in the four multiple cloning sites of the pST39[166] expression vector. I then tested the expression of these proteins by using the same conditions used for bacterial topo VI. The *A*. *thaliana* topo VI genes were codon optimised for

E. coli and so the expression vector was transformed into both Rosetta 2(DE3)pLysS and BL21(DE3)pLysS cells. After a 24-hr and a 48-hr incubation in AIM at 37°C, 30°C, and 20°C, no expression of AtTOP6B, AtSPO11-3, AtRHL1, or AtBIN4 was observed for the Rosetta or BL21 cells (**Figure 4.21**). Furthermore, no expression of these four subunits was observed following IPTG induction in 2YT media at 20°C for 18hr and 37°C for 4 hr for both cell lines tested (**Figure 4.21**). Overall, there was no success in expressing the topo VI complex and accessory proteins from *A. thaliana* in *E. coli*.

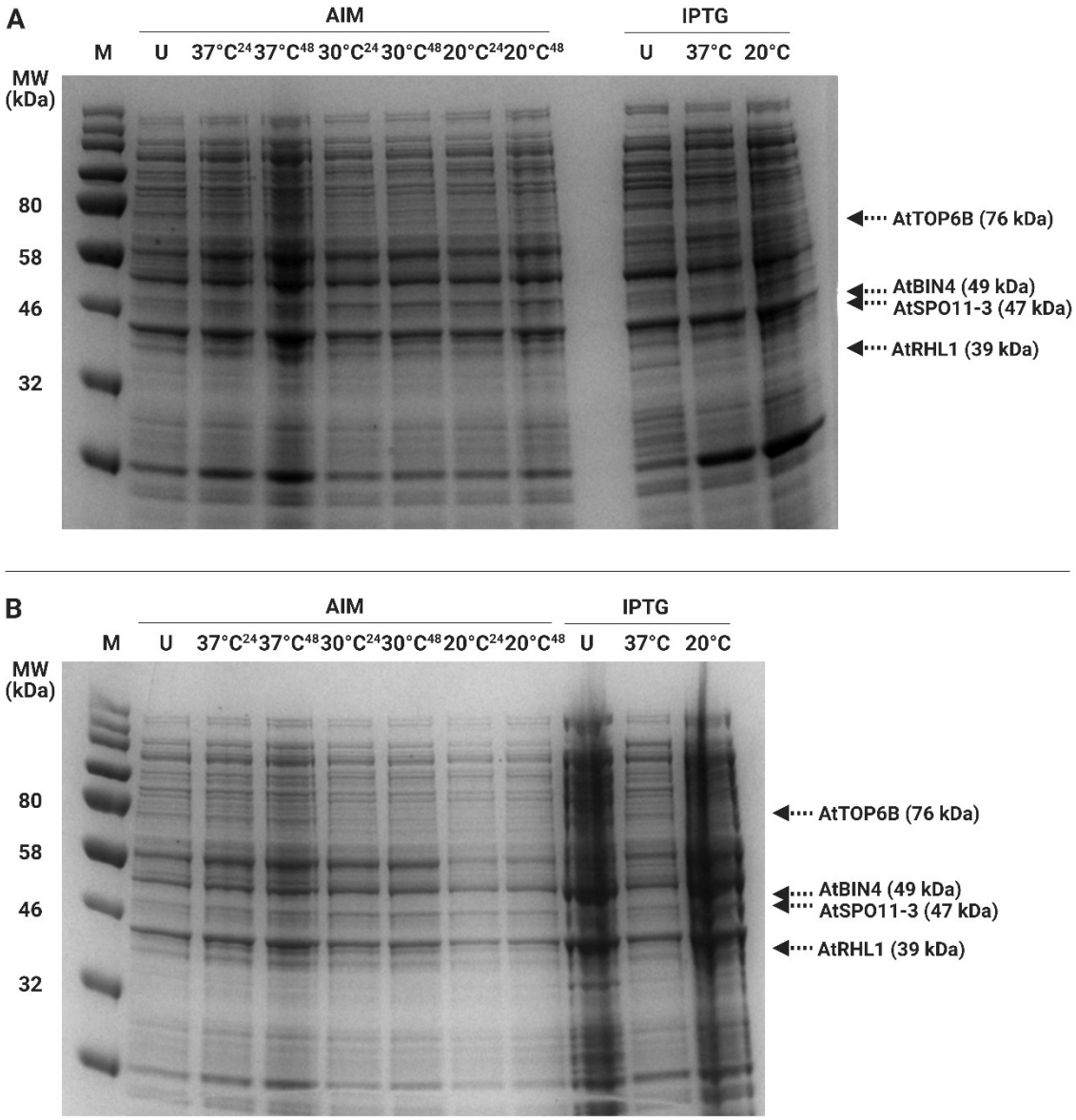

Figure 4.21| *Arabidopsis thaliana* topo VI expression trials in *E. coli*: *The expression of topo VI and its two accessory proteins from **A. thaliana** was tested in (**A**) Rosetta 2(DE3)pLysS cells and (**B**) BL21(DE3)pLysS cells. Cells were induced using at 37°C for 4 hr and 20°C for 16 hr using IPTG or were induced for 24 hr and 48 hr at 37°C, 30°C, and 20°C in AIM. Samples taken from uninduced cells (**U**) and from induced cells were analysed by SDS-PAGE alongside a protein marker (**M**). Dotted arrows indicated the predicted molecular weights of TOP6A, TOP6B, RHL1, and BIN4.*

In Chapter 1, I described a series of studies that demonstrated that SPO11-A and pTOP6B in *Plasmodium* had similar functional and expression profiles in vivo[140-142], suggesting that these two subunits, and not SPO11-B, form a topo VI or topo VI-like complex. In Chapter 3, I then used a combination of phylogenetic analyses and structural modelling to determine that *Plasmodium*, and other members of the Apicomplexa phylum, do not possess topo VI, and that SPO11-A and pTOP6B in these organisms were forming a meiotic topo VI-like complex. However, this result contradicts an earlier study that had reported that *Plasmodium falciparum* SPO11-A (PfSPO11-A) and pTOP6B (PfpTOP6B) could complement a yeast topo II null mutation, and that a crude preparation of yeast cells overexpressing PfSPO11-A and PfpTOP6B possessed decatenation activity[139]. To investigate this discrepancy further, and to help elucidate their identity, I sought to express the subunits for in vitro characterisation.

Significant improvements have been observed when coexpressing the archaeal topo VI subunits, as opposed to expressing them individually[70], and Monica has only successfully expressed the plant enzyme in insect cells when coexpressing the subunits. Therefore, I attempted to isolate the *P. falciparum* putative topo VI (PfpTopo VI) complex by attempting to coexpress its constituent subunits. Although it seems likely that PfSPO11-A and PfpTOP6B constitute a functional complex, I also tested the coexpression of PfpTOP6B and the second Spo11 paralogue in *P. falciparum*, PfSPO11-B. Furthermore, the meiotic topo VI complex in plants comprises a SPO11-1/SPO11-2 heterodimer[112], and so it would not be unusual for the *Plasmodium* SPO11 paralogues to from a heterodimer themselves. Therefore, I also tested the coexpression of all three PfpTopo VI subunits.

Previous attempts in the Maxwell research group to coexpress the PfpTopo VI subunits in *E. coli* and in insect cells have been unsuccessful. Given that many eukaryotic enzymes require expression in eukaryotic cells to acquire their appropriate post-translational modifications, and that Chalapareddy et al have reported the successful expression of an active PfpTopo VI in yeast[139], I also utilised a yeast-expression system to attempt to isolate this complex. Chalapareddy et al reported the coexpression of PfSPO11-A and PfpTOP6B by cloning these constructs into the pESC-HIS vector (Stratagene) and transforming the vector into the *Saccharomyces cerevisiae* strain PJ69-4A[139]. The pESC-HIS vector possesses two multiple cloning sites either side of the bidirectional GAL1/GAL10 promoter, meaning that the expression of one or two genes can be induced by galactose. However, coexpression in yeast

is typically performed by transforming multiple genes into separate equivalent expression vectors that are under the control of the same promoter, but possess different auxotrophic markers, to ensure a more uniform and tightly regulated expression. 12-URA-B is a yeast expression vector that also utilises the GAL1/GAL10 system and has been successfully implemented for the expression of human topo IIα and topo IIβ[208]. Therefore, I cloned PfSPO11-A, PfSPO11-B, and PfpTOP6B into the 12-URA-B, 12-TRP-B, and 12-ADE-B vectors, respectively, gifted by Scott Gradia (QB3 Berkeley, USA), and transformed the vectors into the SY991 yeast strain. Given that PfSPO11-A and PfpTOP6B are likely to be interacting subunit partners, I also transformed their corresponding expression vectors into BMA64-1A and MYA-3332 yeast cells. Following a 7-hr galactose induction at 30°C, I was able to express *S. cerevisiae* topo II from a yeast strain possessing a GAL1 expression plasmid (gifted by Nick Burton, Inspiralis, UK) (**Figure 4.22**). However, under the same conditions I was unable to observe the expression of the individual PfpTopo VI subunits when expressed individually or coexpressed in any combination in any of the three yeast strains tested (**Figure 4.23**). Therefore, my attempts to express PfpTopo VI in yeast were unsuccessful.

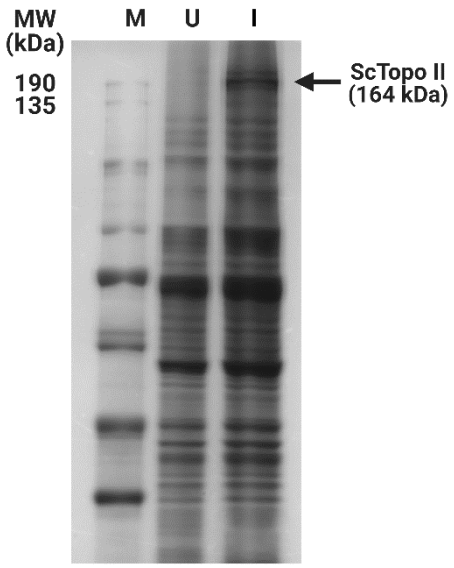

Figure 4.22| Yeast topo II expression control: *Topo II from **Saccharomyces cerevisiae** was expressed in a galactose-inducible expression plasmid in **S. cerevisiae** cells. Samples were taken from uninduced cells (**U**) and induced cells (**I**) and were analysed by SDS-PAGE alongside a protein marker (**M**). The expected band of ~164 kDa for ScTopo II was visible in the **I** sample.*

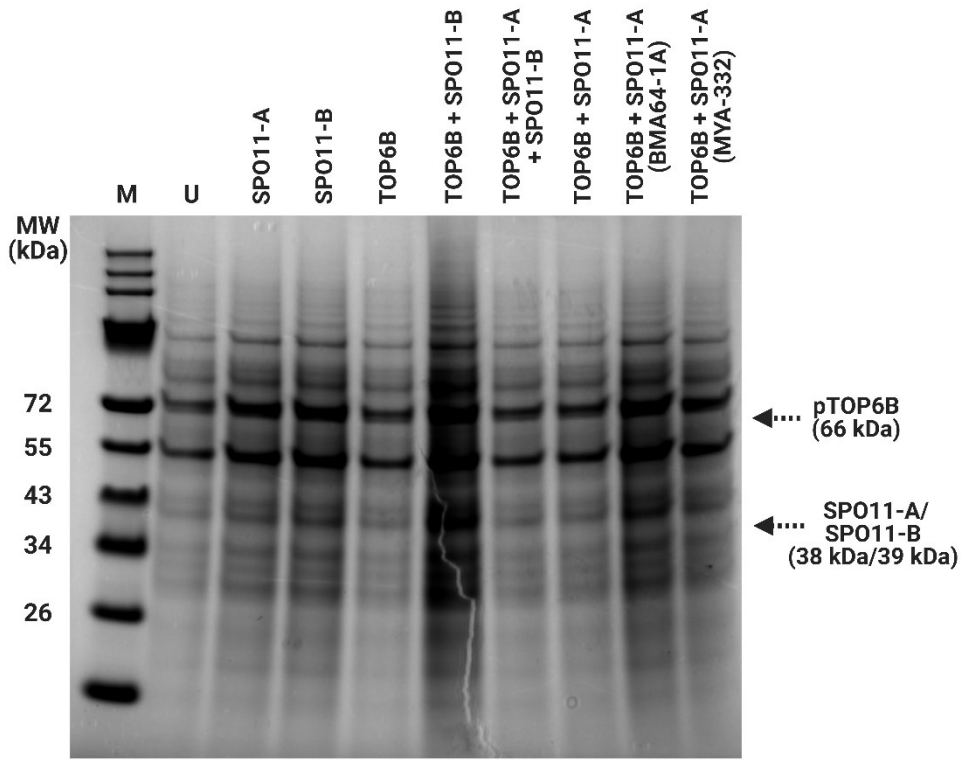

Figure 4.23| *Plasmodium falciparum* putative topo VI expression trials in yeast: *PfSPO11-A, PfSPO11-B, and PfpTOP6B expression was attempted individually and in combination in SY991* ***Saccharomyces cerevisiae*** *cells. Coexpression of PfTOP6B and PfSPO11-A was also attempted in BMA64-1A and MYA-3332* ***S. cerevisiae*** *cells. Samples were taken from uninduced cells (**U**) and were analysed by SDS-PAGE alongside a protein marker (**M**).*

4.2.6 Eukaryotic topo VI truncates expression trials

In this chapter, I have shown that a truncated form of MmTopo VI-B and its GHKL domain can be isolated in high yields to near homogeneity. Given my unsuccessful attempts to express full length topo VI from *A*. *thaliana* and *P*. *falciparum*, I sought to test whether success could be found in expressing the analogous truncates in the eukaryotic enzymes. To do this, I cloned two constructs from each of *A*. *thaliana* TOP6B and *P*. *falciparum* pTOP6B into the pET28-MHL *E*. *coli* expression vector. The AtTOP6B GHKL domain fragment (AtTOP6B-GHKL) and a truncated form of AtTOP6B that ends at the terminus of its long transducing α-helix (AtTOP6B-trans) contained residues 1-306 and 1-576, respectively. The analogous PfpTOP6B constructs, PfpTOP6B-GHKL and PfpTOP6B-trans, contained residues 1-230 and 1-450, respectively. These constructs were designed before the advent of AlphaFold, and so suitable models of AtTOP6B and PfpTOP6B were not available. I therefore designed the eukaryotic topo VI-B

constructs by aligning the sequences with MmTopo VI-B and identifying the analogous regions based off the crystal structure. Although the eukaryotic topo VI-B fragments will not be useful for investigating the relaxation and decatenation activities of their parent enzymes, I intend to use them for ATPase assays and for structural characterisation.

The topo VIB-GHKL fragment from *M*. *mazei* was successfully expressed in Rosetta 2(DE3)pLysS cells after a 24-hr incubation in AIM at 37°C, and so these conditions were replicated to test the expression of the eukaryotic topo VI-B GHKL domains. The AtTOP6B-GHKL and PfpTOP6B-GHKL samples both possessed a post-induction band at the correct MW regions after a 24-hr incubation (**Figure 4.24**), but the expression in both cases was quite poor. Therefore, expression was also tested using an alternative induction method with IPTG in 2YT media. After a 16-hr induction at 20°C with 300 µM IPTG, no improvement in expression was observed for the eukaryotic topo VI GHKL domains, although this condition was very suitable for the expression of MmTopo VI-GHKL (**Figure 4.25**). I next tested to see whether increasing the induction temperature would improve the expression, and indeed AtTOP6B-GHKL was highly expressed following a 4-hr induction using 1 mM IPTG at 37°C, however, the expression of PfpTOP6B-GHKL was still poor (**Figure 4.25**). Finally, I tested whether the induction time-length would improve the expression of PfpTOP6B-GHKL by performing a timecourse expression trial. Remarkably, PfpTOP6B-GHKL was highly expressed after just a 45-min induction using 1 mM IPTG at 37°C, but this expression decreased after 3 hr (**Figure 4.26**). These results show that the GHKL domains of topo VI from *A*. *thaliana* and *P*. *falciparum* can be highly expressed in *E*. *coli* following a 1mM IPTG-induction at 37°C, with optimum induction times of 4 hr and 2.5 hr, respectively.

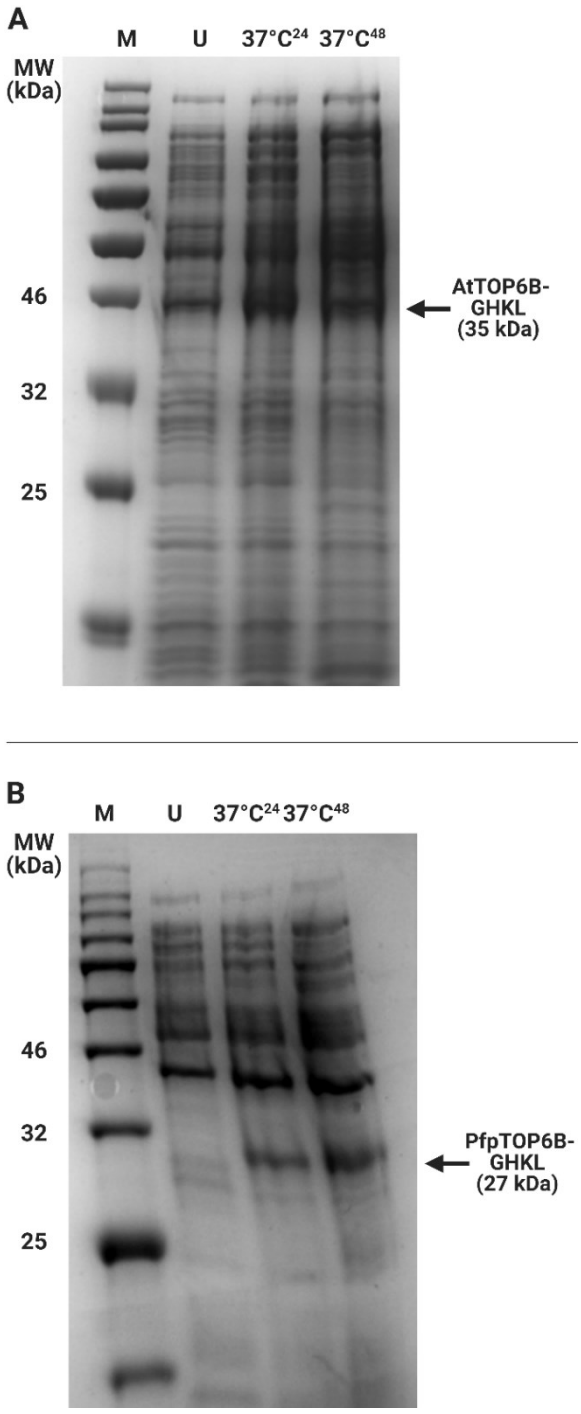

Figure 4.24| Eukaryotic topo VIB GHKL domain expression trials in AIM: *The expression of topo VIB-GHKL from (**A**) **A. thaliana** (**B**) **P. falciparum** was tested in AIM in Rosetta 2(DE3)pLysS cells. Samples taken from uninduced cells (**U**) and cells induced for 24 hr and 48 hr at 37°C were analysed by SDS-PAGE alongside a protein marker (**M**). The expected bands of ~35 kDa and ~27 kDa for AtTOP6B-GHKL and PfpTOP6B-GHKL, respectively, were visible in the induced samples.*

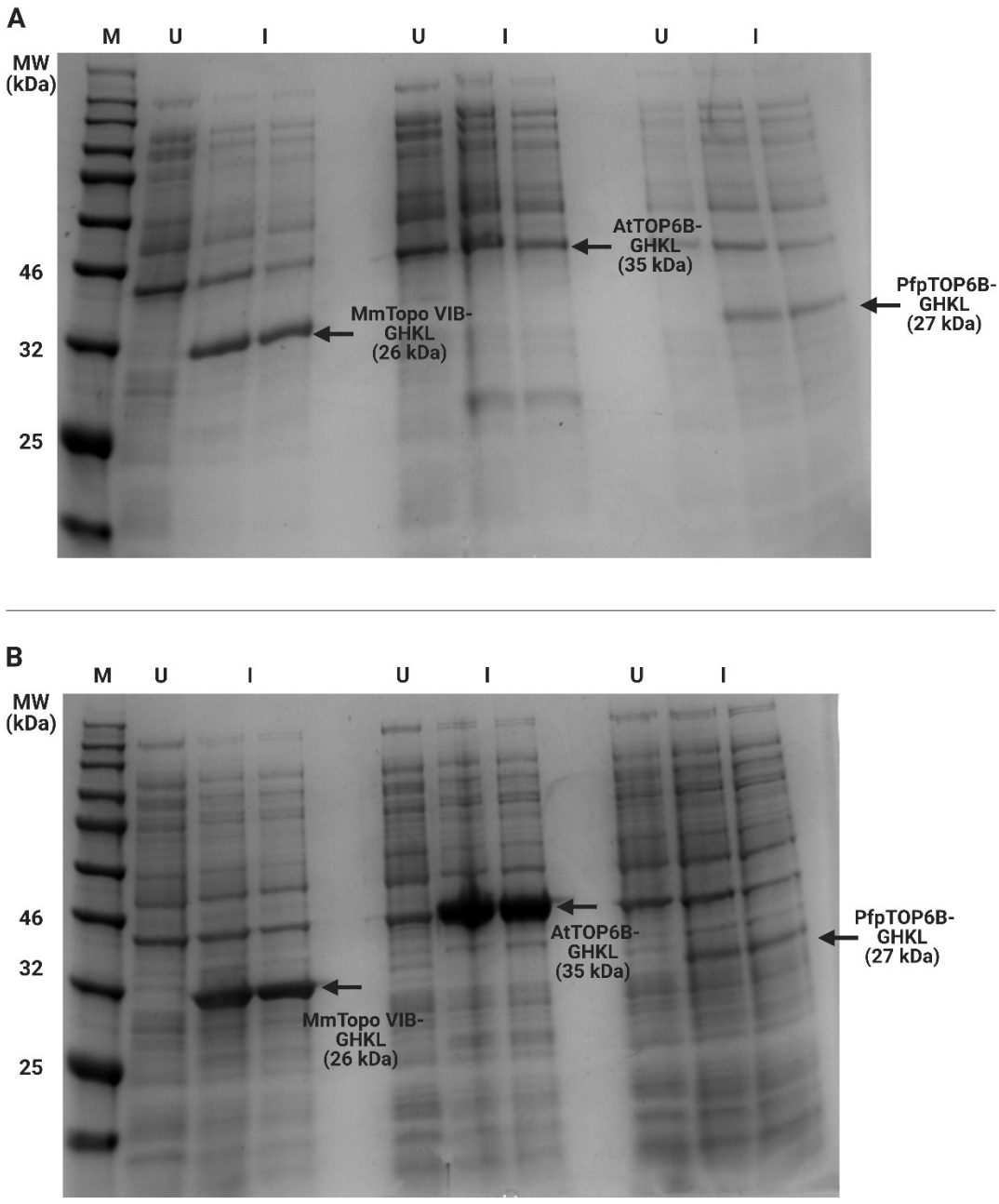

Figure 4.25| Eukaryotic topo VIB GHKL domain expression trials using IPTG: *The expression of topo VIB-GHKL from **A. thaliana** and **P. falciparum** was tested in Rosetta 2(DE3)pLysS cells alongside MmTopo VIB-GHKL as a control. Samples taken from uninduced cells (**U**) and cell cultures induced in duplicate (**I**) for (**A**) 4 hr at 37°C and for (**B**) 16 hr at 20°C were analysed by SDS-PAGE alongside a protein marker (**M**). The expected bands of ~35 kDa and ~27 kDa for AtTOP6B-GHKL, and PfpTOP6B-GHKL respectively, were visible in the induced samples.*

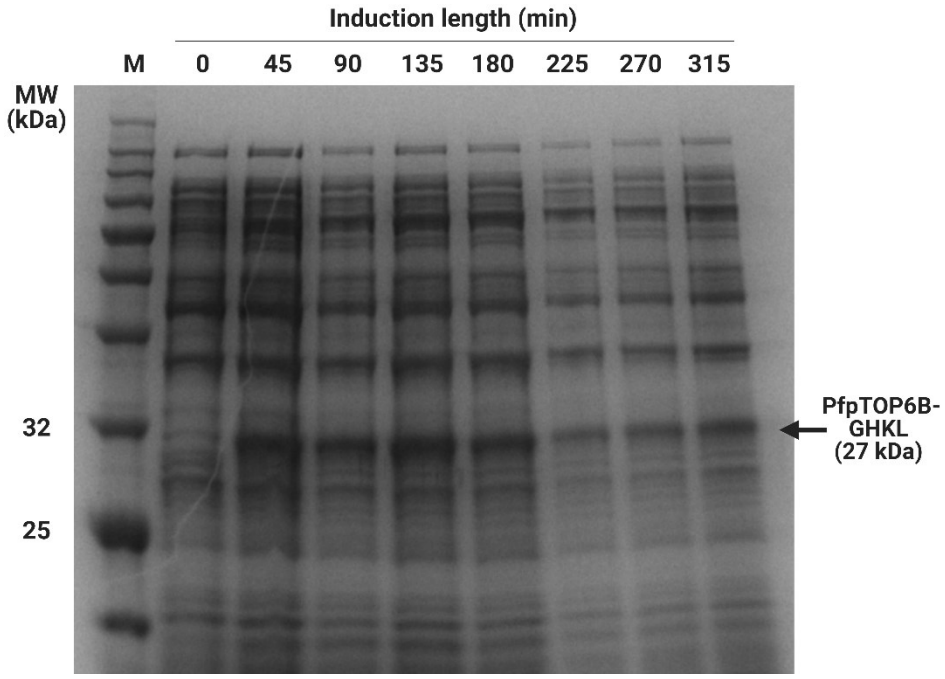

Figure 4.26| *Plasmodium falciparum* putative TOP6B GHKL domain IPTG induction timecourse: *The expression of **P. falciparum** pTOP6B-GHKL was tested by inducing Rosetta 2(DE3)pLysS cells with IPTG at 37°C. Samples were taken from induced cells every 45 min for 315 min and were analysed by SDS-PAGE alongside a protein marker (**M**). The expected band of ~27 kDa for PfpTOP6B-GHKL was after 45 min.*

To test the solubility of the eukaryotic topo VI GHKL fragments expressed under their optimum conditions, I separated the soluble and insoluble lysate fractions by subjecting the induced cells to lysis and centrifugation. Unfortunately, AtTOP6B-GHKL and PfpTOP6B-GHKL were both shown to only be present in the insoluble fractions (**Figure 4.27**). I therefore sought alternative *E. coli* expression strains that could enhance protein solubility. First, I transformed the expression vectors into Tuner(DE3)pLacI (Novagen) *E. coli* cells, which are BL21 derivatives that possess a deletion of lactose permease. IPTG normally acts in a stochastic manner, whereby its entry into cells is not uniform and higher concentrations of IPTG are required to induce protein expression in all cells. The lactose permease deletion prevents the active transport of IPTG into the cell and makes all cells equally-permeable to IPTG[209]. This results in homogeneous protein expression at lower IPTG concentrations, which has been shown to enhance the solubility of proteins[210]. In Tuner cells, a 4-hr incubation in 2YT media using an IPTG concentration range of 50 µM to 1 mM could induce expression of AtTOP6B- GHKL, but these fragments were still insoluble (**Figure 4.28**). However, expression of PfpTOP6B-GHKL

could only be induced at an IPTG concentration of 300 µM following a 2.5-hr incubation in 2YT media, and this fragment was also shown to remain insoluble (**Figure 4.29**).

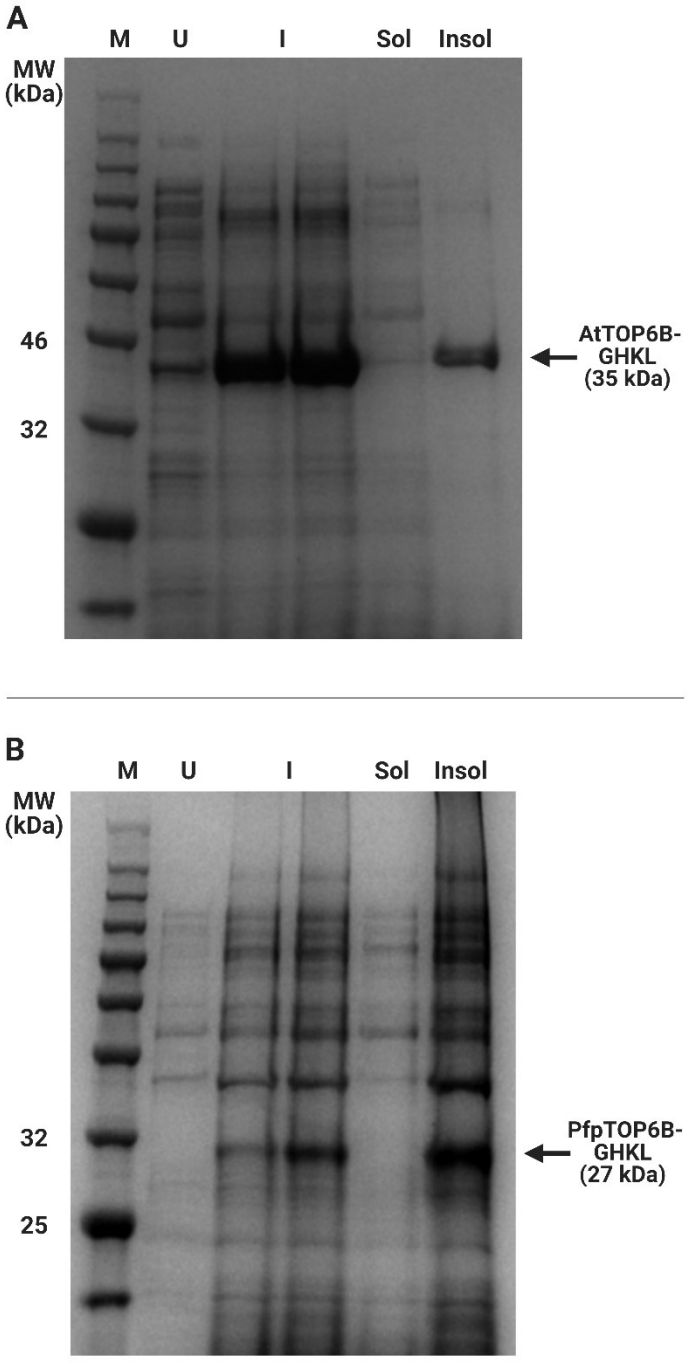

Figure 4.27| Eukaryotic topo VIB GHKL domain soluble expression test: *The expression of topo VIB-GHKL from (**A**) **A. thaliana** and (**B**) **P. falciparum** was tested for solubility. Samples were taken from uninduced cells (**U**) and cells induced using IPTG at 37°C (**I**), and the **I** samples were centrifuged and separated into soluble (**Sol**) and insoluble (**Insol**) fractions. All samples were analysed by SDS-PAGE alongside a protein marker (**M**). AtTOP6B-GHKL and PfpTOP6B-GHKL were absent in the soluble fractions.*

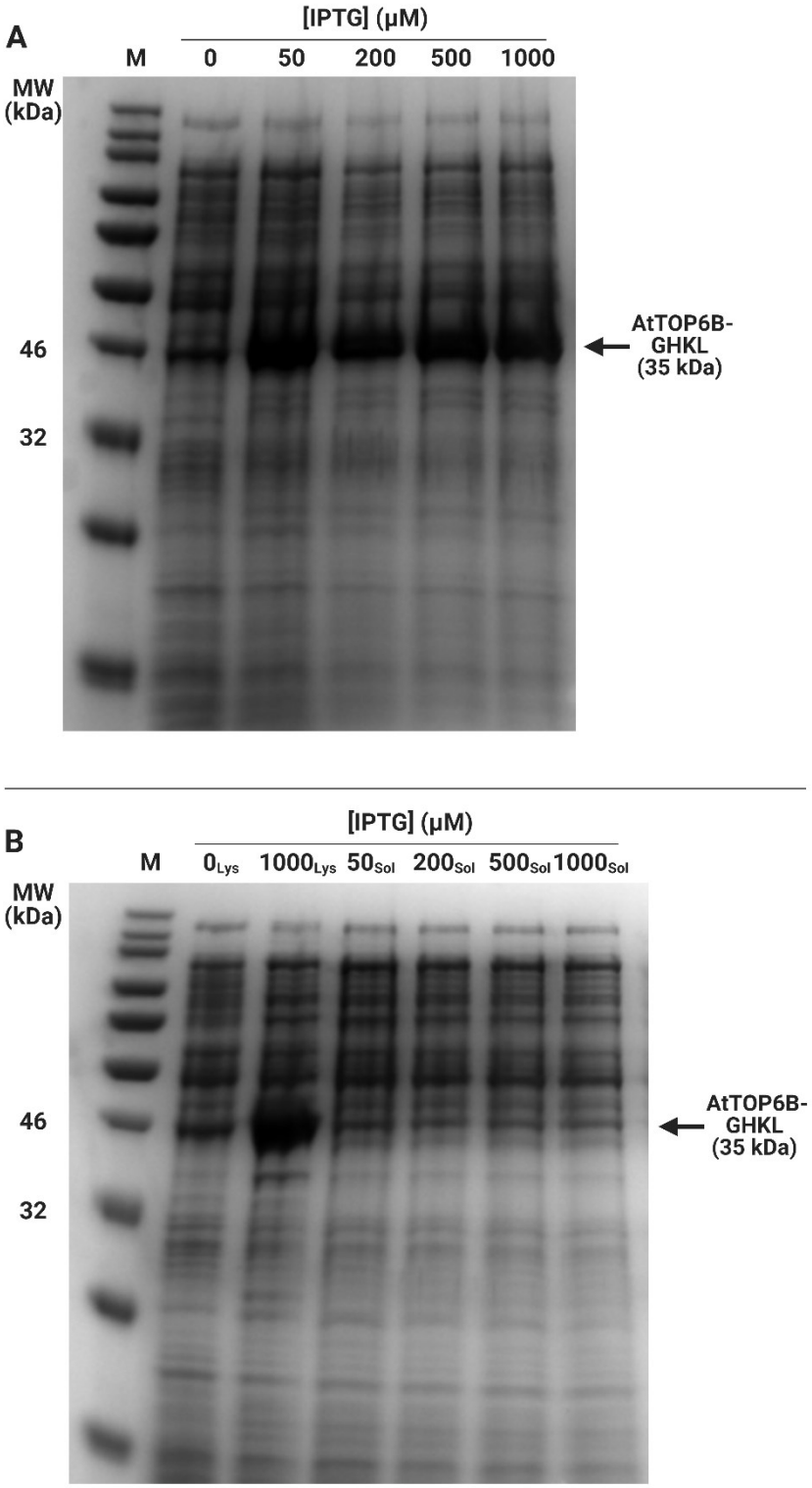

Figure 4.28| *Arabidopsis thaliana* TOP6B GHKL domain soluble expression trial in Tuner cells: *(**A**) AtTOP6B-GHKL expression was induced in Tuner(DE3)pLacI cells by an IPTG titration for 4 hr at 37°C. (**B**) These samples were tested for solubility by isolating the soluble lysate (**Sol**) from the lysate (**Lys**) and all samples were analysed by SDS-PAGE alongside a protein marker (**M**). AtTOP6B-GHKL was absent in the soluble fractions.*

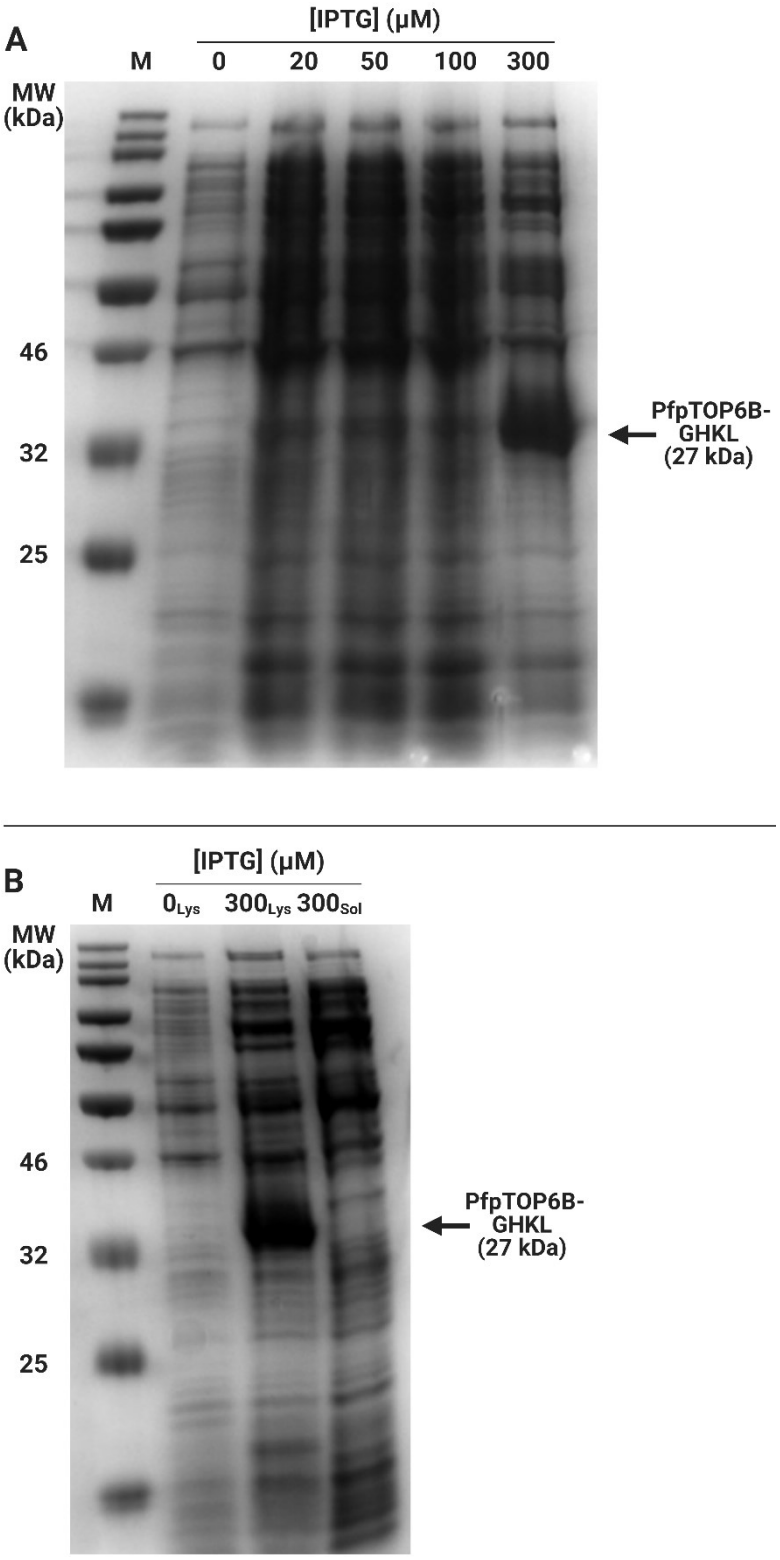

Figure 4.29| *Plasmodium falciparum* putative TOP6B GHKL domain soluble expression trial in Tuner cells: ***(A)*** *PfpTOP6B-GHKL expression was induced in Tuner(DE3)pLacI cells by an IPTG titration for 2.5 hr at 37°C. (**B**) The 300 μM IPTG sample was tested for solubility by isolating the soluble lysate (**Sol**) from the lysate (**Lys**) and all samples were analysed by SDS-PAGE alongside a protein marker (**M**). PfpTOP6B-GHKL was absent in the soluble fractions.*

The second *E. coli* strain I tested was the BL21-derivative Lemo21(DE3) which possesses a natural inhibitor of T7 RNA polymerase, T7 lysozyme, under the control of an L-rhamnose inducible promoter[211]. An L-rhamnose titration in Lemo21 cells can fine-tune the T7-based expression system in pET28-MHL to trial a wide range of expression intensities, and high L-rhamnose concentrations has been shown to reduce the formation of protein aggregates expressed in Lemo21 cells[211]. I therefore transformed the eukaryotic topo VI-GHKL fragment expression vectors into Lemo21 cells and performed L-rhamnose titrations to test their effect on the solubility of the fragments. Following a 4-hr induction using 1 mM IPTG in 2YT medium at 37°C in the presence of 0 µM L-rhamnose, AtTOP6B-GHKL was highly expressed (**Figure 4.30**). PfpTOP6B-GHKL was also highly expressed under the same conditions following a 2.5-hr incubation, however, both eukaryotic GHKL fragments remained insoluble after titrating L-rhamnose to a concentration of 2 mM (**Figure 4.30**).

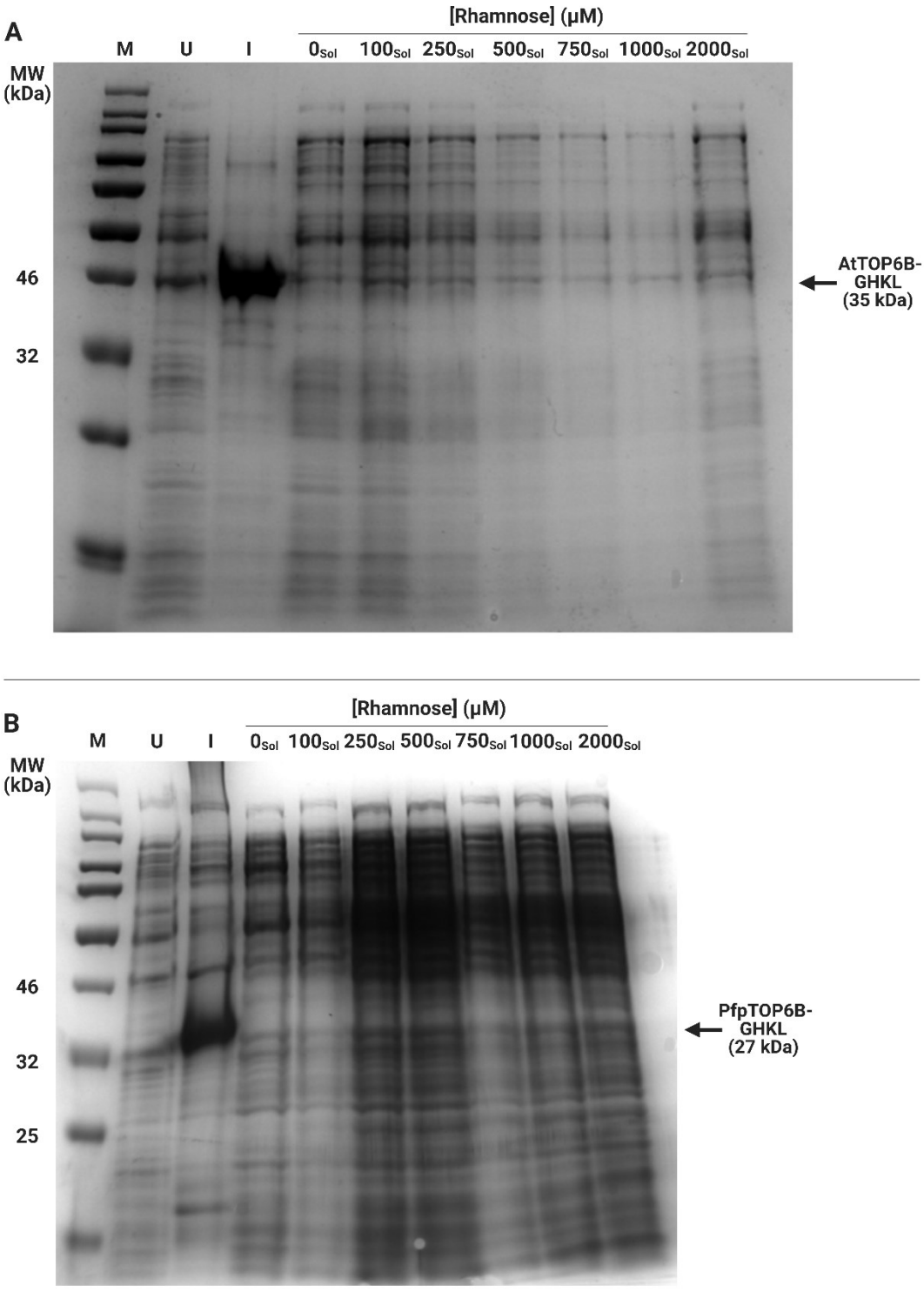

Figure 4.30| Eukaryotic topo VIB GHKL domain soluble expression trials in Lemo21 cells: *The expression of topo VIB-GHKL from (**A**) **A. thaliana** and (**B**) **P. falciparum** was tested for solubility in Lemo21(DE3) cells. AtTOP6B-GHKL and PfpTOP6B-GHKL expression was induced (**I**) by IPTG at 37°C for 4 hr and 2.5 hr, respectively. Samples were also induced in the presence of a titration of rhamnose and were tested for solubility by isolating the soluble lysate (**Sol**). These samples and uninduced lysate (**U**) were analysed by SDS-PAGE alongside a protein marker (**M**). AtTOP6B-GHKL and PfpTOP6B-GHKL were absent in the soluble fractions.*

I was unable to express soluble eukaryotic topo VI-B GHKL domains, but I nonetheless attempted to express the eukaryotic topo VIB-trans fragments. The topo VIB-GHKL fragments from *A*. *thaliana* and *P*. *falciparum* were highly expressed in Rosetta 2(DE3)pLysS cells after a 4-hr and 2.5 hr incubation, respectively, in 2YT media at 37°C. I therefore tested the expression of the eukaryotic topo VIB-trans fragments using similar conditions over a range of induction time lengths. The AtTOP6B-trans and PfpTOP6B-trans samples both possessed a thick post-induction band at the correct MW region after a 2-hr induction, signifying a high level of expression (**Figure 4.31**), however, these two fragments were also insoluble. Overall, although the topo VI-B truncates from *A*. *thaliana* and *P*. *falciparum* can be successfully expressed in *E*. *coli*, attempts to express soluble preparations of these fragments were unsuccessful.

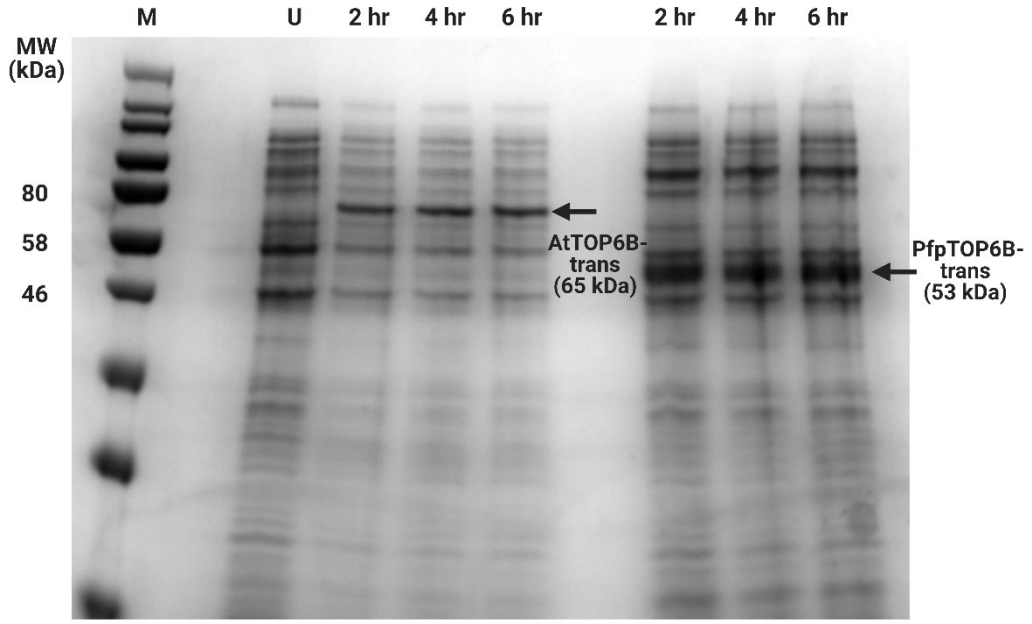

Figure 4.31| Eukaryotic topo VIB trans fragment expression trials: *The expression of topo VIB-trans from **A**. **thaliana** and **P**. **falciparum** was tested in Rosetta 2(DE3)pLysS cells. Samples taken from uninduced cells (**U**) and from cells induced with IPTG for 2 hr, 4 hr, and 6 hr at 37°C were analysed by SDS-PAGE alongside a protein marker (**M**). The expected bands of ~65 kDa and ~53 kDa for AtTOP6B-trans, and PfpTOP6B-trans respectively, were visible in the induced samples.*

4.3 Discussion

4.3.1 General discussion

The recent advent of the field of bioinformatics has revolutionised how we study proteins and has allowed us to mine a wealth of information that was hidden in their amino acid sequences. However, a detailed biochemical study of the in vitro biological activities of an enzyme of interest is necessary for it to be exploited in pharmaceutical or agricultural industry. For topo VI to be validated as an appropriate target for inhibitors, we must acquire vast quantities of soluble preparations of the enzyme, which can then be subject to biochemical characterisation and high-throughput compound screening.

4.3.2 Isolation of *Methanosarcina mazei* topo VI and its mutants and truncates

Topo VI from *M*. *mazei* was first isolated in 2007[70], following recombinant expression in BL21-CodonPlus *E*. *cells* induced by IPTG. These cells possess additional tRNAs for codons that are rarely used in *E*. *coli*, which are particularly useful for the recombinant expression of archaeal enzymes. In 2018, an updated protocol for the expression of MmTopo VI was published which featured an alternative induction procedure following growth in AIM[58]. A further improvement to the MmTopo VI expression protocol was published in 2020 by Shannon McKie[202], whose thesis detailed the use of Rosetta 2(DE3)pLysS cells that contain extra copies of four additional tRNA genes than those found in the BL21-CodonPlus cells. Using Shannon's most recent protocol, I was able to isolate the MmTopo VI enzyme in high yields to near homogeneity with relative ease (**Figure 4.4**). This feat is especially important given that large quantities of enzyme are required for the highly sensitive ATPase assays that I will perform in Chapter 5. Furthermore, the high-yielding and simple purification procedure is necessary if MmTopo VI is to be used for high-throughput compound screening.

Residues that are highly conserved in enzymes from diverse species often play a key role in catalysis or in structural stability. Therefore, it is not surprising that many of the mutants that I attempted to isolate in this chapter were more prone to aggregation, which either reduced their total yields or prevented their purification completely. To improve the expression of MmTopo VI mutants in AIM, one study decreased the incubation temperature from 37°C to 25°C[58] and this may be a better strategy for expressing other MmTopo VI mutants in future work. Substituted residues in mutants isolated with a similar efficiency as the WT enzyme in

this chapter are likely to play limited roles in enzyme stability and may instead perform key catalytic roles and so will be useful probes for mechanistic investigation in Chapter 5.

A truncation of topo VI-B from the archaeon *S. shibatae* has been shown to be more susceptible to crystallisation[61] and this fragment was later shown to be a useful model for investigating the enzyme's ATPase activity[67]. I have shown that truncates of topo VI-B from *M. mazei* can also be successfully isolated and this will pave the way for similar analyses in this enzyme, some of which I will perform myself in Chapter 5. MmTopo VIB-GHKL and MmTopo VIB-trans have theoretical isoelectric points of 5.26 and 6.38, respectively, and so both fragments should exhibit negatively charged surfaces at pH 7.5. However, the fragments could not bind to the positively charged HiTrap Q HP column at this pH, and so their surfaces are likely to be more positively charged than expected. Nonetheless, replacing the Q column with a HiTrap Heparin HP column was sufficient for removing a substantial proportion of the contaminants ready for gel-filtration, demonstrating the ease and flexibility of purifying these fragments.

4.3.3 Challenges in expressing bacterial topo VI

There are only a few topo VI-possessing bacteria mentioned in the literature and so I was astonished to discover in Chapter 3 just how widespread the enzyme is in this domain. Any mechanistic understanding that we have for topo VI mostly stems from the study of the archaeal enzymes, however it is not known whether these insights will also translate to the bacterial and eukaryotic enzymes. For example, members of the type IIA topo family all generate double-strand breaks with four-nucleotide overhangs[7], whereas these breaks are generated with two-nucleotide overhangs by topo VI from the archaeal species *S. shibatae*[47] and *M. mazei*[202]. It would be particularly useful to characterise topo VI from a bacterial species, given the relative ease of isolating prokaryotic enzymes over those from eukaryotes using an *E. coli* expression system, and this would confirm whether the mechanistic properties of topo VI are conserved outside of the archaea domain. However, my attempts to express a bacterial topo VI in this chapter were in vain and so it is worth discussing what alternative strategies could be successful in future work.

The first challenge I faced was choosing suitable bacterial species from which to pull topo VI genes for recombinant expression. The phylogenetic and bioinformatic analysis I performed in Chapter 3 gave me confidence that I was identifying functional topo VI sequences, given

the conservation of key functional residues. However, these enzymes are likely to have adapted to their specific cellular environments and may not be able to fold properly in the alternative conditions presented in an *E*. *coli* expression background. Although I selected topo VI-possessing bacteria that had similar optimum pH and growth temperatures to *E*. *coli*, there are many other factors such as ionic strength and oxidative stress that can heavily influence protein stability. Topo VI was first isolated from the thermophilic archaeon *S*. *shibatae*[46] and so a successful expression strategy may be the identification of a thermostable bacterial enzyme that is likely to be more resilient to misfolding. I have identified topo VI genes in the thermophilic bacterial species *Bythopirellula goksoeyrii*, and these may be more suitable to recombinant expression. It may have been unwise to attempt to express topo VI from *P*. *abortibovis* given that the bacteria has yet to be cultured in vitro and so mimicking the enzyme's natural environment may have been be too much of a stab in the dark.

The second challenge I faced was choosing suitable tools and conditions for inducing protein expression. The pST39 vector was successfully implemented for the expression of *M*. *mazei* topo VI and it seems likely that coexpressing the bacterial topo VI genes is the best strategy. However, expressing bacterial topo VI may require alternative polycistronic expression vectors, such as the Duet vector series (Novagen), or the coexpression of the subunits on separate compatible vectors. Furthermore, there remains the possibility that a successful strategy would involve expressing the two bacterial topo VI subunits individually and reconstituting the enzyme following their purification, as has been performed with *S*. *shibatae* topo VI[185]. Although I tested a variety of induction methods, temperatures, and cell lines, a strategy I neglected to implement is the use of chaperones. These proteins can bind to non-native polypeptides to assist their folding, prevent their aggregation, and promote their refolding following denaturation due to intrinsic instability[212]. Coexpression with different combinations of the chaperone systems GroEL/GroES, DnaK/DnaJ/GrpE, IbpA/IbpB, and ClpB on compatible plasmids has been shown to increase the yields of soluble recombinant protein by up to 42-fold[213]. Furthermore, a heat-shock response can be artificially induced in *E*. *coli* at low temperatures by adding the membrane-fluidising agent benzyl alcohol to the growth medium[214] and this method has been used to increase the yields of soluble recombinant protein by up to 48-fold[215]. In future attempts to express bacterial topo VI these methods should be considered.

4.3.4 Challenges in expressing eukaryotic topo VI

I demonstrated in Chapter 3 that topo VI is widely distributed in eukaryotes, and this enzyme has been heavily investigated in plants. Given that topoisomerases have been successful targets for antibiotics and chemotherapeutic drugs, plant topo VI is a potential target for novel herbicides. For this potential to be realised, the enzyme must be isolated and subjected to extensive biochemical characterisation. The first major hurdle in isolating plant topo VI was overcame by Monica Agarwal (John Innes Centre, UK) in the Maxwell research group, who succeeded in expressing active topo VI from *A*. *thaliana* using an insect-cell expression system (personal communication). However, this expression resulted in an enzyme preparation with an extremely low yield and purity. Monica also showed that AtTopo VI expression could only be achieved by coexpressing the two topo VI subunits and RHL1 and BIN4 on separate vectors. To improve the yield of AtTopo VI, and considering the lessons from Monica's work, I attempted to coexpress all four components of AtTopo VI from a single polycistronic vector in *E*. *coli*, but these attempts were unsuccessful. I have discussed the alternative strategies that could be successful in future attempts to express bacterial topo VI, and it would be useful to have a similar evaluation of the expression trials of the plant enzyme.

Many of the strategies I would consider for future attempts to express topo VI from bacteria are also applicable to the plant enzyme. First of all, I only attempted to express topo VI from the dicotic plant *A*. *thaliana* and future success may be found using alternative plant species, including a monocot. It may also be worth considering the use of chaperones to enhance the assembly of the plant topo VI constituents into their native folds in the *E*. *coli* host. Furthermore, alternative expression vectors may need to be considered, and coexpressing the plant topo VI constituents on separate compatible vectors may yield more success than using a single vector. An additional consideration that is not applicable to bacterial topo VI is that plant topo VI may require eukaryotic post-translational modifications to form a stable complex, and so *E*. *coli* may be an unsuitable host for recombinant expression. Therefore, an alternative eukaryotic expression system, such as yeast, may be required to attain higher yields than the insect-cell system.

Although topo VI-A, topo VI-B, RHL1, and BIN4 are present in close protist relatives, my analyses in Chapter 3 suggest that these proteins are absent in *Plasmodium*. However, PfSPO11-A and PfpTOP6B, which my phylogenetic and bioinformatic analyses have

characterised as subunits of the meiotic topo VI-like complex in *Plasmodium*, have been reported to possess decatenation activity following expression in yeast[139]. Attempts to express combinations of PfSPO11-A, PfSPO11-B, and PtTOP6B in insect cells and *E*. *coli* have previously been unsuccessful in the Maxwell research group, and I too was unable to express these subunits in a yeast expression system. As with topo VI from bacteria and plants, there are many strategies that could be implemented in future attempts to express the *Plasmodium* putative topo VI. However, given the evidence that I have presented in Chapter 3 for *Plasmodium* not possessing topo VI, I suggest that the significant effort required to do this would not be worthwhile. I did not attempt to replicate the exact conditions that were previously reported to successfully express *Plasmodium* topo VI, and these should be repeated and validated before attempting to optimise.

Although I was unable to express AtTopo VI or PfpTopo VI, I was able to show that truncates of their B subunits can be highly expressed in *E*. *coli*. However, these truncates were only present in inclusion bodies and my attempts to express them in soluble form were unsuccessful. Isolating these truncates would be very useful for structural studies, and so it is worth discussing what future improvements can be made. First of all, although the MmTopo VI-B truncates could be designed using a crystal structure, the AtTOP6B and PfpTOP6B truncates were designed by alignment with the *M*. *mazei* truncates. Following the release of AlphaFold, it is now possible to design constructs from accurate models of their structures. Analogous truncates of MmTopo VIB-GHKL and MmTopo VIB-trans in AtTOP6B based on an AlphaFold model would possess residues 1-304 and 1-585, respectively, instead of residues 1-306 and 1-576, respectively. The additional two residues of AtTOP6B-GHKL are disordered, and AtTOP6B-trans ends 14 residues before the end of the long transducing α-helix. Therefore, an improvement in solubility may occur when expressing the correct constructs that terminate at the last residue of a secondary structure element. In Chapter 3, I showed that the AlphaFold model of PfpTOP6B strongly resembles the meiotic topo VIB-like subunit AtMTOPVIB, and so the PfpTOP6B truncates may not be useful in the study of topos. I have already discussed the potential use of chaperones to improve the expression of bacterial and eukaryotic topo VI, however, their use may be particularly useful in refolding the misfolded AtTOP6B truncates. Another strategy would be to purify the insoluble truncates under denaturing conditions. Inclusion body material can be solubilised in the presence of

chaotropic agents, such as urea and guanidine, and refolded into their native states by removing the chaotrope by dilution or dialysis[216].

4.3.5 Conclusion

In this chapter, I have isolated full length topo VI from *M*. *mazei*, as well as a series of mutants and truncates, as a prelude for their biochemical characterisation in Chapter 5. I have also attempted to express novel topo VI species from bacteria and eukaryotes, although these attempts were unsuccessful. Furthermore, I have shown that truncates of the B subunit of TOP6B from *A*. *thaliana* and pTOP6B from *P*. *falciparum* can be highly expressed in *E*. *coli*, albeit in inclusion bodies. Future work should focus on successfully purifying bacterial topo VI, optimising the expression of *A*. *thaliana* topo VI, and purifying the eukaryotic TOP6B truncates.

Chapter 5

Biochemical characterisation of *Methanosarcina mazei* topo VI

5.1 Introduction

5.1.1 General introduction

An enzyme's mechanism must be sufficiently understood for the rational design of inhibitors. Therefore, to explore the potential of DNA topoisomerase (topo) VI as target for inhibitors, the enzyme must be subjected to stringent biochemical characterisation. Considering the difficulties in isolating the eukaryotic enzymes, topo VI from the archaeon *Methanosarcina mazei* (MmTopo VI) is an ideal model due to the ease in which it can be produced recombinantly in *Escherichia coli* in high yields. In Chapter 6, I will conduct preliminary compound screening on MmTopo VI to provide hits for the eukaryotic enzymes, and the current chapter will explore the key uncertainties surrounding the enzyme's mechanism in preparation for this.

5.1.2 Visualisation of DNA topoisomers

A key technique that has been widely employed for characterising topos is agarose gel electrophoresis, which exploits the differential rates of plasmid migration within a gel due to their topology. This technique has been extremely valuable given the roles topos play in manipulating DNA topology. As will be demonstrated in this chapter, supercoiled (sc) topologies induce the fastest migration on an agarose gel, due to plasmid compactness, whereas nicked DNA has the slowest migration, due to the loss of twist and writhe and overall compactness. Depending on gel-conditions, linear plasmids migrate faster than nicked conformations due to their smaller width, however, they migrate slower than sc forms due to their reduced compactness. Titrating scDNA with a topo will progressively slow down the migration of the plasmid and this can be visualised on a gel as a series of bands that correspond to topoisomers that become increasingly more relaxed. Topos are unable to convert DNA into its fully relaxed state and instead a dynamic equilibrium is reached in which various degrees of supercoiling remain. Therefore, instead of a single fully relaxed band being observed on a gel, multiple bands will persist at high topo concentrations that represent an

equilibrium distribution of relaxed topoisomers. The type IIA topos have been shown to perform topological simplification, whereby the DNA reaches a state of relaxation that is beyond what is thermodynamically expected by random strand-passage events[217,218]. Type IA and IIB topos, however, have been shown to not participate in topological simplification and instead reach an equilibrium that is thermodynamically expected[70,217].

5.1.3 *Methanosarcina mazei* topo VI distributive behaviour

For type II topos, processivity describes a mechanism whereby an enzyme binds to a G-segment before capturing a T-segment and remains bound to the G-segment while performing multiple strand-passage events. A distributive activity, however, describes the disengagement of the G-segment after resolving a single crossing, and with MmTopo VI, has been shown to involve the simultaneous binding of a G-segment and a T-segment[58,172]. To measure the processivity of topos, a magnetic tweezers assay has been developed that tethers a DNA duplex between a magnetic bead and a surface[219]. By rotating the magnet, the duplex can be forced into a right- or left-handed superhelical conformation that mimics that of negatively (-) or (+) scDNA, respectively[219]. This assay has been used to show that *E. coli* topo IV relaxes +scDNA by performing rapid bursts of multiple strand-passage events, indicating a processive relaxation, but relaxes -scDNA by performing short bursts of individual strand-passage events, indicating a distributive relaxation[220]. It was also shown that this processivity could permit topo IV to relax +scDNA ~21-fold faster than -scDNA[220]. The same technique has also been used to demonstrate a similar chiral selectivity in MmTopo VI, with the enzyme being shown to relax positive supercoils ~2-3-fold faster than negative supercoils[172]. Unlike topo IV, however, topo VI was shown to act in a distributive manner on both positive and negative supercoils, although was less distributive on negative supercoils[172]. The distributive mechanism of MmTopo VI has also been demonstrated using a plasmid competition assay, where an enzyme is preincubated with a -sc plasmid, and a second competing -sc plasmid is added along with ATP. The initial rate of relaxation of the first plasmid by the highly processive yeast topo II was shown to be unaffected by the addition of the second plasmid, whereas topo VI relaxed both plasmids simultaneously in a highly distributive manner[58]. This distributive behaviour accounts for MmTopo VI having a maximum positive supercoil relaxation rate of ~6 strand-passage events min^{-1}, which is ~23-fold slower than that of *E. coli* topo IV[172,219].

5.1.4 *Methanosarcina mazei* topo VI chiral selectivity

The preference of MmTopo VI for various DNA topologies has been assessed using a competitive binding assay, whereby the enzyme is incubated with a fluorescent-labelled oligonucleotide prior to the addition of competitor DNAs molecules. Using this assay, MmTopo VI was shown to preferentially bind oligonucleotides longer than 40 base pairs (bps) and to exhibit a ~60-fold preference for -scDNA over linear DNA[58], providing further evidence that MmTopo VI is a DNA-crossing sensor. Interestingly, a nitrocellulose-membrane capture technique[221] has shown that MmTopo VI preferentially binds -scDNA over +scDNA, despite having greater relaxation activity on the latter[172]. Using an agarose gel-based assay, MmTopo VI was also shown to maintain higher levels of ADPNP-induced double-strand breaks on -scDNA than +scDNA[172]. This behaviour is also observed in human topo IIα, which relaxes positive supercoils ~10-fold faster than negative supercoils, but preferentially binds and cleaves -scDNA[222].

An assay that couples the hydrolysis of ATP to the oxidation of NADH with the enzymes pyruvate kinase (PK) and lactate dehydrogenase (LDH)[223] has been used to show that the ATPase activity of MmTopo VI is DNA-dependent and ~5-fold greater in the presence of -scDNA than linear DNA[58]. The study also showed that mutating a small region of positively charged residues (K186, R188, R189) on the B subunit of MmTopo VI to alanine or glutamate can increase -scDNA-dependent ATP hydrolysis by the enzyme[58]. This KRR loop lines the path of the T-segment within the central cavity and therefore seems to contribute to the coupling of ATPase activity to strand-passage in MmTopo VI. Another finding from this work was that mutations to the helix-2-turn-helix (H2TH) domain decreased the -sc relaxation activity of MmTopo VI by ~20-30-fold and that mutations in the KRR loop abolished this activity. Mutations to both regions also produced a ~10-20-fold decrease in the enzyme's binding affinity for -scDNA as measured by fluorescence anisotropy[58], which implicates the KRR and H2TH structural elements in the specificity of MmTopo VI for a sc DNA crossing. Furthermore, a gel-based assay has shown that that a 70 bp oligonucleotide is required to promote ADPNP-induced cleavage by MmTopo VI, which is the length required for a G-segment to span the two H2TH domains in a bent conformation around the ATP gate[58]. A bulk FRET approach using a Cy5/Cy5.5-labeled 70 bp oligonucleotide then showed that mutations in the H2TH domain abolished the nucleotide-dependent G-segment bending of MmTopo VI, which therefore

implicates the H2TH region in the coupling of ATP hydrolysis to productive DNA cleavage. Topo VI, unlike the type IIA topos, lacks a C-gate and must therefore utilise an alternative mechanism for tightly controlling G-segment cleavage to ensure that DSBs only occur in the context of strand-passage. These experiments show that topo VI avoids uncontrolled DSBs by sensing DNA crossings and tightly coupling ATP hydrolysis to strand-passage.

An alternative magnetic tweezers assay has been developed whereby two DNA duplexes, instead of one, are tethered to a single magnetic bead and a surface[224]. By rotating the magnet, the duplexes are wrapped around each other and forced into a right- or left-handed braid conformation that mimics that of catenated DNA rather than of scDNA[224]. This assay has been used to show that the rate of DNA decatenation is ~5-fold greater than that of -sc relaxation[172]. The same study also used Monte Carlo simulations of catenane configurations to show that MmTopo VI has a preferred DNA-crossing angle of ~88°. It had previously been shown using a similar method that the most probable crossing angle is 90° in catenanes, 60° in positive supercoils, and 120° in negative supercoils[219], which rationalises the strong preference of MmTopo VI for decatenation. Furthermore, Monte Carlo simulations have shown that a crossing angle of ~88° is ~5-fold more probable in positive DNA crossings than negative DNA crossings, which rationalises the chiral preference of MmTopo VI for relaxing +scDNA[172].

5.1.5 Aims and objectives

To gain a better understanding of the mechanism of topo VI, I will conduct a series of gel-based experiments to assess the enzyme's relaxation, decatenation, and cleavage activities, and conduct a spectrophotometric assay to probe its ATPase activity. I will also investigate key mechanistic phenomena, such as chiral discrimination and the direction of strand-passage, by testing a variety of different DNA substrates and enzyme mutants.

5.2 Results

5.2.1 The DNA relaxation, cleavage, and decatenation activities of *Methanosarcina mazei* topo VI

To test the activity of the MmTopo VI sample that I isolated in Chapter 4, the enzyme was assayed for relaxation, cleavage, and decatenation activity using reaction conditions previously optimised[202]. These reactions were then analysed by the traditional agarose gel-based method, which permits the visualisation of the topological state of DNA substrates. First, I assessed the relaxation and cleavage activities of MmTopo VI on the -sc pBR322 plasmid by performing a sequential enzyme titration. This was revealed that MmTopo VI has an observable ATP-dependent relaxation activity at a concentration of 5 nM, as evidenced by the conversion of a single DNA species into a series of relaxed topoisomers, and more complete relaxation activity at 10 nM (**Figure 5.1**). Replacing ATP with ADPNP, a non-hydrolysable ATP analogue, in the reaction mixture has previously been shown to stabilise the topo VI-DNA cleavage complex and inhibit DNA relaxation[202]. In my hands, incubation with ADPNP granted MmTopo VI an observable cleavage activity at ~10 nM, as indicated by the appearance of a linear DNA band, with increasing cleavage activity up to 80 nM (**Figure 5.2**).

I then assessed the decatenation activity of MmTopo VI on a kinetoplast (k) DNA substrate, which comprises a giant network of catenated DNA maxicircles and minicircles, and a singly-linked catenated DNA substrate (bis-cat), which comprises two catenated -sc plasmids with different molecular weights[174]. In the kDNA assay, the substrate is too large to migrate through the gel and is trapped in the wells, however, the addition of a decatenase can release minicircles from the kDNA that are small enough to migrate through the gel. By performing a serial enzyme dilution, I showed that MmTopo VI has an observable ATP-dependent decatenation activity on kDNA at 5 nM, as illustrated by the appearence of mini-circles (**Figure 5.3**). In the bis-cat DNA assay, the addition of a decatenase can convert the substrate into two -sc plasmids, which can then be converted into their relaxed states at a higher enzyme concentration[174]. Using this assay, I have shown that MmTopo VI also has observable decatenation activity on bis-cat DNA at 5 nM, as evidenced by the appearance of two -sc plasmids (**Figure 5.4**). MmTopo VI can then be observed to relax these -sc plasmids at 10 nM, with more complete relaxation observed at 20 nM, providing further evidence that MmTopo VI is a preferential decatenase (**Figure 5.4**).

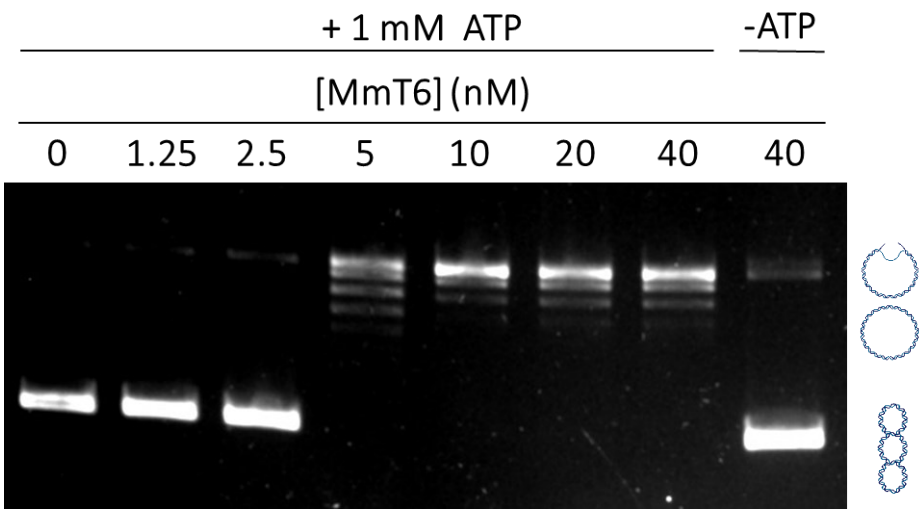

Figure 5.1| *Methanosarcina mazei* topo VI relaxation activity: *MmTopo VI relaxes 2.5 nM negatively-supercoiled DNA, producing a series of relaxed DNA topoisomers. Relaxation activity is observed at ~5 nM and is ATP-dependent. Reactions were run at 37°C for 30 min.*

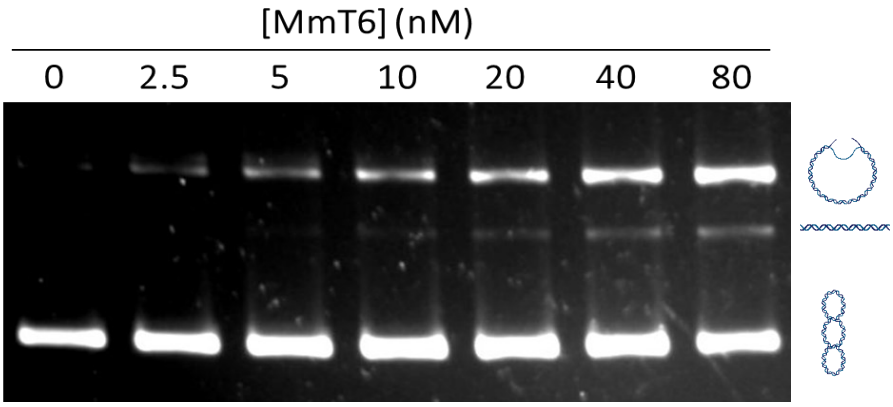

Figure 5.2| *Methanosarcina mazei* topo VI cleavage activity: *MmTopo VI can be converted into an enzyme capable of generating permanent double-stranded breaks by stabilising the enzyme-DNA cleavage complex with a non-hydrolysable ATP analogue, ADPNP. Cleavage activity on 5 nM negatively-supercoiled DNA is observed at ~10 nM. Reactions were run at 37°C*

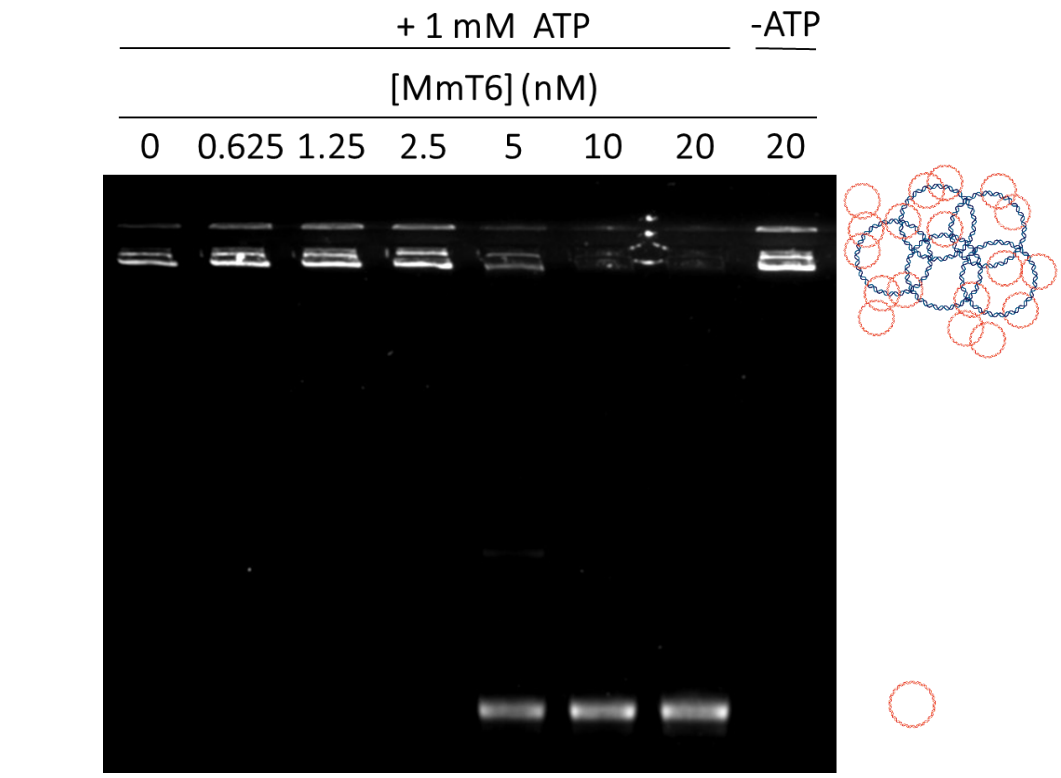

Figure 5.3| *Methanosarcina mazei* topo VI decatenation activity on kDNA: *MmTopo VI can decatenate 6.7 ng µL^{-1} kDNA, which is too large to migrate out of the agarose wells, and release DNA mini-circles. Decatenation activity is observed at ~5 nM and is ATP-dependent. Reactions were run at 37°C for 30 min.*

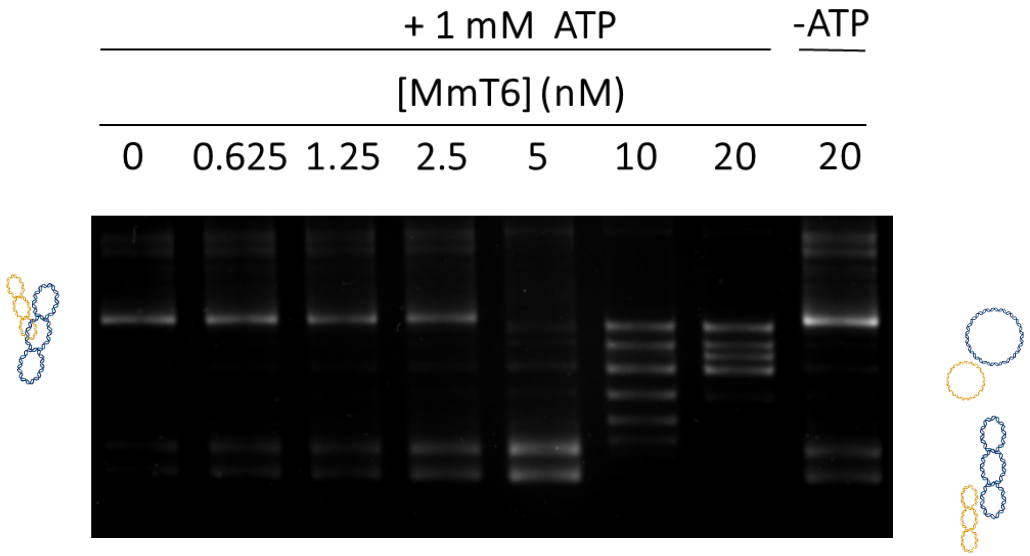

Figure 5.4| *Methanosarcina mazei* topo VI decatenation activity on bis-cat DNA: *MmTopo VI can decatenate 1.3 ng µL^{-1} bis-cat DNA in an ATP-dependent manner at ~5 nM and relax the decatenated products at ~10 nM. Reactions were run at 37°C for 30 min.*

5.2.2 Mutagenic analysis of *Methanosarcina mazei* topo VI G137 and D294 residues

In Chapter 3, I identified two key residues in MmTopo VI that have yet to be investigated for their influence on the biochemical activities of the enzyme. The first residue, G137 in MmTopo VI-B, is conserved in the G3-box of GHKL domain-possessing enzymes and confers flexibility to a disordered loop that stabilises the ATP-binding pocket (**Figure 5.5**)[59]. G137 has also been shown to be conserved in the GHKL-like domains of the topo VIB-like family[186], and in Chapter 3 I showed that this residue is conserved in the GHKL domains of topo VI from archaea, bacteria, and eukaryotes. I also showed that this residue is absent in the GHKL-like domains of MTOPVIB from the Apicomplexa phylum, a subunit family I had characterised for the first time in Chapter 3. The second residue, D294 in MmTopo VI-A, was identified in Chapter 3 while searching for sequence motifs that distinguish the SPO11-A and SPO11-B paralogues in Apicomplexa and is uncharacterised in the literature. D294 is present in the toprim domain of topo VI-A homologues and sits adjacent to the metal-binding DxD motif and so may play a role in coordinating a divalent metal ion (**Figure 5.6**). In Chapter 3, I showed that D294 is conserved in all topo VI-A sequences in the three domains of life and is conserved in most Spo11 sequences. I also showed that D294 is conserved in Apicomplexa SPO11-B, but is absent in SPO11-A. Therefore, in Chapter 4, I isolated mutants of G137 and D294 to investigate the biochemical role of these residues on MmTopo VI activity and to investigate the significance of their absence in the Apicomplexa topo VI-like subunits.

Saccharolobus shibatae **topo VIB-GHKL**

Methanosarcina mazei **topo VIB-GHKL**

Figure 5.5| *Methanosarcina mazei* topo VI-B G137 residue: *The crystal structures of the topo VI-B GHKL domains from* ***Saccharolobus shibatae*** *(PDB code: 1Z59) and* ***M. mazei*** *(PDB code: 2Q2E). Each structure possesses a conserved glycine residue, G137 (orange), in the G3-Box motif that confers flexibility to a disordered loop that stabilises the ATP-binding pocket. The* ***S. shibatae*** *structure contains a molecule of ADP.*

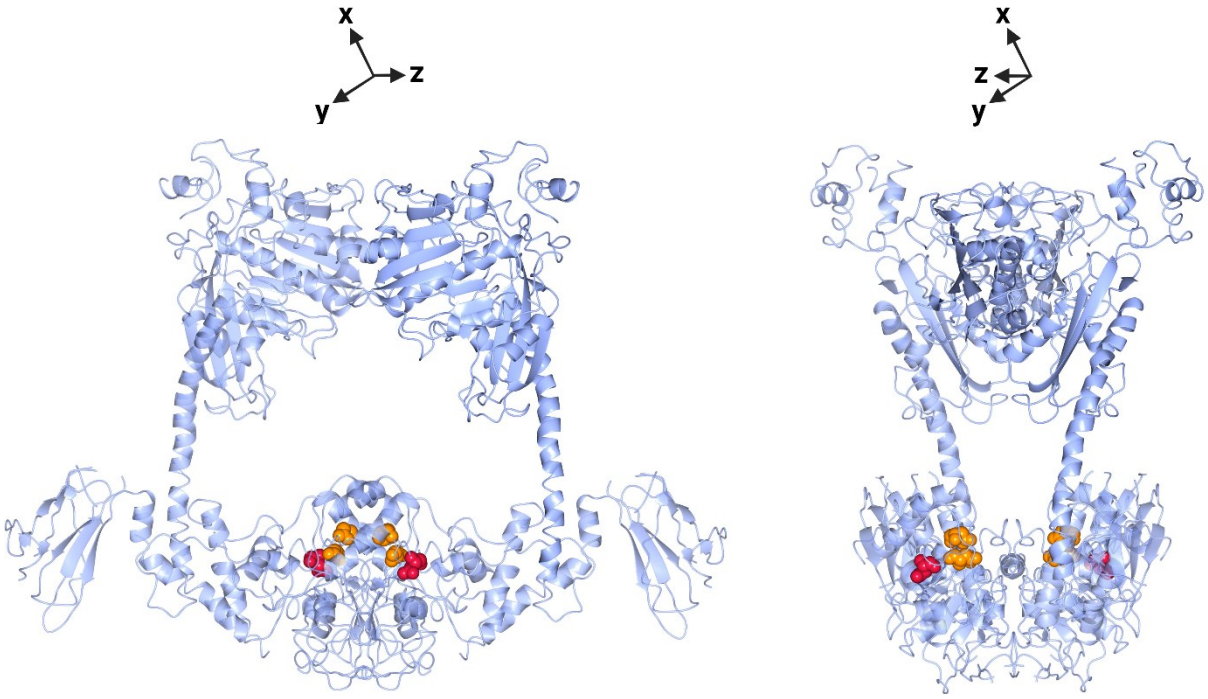

Figure 5.6| *Methanosarcina mazei* topo VI-A D294 residue: *The crystal structure of MmTopo VI (PDB code: 2Q2E) reveals the location of the D294 residue (red) in the toprim domain of topo VI-A. D294 sits adjacent to the metal binding DxD motif (orange).*

The MmTopo VI.G137A mutant was assayed for relaxation, cleavage, and decatenation activity by performing sequential enzyme titrations using the same conditions as the wild type (WT) enzyme. First, I showed that MmTopo VI.G137A has observable relaxation activity at a concentration of 20 nM, with more complete relaxation activity at 40 nM (**Figure 5.7**), which represents an ~4-fold reduction in activity compared to the WT enzyme . MmTopo VI.G137 was also shown to exhibit at least a ~4-fold reduction in decatenation activity, as illustrated by the partial decatenation of kDNA at 20 nM (**Figure 5.8**), but to have observable cleavage activity at ~10 nM, which is comparable to the WT enzyme (**Figure 5.9**). These results suggest that the MmTopo VI.G137A mutant is therefore likely to be defective in performing strand-passage, but fully capable of forming cleavage complexes, which is likely to be due to loss of ATPase activity and not ATP binding.

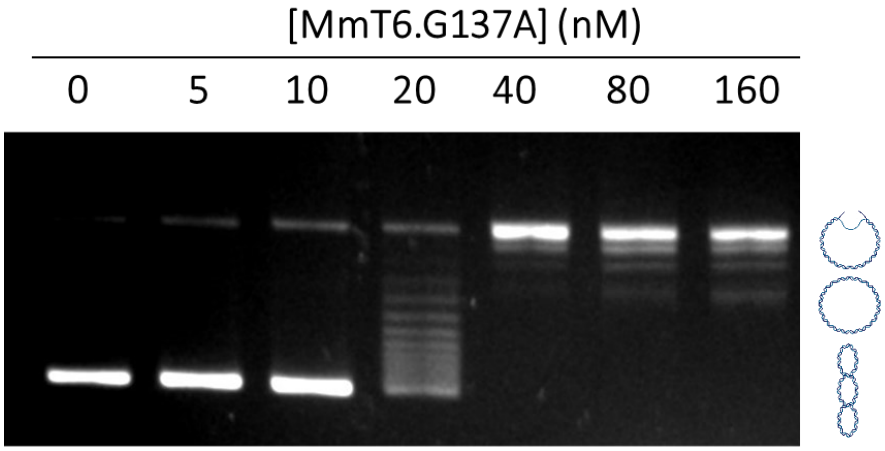

Figure 5.7| *Methanosarcina mazei* G137A mutant relaxation: *MmTopo VI.G137A relaxes 5 nM negatively-supercoiled DNA, producing a series of relaxed DNA topoisomers. Relaxation activity is observed at ~20 nM. Reactions were run at 37°C for 30 min.*

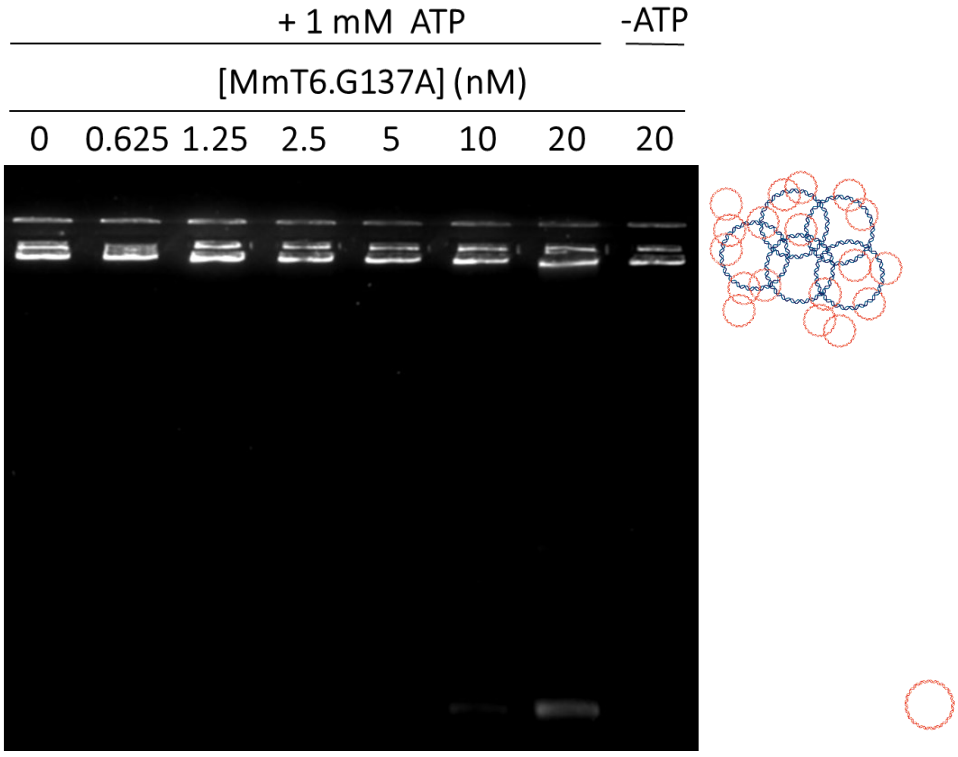

Figure 5.8| *Methanosarcina mazei* G137A mutant decatenation assay: *MmTopo VI.G137A can decatenate 6.7 ng µL^{-1} kDNA, which is too large to migrate out of the agarose wells, and release DNA mini-circles. Decatenation activity is observed at ~20 nM and is ATP-dependent. Reactions were run at 37°C for 30 min.*

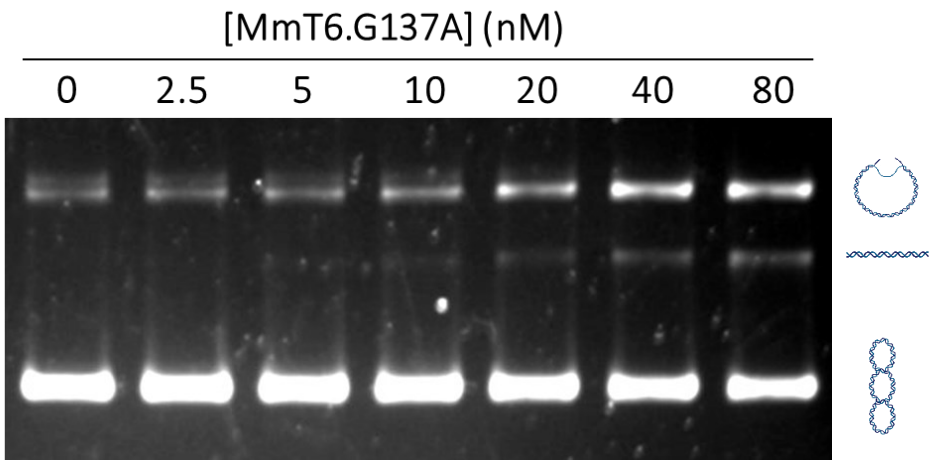

Figure 5.9| *Methanosarcina mazei* G137A mutant cleavage assay: *MmTopo VI can be converted into an enzyme capable of generating permanent double-stranded breaks by stabilising the enzyme-DNA cleavage complex with a non-hydrolysable ATP analogue, ADPNP. The G137A mutant has observable cleavage activity on 5 nM negatively-supercoiled DNA at ~10 nM. Reactions were run at 37°C for 1 hr.*

Next, I assessed the activities of the MmTopo VI.D294E mutant using an analogous series of experiments performed on the G137A mutant. First, I showed that MmTopo VI.D294E has observable relaxation activity at a concentration of 5 nM, with more complete relaxation activity at 10 nM (**Figure 5.10**). This suggests that, unlike G137A, the D294E mutation has no effect on the relaxation of -scDNA. However, like G137A, the D294E mutation also had no effect on cleavage, with a band of linear DNA observed at ~10 nM (**Figure 5.11**). Unexpectedly, MmTopo VI.D294E was shown to have observable decatenation activity on kDNA at a concentration of 625 pM (**Figure 5.12**), which represents an ~8-fold increase of activity compared to the wild-type enzyme. Due to the unusual nature of this result, I also tested the D294E mutant for activity on bis-cat DNA, which is a more direct measurement of decatenation. Using this assay, MmTopo VI was shown to have observable decatenation activity at 1.25 nM, relaxation activity at 2.5 nM, and a more complete relaxation activity at 5 nM (**Figure 5.13**), which represents an ~4-fold increase in decatenation activity compared to the WT enzyme. The ~2-fold increase in relaxation activity of MmTopo VI.D294E on bis-cat DNA compared to -sc pBR322 is likely to be irrelevant due to the plasmids having differing concentrations and superhelicities. The D294E mutation therefore induces an apparent enhancement of decatenation activity that could be attributed to a differential mechanism of strand-passage when resolving catenanes or supercoils.

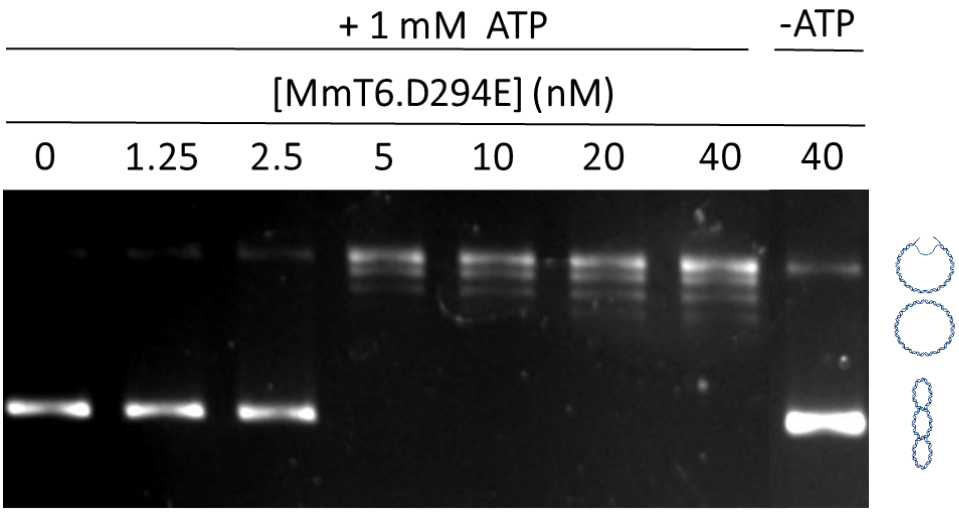

Figure 5.10| *Methanosarcina mazei* D294E mutant relaxation assay: *MmTopo VI.D294E relaxes 5 nM negatively-supercoiled DNA, producing a series of relaxed DNA topoisomers. Maximum relaxation activity is observed at ~5 nM and is ATP-dependent. Reactions were run at 37°C for 30 min.*

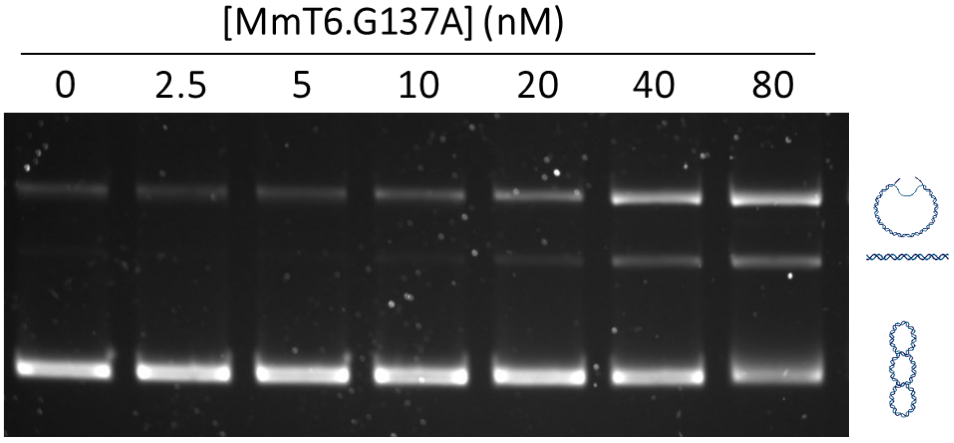

Figure 5.11| *Methanosarcina mazei* D294E mutant cleavage assay: *MmTopo VI can be converted into an enzyme capable of generating permanent double-stranded breaks by stabilising the enzyme-DNA cleavage complex with a non-hydrolysable ATP analogue, ADPNP. The D294E mutant has observable cleavage activity on 5 nM negatively-supercoiled DNA at ~10 nM. Reactions were run at 37°C for 1 hr.*

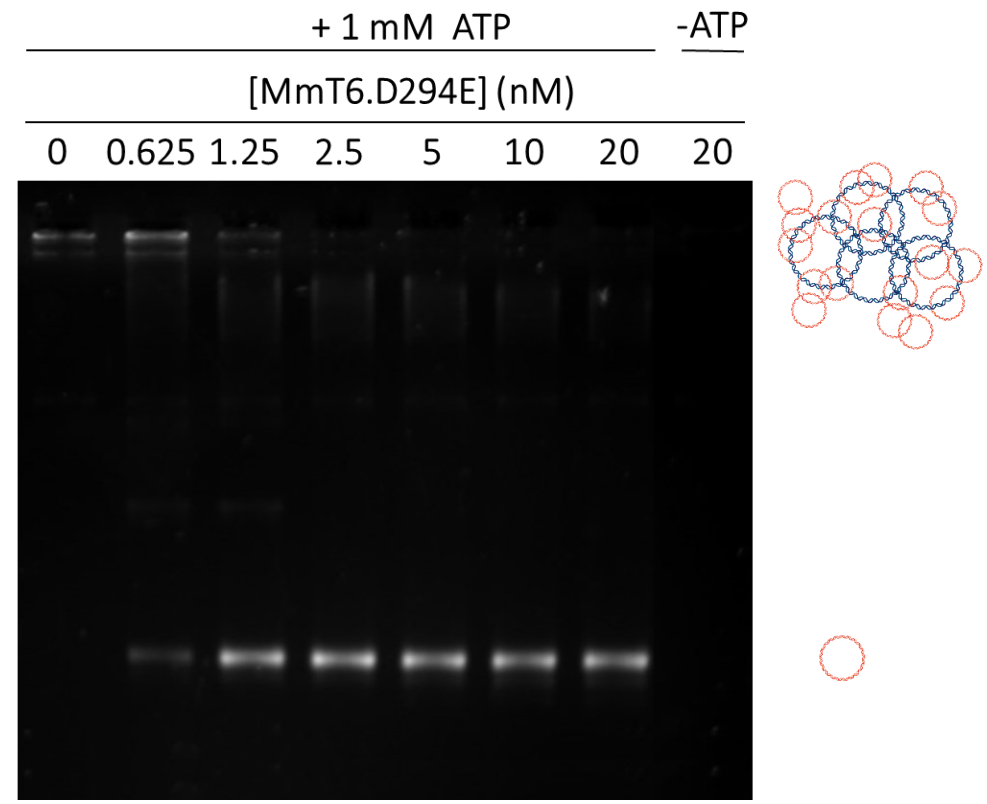

Figure 5.12| *Methanosarcina mazei* D294E mutant kDNA decatenation assay: *MmTopo VI.D294E can decatenate 6.7 ng µL^{-1} kDNA, which is too large to migrate out of the agarose wells, and release DNA mini-circles. Decatenation activity is observed at ~625 pM and is ATP-dependent. Reactions were run at 37°C for 30 min.*

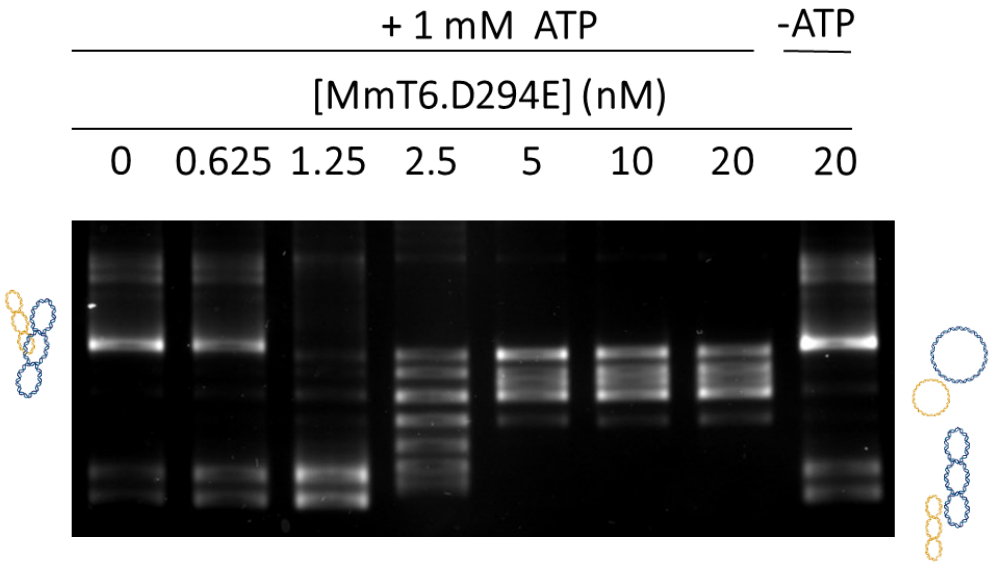

Figure 5.13| *Methanosarcina mazei* D294E mutant bis-cat DNA decatenation assay: *MmTopo VI.D294E can decatenate 1.3 ng µL^{-1} bis-cat DNA in an ATP-dependent manner at ~1.25 nM and relax the decatenated products at ~2.5 nM. Reactions were run at 37°C for 30 min.*

5.2.3 Mutagenic analysis of *Methanosarcina mazei* topo VI DNA-gate residues

A key difference between the type IIA and type IIB topos is the absence of a third dimer interface in the latter. Gyrase can introduce negative supercoils and relax positive supercoils by performing top-down strand-passage, in an analogous manner, in the presence of ATP[225]. An interesting study published in 1999 sought to investigate the role of the gyrase C-gate in determining the direction of strand-passage by cross-linking this interface[226]. A cross-linked C-gate was able to permit very limited strand-passage in the context of ATP-dependent supercoiling and relaxation, whereby a T-segment could enter through the N-gate, pass through a transient break in the G-segment, before getting stuck at the closed C-gate[226]. However, strand-passage was completely abolished in the context of the ATP-independent relaxation of -scDNA, suggesting that this reaction proceeds in a bottom-up direction that is halted at the first dimer interface that is the closed C-gate[226]. It has been proposed that a viable two-gated type II topo could also perform bottom-up strand-passage, whereby ATP-binding does not capture a T-segment but permits the passing of a T-segment through the transiently cleaved G-segment[86].

To test whether MmTopo VI can perform bottom-up strand-passage, I identified six highly conserved positively-charged residues at the bottom entrance of the DNA-gate (**Figure 5.14**) that may play a role in attracting a T-segment upwards. Two of these residues, R137 and R235, are absolutely conserved in topo VI-A sequences from archaea, bacteria, and eukaryotes, while the other residues, K172, K239, R240, and K286 sit in patches of conserved positively charged regions and are often substituted for alternative positive residues (**Figure 5.14**). MmTopo VI preferentially binds -scDNA over +scDNA, despite having greater relaxation activity on the latter[172], and I sought to test whether this discrepancy could be explained by the two substrates being resolved in opposite directions. In this model, strand-passage occurs top-down in the context of relaxing +scDNA and occurs bottom-up in the context of -scDNA, as is the case with gyrase. Therefore, although top-down strand-passage may the favoured route, the MmTopo VI DNA-gate may have a higher affinity for DNA than the ATP-gate. MmTopo VI is a preferential decatenase[172], which would suggest that strand-passage in the context of decatenation would proceed in the same direction as that of the relaxation of +scDNA. Given that the D294E mutant has a disproportionate effect on the decatenation and

-sc relaxation activities of the enzyme, this mutant has also been included in the test for strand-passage directionality.

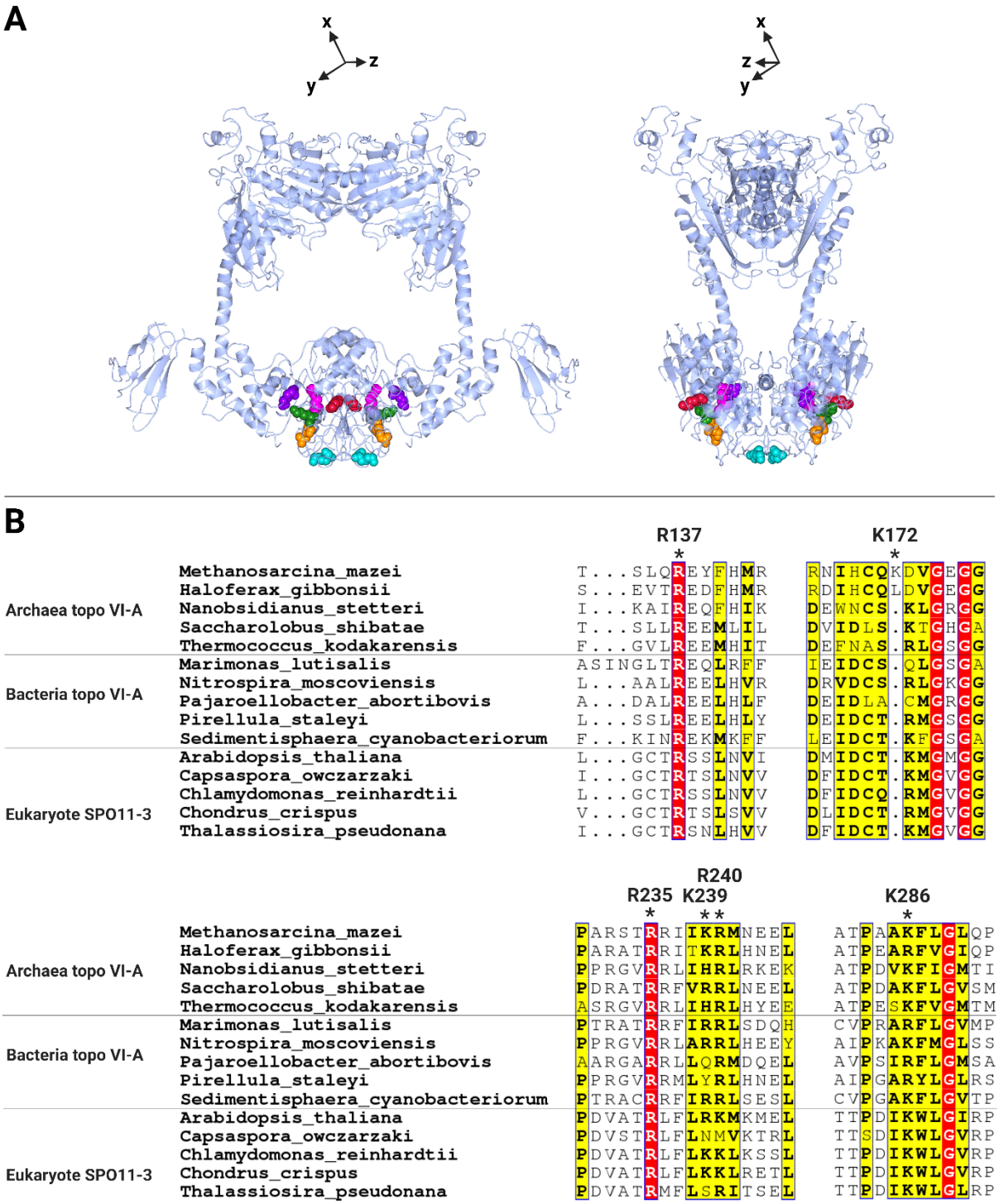

Figure 5.14| *Methanosarcina mazei* topo VI DNA-gate residues: (**A**) *Six positively charged residues were identified at the DNA-gate of MmTopo VI, namely R137 (purple), K172 (turquoise), R235 (magenta), K239 (green), R240 (orange), and K286 (red). (**B**) These DNA-gate residues are highly conserved in topo VI-A from archaea, bacteria, and eukaryotes.*

In Chapter 4, I expressed asparagine mutants of the six DNA-gate residues and obtained soluble preparations of the R137N, K172N, and K239N mutants. Asparagine was chosen to retain residue polarity while removing the positive charges to minimise disruption to the DNA-gate stability. To test whether these substitutions would disturb bottom-up strand-passage, while fully permitting top-down strand-passage, I assessed the ability of the DNA-gate mutants, as well as the D294E mutant, to influence a disproportionate relaxation activity on -scDNA vs +scDNA. A 2-fold enzyme titration in a 30-min reaction was not sufficient to distinguish between the relaxation activity of MmTopo VI on -scDNA and +scDNA (**Figure 5.15**). However, by performing a sequential enzyme titration with intervals of 2.5 nM at higher enzyme concentrations in a 15-min reaction, MmTopo VI was shown to have greater relaxation activity on +scDNA than -scDNA (**Figure 5.16)**. MmTopo VI had observable relaxation activity at 10 nM and 7.5 nM for -scDNA and +scDNA, respectively, while exhibiting more complete relaxation activity at 15 nM and 12.5 nM on -scDNA and +scDNA, respectively (**Figure 5.16**). This subtle enhancement of activity on +scDNA a gel-based assay has previously been demonstrated by lowering the reaction temperature from 37°C to 21°C and performing a timecourse assay[172].

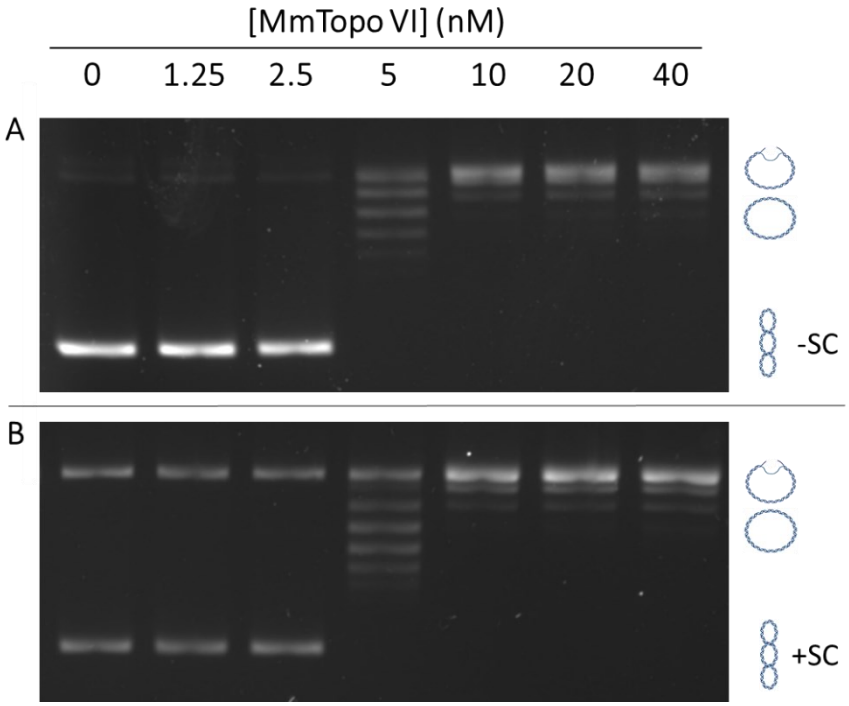

Figure 5.15| *Methanosarcina mazei* topo VI negative vs positive supercoil relaxation 30 min reaction: *MmTopo VI has observable relaxation activity at 5 nM and more complete relaxation at 10 nM on both (**A**) 5 nM negatively-supercoiled (-SC) DNA and (**B**) 5 nM positively-supercoiled (+SC) DNA. Reactions were run at 37°C for 30 min.*

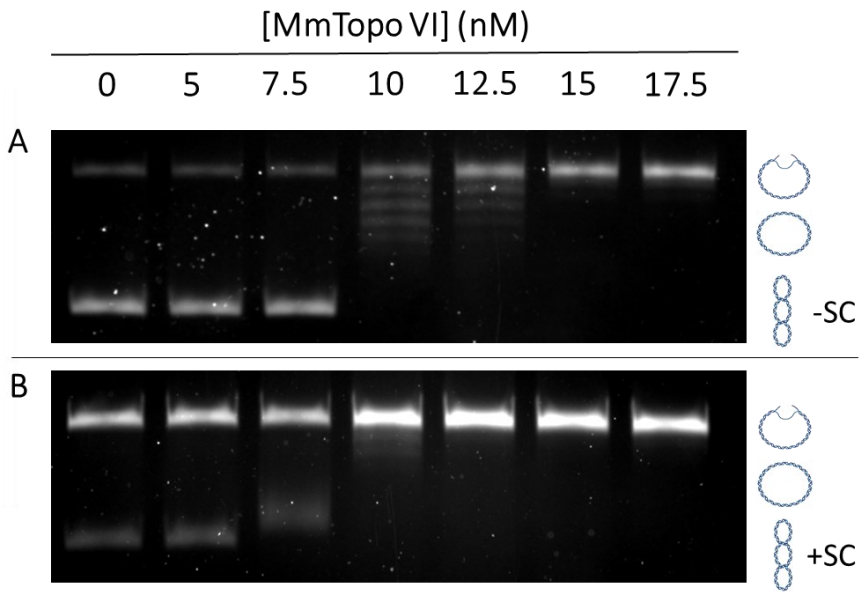

Figure 5.16| *Methanosarcina mazei* topo VI negative vs positive supercoil relaxation 15 min reaction: (**A**) *On 5 nM negatively-supercoiled DNA, MmTopo VI has observable relaxation activity at 10 nM, and more complete relaxation at 15 nM. (**B**) On 5 nM positively-supercoiled DNA, MmTopo VI has observable relaxation activity at 7.5 nM, and more complete relaxation at 12.5 nM. Reactions were run at 37°C for 15 min.*

In the 15-min relaxation assay, the K239N and D294E mutants resolved -scDNA and +scDNA in a comparable manner to the WT enzyme. (**Figures 5.17** and **5.18**), whereas the R137N and K172N mutants both showed a decrease in activity on -scDNA and +scDNA (**Figures 5.19** and **5.20**). R137N exhibited relaxation activity at 12.5 nM and 10 nM for -scDNA and +scDNA, respectively, and K172N exhibited relaxation activity at 15 nM and 12.5 nM for -scDNA and +scDNA, respectively (**Figures 5.19** and **5.20**). Given that R137N, K172N, and the WT enzyme all require a 2.5 nM increase in concentration to exhibit observable relaxation activity on -scDNA than on +scDNA in the 15-min reaction assay, and that R137N and K172N are less active on both -scDNA and +scDNA, then both mutants seem to have a reduction in the enhancement of activity on +scDNA over -scDNA. This result suggests that these DNA-gate mutants have disproportionately affected +scDNA over -scDNA which is the opposite result to what was expected. However, this effect is minimal and could be due to a disproportionate effect on the DNA-binding affinities of -scDNA and +scDNA to the same gate. Overall, none of the DNA-gate mutants, or the D294E mutant, exhibited an exclusive reduction of relaxation activity on either -scDNA or +scDNA, and so it remains inconclusive whether supercoil chirality determines the direction of strand-passage in MmTopo VI.

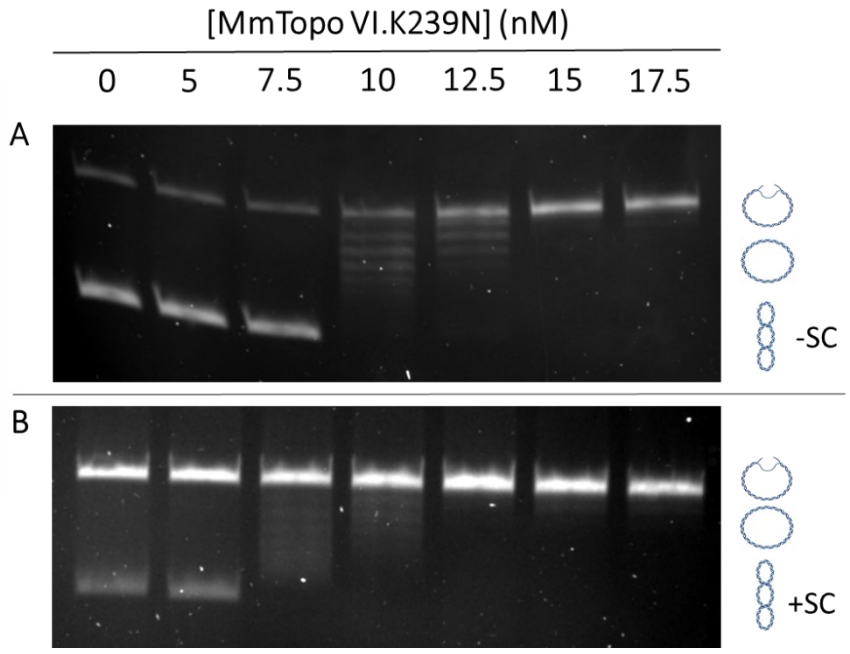

Figure 5.17| *Methanosarcina mazei* topo VI K239N mutant negative vs positive supercoil relaxation: (**A**) *On 5 nM negatively-supercoiled DNA, MmTopo VI.K239N has observable relaxation activity at 10 nM, and more complete relaxation at 15 nM. (**B**) On 5 nM positively-supercoiled DNA, MmTopo VI.K239N has observable relaxation activity at 7.5 nM, and more complete relaxation at 12.5 nM. Reactions were run at 37°C for 15 min.*

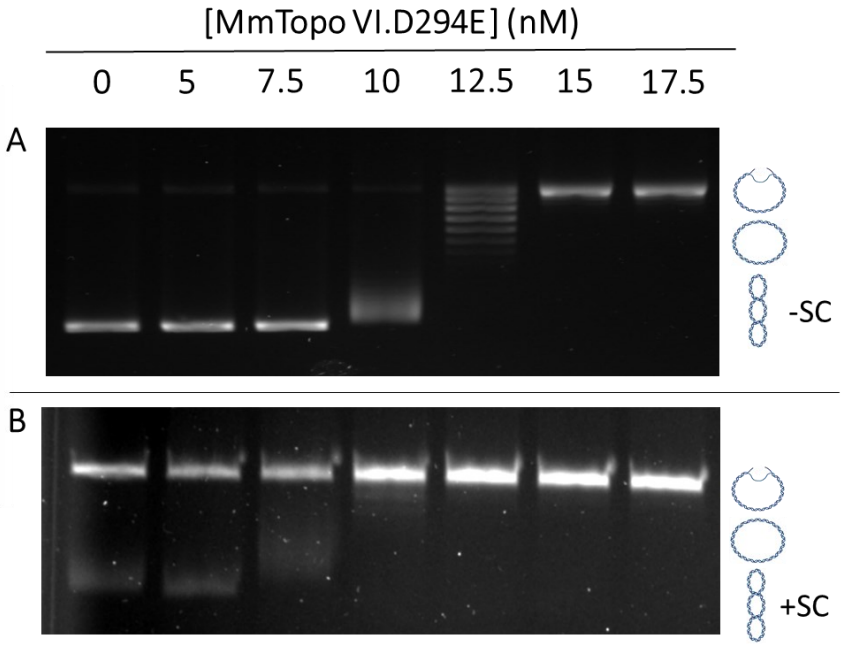

Figure 5.18| *Methanosarcina mazei* topo VI D294E mutant negative vs positive supercoil relaxation: (**A**) *On 5 nM negatively-supercoiled DNA, MmTopo VI. D294E has observable relaxation activity at 10 nM, and more complete relaxation at 15 nM. (**B**) On 5 nM positively-supercoiled DNA, MmTopo VI. D294E has observable relaxation activity at 7.5 nM, and more complete relaxation at 12.5 nM. Reactions were run at 37°C for 15 min.*

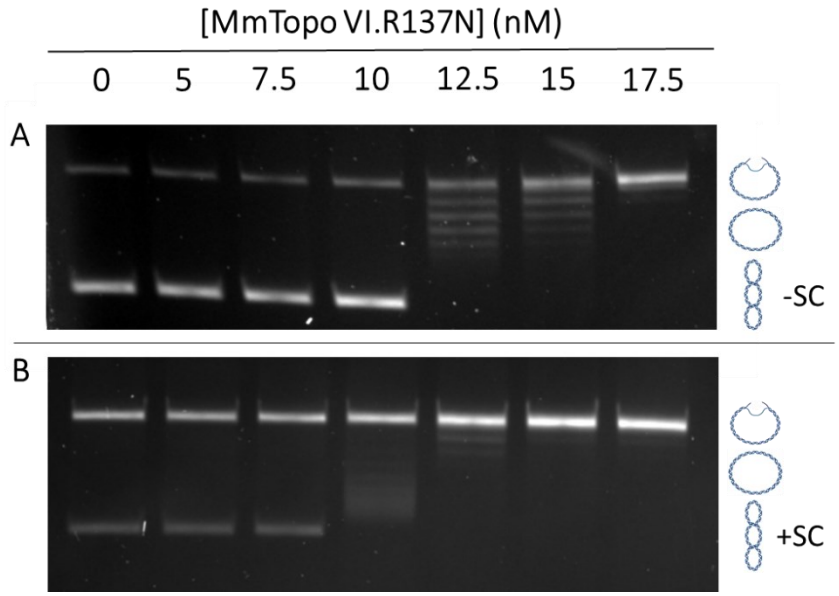

Figure 5.19| *Methanosarcina mazei* topo VI R137N mutant negative vs positive supercoil relaxation: (**A**) *On 5 nM negatively-supercoiled DNA, MmTopo VI.R137N has observable relaxation activity at 12.5 nM, and more complete relaxation at 17.5 nM. (**B**) On 5 nM positively-supercoiled DNA, MmTopo VI.R137N has observable relaxation activity at 10 nM, and more complete relaxation at 15 nM. Reactions were run at 37°C for 15 min.*

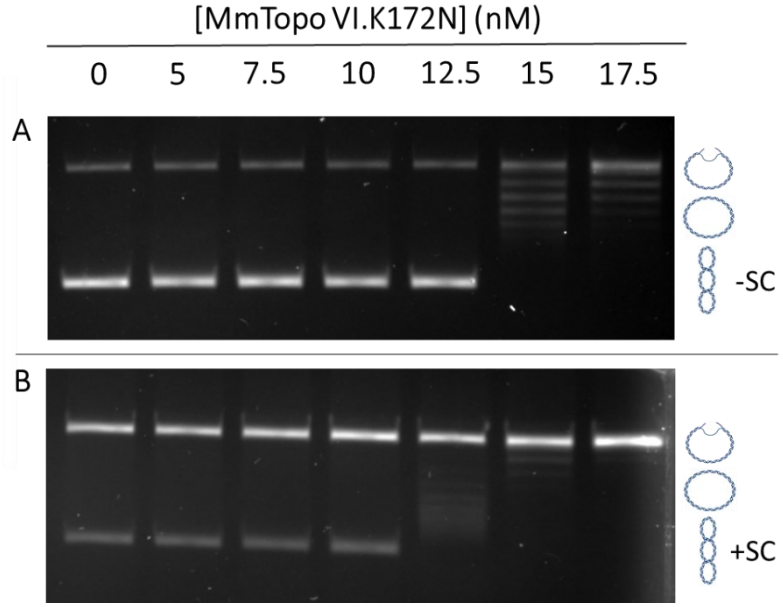

Figure 5.20| *Methanosarcina mazei* topo VI K172N mutant negative vs positive supercoil relaxation: (**A**) *On 5 nM negatively-supercoiled DNA, MmTopo VI.K172N has observable relaxation activity at 15 nM. (**B**) On 5 nM positively-supercoiled DNA, MmTopo VI.K172N has observable relaxation activity at 12.5 nM, and more complete relaxation at 17.5 nM. Reactions were run at 37°C for 15 min.*

5.2.4 The ATPase activity of *Methanosarcina mazei* topo VI is DNA-dependent and chirally selective

MmTopo VI preferentially binds -scDNA over +scDNA, despite relaxing +scDNA ~2-3 fold faster[172]. The PK/LDH coupled assay has shown that the ATPase activity of MmTopo VI is DNA-dependent and ~5-fold greater in the presence of -scDNA than linear DNA, with a maximum rate of ~3 ATP min^{-1} [58]. A dramatic stimulation of MmTopo VI ATPase activity in the presence of -scDNA has also been observed using a radioactive ATPase assay, which determined a maximum rate of ~5 ATP min^{-1} [172]. It is also possible to calculate the rate of ATP hydrolysis from the rate of strand-passage events by MmTopo VI, given that they are tightly coupled[58]. MmTopo VI requires 2 ATP molecules to complete a single round of strand-passage and prepare for a second round[70], and magnetic tweezer experiments have shown that MmTopo VI performs ~1.5-3.5 strand-passage events min^{-1} on -scDNA[172] which would therefore equate to an ATPase rate of ~3-7 ATP min^{-1}. These rates are ~10-20-fold slower than the rates measured for type IIA topos, with *E. coli* gyrase and bovine topo II being shown to turnover 1 ATP s^{-1} in the presence of -scDNA[227,228]. However, even in the absence of DNA these enzymes could still maintain rates of ~3-18 ATP min^{-1} [227,228], which demonstrates just how inefficient a supercoil relaxase MmTopo VI is, and has led to discussion that the enzyme is unable to match the metabolic demands of the cell[58]. Magnetic tweezer experiments have also shown that MmTopo VI performs ~19 strand-passage events min^{-1} on a catenated DNA substrate[172], which would equate to an ATPase rate of ~0.6 ATP s^{-1} bringing the rate closer in line to that of the type IIA topos.

The PK/LDH assay is a useful method that couples the hydrolysis of ATP to the oxidation of NADH (**Figure 5.21**), which absorbs much stronger at 340 nm (A_{340}) than NAD^+. In this assay, an ATPase enzyme converts ATP to ADP, which is then used by pyruvate kinase to convert phosphoenolpyruvate (PEP) to pyruvate and regenerate ATP. Therefore, the rate of ATP hydrolysis in this assay can be measured continuously. The reduction of pyruvate to lactate is then catalysed by lactate dehydrogenase which does so by oxidising NADH, and this process can be measured by a decrease in the spectroscopic absorbance at 340 nm. The loss of NADH is equimolar to the hydrolysis of ATP, and so the Beer-Lambert Law (**Equation 5.1**), can be used to determine the rate of ATP hydrolysis from the change in A_{340}.

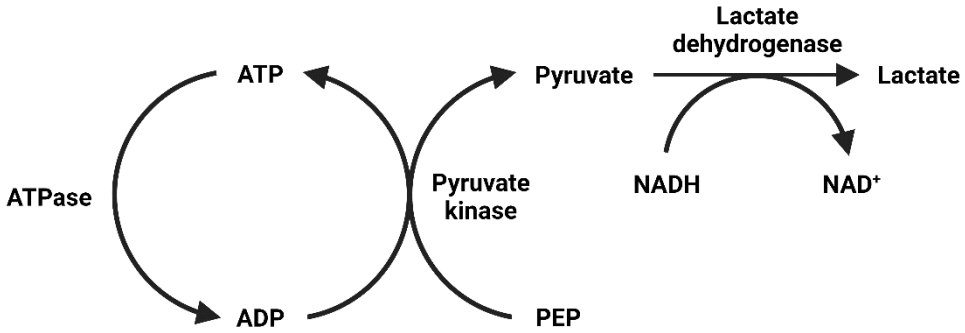

Figure 5.21| Regeneration of ATP in the PK/LDH assay: *An ATPase converts ATP to ADP, and pyruvate kinase uses ADP to convert phosphoenolpyruvate (PEP) to pyruvate and ATP. Lactate dehydrogenase then converts pyruvate to lactate by oxidising NADH.*

$$A = \varepsilon c l$$

Equation 5.1| The Beer-Lambert law: *There is a linear relationship between the absorbance (A), molar absorption coefficient (ε), molar concentration (c) and optical path length (l).*

The ATPase activity of MmTopo VI on a +scDNA substrate has yet to be tested, and so I sought to conduct this experiment using the PK/LDH assay. First, I attempted to establish a control by measuring the rate of ATP hydrolysis by *E. coli* gyrase, for which the rate of ATP hydrolysis is known. I therefore prepared a sample that possessed gyrase and linear DNA; a sample devoid of enzyme; a sample that possessed novobiocin, and a sample devoid of NADH, to measure the background A_{340}. In these samples, ATP was added to the reactions after 10 min, but in a fourth sample that was devoid of enzyme or substrate, ADP was added instead of ATP to confirm the cooperativity between the PK and LDH enzymes. The NADH and ADP controls worked as expected, as seen by a horizontal line and a rapid drop of A_{340}, respectively, and once ATP was added, the gyrase and no enzyme samples exhibited a linear decline of A_{340} (**Figure 5.22**). By subtracting the function of the no-enzyme sample from the enzyme sample, and using the Beer-Lambert law to calculate the concentration change of ATP, it was possible to calculate a rate of ATP hydrolysis of 1.04 sec^{-1} for gyrase, which has previously been described using a radioactive ATPase assay[227]. In the presence of novobiocin, which is a competitive inhibitor of ATP hydrolysis[229,230], the rate of A_{340} decline was comparable to that of the no enzyme control, suggesting that the compound was indeed inhibiting ATP hydrolysis and validating the PK/LDH assay.

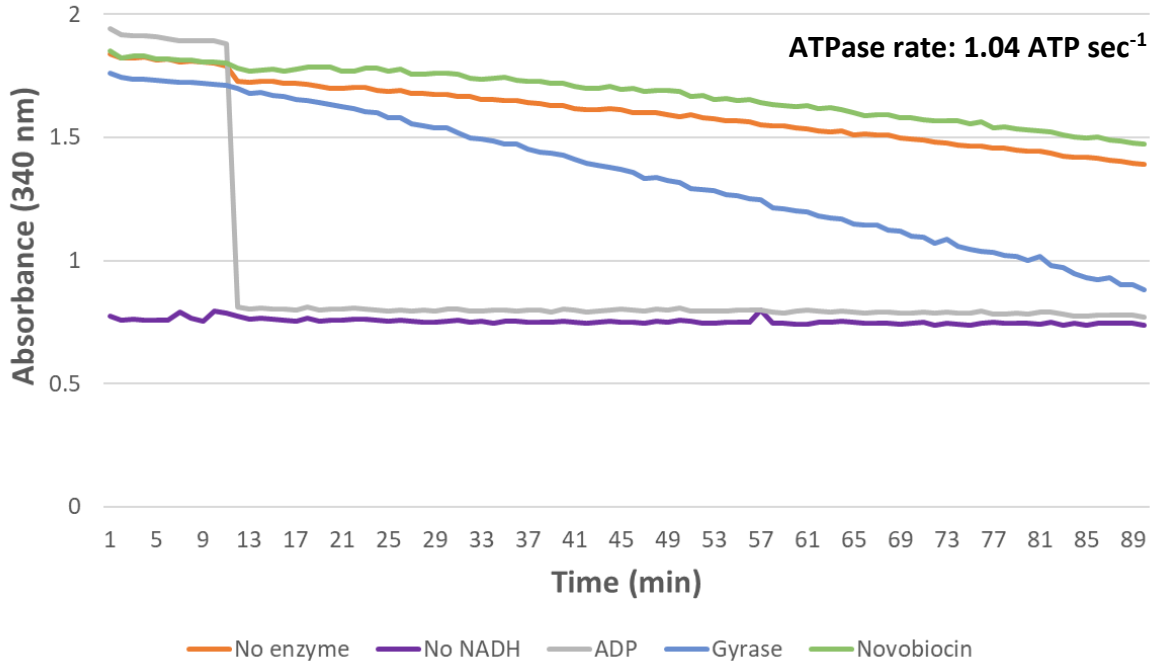

Figure 5.22| Gyrase ATPase activity: *The ATPase activity of **E. coli** gyrase in the presence of linear DNA was measured using the PK/LDH coupled assay at 37°C. Control samples were run in the absence of enzyme, absence of NADH, and replacing ATP with ADP. ATP was added to each sample, except for the ADP control, after 10 min and absorbance at 340 nm was measured every min for 90 min.*

Once a working protocol for the PK/LDH assay had been developed, I shifted my attention to investigating the ATPase activity of MmTopo VI. Previous attempts to perform the PK/LDH assay on MmTopo VI required very high concentrations of DNA, owing to the tight coupling of ATPase activity to the binding of DNA crossovers. These attempts used DNA concentrations ranging from 136 nM to 181 nM, with a DNA:enzyme ratio as high as 1:4[58,202]. I therefore performed analogous attempts using a DNA concentration of 177 nM, and a DNA:enzyme ratio of 1:6, to ensure that the enzyme is in excess and that the maximum rate of ATP hydrolysis is measured. Under these conditions, there was no difference between the reduction of A_{340} in the presence of no enzyme, or in the presence of MmTopo VI with no DNA (**Figure 5.23**). This suggests that MmTopo VI exhibits negligible DNA-independent ATPase activity, as has been previously reported[58].

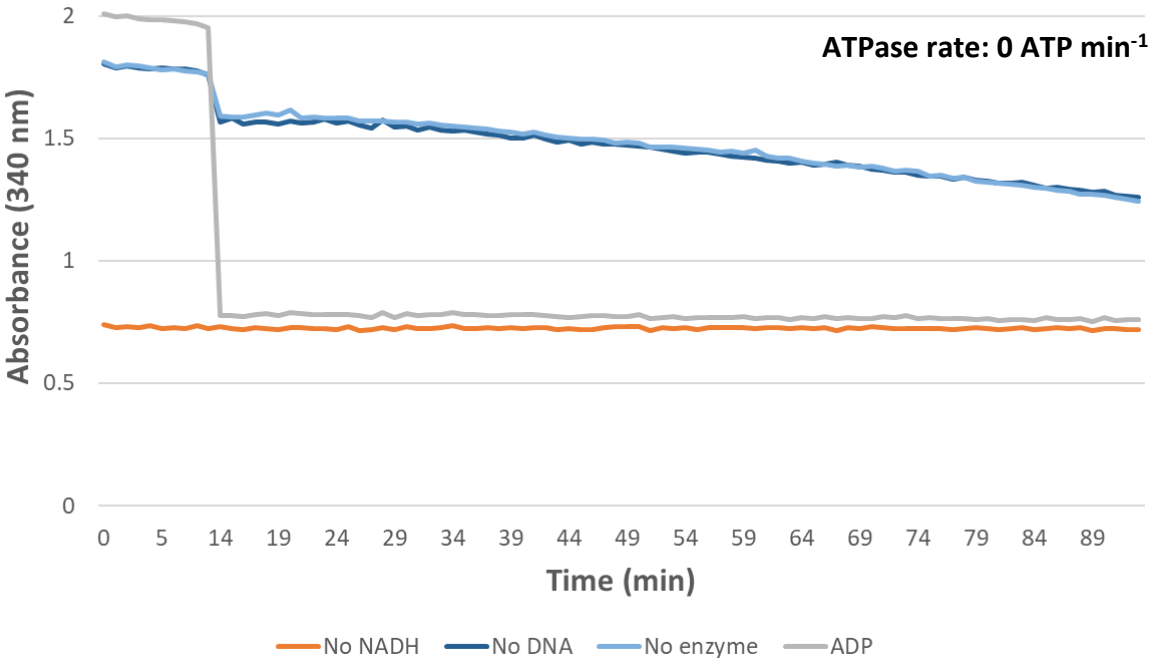

Figure 5.23| *Methanosarcina mazei* topo VI DNA-independent ATPase activity: *The ATPase activity of 1.2 µM MmTopo VI in the absence of DNA was measured using the PK/LDH coupled assay at 37°C. Control samples were run in the absence of enzyme, absence of NADH, and replacing ATP with ADP. ATP was added to each sample, except for the ADP control, after 10 min and absorbance at 340 nm was measured every min for 90 min.*

Next, I tested the DNA-dependent ATPase activity of MmTopo VI in triplicate in the presence of linear DNA, relaxed DNA, -scDNA, and +scDNA. In the presence of all four species, MmTopo VI exhibited a greater rate of A_{340} decline than in the absence of DNA (**Figures 5.24**-**5.27**), confirming that the ATPase activity of MmTopo VI is DNA-dependent. The rate of ATP hydrolysis by MmTopo VI in the presence of linear DNA and relaxed DNA was determined to be 1.55±0.41 ATP min^{-1} and 1.37±0.34 ATP min^{-1}, respectively (**Figure 5.28**), exhibiting a similar level of stimulation by the two DNA species that has previously been shown using a radioactive ATPase assay[172]. The rate of ATPase activity in the presence of linear DNA has previously been determined at ~0.6 ATP min^{-1} using the PK/LDH assay[58], which is line with the rate calculated here. In the presence of -scDNA, the rate of ATP hydrolysis by MmTopo VI was determined at 2.48±0.31 ATP min^{-1} (**Figure 5.28**), which is similar to the value of 3 ATP min^{-1} previously determined by the PK/LDH assay[58], but 2-fold lower than the value of ~5 ATP min^{-1} determined by the radioactive ATPase assay[172]. Furthermore, a less than 2-fold enhancement of ATPase activity in the presence -scDNA compared to linear DNA is calculated

here, which is far less than the 5-fold enhancement previously determined using the same assay[58]. The rate of 2.48 ATP min^{-1} in the presence of -scDNA is also less than the predicted rate of 3-7 ATP min^{-1} determined by molecular tweezer experiments[172]. The rate of ATP hydrolysis by MmTopo VI in the presence of +scDNA was determined to be 4.98±0.692, which is ~2-fold greater than the rate in the presence of -scDNA. Magnetic tweezer experiments have shown that MmTopo VI performs ~2.5-6.5 strand-passage events min^{-1} on +scDNA[172], which equates to an ATPase rate of ~5-13 ATP min^{-1} that is in line with the value calculated here. Overall, the results presented here have confirmed the ATPase activity of MmTopo VI is DNA-dependent and stimulated by superhelicity. Furthermore, I have demonstrated that the chiral-specific activity of MmTopo VI is reflected by an enhancement of ATPase activity in the presence of positive supercoils than with negative supercoils.

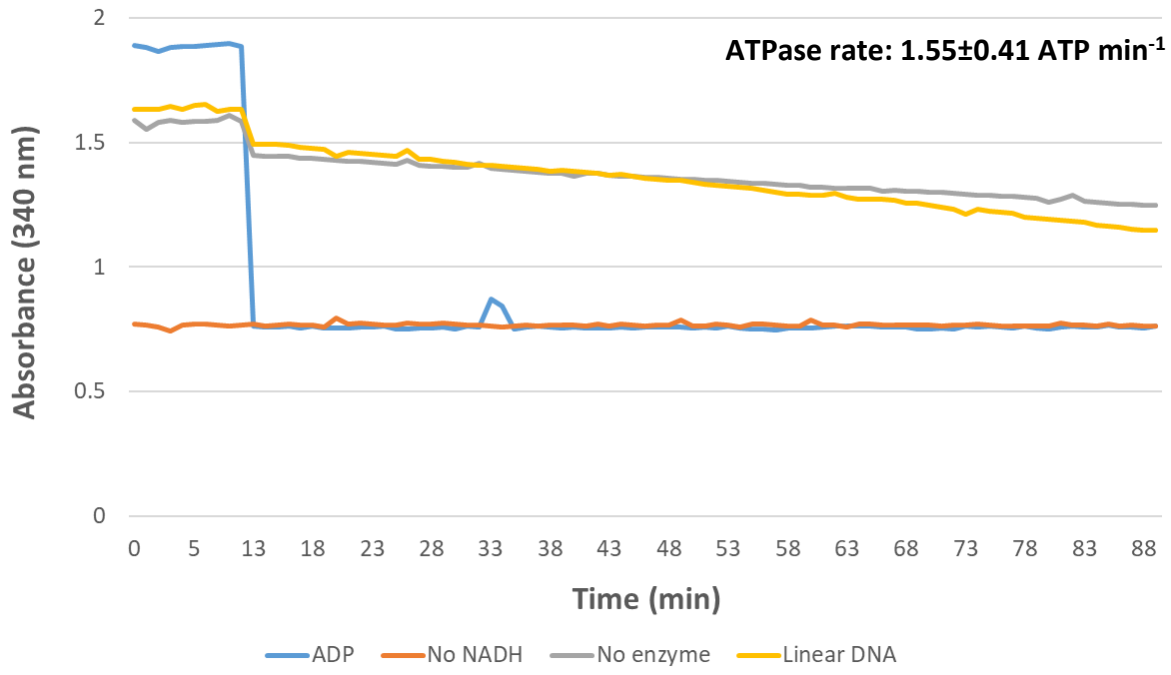

Figure 5.24| *Methanosarcina mazei* topo VI ATPase activity on linear DNA: *The ATPase activity of 1.2 µM MmTopo VI in the presence of 177 nM linear DNA was measured using the PK/LDH coupled assay at 37°C. Control samples were run in the absence of enzyme, absence of NADH, and replacing ATP with ADP. ATP was added to each sample, except for the ADP control, after 10 min and absorbance at 340 nm was measured every min for 90 min.*

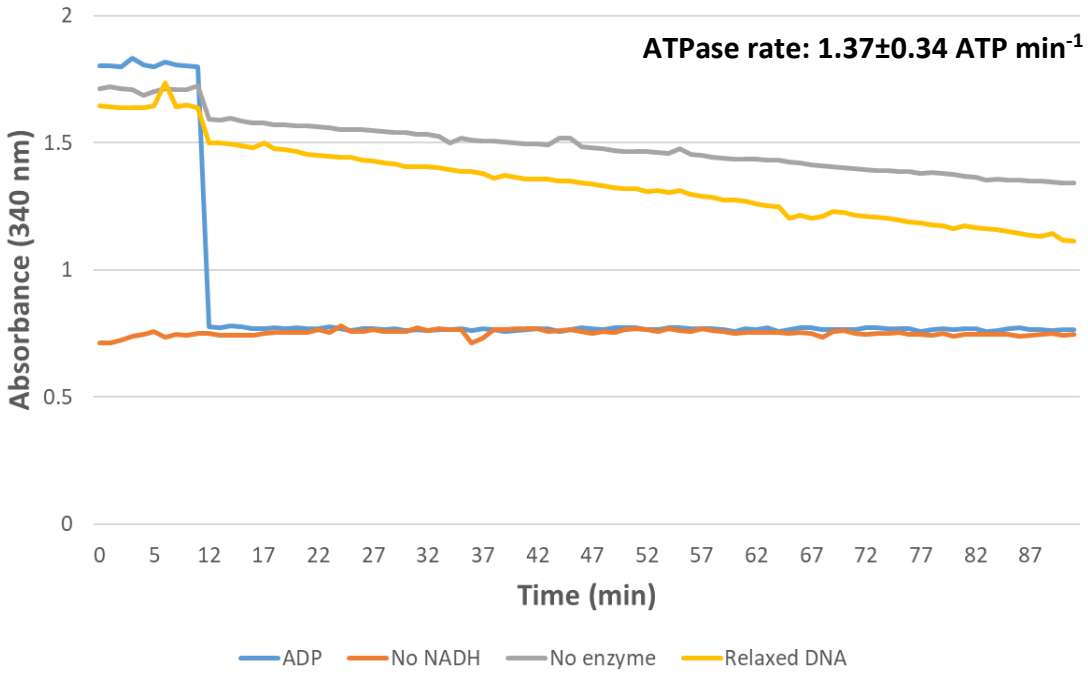

Figure 5.25| *Methanosarcina mazei* topo VI ATPase activity on relaxed DNA: *The ATPase activity of 1.2 µM MmTopo VI in the presence of 177 nM relaxed DNA was measured using the PK/LDH coupled assay at 37°C. Control samples were run in the absence of enzyme, absence of NADH, and replacing ATP with ADP. ATP was added to each sample, except for the ADP control, after 10 min and absorbance at 340 nm was measured every min for 90 min.*

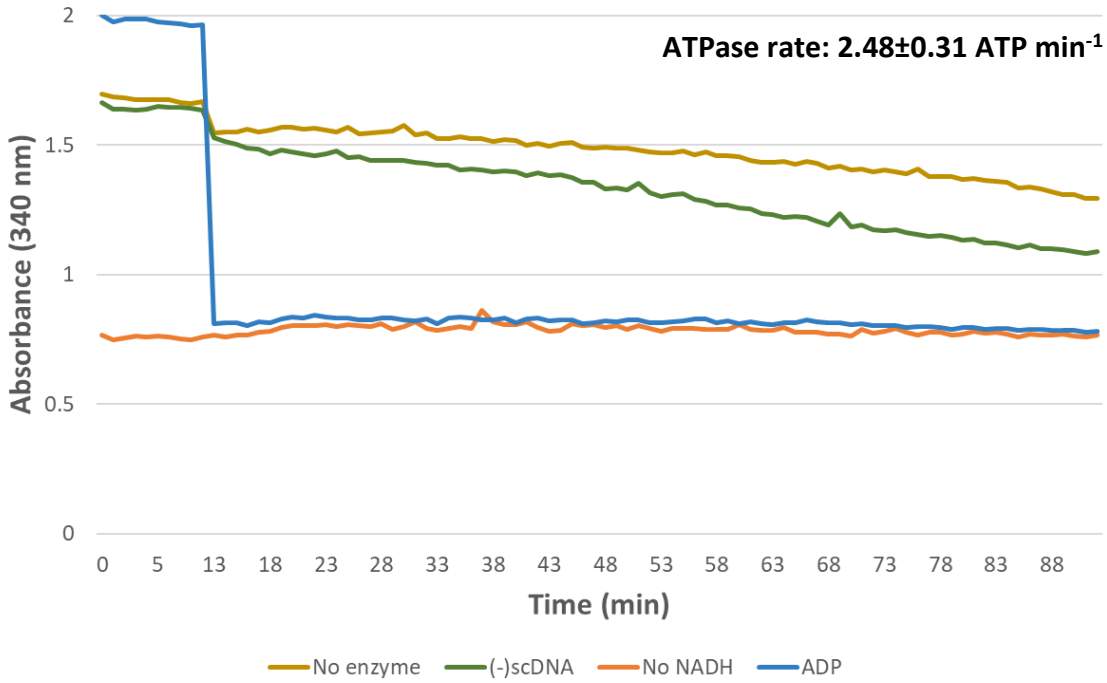

Figure 5.26| *Methanosarcina mazei* topo VI ATPase activity on negatively-supercoiled DNA: *The ATPase activity of 1.2 µM MmTopo VI in the presence of 177 nM negatively-supercoiled DNA was measured using the PK/LDH coupled assay at 37°C. Control samples were run in the absence of enzyme, absence of NADH, and replacing ATP with ADP. ATP was added to each sample, except for the ADP control, after 10 min and absorbance at 340 nm was measured every min for 90 min.*

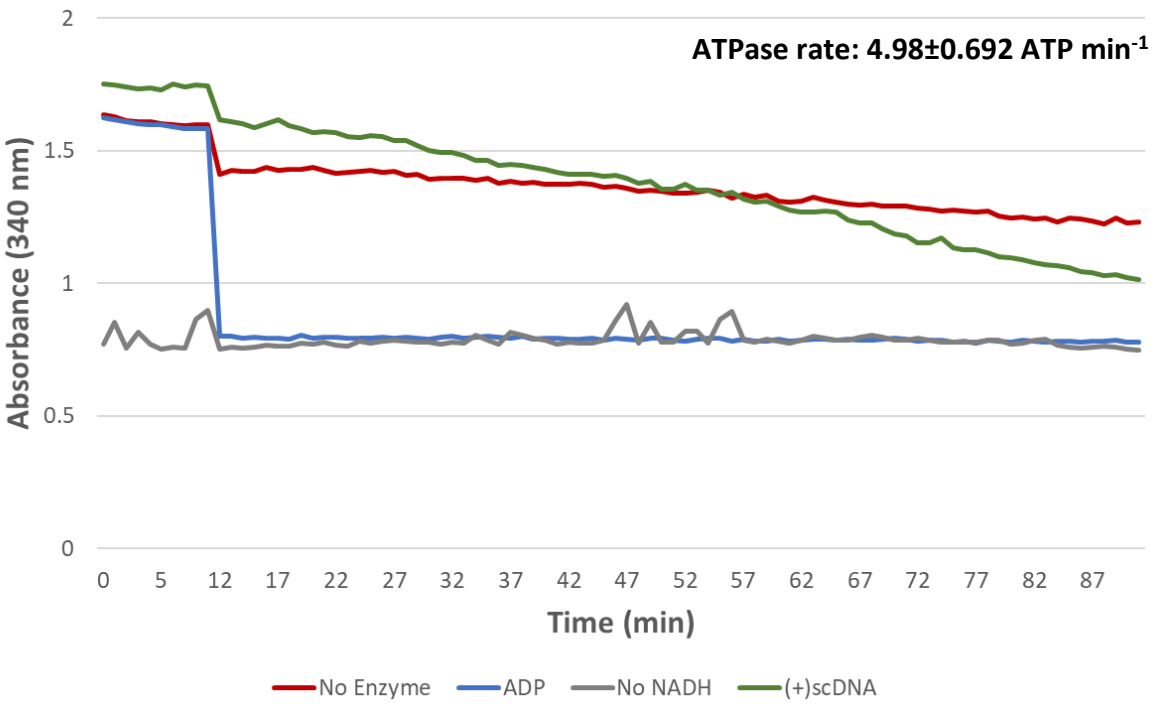

Figure 5.27| *Methanosarcina mazei* topo VI ATPase activity on positively-supercoiled DNA: *The ATPase activity of 1.2 µM MmTopo VI in the presence of 177 nM positively-supercoiled DNA was measured using the PK/LDH coupled assay at 37°C. Control samples were run in the absence of enzyme, absence of NADH, and replacing ATP with ADP. ATP was added to each sample, except for the ADP control, after 10 min and absorbance at 340 nm was measured every min for 90 min.*

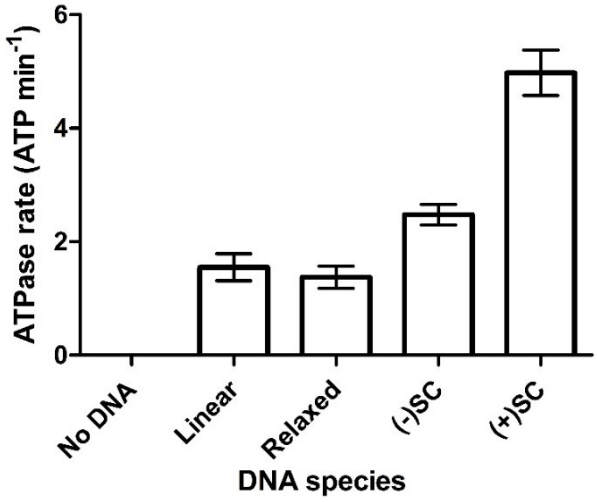

Figure 5.28| *Methanosarcina mazei* topo VI DNA-dependent ATPase activity: *The mean rate of ATP hydrolysis in the presence of different topological species of DNA was calculated from PK/LDH assays performed in triplicate. Error bars represent standard deviation*

Using a radioactive ATPase assay, a previous study had shown that not only is the ATPase activity of MmTopo VI greater in the presence of -scDNA than relaxed DNA, but that this activity decreased as the sc substrate was relaxed[172]. Furthermore, by performing a gel-based relaxation timecourse assay under analogous conditions, it was shown that this decrease in the rate of ATP hydrolysis coincided with the conversion of -scDNA to relaxed DNA at the same timepoints[172]. In the ATPase assays performed in this chapter, however, the decrease in A_{340} was linear for both -scDNA and +scDNA for 90 min, which would suggest that these substrates are in excess and remain in their sc states. However, timecourse relaxation assays performed under analogous conditions show that in the case of both -scDNA and +scDNA the substrate is converted to a maximum relaxed state after 30 min (**Figure 5.29**). Given that a ~4-fold reduction in ATPase activity is observed in the presence of relaxed DNA than with +scDNA, a decrease in the rate of ATP hydrolysis would be expected after 30 min in the PK/LDH assay, instead of the constant rate observed here. These results suggest that, in contradiction to previous studies[58,172], ATP hydrolysis has become uncoupled from strand-passage, and that there is a feature of both -scDNA and +scDNA independent of superhelicity that is stimulating ATP hydrolysis by MmTopo VI.

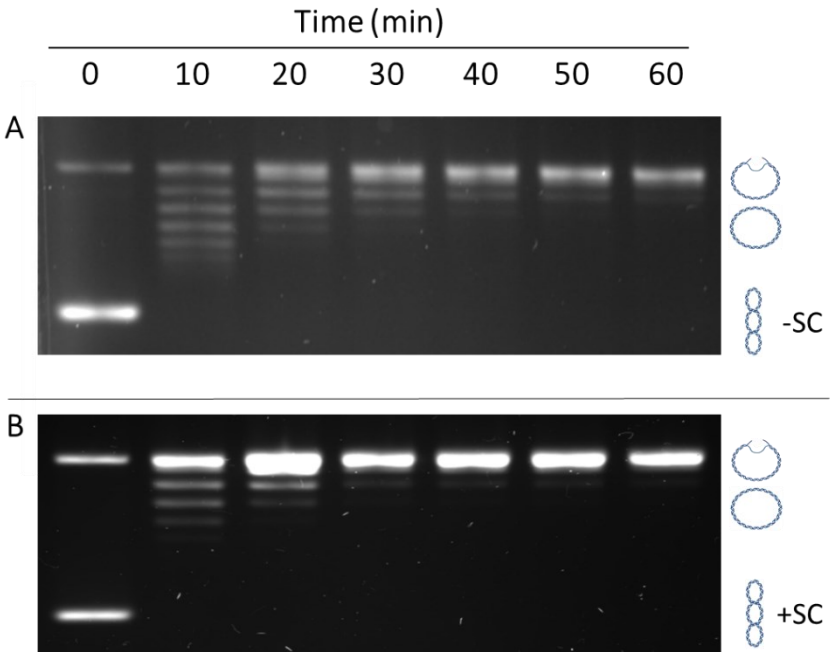

Figure 5.29| Gel-based *Methanosarcina mazei* topo VI relaxation timecourse analogous to PK/LDH assay conditions: *Under analogous conditions of the PK/LDH coupled ATPase assay, 1.2 µM MmTopo VI relaxed both* (**A**) *177 nM negatively-supercoiled DNA and* (***B***) *177 nM positively-supercoiled DNA within 10 min. Reactions were run at 37°C for 1 hr.*

5.2.5 ATPase activity of *Methanosarcina mazei* topo VI-B truncates and the G137A mutant

In section **5.2.4**, I showed that MmTopo VI has a maximum rate of ATP hydrolysis that is ~12-fold lower than the rate previously reported for *E. coli* gyrase and bovine topo II. This low activity seems to be caused by the tight coupling of ATP hydrolysis with strand-passage in MmTopo VI. A series of residues have been identified in MmTopo VI that contribute to this coupling, and mutating these have increased the maximum rate of ATP hydrolysis by the enzyme[58]. Therefore, I sought to test whether a total decoupling of ATP hydrolysis and strand-passage would permit observable ATPase activity. In Chapter 5, I isolated two fragments of MmTopo VI-B: MmTopo VIB-GHKL and MmTopo VIB-trans. The MmTopo VIB-GHKL fragment represents the GHKL domain in isolation, and the MmTopo VIB-trans fragment is a truncated version of MmTopo VI-B that ends at the terminus of its long transducing α-helix (**Figure 5.30**). Both fragments are missing the C-terminal domain (CTD) and removing this module from MmTopo VI has previously been shown to impart a ~5-fold reduction in the enzyme's decatenation and +sc and -sc relaxation activities[70]. However, given its position away from the GHKL domain, it is unlikely that the CTD contributes to ATP hydrolysis. A fragment of mycobacterial gyrase that possesses the GHKL and transducer domains of GyrB in isolation has previously been shown to exhibit ATPase activity, albeit at a rate ~30 times lower than with the full-length enzyme[231]. Therefore, I conducted the PK/LDH assay using the MmTopo VI-B fragments at a concentration of 12 μM, instead of the concentration of 1.2 μM used for the WT enzyme, while retaining a DNA concentration of 177 nM. In the presence of +scDNA, which is the preferred substrate for ATP hydrolysis in WT MmTopo VI, neither MmTopo VIB-GHKL nor MmTopo VIB-trans exhibited an enhanced rate of A_{340} decrease compared to the no-enzyme control (**Figure 5.31**), which suggests that these fragments exhibit negligible ATPase activity. This result also suggests that MmTopo VI-A is required for ATP hydrolysis, although this should be corroborated by testing the ATPase activity of the full-length MmTopo VI-B subunit.

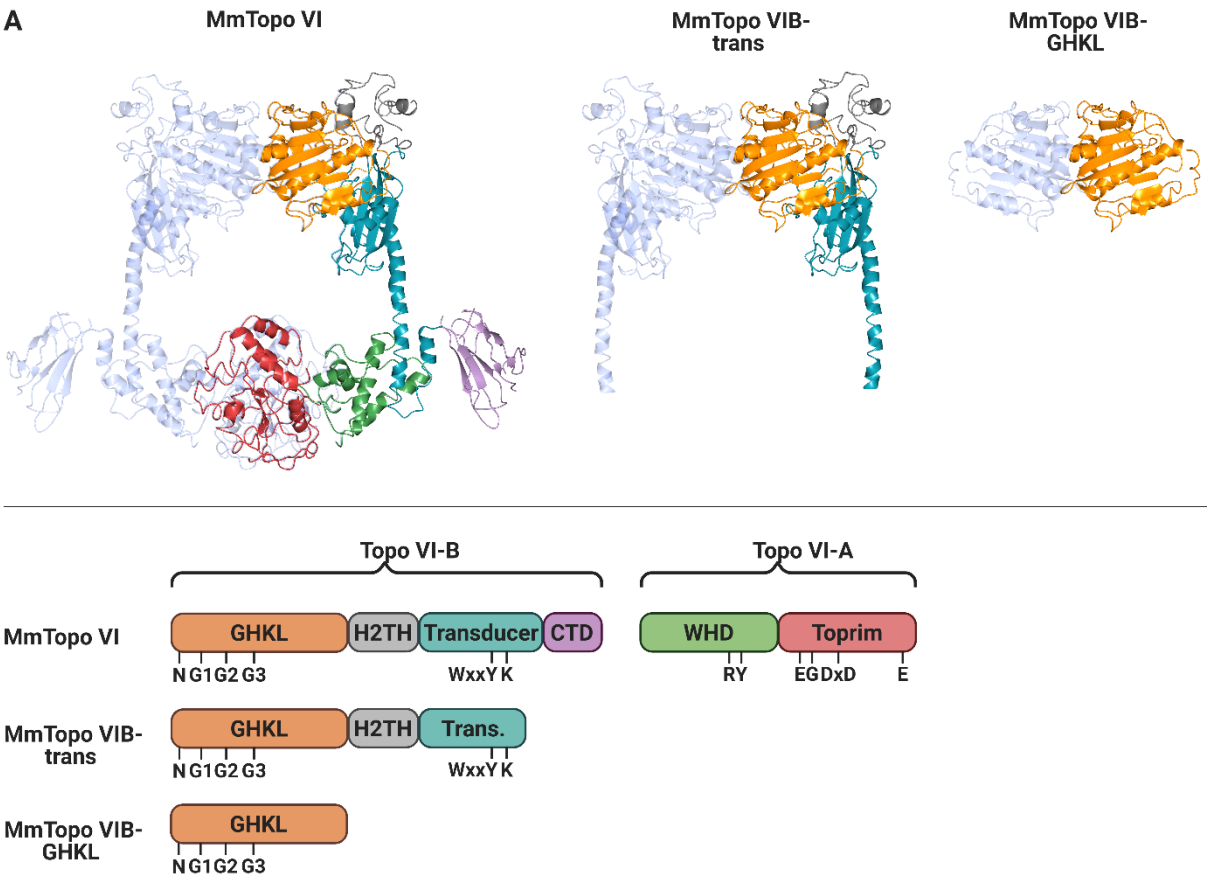

Figure 5.30| *Methanosarcina mazei* topo VI-B ATPase fragments: *A series of ATPase constructs will probe the DNA-dependent ATPase activity of* ***Methanosarcina mazei*** *topo VI.*

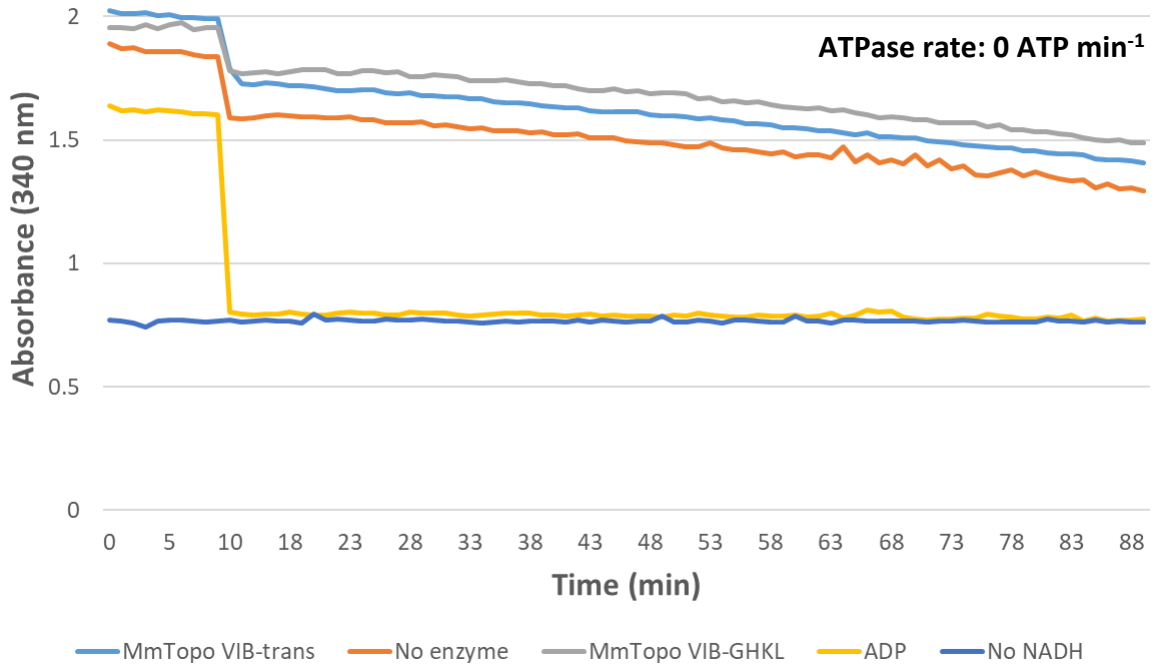

Figure 5.31| *Methanosarcina mazei* topo VI-B truncates ATPase activity: *The ATPase activity of 12 µM MmTopo VIB-GHKL and 12 µM MmTopo VIB-trans in the presence of 177 nM positively-supercoiled DNA was measured using the PK/LDH coupled assay at 37°C. Control samples were run in the absence of enzyme, absence of NADH, and replacing ATP with ADP. ATP was added to each sample, except for the ADP control, after 10 min and absorbance at 340 nm was measured every min for 90 min.*

In section **5.2.2**, I showed that the MmTopo VI.G137A mutant exhibited an ~4-fold reduction in -sc relaxation and decatenation activity compared to the WT enzyme. This residue plays a key role in stabilising the ATP binding site of MmTopo VI, and so the mutation is likely to be influencing strand-passage by impairing the ATPase activity of the enzyme. To confirm this, I performed the PK/LDH assay on MmTopo VI.G137A in triplicate in the presence of +scDNA, and observed a steeper decline of A_{340} in the MmTopo VI.G137A sample than in the no enzyme control (**Figure 5.32**). The difference in rates of these two samples corresponds to an ATP hydrolysis rate of 0.91±0.13 ATP min^{-1}, and this represents an ~5-fold decrease in activity compared to the WT enzyme which is in line with its relaxation and decatenation impairment. This result provides further evidence of the coupling of strand-passage with ATP hydrolysis by MmTopoVI, and the key role the G137 residue plays in ATP utility.

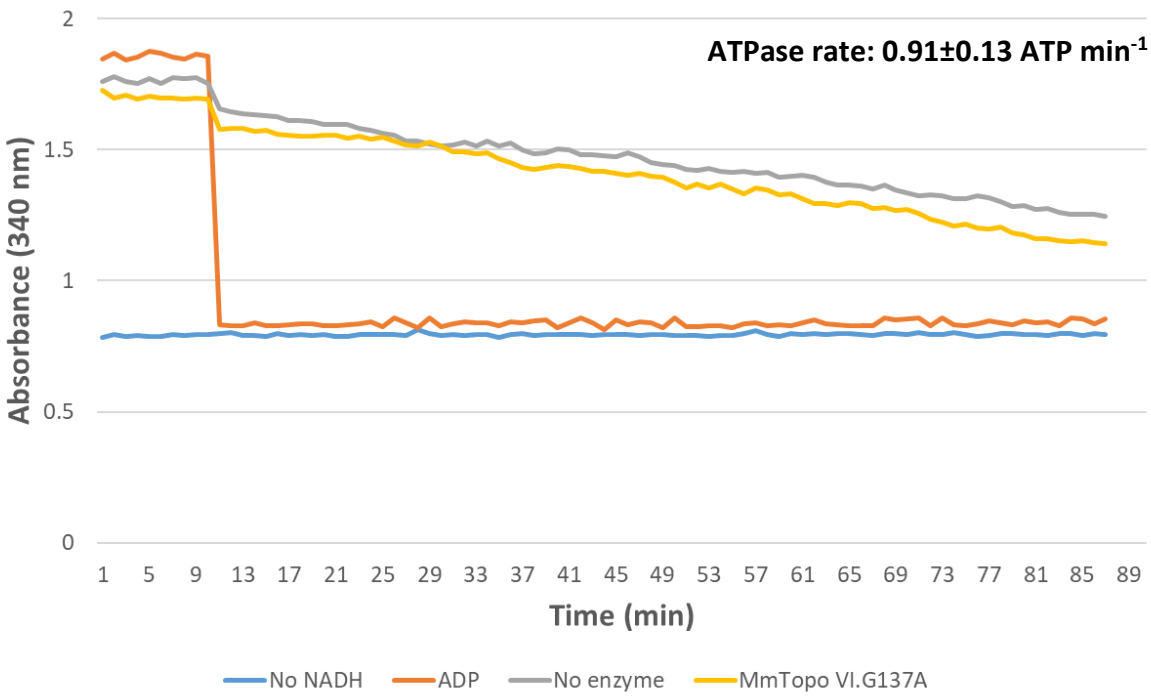

Figure 5.32| *Methanosarcina mazei* topo VI G137A mutant ATPase activity: *The ATPase activity of 1.2 µM MmTopo VI.G137A in the presence of 177 nM positively-supercoiled DNA was measured using the PK/LDH coupled assay at 37°C. Control samples were run in the absence of enzyme, absence of NADH, and replacing ATP with ADP. ATP was added to each sample, except for the ADP control, after 10 min and absorbance at 340 nm was measured every min for 90 min.*

5.3 Discussion

5.3.1 General discussion

The archaeal curiosity that is topo VI has attracted more attention given its prevalence in eukaryotes and the adoption of its scaffold as an essential component of meiosis. MmTopo VI is a useful model for studying both the eukaryotic homologues and the meiotic DSB machinery given the ease at which it can be isolated and biochemically characterised. Indeed, the study of this enzyme is necessary to determine which of its properties have driven its presence in eukaryotes alongside other type II topos, as well as its reduction into a controlled DSB machine during homologous recombination. The investigation of key mechanistic features of MmTopo VI, such as the direction of strand-passage and its coupling to ATP hydrolysis, may provide insights into these uncertainties.

5.3.2 *Methanosarcina mazei* topo VI as a model for the Apicomplexa topo VI-like complex

In Chapter 3, I used AlphaFold modelling to show that putative topo VI-B subunit in the Apicomplexa phylum was a close structural homologue of plant MTOPVIB and a member of the topo VIB-like family. I also showed that the two putative topo VI-A paralogues present in Apicomplexa, SPO11-A and SPO11-B, were bona fide members of the canonical Spo11 family that were unrelated to other Spo11 subfamilies. The GHKL-like domains of plant MTOPVIB only possess four of the eleven key invariant ATP-binding residues present in topo VI-B but have retained the G3-box glycine. The GHKL-like domains of Apicomplexa MTOPVIB, however, can possess between six and ten of these ATP-binding residues but the G3-box glycine is always missing. In the case of the *Plasmodium* MTOPVIB GHKL-like domain, the G3-box glycine is the only key ATP-binding residue that has been lost. Mutating this key glycine residue (G137 in MmTopo VI-B) to alanine in MmTopo VI does not affect ADPNP-induced cleavage, which is the expected result given the predicted function of Apicomplexa MTOPVIB in meiotic DSB formation. However, the G137A mutation imparted at least a 4-fold reduction in the decatenation and relaxation activities of MmTopo VI suggesting that this residue plays key role in strand-passage. The G3-box glycine maintains the structural integrity of GHKL domain, and so it is likely that the G137A mutant is defective in ATP hydrolysis, which is tightly coupled to strand-passage in MmTopo VI. In plant MTOPVIB, it is possible that the loss of 7 key ATP-binding residues was necessary to disable ATP hydrolysis and convert an ancestral topoisomerase into a DSB machine. However, the loss of the G3-box glycine alone may have

been sufficient in the Apicomplexa MTOPVIB to drive the evolution of an incapacitated GHKL-like ATPase fold suitable for supporting meiotic DSBs.

In Chapter 3, I also showed that all canonical Spo11 sequences can classified phylogenetically as either a member of the SPO11-1 or SPO11-2 subfamily. That is, except for the Apicomplexa SPO11-A and SPO11-B sequences, which were classified as members of two unique subfamilies. A series of genome wide knock-out[140-142] and transcriptomics[141,142] studies have implicated SPO11-A as being the functional partner of MTOPVIB in *Plasmodium*, and so the role of SPO11-B is uncharacterised. I have identified an aspartate residue (D294 in MmTopo VI-A) that is conserved in the toprim domains of all topo VI-A sequences, most Spo11 sequences, all SPO11-B sequences, but is absent in SPO11-As. Mutating this invariant residue to alanine in MmTopo VI does not affect ADPNP-induced cleavage, which is the expected result given its absence in canonical Spo11 sequences. Remarkably, however, the D294E mutation had no effect on the relaxation activity of MmTopo VI on +scDNA or -scDNA but induced at least a ~4-fold increase in the decatenation activity of the enzyme on both kDNA and bis-cat DNA. This suggests that the DNA-gate of MmTopo VI plays a key role in specifically regulating the resolution of interlinked chromosomes during mitosis, and the enzyme has evolved a divergent mechanism for strand-passage in the context of sc relaxation. It remains plausible that strand-passage in the context of decatenation and relaxation in MmTopo VI occurs in opposite directions, although there is currently no evidence to support this. Given the absolute conservation of D294 in topo VI species in all three domains of life, it is likely that the enhancement of decatenation activity induced by its substitution is unsuitable for supporting controlled chromosome segregation and may induce mitotic defects.

D294 sits in close proximity to the DxD motif, and so may play a role in metal binding. Substitutions of D294 are rare in the canonical Spo11 sequences, and so it is unusual that the residue is never found in the SPO11-A sequences. It also remains unknown what the effect of these substitutions are in Spo11, and whether the substitutions imparts a functional divergence between SPO11-A and SPO11-B. To investigate the significance of D294 further, it would be interesting to test the effect of other mutations, such as to glutamate, asparagine, or a more neutral glycine. It would also be interesting to test the effect of a D294E mutation on plant endoreduplication, possibly by using a CRISPR/Cas9 system. Furthermore, the effect of D294E on DNA-binding affinities would also uncover novel properties of this key residue.

5.3.3 Top-down vs bottom-up strand-passage in *Methanosarcina mazei* topo VI

Gyrase is the only type II topo that can perform ATP-independent relaxation of -scDNA[56], and does so by performing strand-passage in a bottom-up direction[226] in a mechanism that is the reverse of its ATP-dependent supercoiling and +sc relaxation activities[225]. This unique ability serves to strictly couple the level of supercoiling within the cell to the availability of ATP. It has been postulated that topo VI may also possess the ability to perform bi-directional strand-passage given its lack of a third dimer interface[86]. MmTopo VI preferentially binds -scDNA over +scDNA, despite having greater relaxation activity on the latter[172], and this could be explained by the T-segments moving in opposite directions to each other during their corresponding strand-passage reactions. In this model, strand-passage in one direction would be favoured, however a greater DNA binding affinity would be exhibited by the entrance gate of strand-passage in the opposite direction.

I identified a series of potential DNA-binding residues at the entrance to the DNA-gate, however, introducing single mutations of these residues in MmTopo VI did not differentially affect the relaxation of +scDNA and -scDNA. Therefore, I found no evidence suggesting that bottom-up strand-passage is possible for the relaxation of either of +scDNA or -scDNA. Given that there are many potential DNA-binding residues at the entrance to the DNA-gate and that these positively charged residues occur in clusters, future work should focus on testing more of these residues and introducing double or triple mutations. Furthermore, it may be necessary to mutate the DNA-gate residues into non-polar alternatives, which may have a greater influence on reducing the affinity of DNA to this interface. Another strategy that could be implemented is testing the binding affinities of MmTopo VI to +scDNA and -scDNA using the nitrocellulose-membrane capture technique[221]. By locking the ATP-gate with ADPNP or the DNA-gate by cysteine cross-linking, a differential change in the binding affinities of -scDNA or +scDNA could be observed. MmTopo VI is a preferential decatenase, and so it may also be useful to include bis-cat DNA in these experiments. It remains plausible that strand-passage in the context of both -sc and +sc relaxation occurs in the opposite direction to that of decatenation.

5.3.4 Measuring the ATPase activity of *Methanosarcina mazei* topo VI

The PK/LDH coupled assay is a useful method for accurately measuring the ATPase activity of an enzyme of interest. Given the low ATPase rate exhibited by MmTopo VI the assay is not

straight-forward to perform and requires an enzyme preparation that is very pure and of high concentration. Furthermore, the assay requires a high concentration of DNA which is very expensive and that may suffer from batch inconsistencies. A key problem I faced when designing the ATPase assay, was ensuring that the -sc and +sc pBR322 plasmids were uniform in concentration, purity, and degree of superhelicity. To overcome this, I a performed a phenol/chloroform purification procedure for both substrates and adjusted the DNA concentrations using spectrophotometric analysis. Frustratingly, +scDNA binds ethidium bromide (EtBr) less than -scDNA, which makes direct comparisons between the two substrates challenging on an agarose gel[222]. For example, the +scDNA samples consistently exhibited a higher concentration of nicked plasmid than the -scDNA as observed on an agarose gel, but due to the differential binding affinities of EtBr to these DNA species, it is unclear whether this is a true reflection of the samples themselves. Therefore, I could never be truly confident that differences between the activities of MmTopo VI on -scDNA and +scDNA are solely due to sc chirality.

I performed the PK/LDH assay using MmTopo VI on a variety of topological DNA species and acquired data for the ATPase activity of the enzyme that was largely in agreement with the previously published data. That is, that the ATPase activity of MmTopo VI is DNA-dependent, enhanced by superhelicity, and greater in the presence of +scDNA than -scDNA. These results are consistent with the tight coupling of ATP-hydrolysis with strand-passage, albeit with a greater stimulation by linear and relaxed DNA than expected. However, because the rates of ATP hydrolysis are so low, the PK/LDH assay may not be sensitive enough to capture the true magnitude of ATPase stimulation in MmTopo VI by scDNA. A disappointing feature of the ATPase experiments performed in this chapter is the inconsistency of the apparent topological DNA state-dependent ATPase rate measurements and the change of the topological DNA state as the reaction proceeds.

The results of the PK/LDH assay are largely as expected and in agreement with the literature, however, the tight coupling of ATP hydrolysis with strand-passage should be reflected by a decrease in ATPase activity as the sc substrate is relaxed. The observation that the rate of ATP hydrolysis remains constant for the entire reaction, regardless of substrate turnover, puts doubt on the reliability of these results. It remains plausible that, due to its complexity and indirect measurement of ATP hydrolysis, a feature of the assay itself is maintaining the initial

rate through the whole experiment, although it is difficult to find a rational cause of this. It also remains plausible that a feature of the assay itself has changed the behaviour of MmTopo VI, such that the enzyme is acting more processively, has uncoupled ATP hydrolysis with strand-passage, or has retained its initial rate of ATPase activity granted by the starting state of the DNA substrate. The true nature of this phenomenon should be addressed in future work. It is also worth considering the effect of inhibitors on the ATPase activity of MmTopo VI, in the case that ATP hydrolysis is indeed tightly coupled to strand-passage. Unlike the type IIA topos, it would be difficult to prove that an inhibitor of topo VI was targeting ATP utility given that a strand-passage defective topo VI will also be defective in ATPase activity. Therefore, a structure of topo VI solved in the presence of the inhibitor may be necessary to determine its mechanism of action.

Magnetic tweezer experiments have predicted that MmTopo VI performs ~19 strand-passage events min^{-1} on a catenated DNA substrate[172], which would equate to an ATPase rate of ~0.6 ATP s^{-1}. Therefore, a better characterisation of MmTopo VI ATP hydrolysis may be achieved by performing the PK/LDH assay in the presence of catenanes. The kDNA substrate is unlikely to be suitable for ATPase assays due to its inherently complex and nonuniform structure, and its significant size. Bis-cat DNA, however, would be the ideal substrate for investigating the ATPase activity of MmTopo VI, as one would expect an initial rate of ATP hydrolysis that is ~8-fold greater than in the presence of +scDNA, and then a reduction in the rate of ATP hydrolysis as the bis-cat DNA is converted from supercoiled-catenated, to supercoiled, to relaxed. However, this substrate is currently too expensive and impure to be suitable for the PK/LDH assay. Furthermore, the exclusively linear rates of ATP hydrolysis determined in this chapter in the presence of sc DNA may also apply to the bis-cat substrate.

5.3.5 Conclusion

In this chapter I have performed a series of biochemical analyses on MmTopo VI to attempt to gain a deeper understanding of the poorly characterised type IIB topo family. I have shown that MmTopo VI is a useful model for probing the key conserved residues that are present in the meiotic topo VI-like complex, and that the enzyme has the potential to also serve as a model for its eukaryotic homologues. I was unable to demonstrate that MmTopo VI can perform bottom-up strand-passage but have kick-started a series of experiments that may have more success in the future. Furthermore, I have determined that the preference of

MmTopo VI for +scDNA is reflected in its enhanced rate of ATPase activity, although I have identified peculiarities with the PK/LDH assay that should be investigated further. Future work should focus on probing the significance of the tight-coupling of ATP hydrolysis with strand-passage and determining the direction of strand-passage in the context of decatenation, and -sc and +sc relaxation.

Chapter 6

Identification of *Methanosarcina mazei* topo VI inhibitors

6.1 Introduction

6.1.1 General introduction

The development of novel antibiotics is crucial in the fight against rapidly emerging drug-resistant pathogens and an effective tactic is the targeting of essential enzymes unique to prokaryotes. Similarly, essential cellular processes in humans are often exploited as drug targets to inhibit the growth of cancer cells. DNA topoisomerases (topos) play a fundamental role in preserving genome integrity and have therefore emerged as major molecular targets for clinically important chemotherapeutic and antimicrobial drugs[232]. Topo-targeting antibacterials have shown activity against plants[193] and *Plasmodium*[233], paving the way for the use of topo inhibitors in the development of novel herbicides and antimalarial drugs. In this chapter, I will detail the clinical success story of type IIA topo inhibitors and the high-throughput methods that have been utilised for drug discovery. I will then describe my own work identifying and characterising novel inhibitors of *Methanosarcina mazei* topo VI and discuss the suitability of this archaeal enzyme as a model for eukaryotic topo VI.

6.1.2 Type IIA topoisomerase inhibitors

Compounds that target the type IIA topos have been pursued extensively, giving rise to diverse groups of well-characterised inhibitors that act at different stages of the strand-passage reaction. Typically, these compounds are classified into poisons and catalytic inhibitors and subdivided by their ability or inability to intercalate into DNA. Topo poisons stabilise the cleavage complex and convert the enzyme into a machine capable of inducing irreversible double-strand breaks, which triggers genome instability and cell death[232]. Catalytic inhibitors, however, instigate cellular toxicity by acting on alternative stages of the topo catalytic cycle, such as by preventing DNA binding, or interfering with ATP hydrolysis[232]. DNA intercalators are usually planar, polycyclic, aromatic compounds that insert between nucleic acid base pairs, distort the shape of the DNA helix, and stall the advancement of topos.

6.1.3 Antibacterial type IIA topoisomerase inhibitors

DNA gyrase is essential for bacterial viability and is absent in animals, making this enzyme an attractive antibacterial target. A wide range of compounds that target gyrase have been identified from synthetic and natural sources and, due to a high degree of structural homology with topo IV, many of these inhibitors target both type IIA topos in bacteria[232].

A key group of compounds, the synthetic quinolones, are the most well-characterised family of type IIA topo inhibitors, and have been widely prescribed to treat infection in humans[234]. The quinolones induce bactericidal action by stabilising the DNA cleavage complex of gyrase in gram-negative bacteria and of topo IV in gram-positive bacteria[55,56]. The original quinolone antibiotics have undergone several stages of structural manipulation, giving rise to compounds with improved pharmacokinetic properties and broader spectra of antibacterial activity[234]. The addition of a fluorine atom to the original compound was the first major structural change, giving rise to a new class of drugs called fluoroquinolones[234]. Ciprofloxacin, a second generation fluoroquinolone, is one of the most used antibiotics worldwide[235]. Interestingly, ciprofloxacin has also been shown to target gyrase from *Arabidopsis thaliana*[200] and induce growth defects in the plant by disrupting chloroplast and mitochondrial function[193]. Other compounds that stabilise the gyrase cleavage complex include albicidin, a pathotoxin produced by *Xanthomonas albilineans*[236]; microcin B17, a bactericidal peptide first isolated from *E. coli*[237]; and CcdB, a bacterial toxin produced by the *E. coli* sex factor plasmid (**Table 4.1**)[238].

Table 6.1| Inhibitors of DNA gyrase

Inhibitor	Source	Class	Mode of action
Novobiocin	*Streptomyces*	Aminocoumarin	Competitive inhibition of ATP binding
Ciprofloxacin	Synthetic	Fluoroquinolone	Stabilisation of cleavage complex
Simocyclinone D8	*Streptomyces*	Angucycline	Inhibition of DNA binding
MfpA	*Mycobacterium tuberculosis*	Pentapeptide repeat protein	Competitive inhibition of DNA binding
Qnr	Enterobacteriaceae	Pentapeptide repeat protein	Competitive inhibition of DNA binding
Albicidin	*Xanthomonas albilineans*	Peptide	Stabilisation of cleavage complex
Microcin B17	*E. coli*	Peptide	Stabilisation of cleavage complex
CcdB	*E. coli*	Protein	Stabilisation of cleavage complex

Remarkably, another group of inhibitors known as the pentapeptide repeat protein (PRP) family[239] can protect gyrase against quinolone inhibition[240,241]. Members of this family, such as the quinolone resistance protein Qnr found in Enterobacteriaceae,[241] and *Mycobacterium* fluoroquinolone resistance protein A (MfpA)[240], contain tandemly repeated amino acid

sequences that form a right-handed quadrilateral β-helix capable of mimicking B-form DNA[239,242]. PRPs inhibit the DNA supercoiling and relaxation activities of gyrase in the absence of quinolones but can reverse quinolone-induced DNA cleavage by competing with T-segment binding[231,242,243]. Several groups of catalytic gyrase inhibitors have also been characterised, including the aminocoumarins, such as novobiocin, and the simocyclinones, both of which are produced by *Streptomyces*. The aminocoumarins are competitive inhibitors of ATP hydrolysis which impair the supercoiling activity of gyrase[229,230]. Unlike the aminocoumarins or quinolones, the simocyclinones inhibit the DNA supercoiling and relaxation activities of gyrase by preventing DNA binding and are active against gram-positive bacteria[244,245].

6.1.4 Chemotherapeutic type IIA topoisomerase inhibitors

Several DNA intercalating agents were already in use as chemotherapeutic drugs before being shown to target human topo II. Drugs such as etoposide[246], doxorubicin[247], ellipticine[248] mitoxantrone[247], and amsacrine[249] were known to inhibit DNA replication and eliminate cancer cells and were later identified as topo II poisons and inducers of enzyme-mediated double-strand breaks (**Table 4.2**).

Table 6.2| Inhibitors of human topo II

Inhibitor	Source	Class	Mode of action
Etoposide	American Mandrake	Epipodophyllotoxin	Stabilisation of cleavage complex
Doxorubicin	*Streptomyces*	Anthracycline	Stabilisation of cleavage complex
Ellipticine	Apocynaceae	Pyridocarbazole	Stabilisation of cleavage complex
Mitoxantrone	Synthetic	Anthraquinone	Stabilisation of cleavage complex
Amsacrine	Synthetic	Acridine	Stabilisation of cleavage complex
Dexrazoxane	Synthetic	Bisdioxopiperazine	Inhibition of ATP hydrolysis
Quinacrine	Synthetic	Acridine	Inhibition of DNA binding
9-Aminoacridine	Synthetic	Acridine	Inhibition of DNA binding
Merbarone	Synthetic	Thiobarbituate	Inhibition of DNA binding
Aclarubicin	*Streptomyces*	Anthracycline	Inhibition of DNA binding

Etoposide is a member of a group of anticancer agents called epipodophyllotoxins that are extracted from the roots of the American Mandrake[250], doxorubicin belongs to a group of antineoplastic compounds called anthracyclines which are derived from *Streptomyces*[251], and ellipticine is a pyridocarbazole isolated from Apocynaceae plants. Due to the successes of these anticancer drugs, several classes of synthetic compounds were developed to target topo II, including the doxorubicin analogue mitoxantrone[252] and the acridine amsacrine[253]. Other intercalating acridines, such as the antimalarial quinacrine and the antiseptic 9-aminoacridine, block the binding of DNA to topo II rather than poisoning it[254]. This is also the

case with aclarubicin which, like doxorubicin, is an intercalating anthracycline, but instead of stabilising the topo II cleavage complex it offers protection against cleavage[255]. The thiobarbituate derivative merbarone can also prevent topo II-mediated DNA cleavage[256], and another synthetic compound, the bisdioxopiperazine dexrazoxane, has been shown to inhibit the enzyme's ATPase activity[257]. The African sleeping sickness drug suramin, like quinacrine[258], is a promiscuous inhibitor of many enzymes[259] and has also been found to target human topo II[260]. The majority of topo II-targeting compounds can inhibit both the α and β isoforms of the enzyme[261].

6.1.5 DNA topoisomerase VI inhibitors

The study of type IIA topo inhibitors has set a precedent for the type IIB class, which has largely been neglected. Given the presence of topo VI in plants[49] and Apicomplexa[50], identifying compounds that target this enzyme may aid the discovery of novel herbicides and antiprotozoals. However, a eukaryotic topo VI has yet to be successfully purified and so work has instead been carried out on topo VI preparations from archaea. The first topo VI inhibitor to be characterised was radicicol, an antifungal agent obtained from the fungus *Monosporium bonorden*[262]. Radicicol was shown inhibit the molecular chaperone heat shock protein 90 (Hsp90) by interacting with the Bergerat fold[263] and was subsequently shown to interact with the homologous motif in *Saccharolobus shibatae* topo VI and inhibit its DNA relaxation and decatenation activities[264]. Like novobiocin, radicicol is a competitive ATPase inhibitor and impedes the dimerisation of the *S. shibatae* topo VI GHKL domain to prevent T-segment trapping and stall strand-passage[263,264].

The quinolone CP-115,955, as well as the chemotherapeutic drugs etoposide, doxorubicin, ellipticine, and amsacrine (**Table 4.3**), have all been shown to inhibit the unknotting activity of *S. shibatae* topo VI[46]. Although they are known DNA intercalators, their exact mechanism of action against topo VI was not elucidated. Furthermore, hexylresorcinol, purpurin, and the topo II-targeting compounds quinacrine, 9-aminoacridine, and suramin were all identified as novel inhibitors of *M. mazei* topo VI (MmTopo VI)[191]. Hexylresorcinol is a synthetic compound used as a local anaesthetic and has been shown to induce a dwarf phenotype in *A. thaliana* seedlings[191]. Plants grown on hexylresorcinol-containing medium exhibited a reduction in hypocotyl growth and cell size[191], a phenotype which is observed in topo VI knock-outs[115-117]. Purpurin is an anthraquinone isolated from the Indian Madder plant and is a promising

therapeutic agent with multiple medicinal properties[265]. The chemotherapeutic drug mitoxantrone is a member of the same class of compounds as purpurin and has also been found to inhibit MmTopo VI[191]. Purpurin and suramin both prevent the binding of DNA to MmTopo VI in a native gel-shift assay, while amsacrine, 9-aminoacridine, quinacrine, and mitoxantrone were all shown to inhibit the DNA cleavage reaction of *S. shibatae* topo VI[191].

Table 6.3| Inhibitors of topo VI*: CP-115,955, etoposide, doxorubicin, ellipticine, and amsacrine are DNA intercalators that stabilise the cleavage complex of human topo II. However, their exact mechanism of action against topo VI has not been elucidated.*

Inhibitor	Source	Class	Mode of action
Radicicol	*Monosporium bonorden*	Macrolactone	Competitive inhibition of ATP hydrolysis
CP-115,955	Synthetic	Quinolone	DNA intercalation
Etoposide	American Mandrake	Epipodophyllotoxin	DNA intercalation
Doxorubicin	*Streptomyces*	Anthracycline	DNA intercalation
Ellipticine	Apocynaceae	Pyridocarbazole	DNA intercalation
Amsacrine	Synthetic	Acridine	Prevents DNA cleavage
Hexylresorcinol	Synthetic	Resorcinol	Unknown
Purpurin	Madder plant	Anthraquinone	Blocks DNA binding
Quinacrine	Synthetic	Acridine	Prevents DNA cleavage
9-Aminoacridine	Synthetic	Acridine	Prevents DNA cleavage
Suramin	Synthetic	Phenylurea	Blocks DNA binding
Mitoxantrone	Synthetic	Anthraquinone	Prevents DNA cleavage

6.1.6 High-throughput type II topoisomerase assays

To identify novel compounds targeting type II topos, it is necessary to measure the interconversion of relaxed and supercoiled DNA using an assay compatible with high-throughput screening. Traditional assays that measure topo activity involve the separation of DNA topoisomers by gel electrophoresis, staining by ethidium bromide (EtBr), and the quantification of DNA band intensities. This method is slow and requires intensive sample handing and is therefore poorly suited for the rapid screening of large compound libraries. To this end, many attempts have been made to develop higher-throughput topo assays that reduce sample handling and are faster, more sensitive, and more quantifiable.

One method, which abolishes the need for gel electrophoresis, analyses the decatenation of kinetoplast DNA (kDNA) by flow-injection fluorimetry[266]. In this assay, the decatenated DNA products are removed from the kDNA substrate by centrifugation and stained by the fluorescent dye DAPI. The flow-injection method is faster and more sensitive than its gel-based counterpart but is limited by the laborious centrifugation step and the inability to adapt to a microplate format. Another method exploits the differential fluorescence spectra

exhibited by the commercial dye PicoGreen in response to its binding to negatively-supercoiled (-sc) or relaxed DNA[267]. This assay is more amenable to high-throughput screening and was able to identify 197 novel inhibitors of *E. coli* gyrase but suffers the significant limitation of a very small signal-to-noise ratio[267].

A different approach exploits the observation that DNA supercoiling stabilises the formation of hairpins at sites containing specific inverted repeat sequences[268]. By inserting a fluorophore-quencher pair in close proximity on opposite strands at these sites on a plasmid it is possible to force their separation by promoting hairpin formation[269]. Therefore, the superhelicity of the plasmid can be measured using fluorescence resonance energy transfer (FRET)[269]. The purification of this plasmid, however, results in very low yields making the assay too costly for high-throughput screening. Inspired by this idea, another group designed a similar plasmid that possesses a fluorophore-quencher on the same strand but at opposite ends of the repeat sequences[270]. Upon hairpin formation this pair is brought together, giving rise to an opposing FRET spectrum in response to DNA superhelicity than the original plasmid. The preparation of this plasmid has a superior yield but is hindered by the requirement for an expensive nicking endonuclease. Notwithstanding, this new plasmid was used in a high-throughput screen that identified 155 novel inhibitors of *E. coli* gyrase[271].

A novel tactic for identifying topo inhibitors has been built from the principle of DNA triplex formation, a process which is more efficient in -scDNA than in relaxed DNA under conditions of low pH (5.0) and moderate metal ion concentration (50 mM $MgCl_2$)[272,273]. Intramolecular DNA triplexes form when a third homopyrimidine or homopurine strand associates with the major groove of the DNA duplex via non-canonical hydrogen bonding[274]. By immobilising a triplex-forming oligonucleotide on a microplate surface, double-stranded supercoiled DNA, but not relaxed DNA, can be captured and stained with a fluorescent dye[273]. This method offers a vast improvement on the signal-to-background ratio exhibited by the PicoGreen assay but is limited by the many washing steps required. A variation of this assay uses a triplex-forming oligonucleotide in solution labelled with a fluorescent dye[275]. At a lower pH (3.5) and low metal ion concentration (50 mM $MgCl_2$) triplex formation seems to be favoured by relaxed DNA, which gives rise to an increase in fluorescence anisotropy[275]. Although this method simplifies the immobilised oligonucleotide assay, it requires an additional quenching

step to adjust the pH and high concentrations of plasmid with an expensive purification procedure.

6.1.7 Aims and objectives

Here, a novel fluorescence-based high-throughput relaxation assay was developed to identify inhibitors of eukaryotic topo VI. This method aims to overcome the obstacles with the existing assays and was validated by a proof-of-principle screen using an archaeal topo VI as a model. Novel inhibitors targeting topo VI may aid the development of novel herbicides and antiprotozoals while serving as useful probes for understanding the enzyme's mechanism.

6.2 Results

6.2.1 Validation of the fluorescence-based high-throughput relaxation assay

The commercial H19 dye features a DNA-binding peptide attached to a fluorophore and presents a preferential enhancement of fluorescence in the presence of -scDNA compared to relaxed DNA. By changing the ratio of these DNA isoforms, a greater than two-fold fluorescence differential was established (**Figure 6.1**). This demonstrates the role H19 can play in monitoring the topo-mediated interconversion of supercoiled and relaxed DNA. H19 has been successfully utilised in an assay to show that an extract of the liquorice plant[276] and a derivative of the plant polyphenol resveratrol[277] can inhibit *E. coli* gyrase. H19 has also been used to test the potency of cyclised analogues of a thiophene that targets *E. coli* gyrase[278]. Therefore, this dye will be tested for suitability in a high-throughput assay.

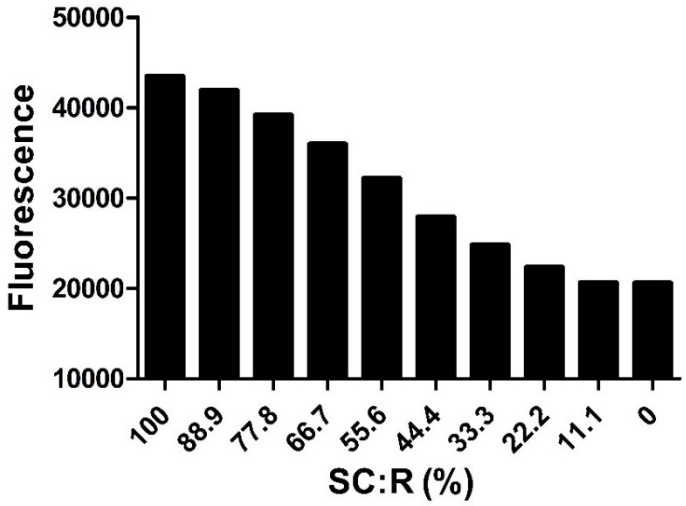

Figure 6.1| Fluorescence intensity of the H19 dye in response to DNA topology. *The H19 dye exhibits an enhancement of fluorescence intensity when bound to negatively-supercoiled (SC) DNA compared to relaxed (R) DNA.*

An enzyme titration was performed to test whether H19 can measure the conversion of -scDNA to its relaxed form by MmTopo VI and to identify the required concentration to do so. The test compounds were dissolved in 100% DMSO and so this assay was carried out in 6.7% DMSO to match the conditions of the high-throughput screen. H19 successfully monitored the relaxation of -scDNA by MmTopo VI, with full relaxation occurring at an enzyme concentration of 2 nM (**Figure 6.2**). The relaxation activity of MmTopo VI is ATP-dependent[58]

and so the ability of the H19 assay to reflect this was also tested. As expected, omitting ATP from the reaction mixture abolished the reduction in fluorescence signal seen when incubating -scDNA with the enzyme (**Figure 6.2**). Together, these results validate H19 as a suitable dye for measuring the DNA relaxation activity of MmTopo VI.

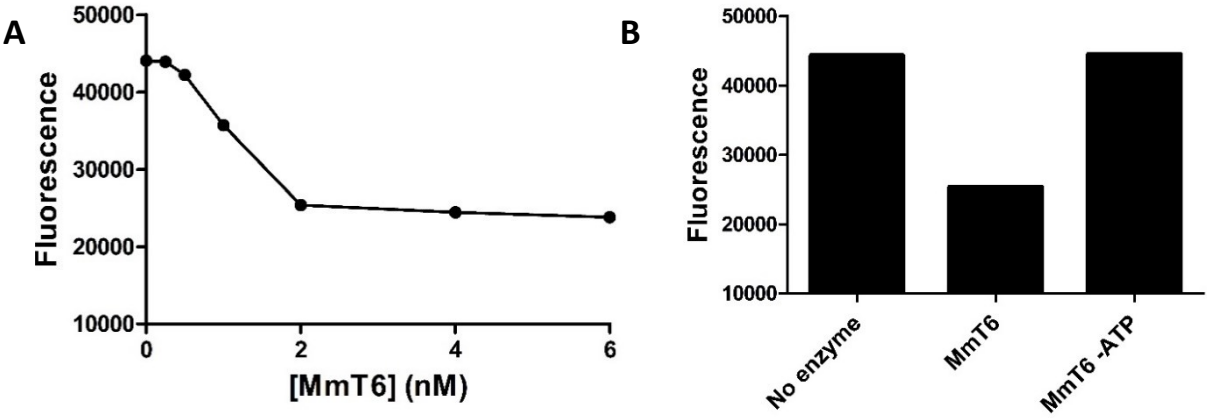

Figure 6.2| *Methanosarcina mazei* topo VI relaxation assay analysed by the H19 dye. *(A)* *The conversion of 25 nM -scDNA to relaxed DNA by **M. mazei** topo VI can be monitored by measuring the fluorescence intensity of the H19 dye. **(B)** The decrease in fluorescence intensity when incubating supercoiled DNA with **M. mazei** topo VI is abolished by omitting ATP. Reactions were run at 37°C for 45 min.*

6.2.2 Assay optimisation

To establish a robust protocol for the high-throughput screening of topo VI inhibitors using the H19 dye, I evaluated the constituents of the assay. The supplier specification of H19 recommends using a DNA concentration of 25 nM, diluting the dye in 1X H19 Dilution Buffer (DB), and adding 62.5 nL of dye in 93.75 µL of 1X H19 DB to each well in a 15 µL reaction. These constituents alone would cost £315 per 384-well plate. Replacing the H19 DB with water increased the fluorescence intensity of the dye in the presence of relaxed DNA and reduced the difference in fluorescence intensities between relaxed and -scDNA by up to 15% (**Figure 6.3**). Therefore, the H19 DB seems necessary for the measurement and will be retained in the high-throughput assay. However, because this change in signal is modest the total volume of H19 dye-DB added to each reaction was reduced from 93.75 µL to 10 µL to simplify the high-throughput handling. Lowering the concentration of DNA or H19 in the assay also reduced the difference in fluorescence signal induced by the presence of -sc vs relaxed DNA (**Figure 6.4**). A large signal-to-noise ratio must be maintained to ensure that the high-

throughput screen identifies real hits. However, using a significantly lower DNA concentration than recommended by the supplier can still offer a large difference in signal. Reducing the DNA concentration from 25 nM to 6.7 nM will save ~£167 per 384-well plate while sacrificing only a 10% decrease in the difference of fluorescence signal between -sc and relaxed DNA. The high-throughput assay can therefore be carried out using 6.7 nM -scDNA.

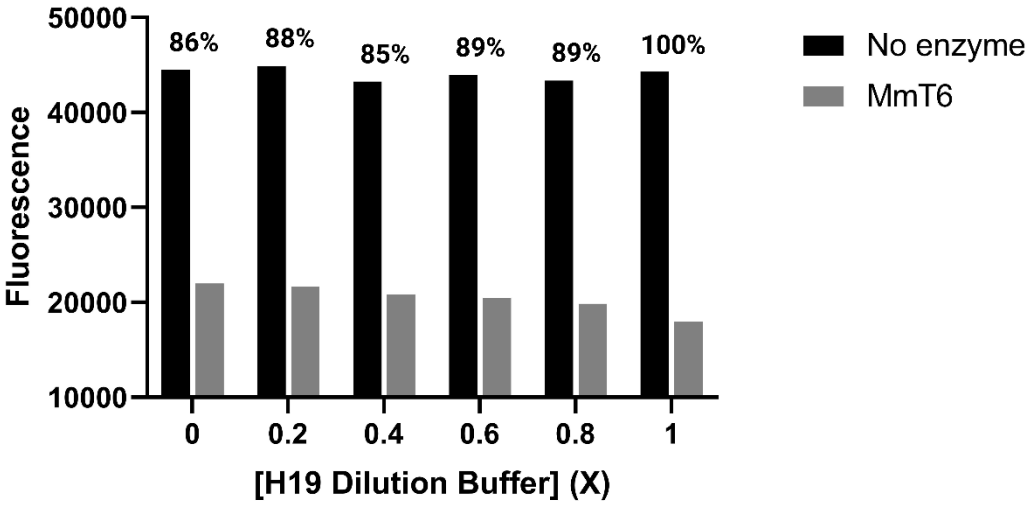

Figure 6.3| Effect of H19 Dilution Buffer on the fluorescence intensity of H19. *Decreasing the concentration of the H19 Dilution Buffer (DB) increases the fluorescence intensity of H19 in the presence of relaxed DNA. The % difference between the relaxed and supercoiled DNA fluorescence intensities compared to the maximum concentration of the H19 DB is shown above the bars. Reactions were run in the presence of 25 nM -scDNA at 37°C for 45 min.*

The concentration of H19 seems to be better optimised for the assay than the concentration of DNA, and a more dramatic difference in the signal-to-noise ratio occurs as it is decreased (**Figure 6.4**). Reducing the volume of H19 from 62.5 nL to 48.6 nL will save ~£13 per 384-well plate but provokes a 27% decrease in the difference of fluorescence signal between -sc and relaxed DNA. Despite this decrease in signal, and given that the dye was in limited supply, the reactions performed here were carried out with 48.6 nL of H19, although the recommended volume would be more suitable.

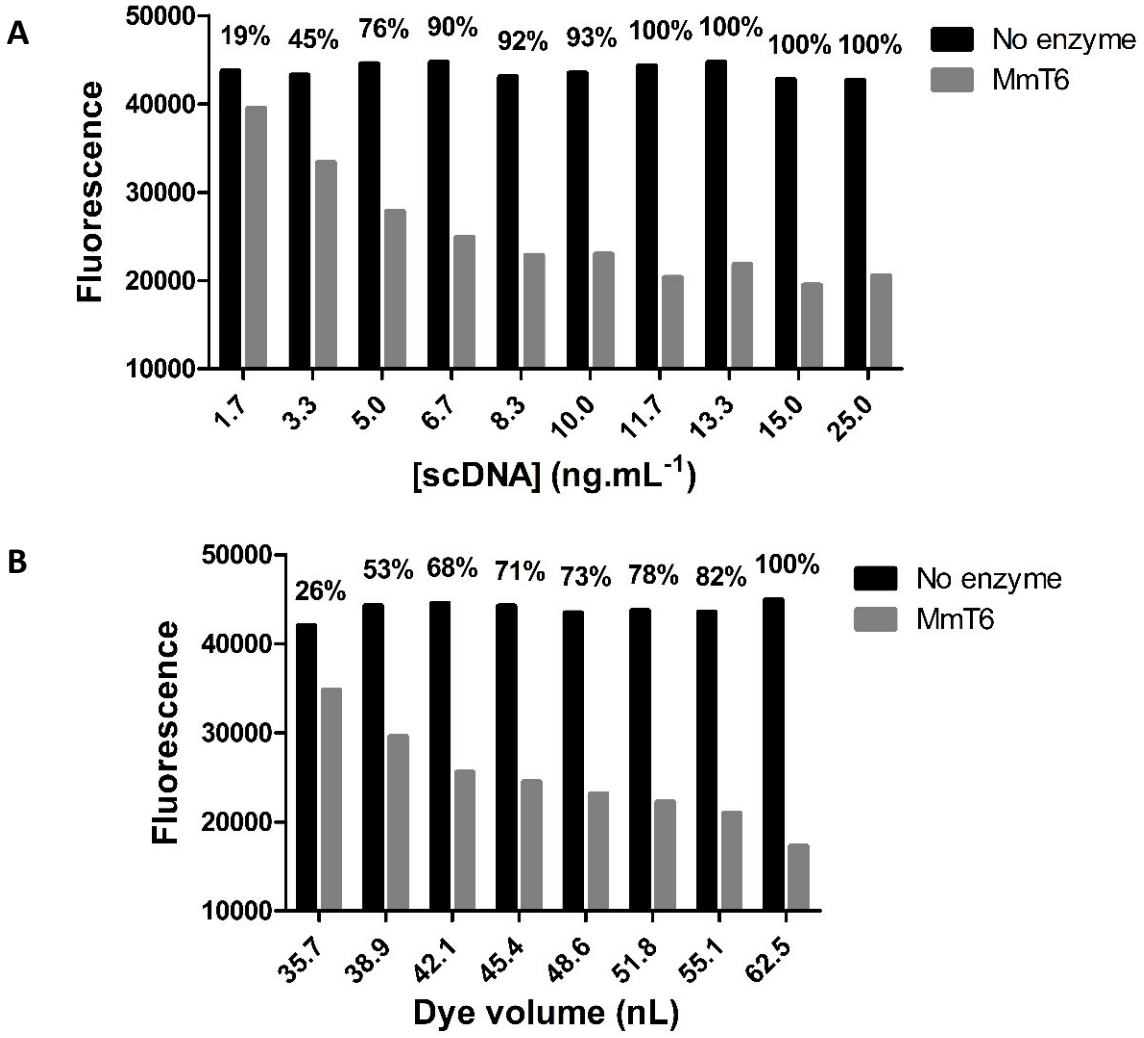

Figure 6.4| Effect of DNA and H19 concentration on the fluorescence intensity of H19. ***(A)*** *Decreasing the concentration of -scDNA increases the fluorescence intensity of H19 in the presence of* ***M. mazei*** *topo VI.* ***(B)*** *Decreasing the concentration of the H19 increases the fluorescence intensity of H19 in the presence of relaxed DNA. The % difference between the relaxed and supercoiled DNA fluorescence intensities compared to the maximum concentration of the H19 DB is shown above the bars. Reactions were run at 37°C for 45 min.*

6.2.3 Identification of *Methanosarcina mazei* topo VI inhibitors by high-throughput screening

The occurrence of topo VI in plants[49] and Apicomplexa[50] affords an exciting new target for herbicides and antiprotozoals, but these eukaryotic enzymes have proven difficult to isolate. Topo VI from the archaeon *M. mazei*, however, can be purified in high yields to near homogeneity and was therefore used as a surrogate for its eukaryotic homologues in the high-throughput assay. The disruption of topo VI in plants stalls the cellular process of endoreduplication which manifests as an extreme dwarf phenotype[115-117]. To exploit this

effect, a selection of Syngenta compounds was chosen that possess unknown modes of action and inflict a stunting symptomology in whole plant assays thereby enriching for potential inhibitors of plant topo VI.

To test the validity of the fluorescence-based high-throughput relaxation assay, I conducted a proof-of-principle screen of 1750 Syngenta compounds with MmTopo VI. A total of 25 compounds reached the 70% inhibition threshold at 10 µg.mL^{-1} and were evaluated under the same conditions in a three-fold dose-response analysis. Of the 25 preliminary hits, 6 inhibited DNA relaxation with a sigmoidal dose-response (**Figure 6.5**), 3 exhibited autofluorescence, and the remaining 16 failed to show any detectable inhibition.

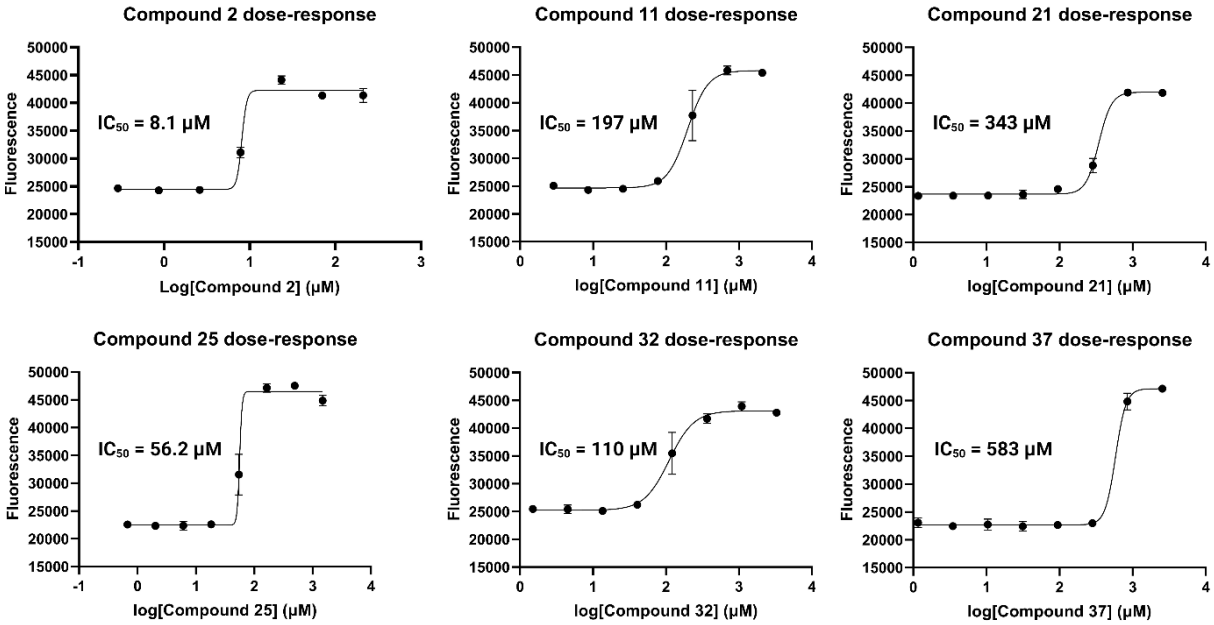

Figure 6.5| Three-fold dose-response curves of preliminary screen hits. *Compounds 2, 11, 21, 25, 32, and 37 were identified in the preliminary screen and produced sigmoidal dose-response curves in the confirmation screen. Compounds were tested in an 8-point three-fold dose-response fashion from a maximum concentration of 10 mg.mL^{-1}. Reactions were run at 37°C for 45 min. Error bars represent standard deviation.*

To improve the accuracy of the IC_{50} values calculated for the two most potent compounds, '2' and '25', I performed a 10-point two-fold dose-response assay. The determined IC_{50} values for compounds 2 and 25 were 6.9 µM and 61.3 µM, respectively (**Figure 6.6**). Compound 2 features a central ether moiety attached to a pyridine, a 4-fluoropyrimidine, and a

chlorobenzene, while compound 25 possesses a phenyl group adjacent to an oxazole attached to a 2-fluoroethanethiol (**Figure 6.6**). An analogue of compound 2 was also tested that shares its ether moiety but is missing its pyridine ring, has replaced fluorine with chlorine in its 4-fluoropyrimidine ring, and has substituted chlorobenzene with 2-Phenyl-iso-butane. This analogue inhibited -scDNA relaxation by MmTopo VI with a sigmoidal dose-response, but with a far less potent IC_{50} value of 112 µM (**Figure 6.6**).

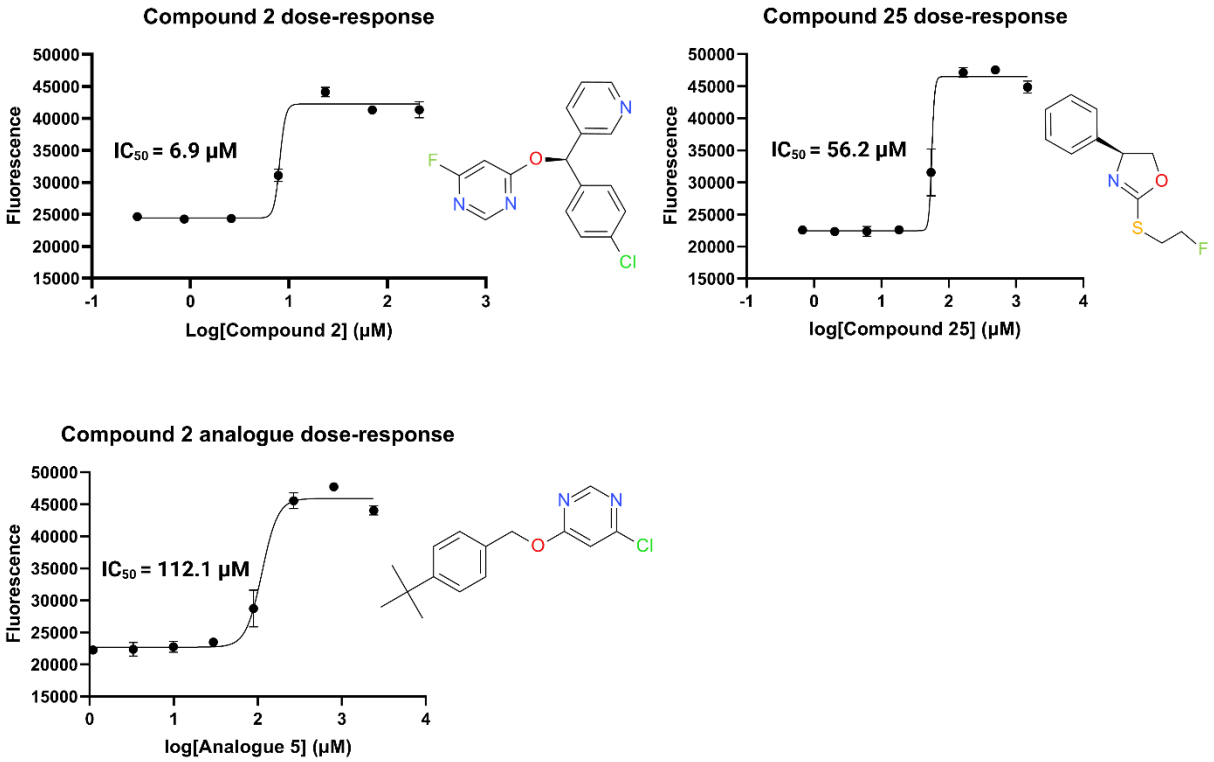

Figure 6.6| Two-fold dose-response curves and structures of compounds 2 and 25 and an analogue of compound 2. *Compounds 2 and 25 were subject to a 10-point two-fold dose-response analysis from a maximum concentration of 10 mg.mL^{-1} yielding IC_{50} values of 6.9 µM and 56.2 µM, respectively. Under the same analysis, an analogue of compound 2 had an IC_{50} value of 112.1 µM. The structures of these compounds are shown adjacent to their curves. Reactions were run at 37°C for 45 min. Error bars represent standard deviation.*

6.2.4 Validation of *Methanosarcina mazei* topo VI inhibitors by agarose gel electrophoresis

It is necessary to perform the dose-response reactions on a gel-based assay under the same conditions as the fluorescence-based method to validate the inhibitory action of the compounds. This ensures that the interconversion of DNA topoisomers can be visualised directly. First, I performed enzyme titration assays to determine the appropriate concentration of MmTopo VI to be used in -sc and positively sc (+sc) DNA relaxation and

kinetoplast DNA (kDNA) decatenation dose-response reactions. Full relaxation and decatenation was observed at an enzyme concentration of 2 nM (**Figure 6.7**), which is the same value obtained from the fluorescence-based assay titration. However, a lower enzyme concentration of 1.6 nM was carried forward as this would be more suitable for running dose-response curves.

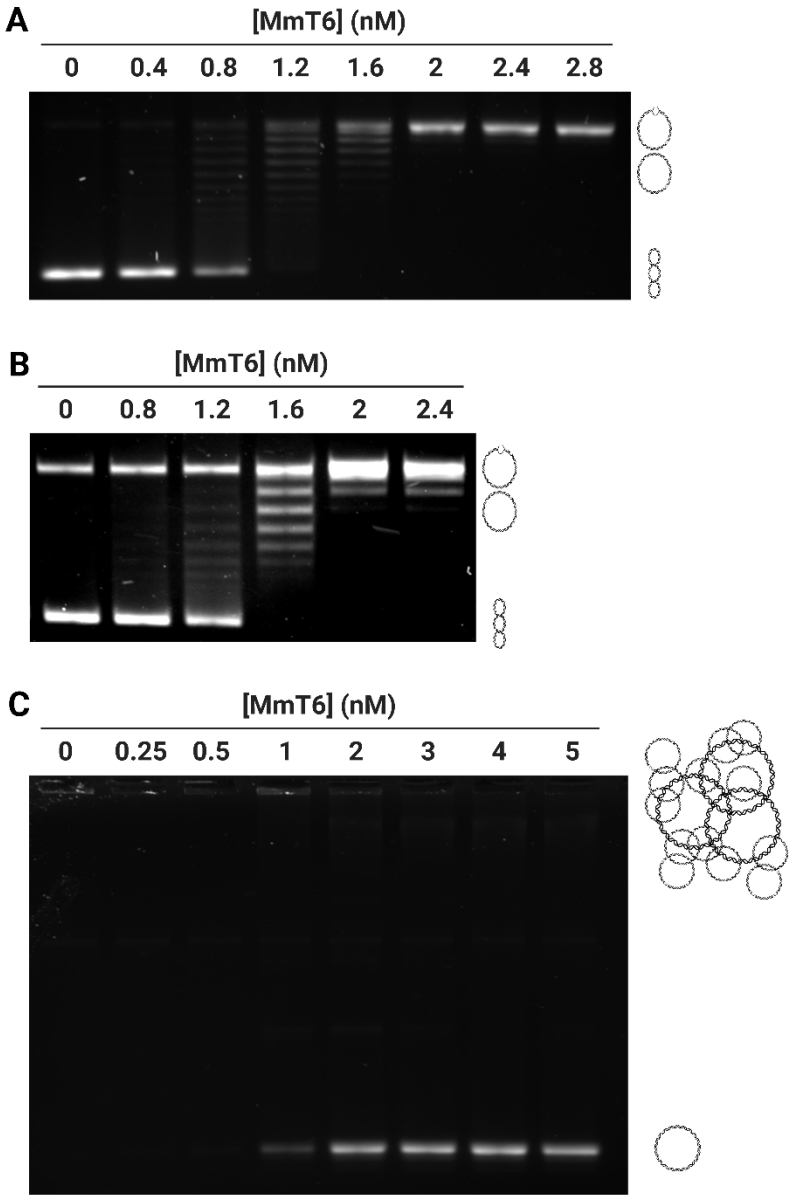

Figure 6.7| *Methanosarcina mazei* topo VI relaxation and decatenation assay titrations. ***M. mazei*** *topo VI was titrated in a -scDNA (6.7 nM) relaxation assay* ***(A),*** *a +scDNA (6.7 nM) relaxation assay (**B**), and a kDNA (6.7 ng µL^{-1}) decatenation assay* ***(C)****. In the relaxation assay,* ***M. mazei*** *topo VI removed all negative supercoils at a concentration of 1.6 nM and fully relaxed the substrate at 2 nM. For decatenation, the release of minicircle products was observed at 1 nM enzyme and full substrate turnover occurred at 2 nM. Reactions were run at 37°C for 45 min.*

I then performed nine-point two-fold dose-response assays for compounds 2 and 25 which were visualised by agarose gel electrophoresis. Both compounds inhibited the relaxation of -scDNA, the relaxation of +scDNA, and the decatenation of kDNA on the gel-based assays (**Figure 6.8**). These reactions were performed in triplicate to generate dose-response curves and to calculate IC_{50} values (**Figure 6.9**). For all substrates tested, compound 2 had an IC_{50} value ranging from 4.2 µM to 7.1 µM which is comparable to the value of 6.9 µM calculated from the fluorescence-based assay. The IC_{50} values for compound 25, however, ranged from 7.2 µM to 10.2 µM which is significantly lower than the fluorescence-based IC_{50} value of 56.2 µM. This phenomenon might reflect the indirect nature of measuring the interconversion of relaxed and supercoiled DNA by the H19 dye compared with the direct gel-based approach.

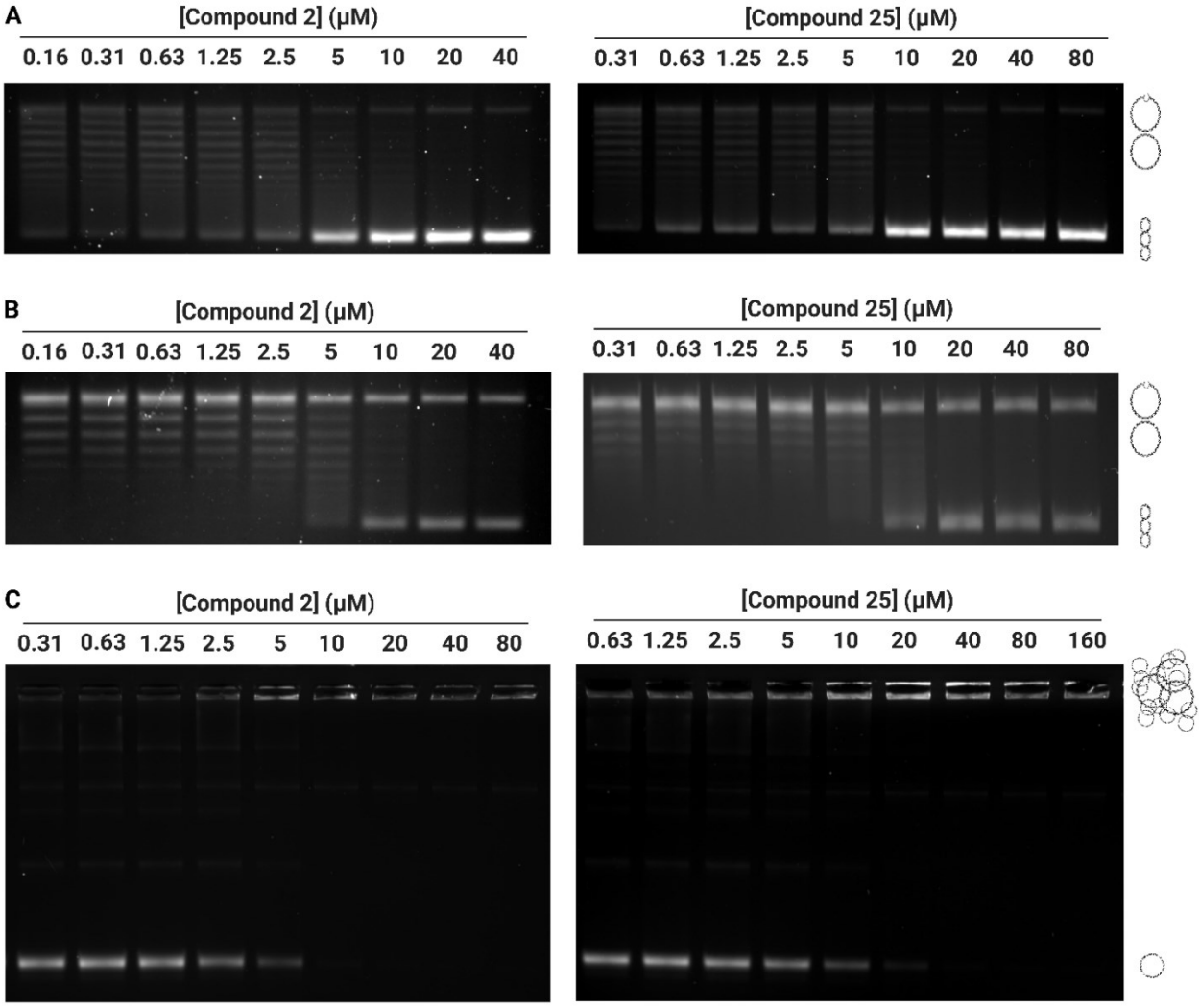

Figure 6.8| *Methanosarcina mazei* topo VI gel-based dose-response assays for compounds 2 and 25. *Dose-response analyses were performed for compounds 2 and 25 on the relaxation of -scDNA (6.7 nM) (**A**), the relaxation of +scDNA (6.7 nM) (**B**) and the decatenation of kDNA (6.7 ng µL^{-1}) (**C**) by 1.6 nM **M. mazei** topo VI. Reactions were run at 37°C for 45 min.*

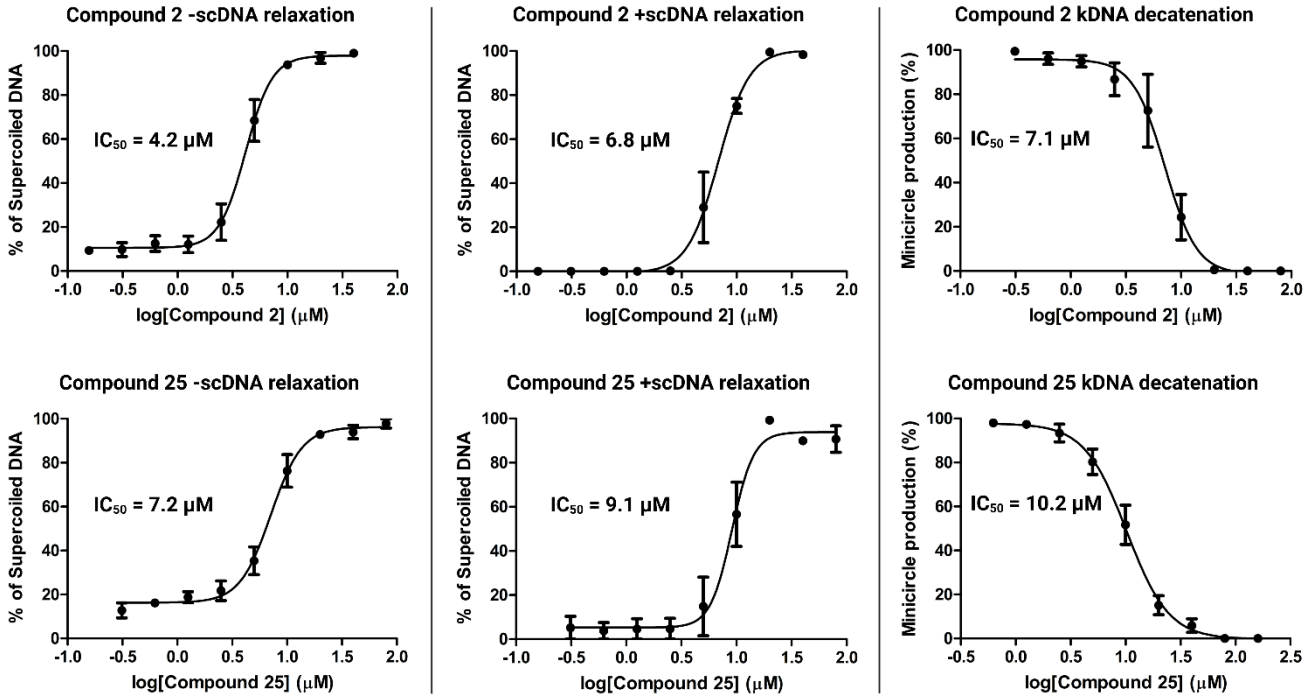

Figure 6.9| *Methanosarcina mazei* topo VI dose-response curves for compounds 2 and 25 from gel-based assays. *Dose-response curves were generated for compounds 2 and 25 for the relaxation of 6.7 nM -scDNA, the relaxation of 6.7 nM +scDNA, and the decatenation of 6.7 ng µL^{-1} kDNA, by 1.6 nM **M. mazei** topo VI from the gel-based assay. Reactions were run at 37°C for 45 min. Gels were run in triplicate and error bars represent standard deviation.*

6.2.5 Timecourse of inhibition of *Methanosarcina mazei* topo VI by Syngenta compounds

I investigated the timecourse effects of compounds 2 and 25 at three different concentrations on the relaxation of -scDNA by MmTopo VI. Reactions were run for 55 min with full substrate turnover occurring after 15 min in the absence of compound (**Figure 6.10**). In the presence of 1.0 µM compound 2 the negative supercoils were removed after 30 min, and at 4.1 µM the supercoils were removed after 40 min. In the presence of compound 25 at concentrations of 2.9 µM and 5.8 µM, the negative supercoils were removed after 25 min and 40 min, respectively. Therefore, not only do compounds 2 and 25 inhibit the relaxation of -scDNA by MmTopo VI in a time-dependent manor, but the inhibition is consistent of the IC_{50} values from the dose-response assays.

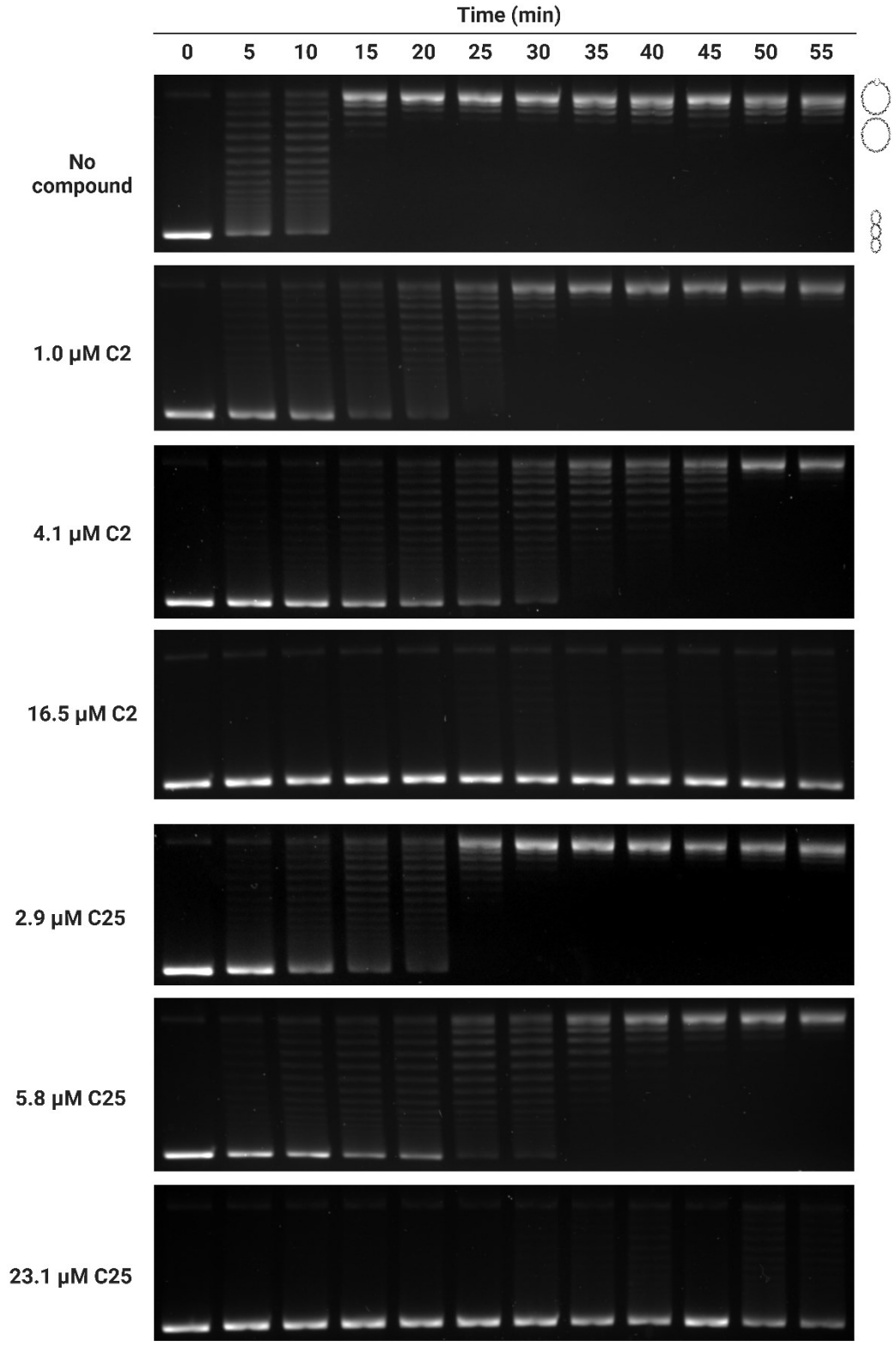

Figure 6.10| Timecourse effects of compounds 2 and 25. *Timecourse analyses were performed for the relaxation of 6.7 nM -scDNA by 1.6 nM* ***M. mazei*** *topo VI. Samples were tested every 5 min for 55 min in the presence of either compound 2 or compound 25 at three different concentrations. Reactions were run at 37°C.*

6.2.6 Effect of the Syngenta compounds on other type II topoisomerases

The type II topos share a degree of structural homology and so it is not unusual for inhibitors to exhibit promiscuity towards different enzymes in this class. To determine whether compounds 2 and 25 are specific to topo VI, I tested the effect of these compounds on other type II topos. First, enzyme titrations were performed for the relaxation of -scDNA by *E. coli* topo IV, yeast topo II, and human topo IIα, and for the supercoiling of relaxed DNA by *E. coli* gyrase to identify the suitable concentrations needed for dose-response assays. All enzymes tested performed their respective reactions as expected (**Figure 6.11**).

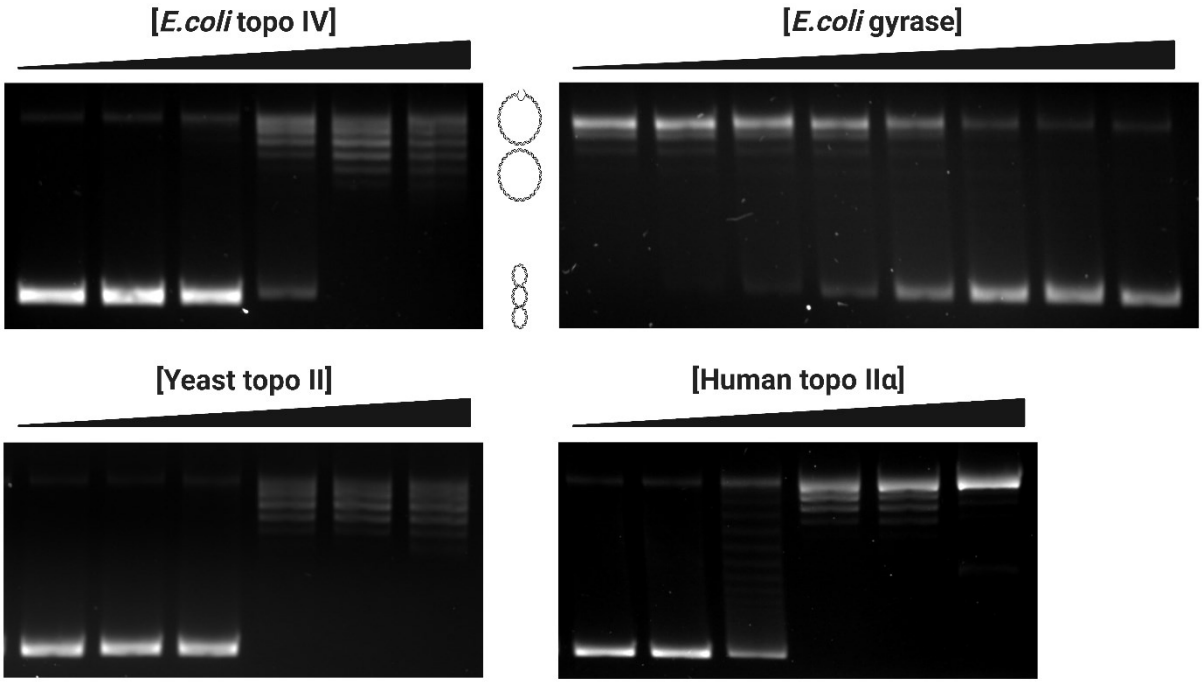

Figure 6.11| Enzyme titrations of various type II topoisomerases. *Enzyme titrations were performed for the relaxation of 6.7 nM -scDNA by **E. coli** topo IV, yeast topo II, and human topo IIα, and for the supercoiling of 6.7 nM relaxed DNA by **E. coli** gyrase. Reactions were run at 37°C for 45 min.*

The conversion of relaxed to supercoiled DNA by *E. coli* gyrase is inhibited by 20 µM novobiocin but is unaffected by compounds 2 and 25 up to 1 mM (**Figure 6.12**). Likewise, the relaxation of supercoiled DNA by *E. coli* topo IV is also inhibited in the presence of 20 µM novobiocin but is unchanged after adding up to 1 mM of compounds 2 and 25. Therefore, both compounds seem to be inactive against bacterial type II topos. To investigate whether

these compounds are also inactive against eukaryotic type II topos, their effect on the removal of DNA supercoils by yeast topo II and human topo IIα was also tested. The human enzyme was inhibited by 300 μM amsacrine but was unaffected by compounds 2 and 25 up to 1 mM, while the yeast enzyme was inhibited by 200 μM etoposide, 10 μM compound 2, and 10 μM compound 25. To calculate accurate IC_{50} values for compounds 2 and 25 for yeast topo II DNA relaxation, nine-point two-fold dose-response reactions were performed in triplicate and these yielded values of 2.9 μM and 13.1 μM, respectively (**Figure 6.13**). Given their inactivity against the human enzyme and their potency against the yeast enzyme, these compounds may have potential as fungicides.

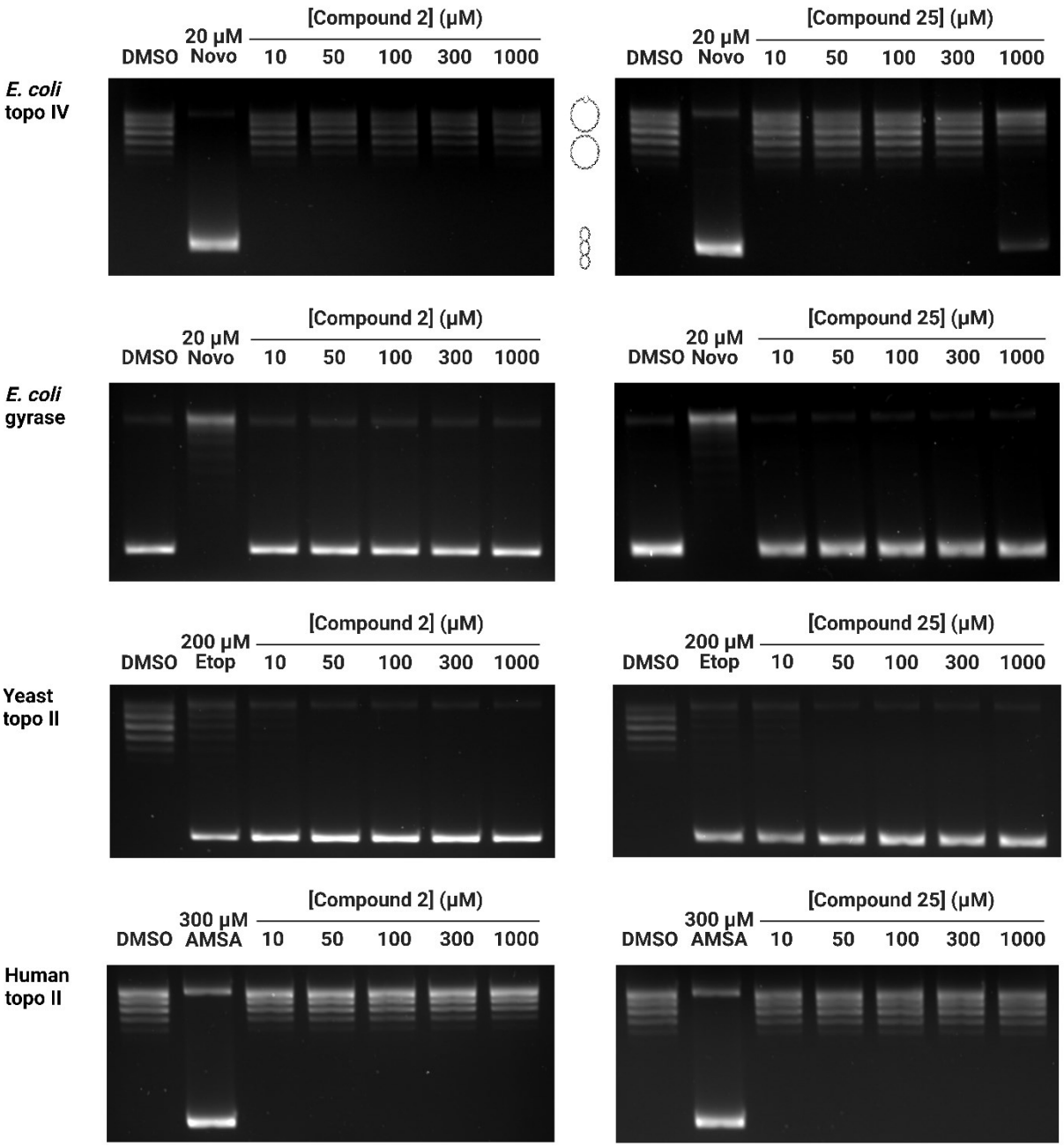

Figure 6.12| The effect of compounds 2 and 25 on type II topoisomerases. *Compounds 2 and 25 were tested for their effect on the relaxation of 6.7 nM -scDNA by **E. coli** topo IV, yeast topo II, and human topo IIα, and for the supercoiling of 6.7 nM relaxed DNA by **E. coli** gyrase. Novobiocin (Novo), Etoposide (Etop) and Amsacrine (AMSA) were used as positive controls. Reactions were run at 37°C for 45 min.*

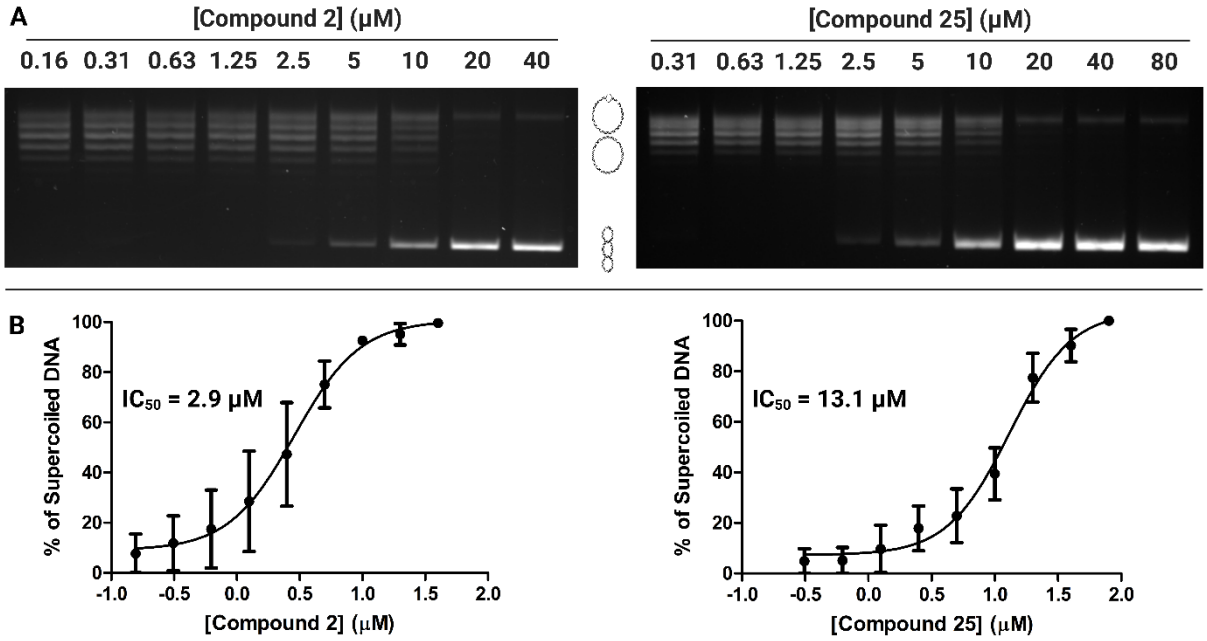

Figure 6.13| Dose-response gels and curves for compounds 2 and 25 against yeast topo II. *(**A**) Nine-point two-fold dose-response reactions were performed for compounds 2 and 25 for the relaxation of 6.7 nM -scDNA by yeast topo II using the gel-based assay. (**B**) Reactions were run at 37°C for 45 min and performed in triplicate. Band intensities were used to generate dose-response curves. Error bars represent standard deviation.*

6.2.7 Investigating the mechanism of action of the Syngenta compounds

It is necessary to understand the mechanism of action of an inhibitor to develop it as a potential drug or herbicide candidate. I therefore tested compounds 2 and 25 for their ability to poison MmTopo VI and intercalate into DNA. To our knowledge, there are no published examples of cleavage-complex stabilising compounds for MmTopo VI. Therefore, the cleavage complex of this enzyme has instead been trapped by incubation with ADPNP[58], a non-hydrolysable ATP analogue, as was first demonstrated with *S. shibatae* topo VI[185]. By titrating MmTopo VI in a cleavage assay, in the presence of 6.7% DMSO, the formation of linear DNA products was observed at an enzyme concentration of 32 nM (**Figure 6.14**). Increasing the concentration of compound 2 or 25 from 10 μM to 500 μM decreased the ADPNP-induced formation of linear products by MmTopo VI at 32 nM (**Figure 6.14**). Under the same conditions, but in the presence of ATP instead of ADPNP, no cleavage by 64 nM MmTopo VI was observed with or without compound 2 or 25 (**Figure 6.14**). Both compounds are therefore unlikely to be acting as topo poisons.

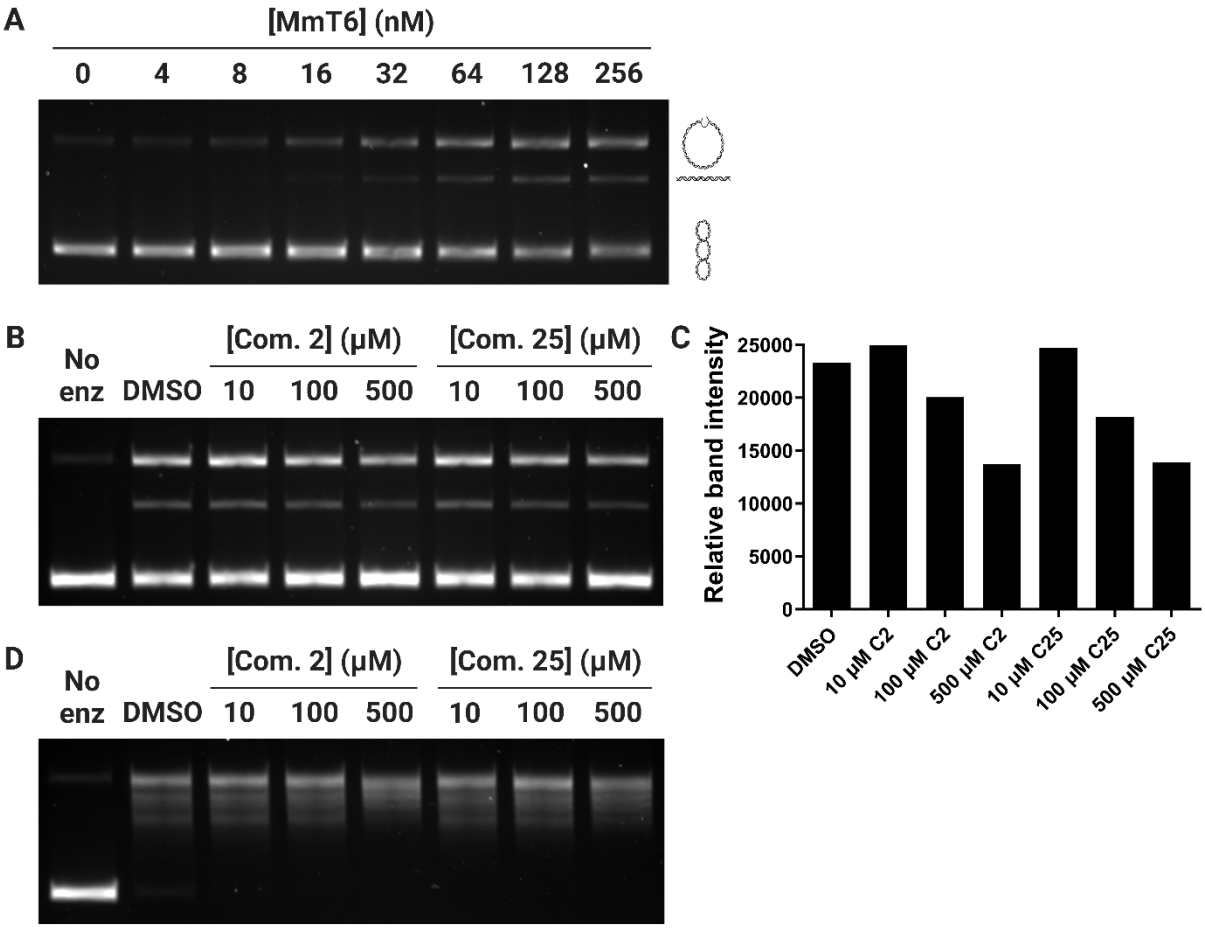

Figure 6.14| *Methanosarcina mazei* topo VI poisoning test. *(**A**) Titrating **M. mazei** topo VI in the presence of ADPNP and 5 nM -scDNA produces linear DNA fragments. (**B**) The effects of compounds 2 and 25 on the formation of these linear products was tested. (**C**) The band intensities of the ADPNP-induced cleavage products in the presence of compounds 2 and 25 were plotted on a bar graph. (**D**) The cleavage assay was also performed in the presence of ATP instead of ADPNP. Reactions were run at 37°C for 1 hr.*

Compounds that intercalate into DNA induce local lengthening and unwinding of the DNA which decreases the twist number. As the linking number must remain fixed, the decrease in twist is compensated for by increasing the writhe number which manifests as the formation of positive supercoils. This phenomenon can be exploited to test whether a compound with an unknown mechanism is a DNA intercalator. In the presence of an intercalating agent, such as ethidium bromide (EtBr) or amsacrine, relaxed DNA will become +sc. If the reaction is incubated with a topo, such as wheatgerm (WG) topo I, the writhe number becomes zero as the supercoils are relaxed and the linking number decreases. By removing the compound and topo from the reaction using an organic solvent, the local DNA unwinding is reversed which increases the twist number and induces a decrease in writhe to keep the linking number

constant. This forces the DNA to become more -sc. Therefore, the presence of an intercalating agent in this assay will cause relaxed DNA to run faster on an agarose gel.

To determine the appropriate concentration of WG topo I required for this assay, I performed an enzyme titration. WG topo I relaxed -scDNA following a 1:225 dilution of the neat enzyme (**Figure 6.15**). The DNA intercalating assay was then employed to demonstrate that 5 $\mu g.mL^{-1}$ EtBr or 300 µM amsacrine can expedite the migration of relaxed DNA on an agarose gel, as expected for an intercalating agent (**Figure 6.15**). However, under the same reaction conditions at a concentration of 50 µM, at least four-fold greater than their highest IC_{50} values, compounds 2 and 25 caused no change in the mobility of relaxed DNA, and therefore do not seem to be intercalating into the DNA.

If a compound is both a DNA intercalator and a WG topo I inhibitor, then the linking, twist, and writhe numbers of the original relaxed DNA will remain unchanged upon removal of the compound and enzyme. In such case, the migration of the relaxed DNA on an agarose gel will be unaffected, giving rise to a false negative. Therefore, it is crucial to test whether compounds 2 and 25 are inhibitors of WG topo I before they can be characterised as non-intercalating agents. At a concentration of 50 µM, compounds 2 and 25 failed to inhibit the relaxation of -scDNA by WG topo I (**Figure 6.15**), confirming the result of the DNA intercalation test. To determine whether the compounds have activity against this enzyme at higher concentrations, I performed triplicate dose-response reactions. Compounds 2 and 25 indeed had activity at a higher concentration range, with IC_{50} values of 148 µM and 164 µM, respectively, for WG topo I -scDNA relaxation (**Figure 6.16**). Therefore, both Syngenta compounds are non-DNA intercalators but are weak inhibitors of WG topo I.

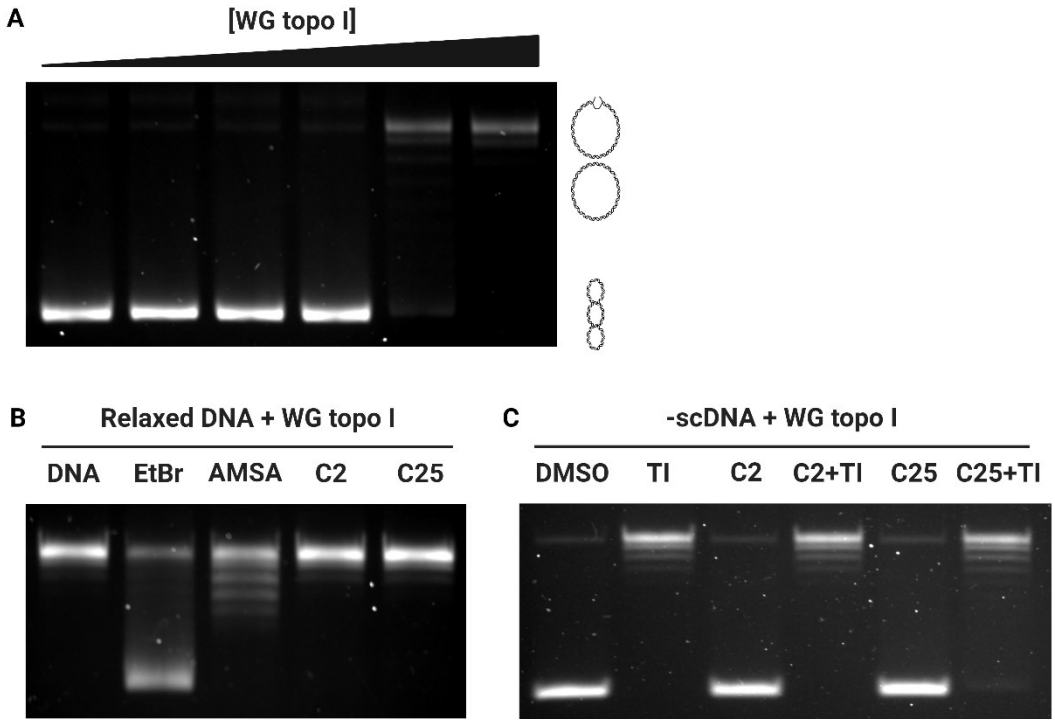

Figure 6.15| Test for DNA intercalation. *(**A**) Wheat germ (WG) topo I (TI) was titrated in a -scDNA (6.7 nM) relaxation assay. (**B**) Compounds 2 and 25 at a concentration of 50 µM were tested for their ability to intercalate into DNA alongside the positive controls ethidium bromide (EtBr) and amsacrine (AMSA). (**C**) Compounds 2 and 25 were also tested for their ability to inhibit the relaxation activity of WG TI. Reactions were run at 37°C for 45 min.*

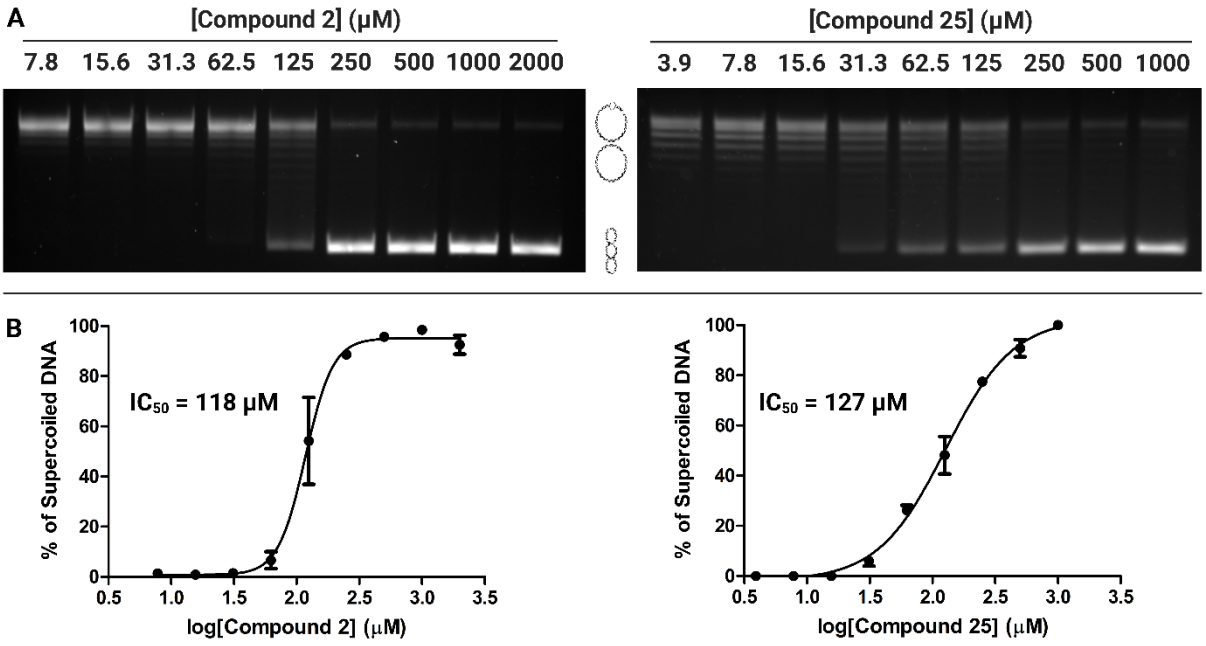

Figure 6.16| Dose-response gels and curves for compounds 2 and 25 against wheat germ topo I. *(**A**) Nine-point two-fold dose-response reactions were performed for compounds 2 and 25 for the relaxation of 6.7 nM -scDNA by wheat germ topo I using the gel-based assay. (**B**) Reactions were run at 37°C for 45 min and performed in triplicate. Band intensities were used to generate dose-response curves. Error bars represent standard deviation.*

6.2.8 The effect of the Syngenta compounds on *Arabidopsis thaliana* topo VI

The high-throughput H19 assay screened a library of compounds that had previously been shown to induce a dwarfing phenotype in whole plant assays to enrich for potential inhibitors of plant topo VI. To prove that topo VI is indeed being targeted by compounds 2 and 25 within the plant, it is necessary to isolate the enzyme for in vitro characterisation. However, there are no published protocols for successfully expressing and purifying topo VI from a plant species and my attempts to express topo VI from *A. thaliana* and *Oryza sativa* in Chapter 2 have been unsuccessful. Nevertheless, Monica Agarwal (John Innes Centre, UK) in the Maxwell research group (personal communication) has been able to express soluble *A. thaliana* topo VI (AtTopo VI), alongside its accessory proteins RHL1 and BIN4, using SuperSf9 insect cells and has performed a single step purification by immobilized-nickel affinity chromatography. This sample of AtTopo VI has been shown to perform relaxation of -scDNA and decatenation of kDNA (data not shown).

Before testing the effect of compounds 2 and 25 on the sample of AtTopo VI, it was first necessary to confirm its enzymatic activity in my hands. Under the same conditions as the relaxation assays performed in this chapter, a titration of AtTopo VI produced nicked and linear products rather than the expected relaxed topoisomers (**Figure 6.17**). If relaxed DNA was present in these samples, they could be observed by running on an agarose gel with EtBr. In the presence of EtBr, the relaxed topoisomers would run faster and as a single band, however this was not observed (**Figure 6.17**). The concentration of DMSO in these reactions was 6.7% (v/v), which is significantly higher than the 0.1% DMSO (v/v) used by Monica. Therefore, a DMSO titration was performed to determine the effect of the solvent on the DNA relaxation activity of AtTopo VI. At low DMSO concentrations, a series of relaxed topoisomers was observed, validating the activity of the preparation of AtTopo VI (**Figure 6.18**). This relaxation activity was observed on the EtBr agarose gel as a small shift when compared to -scDNA and was hampered above a DMSO concentration of 1.7% (v/v). However, even at low DMSO concentrations, there were still substantial nicking and linear products being produced. Therefore, the sample of AtTopo VI used in these assays has very weak topo activity and is heavily contaminated, and it will be difficult to observe the interconversion of relaxed and -sc topo isomers by this enzyme.

Performing AtTopo VI relaxation reactions at a DMSO concentration of 1.6% (v/v) would produce observable relaxed topoisomers on an agarose gel and permit the testing of compounds 2 and 25 at a concentration of up to 500 µM, and so this concentration was carried forward for future assays. To determine the appropriate concentration of AtTopo VI to be used in a relaxation assay at 1.6% (v/v) DMSO, an enzyme titration was performed. The addition of 4 µL of the one-step purified AtTopo VI sample was enough to remove negative supercoils and produce relaxed DNA, as demonstrated by the subtle band shift between the addition of 2 µL and 4 µL of AtTopo VI on the EtBr agarose gel (**Figure 6.19**).

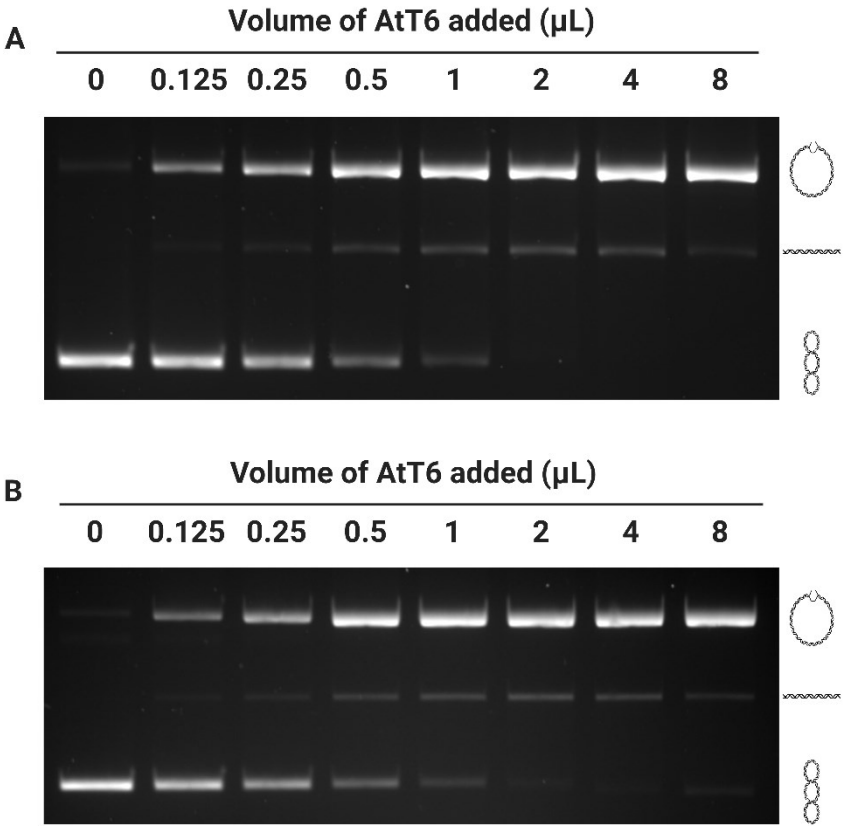

Figure 6.17| *Arabidopsis thaliana* topo VI titration in 6.7% DMSO. *A sample of one-step purified **A. thaliana** topo VI was titrated in a -scDNA (6.7 nM) relaxation assay in 6.7% DMSO. Reactions were analysed by agarose gel electrophoresis in the absence (**A**) and presence (**B**) of ethidium bromide. Reactions were run at 37°C for 45 min.*

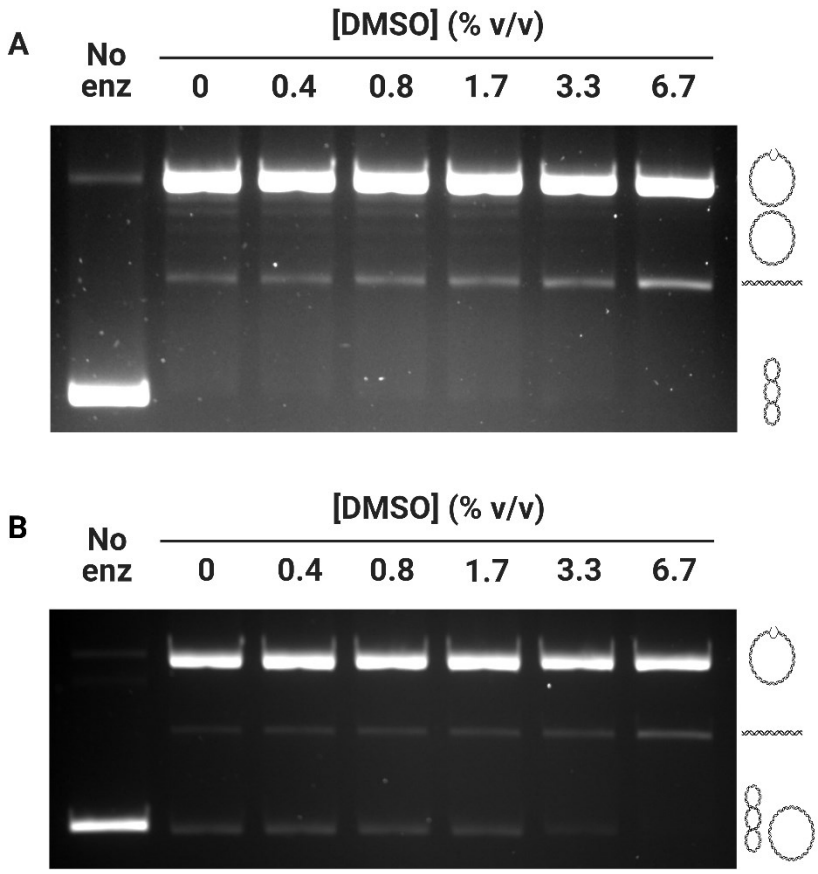

Figure 6.18| DMSO titration in an *Arabidopsis thaliana* topo VI relaxation assay. *DMSO was titrated in an **A. thaliana** topo VI -scDNA (6.7 nM) relaxation assay. 4 µL of the preparation of **A. thaliana** topo VI was added to each reaction, and reactions were analysed by agarose gel electrophoresis in the absence (**A**) and presence (**B**) of ethidium bromide. Reactions were run at 37°C for 45 min.*

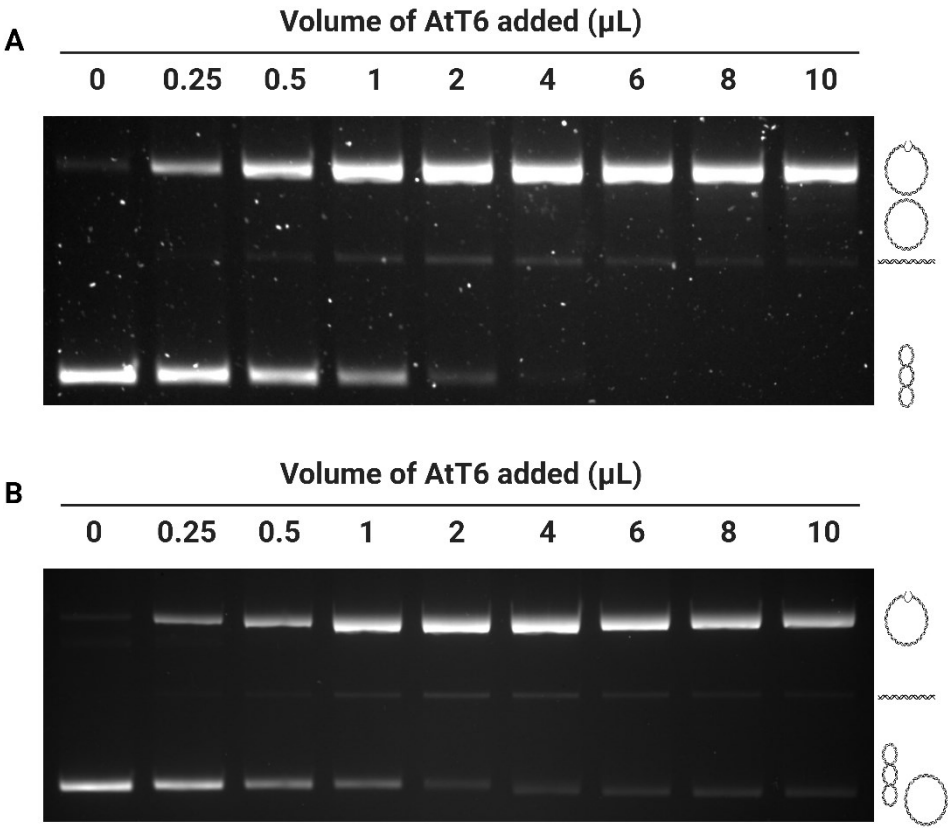

Figure 6.19| *Arabidopsis thaliana* topo VI titration in 1.6% DMSO. *A sample of one-step purified **A. thaliana** topo VI was titrated in a -scDNA (6.7 nM) relaxation assay in 1.6% DMSO. Reactions were analysed by agarose gel electrophoresis in the absence (**A**) and presence (**B**) of ethidium bromide. Reactions were run at 37°C for 45 min.*

Having validated the DNA relaxation activity of the preparation of AtTopo VI, I next tested the effect of compounds 2 and 25 on the enzyme by using 4 µL of AtTopo VI and 1.6% (v/v) DMSO in each reaction mixture. For compound 2, inhibition of the removal of negative supercoils was observed at 100 µM and the production of relaxed topoisomers was inhibited at 500 µM (**Figure 6.20**). On an EtBr agarose gel, a clear band shift was observed between 100 µM and 500 µM that distinguishes between -scDNA and relaxed DNA. Furthermore, compound 2 had no effect on the generation of nicked and linear products by the enzyme sample. For compound 25, however, no inhibition of the removal of negative supercoils was observed up to a concentration of 500 µM. Omitting ATP from the reaction diminished the production of relaxed topoisomers by AtTopo VI but did not abolish this activity completely. Failing to add $MgCl_2$ to the reaction buffer greatly enhanced the production of relaxed topoisomers by the enzyme preparation and significantly reduced the generation of nicked and linear products.

This result was surprising but concrete conclusions cannot be drawn from it given the impure nature of the AtTopo VI sample.

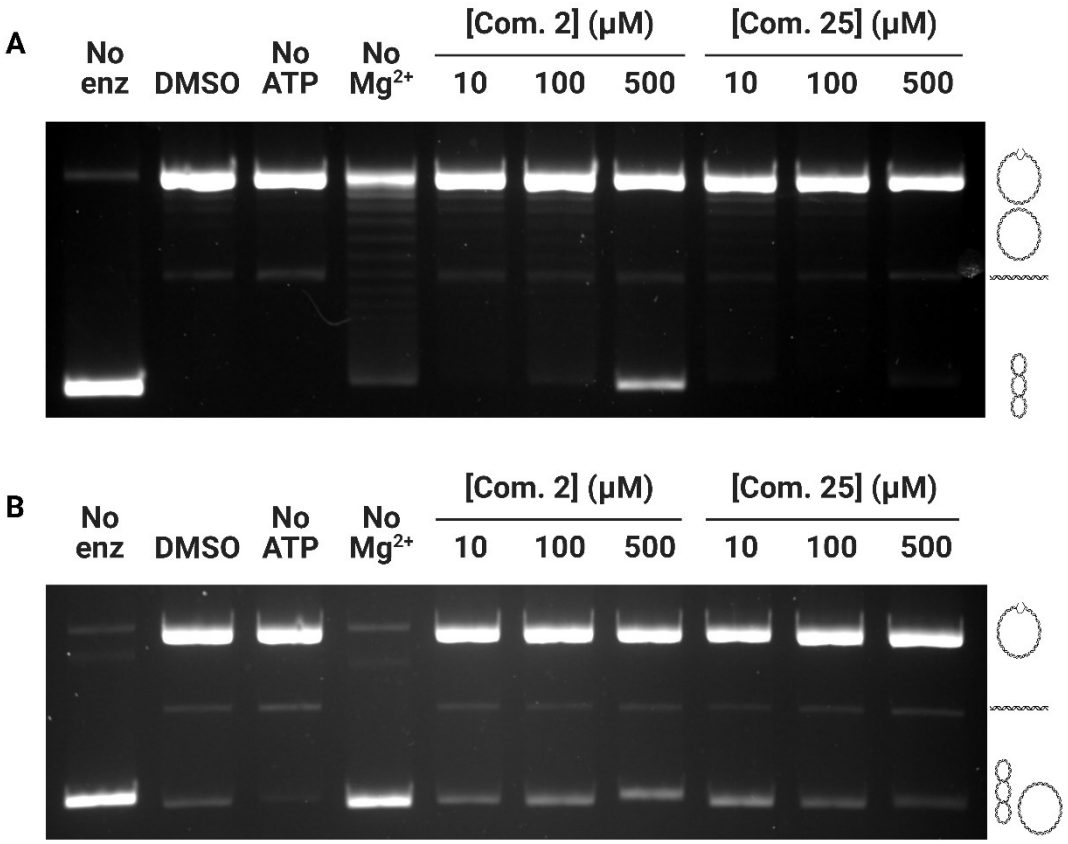

Figure 6.20| The effect of compounds 2 and 25 on *Arabidopsis thaliana* topo VI. *Compounds 2 and 25 were tested for their effect on the relaxation of 6.7 nM -scDNA by **A. thaliana** topo VI. Reactions were analysed by agarose gel electrophoresis in the absence (**A**) and presence (**B**) of ethidium bromide. DMSO was used as a negative control and the effect of omitting ATP or $MgCl_2$ was tested. 4 μL of the preparation of **A. thaliana** topo VI was added to each reaction. Reactions were run at 37°C for 45 min.*

6.3 Discussion

6.3.1 General discussion

The type IIA topos are proven drug targets and have been successfully exploited for anticancer and antibacterial therapies[232]. For many years topo VI was thought to be an enzyme unique to archaea and so was largely neglected as an "archaeal curiosity"[191]. Disinterest in topo VI has left many gaps in our understanding of the enzyme's mechanism and very few inhibitors have been described. However, this oversight was realised following the publications of the complete genomic sequences of the model plant *A. thaliana*[175] and the malaria parasite *Plasmodium falciparum*[138] in 2000 and 2002, respectively. Using these resources it took less than a year to discover that gene products homologous to topo VI-A and topo VI-B were present in both organisms[49,129] and topo VI was later shown to be ubiquitous in plants[49] and Apicomplexa[50], a phylum of parasitic protists. Suddenly, what was just an archaeal curiosity was now an enzyme of agricultural and pharmacological interest. If the clinical success stories of the type IIA topos can be imitated by this type IIB enzyme, then topo VI may become a useful target for the development of novel herbicides and antiprotozoals.

A key component of inhibitor discovery is the isolation of an enzyme of interest, however there are no published protocols for successfully expressing and purifying topo VI from a eukaryotic species. Furthermore, my own attempts to express topo VI from *A. thaliana*, *O. sativa* and *P. falciparum* in Chapter 2 have been unsuccessful. Therefore, the hunt for novel topo VI inhibitors must utilise the archaeal enzymes with the aim of identifying compounds that can later be shown to exhibit cross activity against the eukaryotic homologues. Radicicol was the first topo VI inhibitor to be characterised and was shown to disrupt the DNA relaxation and decatenation activities of the *S. shibatae* enzyme[264], while having very little activity on MmTopo VI[202]. The quinolone CP-115,955 was also identified as an inhibitor of *S. shibatae* topo VI[46], but both CP-115,955[279] and radicicol[280] also target human topo IIα and so are unsuitable for herbicide or drug development.

S. shibatae topo VI has an optimum activity at ~80°C which is challenging to maintain in experiments. Therefore, a better model for eukaryotic topo VI is the enzyme from *M. mazei*, which can be purified in high yields to near homogeneity and performs optimally at just ~37°C[202]. Many of the compounds that have been shown to inhibit MmTopo VI, such as the

antimalarial quinacrine[191,254], the antiseptic 9-aminoacridine[191,254], the African sleeping sickness drug suramin[191,260], and the chemotherapeutic drugs mitoxantrone[191,247] and amsacrine[191,249], have all been shown to target human topo II. The anthraquinone purpurin was also shown to have activity against MmTopo VI[191] but is an inhibitor of human adipocyte-derived leucine aminopeptidase[281], which plays a key role in angiogenesis[282].

Another candidate for the targeting of eukaryotic topo VI is the local anaesthetic hexylresorcinol, which is active against MmTopo VI[191]. Hexylresorcinol was tested for potency against *A. thaliana* seedlings and was shown to induce the same dwarf phenotype observed in topo VI knock-outs[115-117,191]. This raises the possibility that this compound is targeting AtTopo VI. Unfortunately, hexylresorcinol has also shown activity in human cell lines against transglutaminase-2[283] and nuclear factor-κB[284], which are key regulators of apoptosis[285,286]. Therefore, all published inhibitors of MmTopo VI are likely to be cytotoxic and unsuitable for repurposing for herbicide development. Nonetheless, given that quinacrine and hexylresorcinol can inhibit MmTopo VI and that they are active against the topo VI-possessing genera *Plasmodium* and *Arabidopsis*, respectively, it seems plausible that the archaeal enzyme would be a suitable surrogate for the identification of novel inhibitors for eukaryotic topo VI.

6.3.2 Development of a fluorescence-based high-throughput relaxation assay

It is crucial that more topo VI inhibitors are identified for the enzyme to become a recognised target for herbicides and antiprotozoals. To determine whether MmTopo VI is a useful model for identifying novel inhibitors of topo VI from plants and Apicomplexa, it was first necessary to identify a suitable method for compound screening. The traditional gel-based assays are useful for visualising the interconversion of relaxed and supercoiled DNA by topos but are slow and labour-intensive. Higher-throughput topo assays have been developed that permit the rapid screening of large numbers of compounds in a microplate but these methods suffer drawbacks of their own.

The use of hairpin-forming regions on a plasmid to separate a fluorophore-quencher pair[269,270] seems an attractive way to measure the superhelical topological state of a DNA substrate in the presence of topos. These hairpins are stabilised by DNA supercoiling giving rise to differential fluorescence intensities when the plasmid is in a supercoiled or a relaxed state without the need for additional reagents. However, the methods required to generate these

plasmids are either very inefficient[269] or involve the use of expensive components[270]. Other methods have exploited the differential formation of triplex DNA structures in -scDNA and relaxed DNA under different conditions[273,275]. Triplex-forming oligonucleotides immobilised on a microplate[273] or free in solution[275] can preferentially capture supercoiled DNA or relaxed DNA depending on the pH and concentration of metal ions, altering the emission of a fluorescent dye. These methods also suffer weaknesses, such as multiple washing steps[273], additional quenching steps[275], or high concentrations of expensive plasmid[287]. A simpler method involves probing the topological state of DNA by incubation with the commercial dye PicoGreen, which exhibits a differential fluorescence spectra in response to its binding to -sc or relaxed DNA[267]. However, this dye falls short due to a small difference in fluorescence intensity between -sc and relaxed DNA.

To identify inhibitors of MmTopo VI, I sought to develop a novel high-throughput screening assay that overcomes the obstacles with the existing assays. The PicoGreen dye method solves many of the problems with the fluorescent-plasmid and triplex-forming assays by abolishing the need for washing and quenching steps and by utilising a cheaper plasmid substrate but is unreliable for distinguishing between -sc and relaxed DNA. A key component of the existing assays is the use of a fluorescence-based measuring method which allows the scaling-up of an assay to test large amounts of compounds on a single microplate, and this feature should be retained. Although the primary reaction performed by MmTopo VI is decatenation[172], its DNA relaxation activity is more amenable to high-throughput screening as the -sc and relaxed topological states of DNA are easier to monitor. Therefore, a fluorescence-based high-throughput relaxation assay was developed that retains the benefits of the PicoGreen assay while improving the signal-to-noise ratio.

The assay described in this chapter utilises the commercial dye “H19” (ProFoldin), which exhibits a preferential enhancement of fluorescence in the presence of -sc DNA compared to relaxed DNA. No quenching or washing steps are required and the assay only requires a short incubation with the H19 dye upon completion of the topo reaction before measuring the fluorescence. While PicoGreen permits only a 25% difference in fluorescence intensity between -sc and relaxed DNA, H19 offers a greater than two-fold change (**Figure 6.1**). The assay also grants the use of a standard -sc plasmid substrate, which is simpler and cheaper to produce than the fluorescent and triplex-forming plasmids.

The novel fluorescence-based high-throughput relaxation assay described here successfully demonstrated the ATP-dependent relaxation of -scDNA by MmTopo VI (**Figure 6.2**), but before it was utilised for inhibitor screening it was necessary to optimise its components. The -scDNA was the most expensive constituent of the assay and was purchased from a supplier (ProFoldin) but can be produced in-house. The H19 supplier protocol recommends performing reactions with 25 nM DNA but the same difference in fluorescence between -sc and relaxed DNA was maintained as low as 11.7 nM -scDNA which equates to a saving of £107 per 384-well plate (**Figure 6.4**). A further £45 per plate can be saved by reducing the -scDNA concentration to 6.7 nM at a cost of 10% of the signal-to-noise ratio, although this reduction is rapidly enhanced at lower concentrations (**Figure 6.4**). This small change in the difference in fluorescence between -sc and relaxed DNA will save significant amounts of money if screening compounds in the tens or hundreds of thousands, while still permitting the identification of appropriate hits.

The other components considered for optimisation were the H19 dye itself and the H19 dilution buffer (DB). The H19 supplier protocol recommends performing reactions with 62.5 nL of dye per well and using any less than this results in a significant reduction of the signal-to-noise ratio (**Figure 6.4)**. The dye is also relatively cheap in comparison to the cost of the DNA, and so the recommended volume should be retained for this assay. However, due to a limited available quantity of this dye at the time, I performed each reaction with only 48.6 nL of H19 to permit the screening and testing of more compounds. This reduction in volume cost 27% of the difference in fluorescence between -sc and relaxed DNA and would not be recommended for the assay given a greater supply. Replacing the H19 DB with water decreases the signal-noise-ratio by up to 15% and does not reduce the number of steps in the protocol so should be avoided in this assay. The supplier recommends adding 62.5 nL of H19 in 93.75 µL of 1X H19 DB per well, however, with the resources available to me at the time, the dye could only be added to the microplate with a multichannel pipette with a maximum volume of 10 µL. Therefore, 62.5 nL of H19 was added to each well in 10 µL of 1X H19 DB. This reduction of volume would have cost a maximum of 15% of the signal-noise-ratio and could be avoided with alternative sample handing equipment. The other components present in the MmTopo VI reaction buffer were already optimised[202] and so were not investigated.

Although this assay has been optimised for MmTopo VI, it could rapidly be adapted to measure the relaxation of -scDNA by other topos, such as bacterial topo IV or human topo II for antibacterial and chemotherapeutic drug screening, respectively. This assay could also be modified to measure the supercoiling of relaxed DNA by gyrase, but the concentration of relaxed DNA will need to be optimised. The type II topos are essential for removing the positive supercoils that form ahead of replication forks and transcription complexes[3] and topo IV[288], topo II[222], and topo VI[172] have all shown an increase in processivity with +scDNA than with -scDNA. The type IIA enzymes relax positive supercoils 10-20-fold faster than negative supercoils[222,288], while a 2-3-fold increase is seen with the type IIB enzyme[172]. Therefore, it would be reasonable to consider the use of +scDNA as a substrate for the screening of inhibitors of these enzymes. However, given that the rate of decrease of fluorescence by the H19 dye decreases as the DNA writhe number increases towards zero, there is unlikely to be a significant difference in fluorescence as this writhe becomes positive. A +scDNA substrate is also much more difficult and expensive to produce than -scDNA, and so overall would be inappropriate for use in this assay.

Identifying novel inhibitors of MmTopo VI will allow us to gain a better understanding of the enzyme's mechanism, however, compounds that target this enzyme alone will be of no interest to the pharmaceutical and agricultural industries. It is therefore essential that these novel inhibitors also show cross activity against the eukaryotic enzymes. Disrupting topo VI in plants induces a severe dwarf phenotype[115-117] and this phenomenon was exploited in the high-throughput screen. By only screening Syngenta compounds that have been shown to trigger a stunted growth symptomology in whole plant assays but have unknown modes of action, a compound library can be enriched for potential inhibitors of plant topo VI. The aim of this methodology is to increase the chance that novel inhibitors of MmTopo VI identified in the assay will also be active against the plant enzyme. Furthermore, it is often difficult to translate in vitro hits to hits with activity in whole organisms, and this can be mitigated by starting with bioactive compounds. If appropriate cross activity can be observed either with the isolated eukaryotic enzymes or in in vivo studies, then MmTopo VI will be validated as a suitable model for identifying novel inhibitors of eukaryotic topo VI. However, this methodology is unlikely to identify topo VI poisons as compounds capable of stabilising the cleavage complex of plant topo VI would be expected to induce a more severe phenotype.

This is a concern for two reasons: firstly, the absence of published topo VI poisons is hindering the study of the enzyme's cleavage mechanism; and secondly, potent poisons of topo VI are likely to be effective cytotoxic agents and should not be overlooked. Another limitation of enriching for potential inhibitors of plant topo VI is that by doing so may inadvertently reduce the number of identified MmTopo VI inhibitors that have cross activity against the homologous enzymes in Apicomplexa. Therefore, once the *M. mazei* enzyme has been validated as an appropriate model for eukaryotic topo VI, the high-throughput assay described in this chapter should be performed on compound libraries with different biases or no bias at all. Using the Malaria Box compound library[289], for example, will enrich for potential inhibitors of *Plasmodium* topo VI.

6.3.3 Identification of inhibitors of *Methanosarcina mazei* topo VI

The fluorescence-based high-throughput relaxation assay described here identified two potent inhibitors of MmTopo VI, namely 'compound 2' and 'compound 25', with IC_{50} values of 6.9 µM and 61.3 µM, respectively (**Figure 6.6**). The gel-based relaxation assay measures the interconversion of DNA topoisomers more directly and so is trusted more for dose-response analyses. Furthermore, showing inhibition in an orthogonal assay with a more quantitative readout demonstrates that the compounds are not just artefacts of the fluorescence-based assay and that they are genuine inhibitors. For the relaxation of -scDNA by MmTopo VI, the gel-based assay gave IC_{50} values of 4.6 µM and 5.8 µM (**Figure 6.9**). Although the IC_{50} values for compound 2 for the two methods are comparable, the fluorescence method gave a value more than ten-fold greater than the gel-based method for compound 25. This is likely to have been caused by compound deterioration after repeated freeze-thaw cycles, a phenomenon that I eventually discovered affected compound 25 and not compound 2. It is possible that the sulphide of compound 25 is getting oxidised over time, which is decreasing its potency. Also, the enzyme concentration was lowered from 2 nM to 1.6 nM for the gel-based assay, which would also have reduced the IC_{50} values for both compounds.

There is always a concern when performing high-throughput screening assays that compounds that seem to behave normally in dose-response analyses are in fact false positives that have poor specificity and form multiple weak contacts with the target enzyme. These compounds appear as hits in many different screens and have been termed 'promiscuous

inhibitors'[290]. Often these compounds are inactive in their native states but form inhibitory aggregation species at higher concentrations and are thus undesirable as lead compounds. Both compound 2 and compound 25 were inactive against bacterial and human type II topos (**Figure 6.12**) and were shown to inhibit MmTopo VI in a time-dependent manner (**Figure 6.10**) suggesting that these compounds are not promiscuous inhibitors. Furthermore, a structural analogue of compound 2 that is missing one of the six-carbon ring moieties was shown to inhibit MmTopo VI (**Figure 6.6**) which further demonstrates that compound 2 is indeed binding to a specific site on the enzyme. Herbicides and antimicrobial drugs must not target human topo II, or other essential human enzymes, to avoid unwanted cytotoxicity and thus compounds 2 and 25 are the only known inhibitors of MmTopo VI that currently have no known chemotherapeutic properties (**Table 6.4**).

Table 6.4| Inhibitors of *Methanosarcina mazei* topo VI

Inhibitor	IC50 (μM)	Potential chemotherapeutic properties
Purpurin	40	Yes
Amsacrine	30	Yes
Hexylresorcinol	30	Yes
Suramin	30	Yes
Quinacrine	8	Yes
9-Aminoacridine	6	Yes
Compound 25	5.8	Unknown
Compound 2	4.6	Unknown
Mitoxantrone	2	Yes

The topo I enzymes are essential in plants[291] and so the weak activity demonstrated by both compounds on WG topo I (**Figure 6.16**) may contribute to greater herbicidal activity but may also induce undesirable effects on crop germination. Furthermore, the presence of topo I in humans introduces the possibility of off target cytotoxic effects, and so the influence of the compounds on this enzyme should be tested. The potency of compounds 2 and 25 is more than ten-fold less for WG topo I than for MmTopo VI, therefore, these concerns may not materialise if a similar potency is observed for topo VI in plants. Although both compounds were inactive against topo II from a higher eukaryote, they were highly potent against yeast topo II with IC_{50} values of 2.9 μM and 10.8 μM for compound 2 and compound 25, respectively (**Figure 6.13**). Given that topo II is essential in yeast[292], the enzyme has been touted as a potential target for novel antifungal agents but compounds often show similar responses to

topo II from mammals and lower eukaryotes[293,294]. Evidently, enough differences between yeast and human topo II exist that both compounds 2 and 25 can selectively target the former, posing a potential avenue of these compounds in fungicide development. Without knowing the binding site of these compounds, it is difficult to rationalise their selectivity towards the yeast and archaeal enzymes compared to the human enzyme.

The effect of the Syngenta compounds was also tested on the relaxation of +scDNA and the decatenation of kDNA using the gel based-assays. The potencies of the compounds against these activities were a maximum of two-fold greater than against the relaxation of -scDNA (**Figure 6.9**), but it is difficult to draw any conclusions from the comparisons of these substrates. This is due to the difference in preparation quality of the supercoiled substrates and the fact that the decatenation assay is a less direct method of monitoring strand-passage events.

6.3.4 Characterisation of Syngenta compounds

Topo inhibitors are typically classified into poisons or catalytic inhibitors and are subdivided into DNA intercalators and non-intercalators. The compound library screened in this chapter was likely to enrich for catalytic inhibitors and this was corroborated by the reduction of ADPNP-induced cleavage activity of MmTopo VI in the presence of compounds 2 and 25 (**Figure 6.14**). In the absence of ADPNP no cleavage is also observed, confirming that the compounds are not inhibiting the enzyme by poisoning. Catalytic inhibitors of the type IIA topos either cause disruption of activity by binding to a specific site on the enzyme, such as the ATP-binding pocket in the case of compounds such as novobiocin[230] and dexrazoxane[257], or by intercalating into the DNA and inhibiting DNA binding or DNA cleavage, such as aclarubicin[255] and merbarone[256], respectively. Compounds 2 and 25 both failed to demonstrate DNA intercalative properties up to a concentration more than four-fold greater than their highest IC_{50} values (**Figure 6.15**), suggesting that both compounds are non-intercalating catalytic inhibitors. An interesting feature of this chapter is the similar behaviour exhibited by the two most potent identified Syngenta compounds, which share identical target enzyme specificity and comparable IC_{50} values against these enzymes. The similarities suggest that the compounds may have analogous mechanisms of action and it will likely take further biochemical approaches, such as X-ray crystallography, to elucidate them.

6.3.5 The effect of the Syngenta compounds on *Arabidopsis thaliana* topo VI

To validate the methodology of enriching for plant topo VI inhibitors and to endorse MmTopo VI as a suitable model for the screening of eukaryotic topo VI inhibitors, the Syngenta compounds must be tested on the plant enzyme in vitro. The one-step purified preparation of AtTopo VI tested here can relax negative supercoils at low DMSO concentrations but produces substantial nicking and linear products (**Figure 6.18**). The enzyme sample performed much better when fresh but had been through multiple freeze-thaw cycles before being passed to me. Although future preparations of this enzyme will be stored in aliquots, this current sample of AtTopo VI has deteriorated to the extent that agarose gels must be run with EtBr to observe the relaxation of -scDNA. Furthermore, the current method for isolating the enzyme yields an impure sample rife with contamination. Given that the nicking and linear products observed in my hands were not seen with the fresh sample, it is likely that these activities are caused by deterioration of the AtTopo VI complex. The preparation is more sensitive to DMSO than other type II topos tested in this chapter, although this would likely be solved by using a more concentrated enzyme.

Compound 2, and not compound 25, was able to inhibit the production of relaxed topoisomers by the AtTopo VI sample, although needed to be present at 500 µM to do so (**Figure 6.2**). This represents the first distinction of activity between the two compounds. Interestingly, while the relaxation activity of AtTopo VI was ATP-dependent, an enhancement of relaxation activity and a reduction of nicking and cleavage activity was seen when omitting $MgCl_2$ from the reaction buffer (**Figure 6.2**). Due to the large volume of crude enzyme preparation added to each reaction, it cannot be ruled out that there is still ATP and Mg^{2+} contamination present. Nonetheless, the concentration of divalent metal ions clearly has a significant effect on AtTopo VI and should be investigated further. Unfortunately, no concrete conclusions can be drawn from the effects of the Syngenta compounds on the AtTopo VI sample or from analyses of the enzyme itself until a more robust expression and purification protocol is developed.

6.3.6 Conclusion

In this chapter I have described the development and validation of a novel fluorescence-based high-throughput screening assay for topos. This assay is cheaper and quicker than alternative methods and produces an appropriate signal-to-noise ratio for the screening of large

compound libraries. A small library of compounds was tested and two potent catalytic inhibitors of MmTopo VI and yeast topo II were identified. These compounds should be investigated further as potential fungicides and given the presence of topo VI in plants and Apicomplexa, may also serve as useful herbicide and antiprotozoal agents. This work has laid the groundwork for lead compound identification in larger-scale screens for the development of novel pharmacotherapies. Future work should focus on testing the effect of the compounds on the growth of fungi, investigating their effect on DNA binding and ATP hydrolysis, and on co-crystallising the compounds with topo VI.

Chapter 7

Discussion

7.1 Introduction

In the October of 2018, I began my PhD journey by entering the wonderful world of DNA topoisomerases (topos). I had joined the John Innes Centre to study topo VI, a poorly characterised enzyme that had initially generated very little interest from the topo field. Two major discoveries overturned this disinterest: the presence of topo VI in eukaryotes, and the adoption of topo VI by the meiotic Spo11 complex. For the subsequent four years, I was tasked with advancing our knowledge of eukaryotic topo VI to determine its suitability as a target for herbicides and antimalarial compounds. During my PhD, I intended to answer the following key questions: does *Plasmodium*, the causative agent of malaria, possess topo VI? What is the function of plant topo VI? Can a eukaryotic topo VI be isolated? What is the significance of the supercoil chiral discrimination by topo VI? Can we find inhibitors of eukaryotic topo VI? To answer these questions, I embarked on several lines of investigation, including the bioinformatic analysis of the topo VI and Spo11 families, the molecular cloning and expression trials of topo VI, the biochemical characterisation of topo VI from the archaeon *Methanosarcina mazei*, and the development of a novel high-throughput screening assay. In this chapter, I will reflect on the key findings from this work and the implications of the results in a broader perspective, before discussing my recommendations for taking this project forward in the future.

7.2 Conclusions

7.2.1 General introduction

From four years of data collection during my PhD I yielded four results chapters, each one summarising an individual line of inquiry. In this section, I will briefly discuss the rationale for the strategies implemented in each of these chapters and the key conclusions that were drawn from them. I will also reflect on these strategies with hindsight and consider how I would address the aims of my project differently having gained the knowledge that I now have.

7.2.2 *Plasmodium* does not possess topo VI

In Chapter 3, I conducted an extensive phylogenetic analysis of topo VI and Spo11 sequences to determine whether *Plasmodium* possesses topo VI, and if so whether this enzyme would be a suitable target for antimalarial compounds. Using a combination of phylogenetic tree analysis and sequence alignments I showed that *Plasmodium*, and other members of the Apicomplexa phylum, possess two paralogous gene products that are more closely related to the meiotic Spo11 protein than the topo VI A subunit. Furthermore, I showed that the AlphaFold structural models of topo VI-B from bacteria and eukaryotes possess the same characteristic 3D domain architecture exhibited by the crystal structures of archaeal topo VI-B, including a helix-2-turn-helix (H2TH) domain. I then showed that members of the topo VIB-like family, including animal TOPVIBL and plant MTOPVIB, possess divergent AlphaFold structures which do not resemble topo VI-B nor include a H2TH domain. Therefore, I had determined that the H2TH domain was a key structural feature that could distinguish between topo VI-B and topo VIB-like. The AlphaFold structure of the putative topo VI-B subunits from *Plasmodium* and other Apicomplexa were missing this H2TH domain, and strongly resembled those of the MTOPVIB family, and were thus shown to be topo VIB-like subunits. Taken together, these results suggest that *Plasmodium* does not possess topo VI, and that its homologous topo VI subunits assemble into the meiotic Spo11 complex. Topo VI is therefore not an applicable target for antimalarial drugs, as was hoped at the start of this project. It remains a mystery, however, why topo VI has been lost in Apicomplexa lineages, but has been retained in other members of the Stramenopiles-Alveolates-Rhizaria (SAR) taxonomic supergroup.

7.2.3 Topo VI is widely distributed in bacteria and eukaryotes

While performing a phylogenetic analysis in the context of characterising the putative topo VI gene products in *Plasmodium*, I considered it necessary to obtain a comprehensive collection of topo VI and Spo11 sequences that would fully capture the landscape of these subunits across the three domains of life. In doing so, I had characterised for the first time the widespread distribution of topo VI in eukaryotes and bacteria and had lost the mindset that topo VI was just an archaeal enzyme that happened to appear in plants. Topo VI is indeed ubiquitous in archaea, however, I have shown that the enzyme is also ubiquitous in many eukaryotic lineages including a diverse array of protists. Furthermore, I showed that these eukaryotic topo VI sequences possess all the key conserved functional residues that have been characterised in the archaeal enzymes, suggesting that they are likely to be active complexes in these organisms. In bacteria, however, I showed that topo VI was widely distributed across numerous superphyla but only appears sporadically within them. This pattern illustrates the evolutionary retainment of topo VI in eukaryotes following their divergence from archaea, as opposed to the acquisition of topo VI in bacteria from archaea via horizontal gene transfer. Furthermore, topo VI can serve as a useful tool for tracking the evolutionary divergence of different eukaryotic lineages. An interesting finding of this work was that topo VI-possessing bacterial species usually lacked topo IV, however, I also discovered a few cases of bacterial species where both enzymes as well as gyrase were present. These organisms were therefore the first examples of bacteria that had been shown to possess three type II topos, although it remains unclear as to why this redundancy has occurred. The distribution of topo VI in bacteria and eukaryotes highlights a greater need to understand the function of this enzyme in these domains, and I will discuss this topic later in the chapter.

7.2.4 Eukaryotic topo VI requires two structurally homologues accessory proteins.

To investigate the function of eukaryotic topo VI, I first sought to characterise the two accessory proteins, RHL1 and BIN4, that were thought to be unique to the plant enzyme. Although the plant topo VI accessory proteins seemed to be essential for the enzyme[123,124], very little was known about their role, and I therefore used my newly acquired bioinformatic skills to determine what this role was. First, I searched for homologues of RHL1 and BIN4 outside of the plant kingdom but could not find them in the bacteria or archaea domains.

However, I did find RHL1 and BIN4 homologues in all eukaryotic lineages that I had previously shown to possess topo VI. Furthermore, I could not find these accessory proteins in any eukaryotic lineage where I found topo VI to be absent. Therefore, I had discovered that the presence of RHL1 and BIN4 in a eukaryotic species was sufficient to determine whether the organism possessed topo VI. Next, I used AlphaFold modelling to investigate the predicted structures of the accessory proteins and found that both proteins are largely disordered but possess a structurally conserved core domain. This suggests that RHL1 and BIN4 may be performing analogous roles within the eukaryotic topo VI complex, although this role remains undetermined. It is worth speculating that structural homologues of RHL1 and BIN4 also act as accessory proteins to the archaeal and bacterial topo VI complexes, although such sequences would be difficult to search for.

7.2.5 The expression of bacterial and eukaryotic topo VI is non-trivial

In Chapter 4, I attempted to address the major aim of my PhD project: isolating eukaryotic topo VI. Previously, the best attempt to do so involved using an insect-cell expression system to express topo VI and its accessory proteins from *Arabidopsis thaliana*, however, this system produced only modest yields that were not appropriate for multiple purification steps. Previous attempts had also failed to express *A*. *thaliana* topo VI (AtTopo VI) using a bacterial expression system, although here the accessory proteins were neglected. I therefore attempted to coexpress AtTopo VI in series with its two accessory proteins in *E*. *coli*, but this too yielded no success. I then shifted my attention to the expression of what I had previously considered to be the *Plasmodium falciparum* topo VI complex using a yeast expression system. This too yielded no success and given that I now think that this organism does not possess topo VI, I would have diverted my attention to using the yeast expression system on the *A*. *thaliana* enzyme instead. I was able, however, to express fragments of the AtTopo VI B subunit in *E*. *coli* cells, but these proved to be insoluble. I also attempted to express fragments of the *P*. *falciparum* topo VI-B homologue, but I would not have done so had I known at the time that it was a member of the MTOPVIB family. Although the successful isolation of eukaryotic topo VI would be very valuable, a considerable effort in the future would be required to do so.

I discovered in Chapter 3 that topo VI was much more widely distributed in the bacterial domain than first thought, and this led me to ponder whether a bacterial topo VI would be a

useful entity for the characterisation of this enzyme class. I therefore chose three bacterial species from three different superphyla with comparable growth conditions to *E*. *coli* that would be suitable for pulling topo VI genes from for recombinant expression in *E*. *coli*. However, following the theme of the eukaryotic enzymes, I was unable to express any of the bacterial topo VI subunits. It is worth considering whether future success could be obtained by expressing the A and B subunits in isolation, given that I had only attempted their coexpression. It may be the case that although coexpression is required for the archaeal and plant enzymes, bacteria topo VI may require individual subunit isolation and reconstitution. Given that bacterial topo VI was acquired from archaea via horizontal gene transfer recently in its evolutionary history[27], the enzymes show limited divergence from the archaeal enzymes. Therefore, in hindsight, the study of bacterial topo VI may not provide enough useful information that has not already been generated from the archaeal enzymes that would justify the effort required to isolate them. Future efforts should focus instead on the more valuable commodities that are the eukaryotic topo VI enzymes.

7.2.6 *Methanosarcina mazei* topo VI is a useful model for the study of eukaryotic topo VI and the meiotic topo VI-like complex

Archaeal biology remains largely unattractive due to its limited economic potential and is often disregarded. However, given the difficulties in isolating eukaryotic topo VI, archaeal homologues may serve as useful models. In Chapter 4, I demonstrated the easy with which high yields of topo VI from the archaeon *Methanosarcina mazei* can be isolated to near homogeneity, as well as the ease with which mutants and fragments of the enzyme can be isolated. Therefore, *M*. *mazei* topo VI (MmTopo VI) can serve as a model enzyme for eukaryotic topo VI, and the characterisation of this enzyme is likely to shed important insights into eukaryotic biology. Furthermore, given the homology between topo VI and the Spo11 complex, MmTopo VI can also serve as a model for meiotic double-strand break formation. Many key residues are conserved in eukaryotic topo VI, the Spo11 complex, and archaeal topo VI, and the biochemical significance of these residues can be easily tested by mutagenic analysis using MmTopo VI.

7.2.7 Positively-supercoiled DNA is the preferred substrate for ATPase stimulation in *Methanosarcina mazei* topo VI

In Chapter 5, I performed the pyruvate kinase/lactate dehydrogenase (PK/LDH) coupled assay to demonstrate that the ATPase activity of MmTopo VI was DNA-dependent. I also showed that an enhancement of ATPase activity was exhibited by MmTopo VI in the presence of supercoiled (sc) DNA compared to both linear and relaxed DNA, and that a ~2-fold enhancement is seen in the presence of positively (+)-sc DNA compared with negatively (-) sc DNA. These results were largely consistent with previous studies[58,172] and showed for the first time that the chiral selectivity of MmTopo VI for +scDNA was reflected in its ATPase activity. A concerning feature of these results was the exclusively linear rates of ATP hydrolysis that were determined in the presence of scDNA. MmTopoVI is unable to change the topology of the relaxed and linear DNA substrates under the conditions used in the PK/LDH assay, and so it is rational for the ATPase rates in their presence to remain constant. However, given that the rate of ATPase stimulation in the presence of sc DNA is greater than in the presence of relaxed DNA, then it would be expected that a decrease in the rate of ATPase activity be observed as the sc DNA substrate is relaxed. The observations in Chapter 5 that the rate of MmTopo VI ATPase activity remains constant as the sc DNA is relaxed, is inconsistent and puts doubt on the validity of the results. This phenomenon should be addressed before the stimulation of ATPase activity by +scDNA is fully accepted.

Another vulnerability with these results is the lack of an inhibitor to use as a control. Ideally, the rate of ATP hydrolysis would be performed alongside a known ATPase inhibitor, however, such a compound has not been identified for MmTopo VI. ATP hydrolysis is tightly coupled to strand-passage in this enzyme, and so any inhibitor of MmTopo VI strand-passage is likely to also inhibit ATP hydrolysis. Although there was no availability of a suitable MmTopo VI inhibitor to use as a control while performing these ATPase assays, I later identified two novel potent inhibitors of MmTopo VI in Chapter 6. With more time I would have determined the effect of these inhibitors on ATP hydrolysis by MmTopo VI using the PK/LDH assay which would have provided support for the results.

7.2.8 A novel high-throughput relaxation assay can identify novel inhibitors of *Methanosarcina mazei* topo VI

In Chapter 6, I discussed the development of a novel high-throughput relaxation assay for the identification of topo VI inhibitors that was conducted while on placement at the agrichemical company Syngenta. The aim of this chapter was to determine whether topo VI was a suitable target for herbicides, and MmTopo VI was used as a model in the absence of sufficient preparations of the plant enzyme. To enrich for potential inhibitors of plant topo VI, compounds were selected based on inducing an endoreduplication phenotype in vivo and having unknown mechanisms of action. This method was useful for determining whether compounds could target both MmTopo VI and plant topo VI, which would validate MmTopo VI as a model for herbicide discovery. However, this method also discounted potential poisons of plant topo VI, which would yield more severe phenotypes unrecognisable as defects in endoreduplication. Therefore, future compound screening should avoid enriching for potential targets of plant topo VI, to ensure that poisons are identified. This high-throughput assay utilised a commercial dye that exhibited a preferential enhancement of fluorescence in the presence of -scDNA compared to relaxed DNA. I optimised many components of the assay, including the DNA and dye concentrations, to reduce the cost of screening a 384-well plate from £315 to £150 while maintaining at least a 2-fold difference in the signal-background ratio. The key consideration for cost reduction was the DNA concentration, however I also used a lower dye concentration due to its limited supply. Future use of the assay should use the supplier-recommended dye concentration, as the reduction in the signal-noise background by reducing it is not compensated by significant cost savings.

In total, I screened 1750 Syngenta compounds with MmTopo VI, and identified 25 hits that reached a threshold of 70% inhibition at a concentration of 10 µg mL^{-1}. The two most potent compounds were shown to inhibit both MmTopo VI and yeast topo II, with IC_{50} values ranging from 2.9 µM to 13.1 µM, but to be inactive up to 1 mM on human and bacterial type II topos. I also showed that these inhibitors were not topo poisons or DNA intercalators, and so were likely to be catalytic inhibitors that interfere with DNA binding or ATP utility. Future work should focus on further characterising the mechanism of action of these compounds, such as their effect on the PK/LDH assay, as previously discussed, or the use of the electrophoretic mobility shift assay to assess their effect on DNA-binding. Ideally, I would have tested the two

compounds on the preparation of AtTopo VI made in the lab, however its low yield and purity left the effect of the inhibitors unverified. Once the isolation of AtTopo VI has been optimised, the enzyme should be implemented straight into this assay.

7.3 Future directions

7.3.1 General introduction

I was unable to achieve all the aims of my PhD project and there remain several outstanding questions that should be addressed. Furthermore, while accumulating data for my four results chapters, I discovered novel problems that I was unable to solve. In this section, I will give my recommendations for the key lines of inquiry that should be pursued for the future characterisation of topo VI and discuss the best strategies for optimising the development of this project.

7.3.2 The role of bacterial and eukaryotic topo VI

In archaea, topo VI serves as the enzyme responsible for decatenating newly replicated intertwined DNA molecules to permit chromosome segregation during mitosis[46]. Some archaea also possess a second type II topo, gyrase, which serves to maintain a -sc genome in these organisms[51-53]. In bacteria, the archetypal type II topo responsible for chromosome segregation is topo IV[7], and so it is not surprising that the presence of topo VI in bacteria is usually indicative of the absence of topo IV. In these organisms, topo VI has replaced topo IV as the chief double-stranded DNA decatenase. This is particularly interesting given that topo IV has not been transferred the other way to archaea and leads me to speculate that topo VI is the more efficient decatenase. Therefore, I believe a stringent comparison of the decatenation activities of topo VI and topo IV in the context of prokaryotic cellular biology would be very valuable, although difficult to put into practice. It would be more feasible, however, to test the complementation of topo VI and topo IV with the corresponding null mutants in bacteria and archaea, to determine whether topo VI is indeed the only enzyme capable of replacing the other. Unusually, a handful of bacteria possess topo IV and topo VI, and the biochemical characterisation of these enzymes specifically would be incredibly valuable in investigating the core differences between the two enzyme classes.

Disregarding the confusing nomenclature, the archetypal type II topo in eukaryotes is topo II[7], and like some archaea, eukaryotes can also possess gyrase which they use to maintain a -sc genome within plastids[193]. However, unlike in bacteria, the type II topo redundancy implicated by the presence of topo VI in plants can be explained by its apparent function in the cellular process of endoreduplication. Therefore, topo VI clearly possesses a specific feature that distinguishes it from topo II such that it has become specialised to function within endoreduplicating cells. Given that topo II has likely evolved from topo IV[8], it is worth speculating that topo VI is also a more efficient decatenase then topo II. However, what I find most perplexing is the fact that endoreplicating cells replicate their DNA in the absence of chromosome segregation, and so there does not seem to be the need for a specialised decatenase in these cells. Furthermore, not only have there been no unique DNA replication intermediates identified in endoreduplicating plant cells that topo VI could be required to resolve, but endoreduplication is also prevalent in eukaryotes that do not possess topo VI. It is also worth noting that topo VI is ubiquitous in many eukaryotic lineages that do not maintain the endoreduplicative cycles seen in plants, which may suggest that topo VI has a yet-to-be-identified function in eukaryotes that is independent of endoreduplication. Elucidating the function of eukaryotic topo VI remains a key aim that should be addressed, and future work should pursue this by optimising the isolation of AtTopo VI and investigating its mechanism in relation to that of MmTopo VI, topo IV, and topo II.

The eukaryotic topo VI complex requires two accessory proteins[123,124] and it is tempting to speculate that they play a role in further specialising topo VI and separating its activities from topo II. This may be by interfering with the enzyme's mechanism, or by targeting the enzyme to specific locations within the cell. It may be for this reason that the eukaryotic topo VI accessory proteins are not present in archaea and bacteria, as they may function in removing a type II topo redundancy that is unique to eukaryotes. Elucidating the role of the eukaryotic topo VI accessory proteins is another priority for the future development of this project, and insights may be gained from the biochemical characterisation of the AtTopo VI complex itself, or from expressing the accessory proteins alongside MmTopo VI.

7.3.3 The adoption of the topo VI scaffold by the meiotic double-strand break machinery

Type II topos introduce transient double-stranded breaks (DSBs) to manipulate the topology of DNA, and these breaks must be stringently controlled to avoid deleterious DNA damage. It

is therefore likely that a type II topo scaffold has been adopted by the meiotic DSB machinery for its tightly regulated mechanism of DNA cleavage. An important question that has yet to be resolved is why eukaryotes have chosen the topo VI template, and not that of another type II topo, for meiotic recombination. The major difference between the type IIA and type IIB topos is the number of dimer interfaces, with the type IIA and type IIB enzymes possessing three and two gates, respectively[7]. Gyrase is likely to have been the first type II topo that evolved in cells[8], suggesting that three gates are necessary for introducing negative sc into DNA. Topo IV is a degenerate gyrase that has lost the capacity to supercoil[194] while retaining three gates, and topo II is a fusion of the A and B subunits of topo II so has also retained three gates[7]. Having evolved from the context of supercoiling, it is plausible that type IIA enzymes other than gyrase possess an unnecessary third gate, while the type IIB enzymes, which evolved independent of supercoiling activity, evolved exclusively as a double-stranded DNA relaxase and decatenase for which a third gate would be a hindrance. This may explain why topo VI is the only non-supercoiling type II topo that is present in all three domains of life, why the enzyme can replace topo IV in bacteria, and why eukaryotes have acquired the enzyme in addition to topo II: a two-gate type II topo may be a more efficient supercoil relaxase and decatenase. However, MmTopo VI has been shown to be an extremely slow relaxase[58] and a less efficient decatenase than the type IIA topos[219,295]. It is possible that this inefficiency is unique to *M*. *mazei* and may not be true of topo VI enzymes from other species.

In the absence of a third dimer interface, a two-gated type II topo must have evolved a novel mechanism for controlling DSB formation and this may have made the complex more suitable for adoption by the meiotic DSB machinery. Furthermore, a third dimer interface is likely to be unnecessary for an enzyme that does not perform strand-passage and would therefore be a poor choice given the greater cost of producing and maintaining a significantly larger complex. It may just be the case that topo VI, and not topo IV nor topo II, was adopted by the meiotic DSB machinery due to its simplicity, which may have favoured an easier route into its dissolution as a topo. Loss of the helix-2-turn helix domain from the topo VI B subunit and a reduction of its capacity for ATP hydrolysis seems to have been enough to convert the complex into an enzyme capable of initiating homologous recombination. It remains unknown whether the B subunit of Spo11 has completely lost its capability for hydrolysing ATP, or whether it has evolved a slower mechanism of ADP and P_i release to control meiotic DSB

formation. It is plausible that the B subunit of topo VI provides an important mechanistic role independent of ATP hydrolysis that has been retained by the Spo11 complex. The significance of the homology between the topo VI and meiotic DSB machinery is an important occurrence that should be investigated more thoroughly, and MmTopo VI may be an ideal model for doing so. Future work could focus on converting the archaeal enzyme into a degenerate topo by mutagenesis and investigating its biochemistry in the context of meiotic DSB formation.

7.3.4 The suitability of topo VI as a herbicide target

The rise of herbicide-resistant crops has increased the urgency at which herbicidal compounds with novel modes of action must be identified to protect global food security. Given the success of topos as antibiotic and chemotherapeutic drug targets, the validation of plant topo VI as a target for inhibitors should be a major aim for future work. Clearly, the isolation of a plant topo VI would be ideal for high-throughput compound screening, but MmTopo VI can serve as a suitable model in its absence. As more inhibitors of MmTopo VI are identified, their effect on plant growth should always be tested, especially compounds deemed to have a poisoning mechanism of action. Although topo VI is absent in humans, a potential concern is any off-target effects that topo VI inhibitors may have on human topo II or the meiotic topo VI-like complex given their close homology. However, these effects can easily be tested for and should not discourage the pursuit of topo VI as a herbicide target.

7.3.5 Top-down vs bottom-up strand-passage

It is plausible that a two-gated type II topo could perform strand-passage in both the top-down and bottom-up directions. Such a phenomenon may seem widely speculative but investigating this mechanism would be simple with a remarkable pay off. A short project could pursue this hypothesis with a series of experiments, including a more thorough mutagenic investigation of the positively-charged entrance to the MmTopo VI DNA gate, and an examination of the DNA-binding affinities of MmTopo VI by independently locking the ATP and DNA gates. The direction of strand-passage in MmTopo VI may explain the discrepancies between its differential binding affinities and catalytic activities on -scDNA and +scDNA and may have implications on the mechanism of Spo11.

7.4 Final conclusion

The research presented in this thesis has featured a series of bioinformatic and biochemical approaches that have improved our knowledge of topo VI. I have shown that the enzyme is widely distributed across all three domains of life but is not present in the malaria parasite *Plasmodium*. I have also shown that the ATPase activity of MmTopo VI is enhanced ~2-fold in the presence of +scDNA over that in the presence of -scDNA and I have developed a novel high-throughput relaxation assay that has identified two potent inhibitors of MmTopo VI and yeast topo II. Furthermore, I have shown that MmTopo VI is a useful model for the study of eukaryotic topo VI and the meiotic DSB machinery, and that topo VI is a suitable target for herbicides.

References

1 Liu, L. F. & Wang, J. C. Supercoiling of the DNA template during transcription. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 7024-7027, doi:10.1073/pnas.84.20.7024 (1987).

2 Gamper, H. B. & Hearst, J. E. A topological model for transcription based on unwinding angle analysis of E. coli RNA polymerase binary, initiation and ternary complexes. *Cell* **29**, 81-90, doi:https://doi.org/10.1016/0092-8674(82)90092-7 (1982).

3 Postow, L., Crisona, N. J., Peter, B. J., Hardy, C. D. & Cozzarelli, N. R. Topological challenges to DNA replication: Conformations at the fork. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 8219-8226, doi:10.1073/pnas.111006998 (2001).

4 Sundin, O. & Varshavsky, A. Terminal stages of SV40 DNA replication proceed via multiply intertwined catenated dimers. *Cell* **21**, 103-114, doi:10.1016/0092-8674(80)90118-x (1980).

5 Pollock, T. J. & Nash, H. A. Knotting of DNA caused by a genetic rearrangement: Evidence for a nucleosome-like structure in site-specific recombination of bacteriophage lambda. *Journal of molecular biology* **170**, 1-18, doi:https://doi.org/10.1016/S0022-2836(83)80224-1 (1983).

6 Spengler, S. J., Stasiak, A. & Cozzarelli, N. R. The stereostructure of knots and catenanes produced by phage lambda integrative recombination: implications for mechanism and DNA structure. *Cell* **42**, 325-334, doi:10.1016/s0092-8674(85)80128-8 (1985).

7 Bush, N. G., Evans-Roberts, K. & Maxwell, A. DNA Topoisomerases. *EcoSal Plus* **6**, doi:10.1128/ecosalplus.ESP-0010-2014 (2015).

8 Gadelle, D., Filee, J., Buhler, C. & Forterre, P. Phylogenomics of type II DNA topoisomerases. *Bioessays* **25**, 232-242, doi:10.1002/bies.10245 (2003).

9 Bates, A. D. & Maxwell, A. *DNA topology*. (Oxford University Press, USA, 2005).

10 Fuller, F. B. Decomposition of the linking number of a closed ribbon: A problem from molecular biology. *Proc Natl Acad Sci U S A* **75**, 3557-3561, doi:10.1073/pnas.75.8.3557 (1978).

11 Wang, J. C. Interaction between DNA and an Escherichia coli protein omega. *Journal of molecular biology* **55**, 523-533, doi:10.1016/0022-2836(71)90334-2 (1971).

12 Forterre, P. & Gadelle, D. Phylogenomics of DNA topoisomerases: their origin and putative roles in the emergence of modern organisms. *Nucleic Acids Res.* **37**, 679-692, doi:10.1093/nar/gkp032 (2009).

13 Tse, Y. & Wang, J. C. E. coli and M. luteus DNA topoisomerase I can catalyze catenation of decatenation of double-stranded DNA rings. *Cell* **22**, 269-276, doi:10.1016/0092-8674(80)90174-9 (1980).

14 Champoux, J. J. Type IA DNA topoisomerases: strictly one step at a time. *Proc Natl Acad Sci U S A* **99**, 11998-12000, doi:10.1073/pnas.202483499 (2002).

15 Champoux, J. J. & Dulbecco, R. An activity from mammalian cells that untwists superhelical DNA--a possible swivel for DNA replication (polyoma-ethidium bromide-mouse-embryo cells-dye binding assay). *Proc Natl Acad Sci U S A* **69**, 143-146, doi:10.1073/pnas.69.1.143 (1972).

16 Stewart, L., Redinbo, M. R., Qiu, X., Hol, W. G. & Champoux, J. J. A model for the mechanism of human topoisomerase I. *Science* **279**, 1534-1541, doi:10.1126/science.279.5356.1534 (1998).

17 Morham, S. G., Kluckman, K. D., Voulomanos, N. & Smithies, O. Targeted disruption of the mouse topoisomerase I gene by camptothecin selection. *Mol Cell Biol* **16**, 6804-6809, doi:10.1128/MCB.16.12.6804 (1996).

18 Pommier, Y. Topoisomerase I inhibitors: camptothecins and beyond. *Nat Rev Cancer* **6**, 789-802, doi:10.1038/nrc1977 (2006).

19 Mondragon, A. & DiGate, R. The structure of Escherichia coli DNA topoisomerase III. *Structure (London, England : 1993)* **7**, 1373-1383, doi:10.1016/s0969-2126(00)80027-1 (1999).

20 Terekhova, K., Gunn, K. H., Marko, J. F. & Mondragon, A. Bacterial topoisomerase I and topoisomerase III relax supercoiled DNA via distinct pathways. *Nucleic Acids Res* **40**, 10432-10440, doi:10.1093/nar/gks780 (2012).
21 DiGate, R. J. & Marians, K. J. Escherichia coli topoisomerase III-catalyzed cleavage of RNA. *The Journal of biological chemistry* **267**, 20532-20535 (1992).
22 Nurse, P., Levine, C., Hassing, H. & Marians, K. J. Topoisomerase III can serve as the cellular decatenase in Escherichia coli. *The Journal of biological chemistry* **278**, 8653-8660, doi:10.1074/jbc.M211211200 (2003).
23 Li, Z., Mondragon, A. & DiGate, R. J. The mechanism of type IA topoisomerase-mediated DNA topological transformations. *Mol Cell* **7**, 301-307, doi:10.1016/s1097-2765(01)00178-2 (2001).
24 Bizard, A. H. & Hickson, I. D. The many lives of type IA topoisomerases. *The Journal of biological chemistry* **295**, 7138-7153, doi:10.1074/jbc.REV120.008286 (2020).
25 Xu, D. *et al.* Top3beta is an RNA topoisomerase that works with fragile X syndrome protein to promote synapse formation. *Nat Neurosci* **16**, 1238-1247, doi:10.1038/nn.3479 (2013).
26 Forterre, P., Mirambeau, G., Jaxel, C., Nadal, M. & Duguet, M. High positive supercoiling in vitro catalyzed by an ATP and polyethylene glycol-stimulated topoisomerase from Sulfolobus acidocaldarius. *The EMBO journal* **4**, 2123-2128, doi:10.1002/j.1460-2075.1985.tb03902.x (1985).
27 Forterre, P., Gribaldo, S., Gadelle, D. & Serre, M. C. Origin and evolution of DNA topoisomerases. *Biochimie* **89**, 427-446, doi:10.1016/j.biochi.2006.12.009 (2007).
28 Confalonieri, F. *et al.* Reverse gyrase: a helicase-like domain and a type I topoisomerase in the same polypeptide. *Proc Natl Acad Sci U S A* **90**, 4753-4757, doi:10.1073/pnas.90.10.4753 (1993).
29 Slesarev, A. I. *et al.* DNA topoisomerase V is a relative of eukaryotic topoisomerase I from a hyperthermophilic prokaryote. *Nature* **364**, 735-737, doi:10.1038/364735a0 (1993).
30 Belova, G. I. *et al.* A type IB topoisomerase with DNA repair activities. *Proc Natl Acad Sci U S A* **98**, 6015-6020, doi:10.1073/pnas.111040498 (2001).
31 Taneja, B., Patel, A., Slesarev, A. & Mondragon, A. Structure of the N-terminal fragment of topoisomerase V reveals a new family of topoisomerases. *The EMBO journal* **25**, 398-408, doi:10.1038/sj.emboj.7600922 (2006).
32 Gellert, M., Mizuuchi, K., O'Dea, M. H. & Nash, H. A. DNA gyrase: an enzyme that introduces superhelical turns into DNA. *Proc Natl Acad Sci U S A* **73**, 3872-3876, doi:10.1073/pnas.73.11.3872 (1976).
33 Drake, F. H. *et al.* Purification of topoisomerase II from amsacrine-resistant P388 leukemia cells. Evidence for two forms of the enzyme. *The Journal of biological chemistry* **262**, 16739-16747 (1987).
34 Drake, F. H. *et al.* Biochemical and pharmacological properties of p170 and p180 forms of topoisomerase II. *Biochemistry* **28**, 8154-8160, doi:10.1021/bi00446a029 (1989).
35 Capranico, G., Tinelli, S., Austin, C. A., Fisher, M. L. & Zunino, F. Different patterns of gene expression of topoisomerase II isoforms in differentiated tissues during murine development. *Biochim Biophys Acta* **1132**, 43-48, doi:10.1016/0167-4781(92)90050-a (1992).
36 Dong, K. C. & Berger, J. M. Structural basis for gate-DNA recognition and bending by type IIA topoisomerases. *Nature* **450**, 1201-1205, doi:10.1038/nature06396 (2007).
37 Watt, P. M. & Hickson, I. D. Structure and function of type II DNA topoisomerases. *Biochem J* **303 (Pt 3)**, 681-695, doi:10.1042/bj3030681 (1994).
38 Kampranis, S. C. & Maxwell, A. Conversion of DNA gyrase into a conventional type II topoisomerase. *Proc Natl Acad Sci U S A* **93**, 14416-14421, doi:10.1073/pnas.93.25.14416 (1996).
39 Corbett, K. D., Schoeffler, A. J., Thomsen, N. D. & Berger, J. M. The structural basis for substrate specificity in DNA topoisomerase IV. *Journal of molecular biology* **351**, 545-561, doi:10.1016/j.jmb.2005.06.029 (2005).

40 Ward, D. & Newton, A. Requirement of topoisomerase IV parC and parE genes for cell cycle progression and developmental regulation in Caulobacter crescentus. *Molecular microbiology* **26**, 897-910, doi:10.1046/j.1365-2958.1997.6242005.x (1997).
41 Kramlinger, V. M. & Hiasa, H. The "GyrA-box" is required for the ability of DNA gyrase to wrap DNA and catalyze the supercoiling reaction. *The Journal of biological chemistry* **281**, 3738-3742, doi:10.1074/jbc.M511160200 (2006).
42 Lanz, M. A. & Klostermeier, D. The GyrA-box determines the geometry of DNA bound to gyrase and couples DNA binding to the nucleotide cycle. *Nucleic Acids Res* **40**, 10893-10903, doi:10.1093/nar/gks852 (2012).
43 Hsieh, T. J., Farh, L., Huang, W. M. & Chan, N. L. Structure of the topoisomerase IV C-terminal domain: a broken beta-propeller implies a role as geometry facilitator in catalysis. *The Journal of biological chemistry* **279**, 55587-55593, doi:10.1074/jbc.M408934200 (2004).
44 Corbett, K. D., Shultzaberger, R. K. & Berger, J. M. The C-terminal domain of DNA gyrase A adopts a DNA-bending beta-pinwheel fold. *Proc Natl Acad Sci U S A* **101**, 7293-7298, doi:10.1073/pnas.0401595101 (2004).
45 Lanz, M. A. & Klostermeier, D. Guiding strand passage: DNA-induced movement of the gyrase C-terminal domains defines an early step in the supercoiling cycle. *Nucleic Acids Res.* **39**, 9681-9694, doi:10.1093/nar/gkr680 (2011).
46 Bergerat, A., Gadelle, D. & Forterre, P. Purification of a DNA topoisomerase II from the hyperthermophilic archaeon Sulfolobus shibatae. A thermostable enzyme with both bacterial and eucaryal features. *The Journal of biological chemistry* **269**, 27663-27669 (1994).
47 Buhler, C., Lebbink, J. H. G., Bocs, C., Ladenstein, R. & Forterre, P. DNA topoisomerase VI generates ATP-dependent double-strand breaks with two-nucleotide overhangs. *J. Biol. Chem.* **276**, 37215-37222, doi:10.1074/jbc.M101823200 (2001).
48 Villain, P. *et al.* Expanded Dataset Reveals the Emergence and Evolution of DNA Gyrase in Archaea. *Molecular biology and evolution* **39**, doi:10.1093/molbev/msac155 (2022).
49 Hartung, F. & Puchta, H. Molecular characterization of homologues of both subunits A (SPO11) and B of the archaebacterial topoisomerase 6 in plants. *Gene* **271**, 81-86, doi:10.1016/s0378-1119(01)00496-6 (2001).
50 Malik, S. B., Ramesh, M. A., Hulstrand, A. M. & Logsdon, J. M. Protist homologs of the meiotic Spo11 gene and topoisomerase VI reveal an evolutionary history of gene duplication and lineage-specific loss. *Molecular biology and evolution* **24**, 2827-2841, doi:10.1093/molbev/msm217 (2007).
51 Charbonnier, F. & Forterre, P. Comparison of plasmid DNA topology among mesophilic and thermophilic eubacteria and archaebacteria. *Journal of bacteriology* **176**, 1251-1259, doi:10.1128/jb.176.5.1251-1259.1994 (1994).
52 Sioud, M., Possot, O., Elie, C., Sibold, L. & Forterre, P. Coumarin and quinolone action in archaebacteria: evidence for the presence of a DNA gyrase-like enzyme. *J Bacteriol* **170**, 946-953, doi:10.1128/jb.170.2.946-953.1988 (1988).
53 López-García, P., Forterre, P., van der Oost, J. & Erauso, G. Plasmid pGS5 from the hyperthermophilic archaeon Archaeoglobus profundus is negatively supercoiled. *J Bacteriol* **182**, 4998-5000, doi:10.1128/jb.182.17.4998-5000.2000 (2000).
54 Gadelle, D., Krupovic, M., Raymann, K., Mayer, C. & Forterre, P. DNA topoisomerase VIII: a novel subfamily of type IIB topoisomerases encoded by free or integrated plasmids in Archaea and Bacteria. *Nucleic Acids Res.* **42**, 8578-8591, doi:10.1093/nar/gku568 (2014).
55 Sugino, A., Peebles, C. L., Kreuzer, K. N. & Cozzarelli, N. R. Mechanism of action of nalidixic acid: purification of Escherichia coli nalA gene product and its relationship to DNA gyrase and a novel nicking-closing enzyme. *Proc Natl Acad Sci U S A* **74**, 4767-4771 (1977).
56 Gellert, M., Mizuuchi, K., O'Dea, M. H., Itoh, T. & Tomizawa, J. I. Nalidixic acid resistance: a second genetic character involved in DNA gyrase activity. *Proc Natl Acad Sci U S A* **74**, 4772-4776 (1977).

57 Takahashi, T. S. *et al.* Expanding the type IIB DNA topoisomerase family: identification of new topoisomerase and topoisomerase-like proteins in mobile genetic elements. *NAR Genomics and Bioinformatics* **2**, doi:10.1093/nargab/lqz021 (2020).
58 Wendorff, T. J. & Berger, J. M. Topoisomerase VI senses and exploits both DNA crossings and bends to facilitate strand passage. *eLife* **7**, doi:10.7554/eLife.31724 (2018).
59 Dutta, R. & Inouye, M. GHKL, an emergent ATPase/kinase superfamily. *Trends in biochemical sciences* **25**, 24-28, doi:10.1016/s0968-0004(99)01503-0 (2000).
60 Bergerat, A. *et al.* An atypical topoisomerase II from Archaea with implications for meiotic recombination. *Nature* **386**, 414-417, doi:10.1038/386414a0 (1997).
61 Corbett, K. D. & Berger, J. M. Structure of the topoisomerase VI-B subunit: implications for type II topoisomerase mechanism and evolution. *Embo J.* **22**, 151-163, doi:10.1093/emboj/cdg008 (2003).
62 Wigley, D. B., Davies, G. J., Dodson, E. J., Maxwell, A. & Dodson, G. Crystal structure of an N-terminal fragment of the DNA gyrase B protein. *Nature* **351**, 624, doi:10.1038/351624a0 (1991).
63 Jackson, A. P. & Maxwell, A. Identifying the catalytic residue of the ATPase reaction of DNA gyrase. *Proc Natl Acad Sci U S A* **90**, 11232-11236 (1993).
64 Marina, A., Mott, C., Auyzenberg, A., Hendrickson, W. A. & Waldburger, C. D. Structural and mutational analysis of the PhoQ histidine kinase catalytic domain. Insight into the reaction mechanism. *The Journal of biological chemistry* **276**, 41182-41190, doi:10.1074/jbc.M106080200 (2001).
65 Panaretou, B. *et al.* ATP binding and hydrolysis are essential to the function of the Hsp90 molecular chaperone in vivo. *The EMBO journal* **17**, 4829-4836, doi:10.1093/emboj/17.16.4829 (1998).
66 Stewart, R. C., VanBruggen, R., Ellefson, D. D. & Wolfe, A. J. TNP-ATP and TNP-ADP as Probes of the Nucleotide Binding Site of CheA, the Histidine Protein Kinase in the Chemotaxis Signal Transduction Pathway of Escherichia coli. *Biochemistry* **37**, 12269-12279, doi:10.1021/bi980970n (1998).
67 Corbett, K. D. & Berger, J. M. Structural dissection of ATP turnover in the prototypical GHL ATPase TopoVI. *Structure (London, England : 1993)* **13**, 873-882, doi:10.1016/j.str.2005.03.013 (2005).
68 Ban, C., Junop, M. & Yang, W. Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. *Cell* **97**, 85-97 (1999).
69 Meyer, P. *et al.* Structural and functional analysis of the middle segment of hsp90: implications for ATP hydrolysis and client protein and cochaperone interactions. *Mol Cell* **11**, 647-658, doi:10.1016/s1097-2765(03)00065-0 (2003).
70 Corbett, K. D., Benedetti, P. & Berger, J. M. Holoenzyme assembly and ATP-mediated conformational dynamics of topoisomerase VI. *Nat. Struct. Mol. Biol.* **14**, 611-619, doi:10.1038/nsmb1264 (2007).
71 Graille, M. *et al.* Crystal structure of an intact type II DNA topoisomerase: insights into DNA transfer mechanisms. *Structure (London, England : 1993)* **16**, 360-370, doi:10.1016/j.str.2007.12.020 (2008).
72 Vrielynck, N. *et al.* A DNA topoisomerase VI-like complex initiates meiotic recombination. *Science* **351**, 939-943, doi:10.1126/science.aad5196 (2016).
73 Smith, C. V. & Maxwell, A. Identification of a Residue Involved in Transition-State Stabilization in the ATPase Reaction of DNA Gyrase. *Biochemistry* **37**, 9658-9667, doi:10.1021/bi9801309 (1998).
74 Brennan, R. G. The winged-helix DNA-binding motif: another helix-turn-helix takeoff. *Cell* **74**, 773-776, doi:10.1016/0092-8674(93)90456-z (1993).

75 Liu, Q. & Wang, J. C. Identification of active site residues in the "GyrA" half of yeast DNA topoisomerase II. *The Journal of biological chemistry* **273**, 20252-20260, doi:10.1074/jbc.273.32.20252 (1998).
76 Aravind, L., Leipe, D. D. & Koonin, E. V. Toprim - a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Res.* **26**, 4205-4213, doi:10.1093/nar/26.18.4205 (1998).
77 Chen, S. J. & Wang, J. C. Identification of active site residues in Escherichia coli DNA topoisomerase I. *The Journal of biological chemistry* **273**, 6050-6056, doi:10.1074/jbc.273.11.6050 (1998).
78 Zhu, C. X., Roche, C. J., Papanicolaou, N., DiPietrantonio, A. & Tse-Dinh, Y. C. Site-directed mutagenesis of conserved aspartates, glutamates and arginines in the active site region of Escherichia coli DNA topoisomerase I. *The Journal of biological chemistry* **273**, 8783-8789, doi:10.1074/jbc.273.15.8783 (1998).
79 Nichols, M. D., DeAngelis, K., Keck, J. L. & Berger, J. M. Structure and function of an archaeal topoisomerase VI subunit with homology to the meiotic recombination factor Spo11. *Embo J.* **18**, 6177-6188, doi:10.1093/emboj/18.21.6177 (1999).
80 Diaz, R. L., Alcid, A. D., Berger, J. M. & Keeney, S. Identification of residues in yeast Spo11p critical for meiotic DNA double-strand break formation. *Mol Cell Biol* **22**, 1106-1115, doi:10.1128/mcb.22.4.1106-1115.2002 (2002).
81 DeWall, K. M., Davidson, M. K., Sharif, W. D., Wiley, C. A. & Wahls, W. P. A DNA binding motif of meiotic recombinase Rec12 (Spo11) defined by essential glycine-202, and persistence of Rec12 protein after completion of recombination. *Gene* **356**, 77-84, doi:10.1016/j.gene.2005.04.039 (2005).
82 Roca, J. & Wang, J. C. The capture of a DNA double helix by an ATP-dependent protein clamp: A key step in DNA transport by type II DNA topoisomerases. *Cell* **71**, 833-840, doi:10.1016/0092-8674(92)90558-t (1992).
83 Roca, J. & Wang, J. C. DNA transport by a type II DNA topoisomerase: evidence in favor of a two-gate mechanism. *Cell* **77**, 609-616 (1994).
84 Roca, J., Berger, J. M., Harrison, S. C. & Wang, J. C. DNA transport by a type II topoisomerase: direct evidence for a two-gate mechanism. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 4057-4062, doi:10.1073/pnas.93.9.4057 (1996).
85 Roca, J. The path of the DNA along the dimer interface of topoisomerase II. *The Journal of biological chemistry* **279**, 25783-25788, doi:10.1074/jbc.M402555200 (2004).
86 Bates, A. D., Berger, J. M. & Maxwell, A. The ancestral role of ATP hydrolysis in type II topoisomerases: prevention of DNA double-strand breaks. *Nucleic Acids Res* **39**, 6327-6339, doi:10.1093/nar/gkr258 (2011).
87 Baird, C. L., Harkins, T. T., Morris, S. K. & Lindsley, J. E. Topoisomerase II drives DNA transport by hydrolyzing one ATP. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 13685-13690 (1999).
88 Osheroff, N., Shelton, E. R. & Brutlag, D. L. DNA topoisomerase II from Drosophila melanogaster. Relaxation of supercoiled DNA. *The Journal of biological chemistry* **258**, 9536-9543 (1983).
89 Thomson, N. H. *et al.* DNA G-segment bending is not the sole determinant of topology simplification by type II DNA topoisomerases. *Scientific reports* **4**, 6158-6158, doi:10.1038/srep06158 (2014).
90 Sun, H., Treco, D., Schultes, N. P. & Szostak, J. W. Double-strand breaks at an initiation site for meiotic gene conversion. *Nature* **338**, 87-90, doi:10.1038/338087a0 (1989).
91 Cao, L., Alani, E. & Kleckner, N. A pathway for generation and processing of double-strand breaks during meiotic recombination in S. cerevisiae. *Cell* **61**, 1089-1101 (1990).
92 Keeney, S. in *Current Topics in Developmental Biology* Vol. 52 1-53 (Academic Press, 2001).

93 Liu, J., Wu, T. C. & Lichten, M. The location and structure of double-strand DNA breaks induced during yeast meiosis: evidence for a covalently linked DNA-protein intermediate. *The EMBO journal* **14**, 4599-4608 (1995).
94 Keeney, S. & Kleckner, N. Covalent protein-DNA complexes at the 5' strand termini of meiosis-specific double-strand breaks in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 11274-11278, doi:10.1073/pnas.92.24.11274 (1995).
95 de Massy, B., Rocco, V. & Nicolas, A. The nucleotide mapping of DNA double-strand breaks at the CYS3 initiation site of meiotic recombination in Saccharomyces cerevisiae. *The EMBO journal* **14**, 4589-4598, doi:10.1002/j.1460-2075.1995.tb00138.x (1995).
96 de Massy, B. Initiation of meiotic recombination: how and where? Conservation and specificities among eukaryotes. *Annual review of genetics* **47**, 563-599, doi:10.1146/annurev-genet-110711-155423 (2013).
97 Pâques, F. & Haber, J. E. Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. *Microbiol Mol Biol Rev* **63**, 349-404 (1999).
98 Allers, T. & Lichten, M. Differential Timing and Control of Noncrossover and Crossover Recombination during Meiosis. *Cell* **106**, 47-57, doi:10.1016/s0092-8674(01)00416-0 (2001).
99 McMahill, M. S., Sham, C. W. & Bishop, D. K. Synthesis-dependent strand annealing in meiosis. *PLoS Biol* **5**, e299-e299, doi:10.1371/journal.pbio.0050299 (2007).
100 Nassif, N., Penney, J., Pal, S., Engels, W. R. & Gloor, G. B. Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. *Molecular and cellular biology* **14**, 1613-1625, doi:10.1128/mcb.14.3.1613 (1994).
101 Sasanuma, H. *et al.* Meiotic association between Spo11 regulated by Rec102, Rec104 and Rec114. *Nucleic Acids Res* **35**, 1119-1133, doi:10.1093/nar/gkl1162 (2007).
102 Eichinger, L. *et al.* The genome of the social amoeba Dictyostelium discoideum. *Nature* **435**, 43-57, doi:10.1038/nature03481 (2005).
103 Heidel, A. J. *et al.* Phylogeny-wide analysis of social amoeba genomes highlights ancient origins for complex intercellular communication. *Genome Res* **21**, 1882-1891, doi:10.1101/gr.121137.111 (2011).
104 Urushihara, H. *et al.* Comparative genome and transcriptome analyses of the social amoeba Acytostelium subglobosum that accomplishes multicellular development without germ-soma differentiation. *BMC Genomics* **16**, 80, doi:10.1186/s12864-015-1278-x (2015).
105 Bloomfield, G. Atypical ploidy cycles, Spo11, and the evolution of meiosis. *Semin Cell Dev Biol* **54**, 158-164, doi:10.1016/j.semcdb.2016.01.026 (2016).
106 Boateng, K. A., Bellani, M. A., Gregoretti, I. V., Pratto, F. & Camerini-Otero, R. D. Homologous pairing preceding SPO11-mediated double-strand breaks in mice. *Dev Cell* **24**, 196-205, doi:10.1016/j.devcel.2012.12.002 (2013).
107 Heap, I. *The International Herbicide-Resistant Weed Database*, <www.weedscience.org> (2022).
108 Duke, S. O. & Powles, S. B. Glyphosate: a once-in-a-century herbicide. *Pest management science* **64**, 319-325, doi:10.1002/ps.1518 (2008).
109 Powles, S. B. Evolved glyphosate-resistant weeds around the world: lessons to be learnt. *Pest management science* **64**, 360-365, doi:10.1002/ps.1525 (2008).
110 Grelon, M., Vezon, D., Gendrot, G. & Pelletier, G. AtSPO11-1 is necessary for efficient meiotic recombination in plants. *The EMBO journal* **20**, 589-600, doi:10.1093/emboj/20.3.589 (2001).
111 Stacey, N. J. *et al.* Arabidopsis SPO11-2 functions with SPO11-1 in meiotic recombination. *Plant J* **48**, 206-216, doi:10.1111/j.1365-313X.2006.02867.x (2006).
112 Hartung, F. *et al.* The catalytically active tyrosine residues of both SPO11-1 and SPO11-2 are required for meiotic double-strand break induction in Arabidopsis. *Plant Cell* **19**, 3090-3099, doi:10.1105/tpc.107.054817 (2007).
113 Sugimoto-Shirasu, K. & Roberts, K. "Big it up": endoreduplication and cell-size control in plants. *Current opinion in plant biology* **6**, 544-553, doi:10.1016/j.pbi.2003.09.009 (2003).

114 De Veylder, L., Larkin, J. C. & Schnittger, A. Molecular control and function of endoreplication in development and physiology. *Trends in plant science* **16**, 624-634, doi:10.1016/j.tplants.2011.07.001 (2011).
115 Hartung, F. *et al.* An archaebacterial topoisomerase homolog not present in other eukaryotes is indispensable for cell proliferation of plants. *Curr Biol* **12**, 1787-1791 (2002).
116 Sugimoto-Shirasu, K., Stacey, N. J., Corsar, J., Roberts, K. & McCann, M. C. DNA topoisomerase VI is essential for endoreduplication in Arabidopsis. *Current Biology* **12**, 1782-1786, doi:10.1016/s0960-9822(02)01198-3 (2002).
117 Yin, Y. H. *et al.* A crucial role for the putative Arabidopsis topoisomerase VI in plant growth and development. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 10191-10196, doi:10.1073/pnas.152337599 (2002).
118 Li, J. & Chory, J. Brassinosteroid actions in plants. *Journal of experimental botany* **50**, 275-282, doi:10.1093/jxb/50.332.275 (1999).
119 Tian, Y. *et al.* Role of a cotton endoreduplication-related gene, GaTOP6B, in response to drought stress. *Planta*, doi:10.1007/s00425-018-3067-7 (2018).
120 Jain, M., Tyagi, A. K. & Khurana, J. P. Overexpression of putative topoisomerase 6 genes from rice confers stress tolerance in transgenic Arabidopsis plants. *The FEBS journal* **273**, 5245-5260, doi:10.1111/j.1742-4658.2006.05518.x (2006).
121 McKie, S. J., Neuman, K. C. & Maxwell, A. DNA topoisomerases: Advances in understanding of cellular roles and multi-protein complexes via structure-function analysis. *Bioessays* **43**, e2000286, doi:10.1002/bies.202000286 (2021).
122 El Sayyed, H. *et al.* Mapping Topoisomerase IV Binding and Activity Sites on the E. coli Genome. *PLoS Genet* **12**, e1006025, doi:10.1371/journal.pgen.1006025 (2016).
123 Sugimoto-Shirasu, K. *et al.* RHL1 is an essential component of the plant DNA topoisomerase VI complex and is required for ploidy-dependent cell growth. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 18736-18741, doi:10.1073/pnas.0505883102 (2005).
124 Breuer, C. *et al.* BIN4, a novel component of the plant DNA topoisomerase VI complex, is required for endoreduplication in Arabidopsis. *Plant Cell* **19**, 3655-3668, doi:10.1105/tpc.107.054833 (2007).
125 Schneider, K., Wells, B., Dolan, L. & Roberts, K. Structural and genetic analysis of epidermal cell differentiation in Arabidopsis primary roots. *Development* **124**, 1789-1798 (1997).
126 Schneider, K. *et al.* The ROOT HAIRLESS 1 gene encodes a nuclear protein required for root hair initiation in Arabidopsis. *Genes Dev* **12**, 2013-2021, doi:10.1101/gad.12.13.2013 (1998).
127 Reeves, R. & Nissen, M. S. The A.T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. *The Journal of biological chemistry* **265**, 8573-8582 (1990).
128 Jensen, S. *et al.* Analysis of functional domain organization in DNA topoisomerase II from humans and Saccharomyces cerevisiae. *Mol Cell Biol* **16**, 3866-3877, doi:10.1128/mcb.16.7.3866 (1996).
129 Aravind, L., Iyer, L. M., Wellems, T. E. & Miller, L. H. Plasmodium biology: Genomic gleanings. *Cell* **115**, 771-785, doi:10.1016/s0092-8674(03)01023-7 (2003).
130 World Health, O. World malaria report 2021. (World Health Organization, Geneva, 2021).
131 Eastman, R. T. & Fidock, D. A. Artemisinin-based combination therapies: a vital tool in efforts to eliminate malaria. *Nat Rev Microbiol* **7**, 864-874, doi:10.1038/nrmicro2239 (2009).
132 Dondorp, A. M. *et al.* Artemisinin Resistance in Plasmodium falciparum Malaria. **361**, 455-467, doi:10.1056/NEJMoa0808859 (2009).
133 Tuteja, R. Malaria - an overview. *The FEBS journal* **274**, 4670-4679, doi:10.1111/j.1742-4658.2007.05997.x (2007).

134 Delves, M. *et al.* The activities of current antimalarial drugs on the life cycle stages of Plasmodium: a comparative study with human and rodent parasites. *PLoS Med* **9**, e1001169-e1001169, doi:10.1371/journal.pmed.1001169 (2012).
135 Kuehn, A. & Pradel, G. The coming-out of malaria gametocytes. *J Biomed Biotechnol* **2010**, 976827, doi:10.1155/2010/976827 (2010).
136 Gerald, N., Mahajan, B. & Kumar, S. Mitosis in the human malaria parasite Plasmodium falciparum. *Eukaryotic cell* **10**, 474-482, doi:10.1128/EC.00314-10 (2011).
137 Francia, M. E. & Striepen, B. Cell division in apicomplexan parasites. *Nature Reviews Microbiology* **12**, 125-136, doi:10.1038/nrmicro3184 (2014).
138 Gardner, M. J. *et al.* Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* **419**, 498-511, doi:10.1038/nature01097 (2002).
139 Chalapareddy, S., Chakrabarty, S., Bhattacharyya, M. K. & Bhattacharyya, S. Radicicol-Mediated Inhibition of Topoisomerase VIB-VIA Activity of the Human Malaria Parasite Plasmodium falciparum. *Msphere* **1**, 17, doi:10.1128/mSphere.00025-15 (2016).
140 Zhang, M. *et al.* Uncovering the essential genes of the human malaria parasite Plasmodium falciparum by saturation mutagenesis. *Science* **360**, 506-+, doi:10.1126/science.aap7847 (2018).
141 Bushell, E. *et al.* Functional Profiling of a Plasmodium Genome Reveals an Abundance of Essential Genes. *Cell* **170**, 260-272.e268, doi:10.1016/j.cell.2017.06.030 (2017).
142 Hall, N. *et al.* A comprehensive survey of the Plasmodium life cycle by genomic, transcriptomic, and proteomic analyses. *Science* **307**, 82-86, doi:10.1126/science.1103717 (2005).
143 Melaragno, J. E., Mehrotra, B. & Coleman, A. W. Relationship between Endopolyploidy and Cell Size in Epidermal Tissue of Arabidopsis. *Plant Cell* **5**, 1661-1668, doi:10.1105/tpc.5.11.1661 (1993).
144 Orr-Weaver, T. L. When bigger is better: the role of polyploidy in organogenesis. *Trends Genet* **31**, 307-315, doi:10.1016/j.tig.2015.03.011 (2015).
145 Calvi, B. R., Lilly, M. A. & Spradling, A. C. Cell cycle control of chorion gene amplification. *Genes Dev* **12**, 734-744, doi:10.1101/gad.12.5.734 (1998).
146 Barlow, P. W. & Sherman, M. I. Cytological studies on the organization of DNA in giant trophoblast nuclei of the mouse and the rat. *Chromosoma* **47**, 119-131, doi:10.1007/bf00331800 (1974).
147 De Souza, C. P. C. & Osmani, S. A. Mitosis, Not Just Open or Closed. *Eukaryotic Cell* **6**, 1521, doi:10.1128/EC.00178-07 (2007).
148 D'Amato, F. Polyploidy in Cell Differentiation. *Caryologia* **42**, 183-211, doi:10.1080/00087114.1989.10796966 (1989).
149 Reilly, H. B., Wang, H., Steuter, J. A., Marx, A. M. & Ferdig, M. T. Quantitative dissection of clone-specific growth rates in cultured malaria parasites. *Int J Parasitol* **37**, 1599-1607, doi:10.1016/j.ijpara.2007.05.003 (2007).
150 Shortt, H. E., Fairley, N. H., Covell, G., Shute, P. G. & Garnham, P. C. The pre-erythrocytic stage of Plasmodium falciparum. *Trans R Soc Trop Med Hyg* **44**, 405-419, doi:10.1016/s0035-9203(51)80019-1 (1951).
151 Vaughan, J. A., Noden, B. H. & Beier, J. C. Population dynamics of Plasmodium falciparum sporogony in laboratory-infected Anopheles gambiae. *J Parasitol* **78**, 716-724 (1992).
152 Bush, N., Agarwal, M., Henderson, S., Waraich, N. & Maxwell, A. DNA in a twist? How topoisomerases solve topological problems in DNA. *Biochemist* **40**, 26-31, doi:10.1042/BIO04002026 (2018).
153 Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402, doi:10.1093/nar/25.17.3389 (1997).
154 Schoch, C. L. *et al.* NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database (Oxford)* **2020**, doi:10.1093/database/baaa062 (2020).

155 Chojnacki, S., Cowley, A., Lee, J., Foix, A. & Lopez, R. Programmatic access to bioinformatics tools from EMBL-EBI update: 2017. *Nucleic Acids Res* **45**, W550-w553, doi:10.1093/nar/gkx273 (2017).
156 Notredame, C., Higgins, D. G. & Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of molecular biology* **302**, 205-217, doi:10.1006/jmbi.2000.4042 (2000).
157 Trifinopoulos, J., Nguyen, L. T., von Haeseler, A. & Minh, B. Q. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res* **44**, W232-235, doi:10.1093/nar/gkw256 (2016).
158 Nguyen, L. T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution* **32**, 268-274, doi:10.1093/molbev/msu300 (2015).
159 Minh, B. Q., Nguyen, M. A. & von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. *Molecular biology and evolution* **30**, 1188-1195, doi:10.1093/molbev/mst024 (2013).
160 Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England)* **28**, 1647-1649, doi:10.1093/bioinformatics/bts199 (2012).
161 Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-324, doi:10.1093/nar/gku316 (2014).
162 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589, doi:10.1038/s41586-021-03819-2 (2021).
163 Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679-682, doi:10.1038/s41592-022-01488-1 (2022).
164 Potterton, E., McNicholas, S., Krissinel, E., Cowtan, K. & Noble, M. The CCP4 molecular-graphics project. *Acta Crystallogr D Biol Crystallogr* **58**, 1955-1957, doi:10.1107/s0907444902015391 (2002).
165 Krissinel, E. Enhanced fold recognition using efficient short fragment clustering. *J Mol Biochem* **1**, 76-85 (2012).
166 Tan, S. A modular polycistronic expression system for overexpressing protein complexes in Escherichia coli. *Protein expression and purification* **21**, 224-234, doi:10.1006/prep.2000.1363 (2001).
167 Burke, D., Dawson, D. & Stearns, T. Methods in yeast genetics: a Cold Spring Harbor Laboratory course manual (2000 edition). *Plainview, NY: Cold Spring Harbor Laboratory Press.[Google Scholar]* (2000).
168 Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein expression and purification* **41**, 207-234, doi:10.1016/j.pep.2005.01.016 (2005).
169 Giaever, G. N., Snyder, L. & Wang, J. C. DNA supercoiling in vivo. *Biophys Chem* **29**, 7-15, doi:10.1016/0301-4622(88)87020-0 (1988).
170 Tomlin, G. C., Wixon, J. L., Bolotin-Fukuhara, M. & Oliver, S. G. A new family of yeast vectors and S288C-derived strains for the systematic analysis of gene function. *Yeast* **18**, 563-575, doi:10.1002/yea.703 (2001).
171 Entian, K.-D. & Kötter, P. 25 yeast genetic strain and plasmid collections. *Methods in microbiology* **36**, 629-666 (2007).
172 McKie, S. J. *et al.* Topoisomerase VI is a chirally-selective, preferential DNA decatenase. *eLife* **11**, doi:10.7554/eLife.67021 (2022).
173 Wilkins, M. R. *et al.* Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* **112**, 531-552, doi:10.1385/1-59259-584-7:531 (1999).
174 Waraich, N. F. *et al.* A novel decatenation assay for DNA topoisomerases using a singly-linked catenated substrate. *Biotechniques* **69**, 356-362, doi:10.2144/btn-2020-0059 (2020).

175 Arabidopsis Genome, I. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* **408**, 796-815, doi:10.1038/35048692 (2000).
176 Hartung, F. & Puchta, H. Molecular characterisation of two paralogous SPO11 homologues in Arabidopsis thaliana. *Nucleic Acids Res.* **28**, 1548-1554, doi:10.1093/nar/28.7.1548 (2000).
177 Keeney, S. Mechanism and control of meiotic recombination initiation. *Curr Top Dev Biol* **52**, 1-53, doi:10.1016/s0070-2153(01)52008-6 (2001).
178 Arora, C., Kee, K., Maleki, S. & Keeney, S. Antiviral Protein Ski8 Is a Direct Partner of Spo11 in Meiotic DNA Break Formation, Independent of Its Cytoplasmic Role in RNA Metabolism. *Molecular Cell* **13**, 549-559, doi:10.1016/s1097-2765(04)00063-2 (2004).
179 Kee, K. & Keeney, S. Functional interactions between SPO11 and REC102 during initiation of meiotic recombination in Saccharomyces cerevisiae. *Genetics* **160**, 111-122 (2002).
180 Jiao, K., Salem, L. & Malone, R. Support for a meiotic recombination initiation complex: interactions among Rec102p, Rec104p, and Spo11p. *Mol Cell Biol* **23**, 5928-5938, doi:10.1128/mcb.23.16.5928-5938.2003 (2003).
181 Li, J., Hooker, G. W. & Roeder, G. S. Saccharomyces cerevisiae Mer2, Mei4 and Rec114 form a complex required for meiotic double-strand break formation. *Genetics* **173**, 1969-1981, doi:10.1534/genetics.106.058768 (2006).
182 Maleki, S., Neale, M. J., Arora, C., Henderson, K. A. & Keeney, S. Interactions between Mei4, Rec114, and other proteins required for meiotic DNA double-strand break formation in Saccharomyces cerevisiae. *Chromosoma* **116**, 471-486, doi:10.1007/s00412-007-0111-y (2007).
183 Prieler, S., Penkner, A., Borde, V. & Klein, F. The control of Spo11's interaction with meiotic recombination hotspots. *Genes Dev* **19**, 255-269, doi:10.1101/gad.321105 (2005).
184 Panizza, S. *et al.* Spo11-accessory proteins link double-strand break sites to the chromosome axis in early meiotic recombination. *Cell* **146**, 372-383, doi:10.1016/j.cell.2011.07.003 (2011).
185 Buhler, C., Gadelle, D., Forterre, P., Wang, J. C. & Bergerat, A. Reconstitution of DNA topoisomerase VI of the thermophilic archaeon Sulfolobus shibatae from subunits separately overexpressed in Escherichia coli. *Nucleic Acids Res.* **26**, 5157-5162, doi:10.1093/nar/26.22.5157 (1998).
186 Robert, T. *et al.* The TopoVIB-Like protein family is required for meiotic DNA double-strand break formation. *Science* **351**, 943-949, doi:10.1126/science.aad5309 (2016).
187 Claeys Bouuaert, C. *et al.* Structural and functional characterization of the Spo11 core complex. *Nat Struct Mol Biol* **28**, 92-102, doi:10.1038/s41594-020-00534-w (2021).
188 Burki, F. *et al.* Phylogenomics reshuffles the eukaryotic supergroups. *PloS one* **2**, e790, doi:10.1371/journal.pone.0000790 (2007).
189 Burki, F., Roger, A. J., Brown, M. W. & Simpson, A. G. B. The New Tree of Eukaryotes. *Trends Ecol Evol* **35**, 43-55, doi:10.1016/j.tree.2019.08.008 (2020).
190 Tretter, E. M., Lerman, J. C. & Berger, J. M. A naturally chimeric type IIA topoisomerase in Aquifex aeolicus highlights an evolutionary path for the emergence of functional paralogs. *Proc Natl Acad Sci U S A* **107**, 22055-22059, doi:10.1073/pnas.1012938107 (2010).
191 Taylor, J. A., Mitchenall, L. A., Rejzek, M., Field, R. A. & Maxwell, A. Application of a novel microtitre plate-based assay for the discovery of new inhibitors of DNA gyrase and DNA topoisomerase VI. *PloS one* **8**, e58010, doi:10.1371/journal.pone.0058010 (2013).
192 Pereira, J. *et al.* High-accuracy protein structure prediction in CASP14. *Proteins* **89**, 1687-1699, doi:10.1002/prot.26171 (2021).
193 Wall, M. K., Mitchenall, L. A. & Maxwell, A. Arabidopsis thaliana DNA gyrase is targeted to chloroplasts and mitochondria. *Proc Natl Acad Sci U S A* **101**, 7821-7826, doi:10.1073/pnas.0400836101 (2004).
194 Huang, W. M. Bacterial diversity based on type II DNA topoisomerase genes. *Annual review of genetics* **30**, 79-107, doi:10.1146/annurev.genet.30.1.79 (1996).

195 Mizuuchi, K., Mizuuchi, M., O'Dea, M. H. & Gellert, M. Cloning and simplified purification of Escherichia coli DNA gyrase A and B proteins. *The Journal of biological chemistry* **259**, 9199-9201 (1984).
196 Kato, J.-i. *et al.* New topoisomerase essential for chromosome segregation in E. coli. *Cell* **63**, 393-404, doi:10.1016/0092-8674(90)90172-b (1990).
197 Worland, S. T. & Wang, J. C. Inducible overexpression, purification, and active site mapping of DNA topoisomerase II from the yeast Saccharomyces cerevisiae. *J. Biol. Chem.* **264**, 4412-4416, doi:10.1016/s0021-9258(18)83757-7 (1989).
198 Chene, P. *et al.* Catalytic inhibition of topoisomerase II by a novel rationally designed ATP-competitive purine analogue. *BMC Chem Biol* **9**, 1, doi:10.1186/1472-6769-9-1 (2009).
199 Makarevitch, I. & Somers, D. A. Purification and characterization of topoisomerase IIA from Arabidopsis thaliana. *Plant Science* **168**, 1023-1033, doi:10.1016/j.plantsci.2004.11.019 (2005).
200 Evans-Roberts, K. M. *et al.* DNA Gyrase Is the Target for the Quinolone Drug Ciprofloxacin in Arabidopsis thaliana. *The Journal of biological chemistry* **291**, 3136-3144, doi:10.1074/jbc.M115.689554 (2016).
201 Riou, J. F., Gabillot, M., Philippe, M., Schrevel, J. & Riou, G. Purification and characterization of Plasmodium berghei DNA topoisomerases I and II: drug action, inhibition of decatenation and relaxation, and stimulation of DNA cleavage. *Biochemistry* **25**, 1471-1479, doi:10.1021/bi00355a001 (1986).
202 Mckie, S. *The Characterisation of DNA Topoisomerase VI from Methanosarcina mazei Using Ensemble and Single-Molecule Methods*, University of East Anglia, (2020).
203 Ehrich, S., Behrens, D., Lebedeva, E., Ludwig, W. & Bock, E. A new obligately chemolithoautotrophic, nitrite-oxidizing bacterium, Nitrospira moscoviensis sp. nov. and its phylogenetic relationship. *Arch Microbiol* **164**, 16-23, doi:10.1007/BF02568729 (1995).
204 Spring, S., Bunk, B., Sproer, C., Rohde, M. & Klenk, H. P. Genome biology of a novel lineage of planctomycetes widespread in anoxic aquatic environments. *Environ Microbiol* **20**, 2438-2455, doi:10.1111/1462-2920.14253 (2018).
205 Brooks, R. S. *et al.* Characterization of Pajaroellobacter abortibovis, the etiologic agent of epizootic bovine abortion. *Vet Microbiol* **192**, 73-80, doi:10.1016/j.vetmic.2016.07.001 (2016).
206 Lopez-Albors, O., Llamas-Lopez, P. J., Ortuno, J. A., Latorre, R. & Garcia-Vazquez, F. A. In vivo measurement of pH and CO2 levels in the uterus of sows through the estrous cycle and after insemination. *Scientific reports* **11**, 3194, doi:10.1038/s41598-021-82620-7 (2021).
207 Kaur, J., Kumar, A. & Kaur, J. Strategies for optimization of heterologous protein expression in E. coli: Roadblocks and reinforcements. *Int J Biol Macromol* **106**, 803-822, doi:10.1016/j.ijbiomac.2017.08.080 (2018).
208 Sun, Y. *et al.* A conserved SUMO pathway repairs topoisomerase DNA-protein cross-links by engaging ubiquitin-mediated proteasomal degradation. *Sci Adv* **6**, doi:10.1126/sciadv.aba6290 (2020).
209 Marbach, A. & Bettenbrock, K. lac operon induction in Escherichia coli: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA. *J Biotechnol* **157**, 82-88, doi:10.1016/j.jbiotec.2011.10.009 (2012).
210 Martinez-Alonso, M., Garcia-Fruitos, E. & Villaverde, A. Yield, solubility and conformational quality of soluble proteins are not simultaneously favored in recombinant Escherichia coli. *Biotechnol Bioeng* **101**, 1353-1358, doi:10.1002/bit.21996 (2008).
211 Wagner, S. *et al.* Tuning Escherichia coli for membrane protein overexpression. *Proc Natl Acad Sci U S A* **105**, 14371-14376, doi:10.1073/pnas.0804090105 (2008).
212 Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M. & Hartl, F. U. Molecular chaperone functions in protein folding and proteostasis. *Annu Rev Biochem* **82**, 323-355, doi:10.1146/annurev-biochem-060208-092442 (2013).

213 de Marco, A., Deuerling, E., Mogk, A., Tomoyasu, T. & Bukau, B. Chaperone-based procedure to increase yields of soluble recombinant proteins produced in E. coli. *BMC Biotechnol* **7**, 32, doi:10.1186/1472-6750-7-32 (2007).

214 Horvath, I. *et al.* Membrane physical state controls the signaling mechanism of the heat shock response in Synechocystis PCC 6803: identification of hsp17 as a "fluidity gene". *Proc Natl Acad Sci U S A* **95**, 3513-3518, doi:10.1073/pnas.95.7.3513 (1998).

215 de Marco, A., Vigh, L., Diamant, S. & Goloubinoff, P. Native folding of aggregation-prone recombinant proteins in Escherichia coli by osmolytes, plasmid- or benzyl alcohol-overexpressed molecular chaperones. *Cell Stress Chaperones* **10**, 329-339, doi:10.1379/csc-139r.1 (2005).

216 Singh, S. M. & Panda, A. K. Solubilization and refolding of bacterial inclusion body proteins. *J Biosci Bioeng* **99**, 303-310, doi:10.1263/jbb.99.303 (2005).

217 Stuchinskaya, T. *et al.* How do type II topoisomerases use ATP hydrolysis to simplify DNA topology beyond equilibrium? Investigating the relaxation reaction of nonsupercoiling type II topoisomerases. *Journal of molecular biology* **385**, 1397-1408, doi:10.1016/j.jmb.2008.11.056 (2009).

218 Rybenkov, V. V., Ullsperger, C., Vologodskii, A. V. & Cozzarelli, N. R. Simplification of DNA topology below equilibrium values by type II topoisomerases. *Science* **277**, 690-693, doi:10.1126/science.277.5326.690 (1997).

219 Stone, M. D. *et al.* Chirality sensing by Escherichia coli topoisomerase IV and the mechanism of type II topoisomerases. *Proc Natl Acad Sci U S A* **100**, 8654-8659, doi:10.1073/pnas.1133178100 (2003).

220 Neuman, K. C., Charvin, G., Bensimon, D. & Croquette, V. Mechanisms of chiral discrimination by topoisomerase IV. *Proc Natl Acad Sci U S A* **106**, 6986-6991, doi:10.1073/pnas.0900574106 (2009).

221 Litwin, T. R., Sola, M., Holt, I. J. & Neuman, K. C. A robust assay to measure DNA topology-dependent protein binding affinity. *Nucleic Acids Res* **43**, e43, doi:10.1093/nar/gku1381 (2015).

222 McClendon, A. K., Rodriguez, A. C. & Osheroff, N. Human topoisomerase IIalpha rapidly relaxes positively supercoiled DNA: implications for enzyme action ahead of replication forks. *The Journal of biological chemistry* **280**, 39337-39345, doi:10.1074/jbc.M503320200 (2005).

223 Morrical, S. W., Lee, J. & Cox, M. M. Continuous association of Escherichia coli single-stranded DNA binding protein with stable complexes of recA protein and single-stranded DNA. *Biochemistry* **25**, 1482-1494, doi:10.1021/bi00355a003 (1986).

224 Charvin, G., Bensimon, D. & Croquette, V. Single-molecule study of DNA unlinking by eukaryotic and prokaryotic type-II topoisomerases. *Proc Natl Acad Sci U S A* **100**, 9820-9825, doi:10.1073/pnas.1631550100 (2003).

225 Bates, A. D., O'Dea, M. H. & Gellert, M. Energy coupling in Escherichia coli DNA gyrase: the relationship between nucleotide binding, strand passage, and DNA supercoiling. *Biochemistry* **35**, 1408-1416, doi:10.1021/bi952433y (1996).

226 Williams, N. L. & Maxwell, A. Probing the two-gate mechanism of DNA gyrase using cysteine cross-linking. *Biochemistry* **38**, 13502-13511, doi:10.1021/bi9912488 (1999).

227 Maxwell, A. & Gellert, M. The DNA dependence of the ATPase activity of DNA gyrase. *The Journal of biological chemistry* **259**, 14472-14480 (1984).

228 Halligan, B. D., Edwards, K. A. & Liu, L. F. Purification and characterization of a type II DNA topoisomerase from bovine calf thymus. *The Journal of biological chemistry* **260**, 2475-2482 (1985).

229 Gellert, M., O'Dea, M. H., Itoh, T. & Tomizawa, J. Novobiocin and coumermycin inhibit DNA supercoiling catalyzed by DNA gyrase. *Proceedings of the National Academy of Sciences of the United States of America* **73**, 4474-4478, doi:10.1073/pnas.73.12.4474 (1976).

230 Sugino, A., Higgins, N. P., Brown, P. O., Peebles, C. L. & Cozzarelli, N. R. Energy coupling in DNA gyrase and the mechanism of action of novobiocin. *Proc Natl Acad Sci U S A* **75**, 4838-4842, doi:10.1073/pnas.75.10.4838 (1978).
231 Feng, L. *et al.* The pentapeptide-repeat protein, MfpA, interacts with mycobacterial DNA gyrase as a DNA T-segment mimic. *Proc Natl Acad Sci U S A* **118**, doi:10.1073/pnas.2016705118 (2021).
232 Pommier, Y., Leo, E., Zhang, H. & Marchand, C. DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chemistry & biology* **17**, 421-433, doi:10.1016/j.chembiol.2010.04.012 (2010).
233 Ryley, J. F. & Peters, W. The antimalarial activity of some quinolone esters. *Annals of Tropical Medicine & Parasitology* **64**, 209-222, doi:10.1080/00034983.1970.11686683 (2016).
234 Wolfson, J. S. & Hooper, D. C. Fluoroquinolone antimicrobial agents. *Clin Microbiol Rev* **2**, 378-424, doi:10.1128/cmr.2.4.378 (1989).
235 Appelbaum, P. C. & Hunter, P. A. The fluoroquinolone antibacterials: past, present and future perspectives. *International Journal of Antimicrobial Agents* **16**, 5-15, doi:10.1016/s0924-8579(00)00192-8 (2000).
236 Hashimi, S. M., Wall, M. K., Smith, A. B., Maxwell, A. & Birch, R. G. The phytotoxin albicidin is a novel inhibitor of DNA gyrase. *Antimicrobial agents and chemotherapy* **51**, 181-187, doi:10.1128/aac.00918-06 (2007).
237 Vizan, J. L., Hernandez-Chico, C., del Castillo, I. & Moreno, F. The peptide antibiotic microcin B17 induces double-strand cleavage of DNA mediated by E. coli DNA gyrase. *The EMBO journal* **10**, 467-476 (1991).
238 Bernard, P. & Couturier, M. Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *Journal of molecular biology* **226**, 735-745 (1992).
239 Bateman, A., Murzin, A. G. & Teichmann, S. A. Structure and distribution of pentapeptide repeats in bacteria. *Protein Sci* **7**, 1477-1480, doi:10.1002/pro.5560070625 (1998).
240 Montero, C., Mateu, G., Rodriguez, R. & Takiff, H. Intrinsic resistance of Mycobacterium smegmatis to fluoroquinolones may be influenced by new pentapeptide protein MfpA. *Antimicrobial agents and chemotherapy* **45**, 3387-3392, doi:10.1128/aac.45.12.3387-3392.2001 (2001).
241 Martínez-Martínez, L., Pascual, A., García, I., Tran, J. & Jacoby, G. A. Interaction of plasmid and host quinolone resistance. *Journal of Antimicrobial Chemotherapy* **51**, 1037-1039, doi:10.1093/jac/dkg157 (2003).
242 Hegde, S. S. *et al.* A fluoroquinolone resistance protein from Mycobacterium tuberculosis that mimics DNA. *Science* **308**, 1480-1483, doi:10.1126/science.1110699 (2005).
243 Tran, J. H., Jacoby, G. A. & Hooper, D. C. Interaction of the plasmid-encoded quinolone resistance protein Qnr with Escherichia coli DNA gyrase. *Antimicrobial agents and chemotherapy* **49**, 118-125, doi:10.1128/aac.49.1.118-125.2005 (2005).
244 Schimana, J. *et al.* Simocyclinones, novel cytostatic angucyclinone antibiotics produced by Streptomyces antibioticus Tu 6040. I. Taxonomy, fermentation, isolation and biological activities. *The Journal of antibiotics* **53**, 779-787 (2000).
245 Flatman, R. H., Howells, A. J., Heide, L., Fiedler, H. P. & Maxwell, A. Simocyclinone D8, an inhibitor of DNA gyrase with a novel mode of action. *Antimicrobial agents and chemotherapy* **49**, 1093-1100, doi:10.1128/aac.49.3.1093-1100.2005 (2005).
246 Ross, W., Rowe, T., Glisson, B., Yalowich, J. & Liu, L. Role of topoisomerase II in mediating epipodophyllotoxin-induced DNA cleavage. *Cancer Res* **44**, 5857-5860 (1984).
247 Tewey, K. M., Rowe, T. C., Yang, L., Halligan, B. D. & Liu, L. F. Adriamycin-induced DNA damage mediated by mammalian DNA topoisomerase II. *Science* **226**, 466-468, doi:10.1126/science.6093249 (1984).

248 Tewey, K. M., Chen, G. L., Nelson, E. M. & Liu, L. F. Intercalative antitumor drugs interfere with the breakage-reunion reaction of mammalian DNA topoisomerase II. *J. Biol. Chem.* **259**, 9182-9187, doi:10.1016/s0021-9258(17)47282-6 (1984).
249 Nelson, E. M., Tewey, K. M. & Liu, L. F. Mechanism of antitumor drug action: poisoning of mammalian DNA topoisomerase II on DNA by 4'-(9-acridinylamino)-methanesulfon-m-anisidide. *Proc Natl Acad Sci U S A* **81**, 1361-1365, doi:10.1073/pnas.81.5.1361 (1984).
250 Sinkule, J. A. Etoposide: a semisynthetic epipodophyllotoxin. Chemistry, pharmacology, pharmacokinetics, adverse effects and use as an antineoplastic agent. *Pharmacotherapy* **4**, 61-73, doi:10.1002/j.1875-9114.1984.tb03318.x (1984).
251 Grein, A. Antitumor anthracyclines produced by Streptomyces peucetius. *Adv Appl Microbiol* **32**, 203-214, doi:10.1016/s0065-2164(08)70081-9 (1987).
252 Faulds, D., Balfour, J. A., Chrisp, P. & Langtry, H. D. Mitoxantrone. A review of its pharmacodynamic and pharmacokinetic properties, and therapeutic potential in the chemotherapy of cancer. *Drugs* **41**, 400-449, doi:10.2165/00003495-199141030-00007 (1991).
253 Cain, B. F. & Atwell, G. J. The experimental antitumour properties of three congeners of the acridylmethanesulphonanilide (AMSA) series. *European Journal of Cancer (1965)* **10**, 539-549, doi:10.1016/0014-2964(74)90079-6 (1974).
254 Finlay, G. J., Wilson, W. R. & Baguley, B. C. Chemoprotection by 9-aminoacridine derivatives against the cytotoxicity of topoisomerase II-directed drugs. *European Journal of Cancer and Clinical Oncology* **25**, 1695-1701, doi:10.1016/0277-5379(89)90337-4 (1989).
255 Sørensen, B. S. *et al.* Mode of action of topoisomerase II-targeting agents at a specific DNA sequence. *Journal of molecular biology* **228**, 778-786, doi:10.1016/0022-2836(92)90863-f (1992).
256 Fortune, J. M. & Osheroff, N. Merbarone inhibits the catalytic activity of human topoisomerase IIalpha by blocking DNA cleavage. *The Journal of biological chemistry* **273**, 17643-17650, doi:10.1074/jbc.273.28.17643 (1998).
257 Roca, J., Ishida, R., Berger, J. M., Andoh, T. & Wang, J. C. Antitumor bisdioxopiperazines inhibit yeast DNA topoisomerase II by trapping the enzyme in the form of a closed protein clamp. *Proc Natl Acad Sci U S A* **91**, 1781-1785, doi:10.1073/pnas.91.5.1781 (1994).
258 Wallace, D. J. The use of quinacrine (Atabrine) in rheumatic diseases: A reexamination. *Seminars in Arthritis and Rheumatism* **18**, 282-296, doi:10.1016/0049-0172(89)90050-4 (1989).
259 Wiedemar, N., Hauser, D. A. & Maser, P. 100 Years of Suramin. *Antimicrobial agents and chemotherapy* **64**, doi:10.1128/AAC.01168-19 (2020).
260 Rubio, G. J., Pinedo, H. M., Virizuela, J., van Ark-Otte, J. & Giaccone, G. Effects of suramin on human lung cancer cell lines. *European Journal of Cancer* **31**, 244-251, doi:10.1016/0959-8049(94)00444-a (1995).
261 Perrin, D., van Hille, B. t. & Hill, B. T. Differential sensitivities of recombinant human topoisomerase IIα and β to various classes of topoisomerase II-interacting agents. *Biochemical pharmacology* **56**, 503-507, doi:10.1016/s0006-2952(98)00082-3 (1998).
262 Delmotte, P. & Delmotte-Plaque, J. A new antifungal substance of fungal origin. *Nature* **171**, 344, doi:10.1038/171344a0 (1953).
263 Roe, S. M. *et al.* Structural basis for inhibition of the Hsp90 molecular chaperone by the antitumor antibiotics radicicol and geldanamycin. *Journal of medicinal chemistry* **42**, 260-266, doi:10.1021/jm980403y (1999).
264 Gadelle, D., Bocs, C., Graille, M. & Forterre, P. Inhibition of archaeal growth and DNA topoisomerase VI activities by the Hsp90 inhibitor radicicol. *Nucleic Acids Res.* **33**, 2310-2317, doi:10.1093/nar/gki526 (2005).
265 Singh, J., Hussain, Y., Luqman, S. & Meena, A. Purpurin: A natural anthraquinone with multifaceted pharmacological activities. *Phytother Res*, doi:10.1002/ptr.6965 (2020).

266 Barnabé, N. & Hasinoff, B. B. High-throughput fluorescence flow-injection topoisomerase II inhibition assay. *Journal of Chromatography B: Biomedical Sciences and Applications* **760**, 263-269, doi:10.1016/s0378-4347(01)00274-2 (2001).
267 Roychoudhury, S. *et al.* Development and use of a high-throughput bacterial DNA gyrase assay to identify mammalian topoisomerase II inhibitors with whole-cell anticancer activity. *J Biomol Screen* **8**, 157-163, doi:10.1177/1087057103252302 (2003).
268 Lilley, D. M. The inverted repeat as a recognizable structural feature in supercoiled DNA molecules. *Proc Natl Acad Sci U S A* **77**, 6468-6472, doi:10.1073/pnas.77.11.6468 (1980).
269 Jude, K. M., Hartland, A. & Berger, J. M. Real-time detection of DNA topological changes with a fluorescently labeled cruciform. *Nucleic Acids Res* **41**, e133, doi:10.1093/nar/gkt413 (2013).
270 Gu, M. *et al.* Fluorescently labeled circular DNA molecules for DNA topology and topoisomerases. *Scientific reports* **6**, 36006, doi:10.1038/srep36006 (2016).
271 Alfonso, E. E. *et al.* Novel and Structurally Diversified Bacterial DNA Gyrase Inhibitors Discovered through a Fluorescence-Based High-Throughput Screening Assay. *ACS Pharmacology & Translational Science*, doi:10.1021/acsptsci.2c00113 (2022).
272 Hanvey, J. C., Shimizu, M. & Wells, R. D. Intramolecular DNA triplexes in supercoiled plasmids. *Proc Natl Acad Sci U S A* **85**, 6292-6296, doi:10.1073/pnas.85.17.6292 (1988).
273 Maxwell, A., Burton, N. P. & O'Hagan, N. High-throughput assays for DNA gyrase and other topoisomerases. *Nucleic Acids Res* **34**, e104, doi:10.1093/nar/gkl504 (2006).
274 Gilbert, D. E. & Feigon, J. Multistranded DNA structures. *Current Opinion in Structural Biology* **9**, 305-314, doi:10.1016/s0959-440x(99)80041-4 (1999).
275 Shapiro, A. *et al.* A homogeneous, high-throughput fluorescence anisotropy-based DNA supercoiling assay. *J Biomol Screen* **15**, 1088-1098, doi:10.1177/1087057110378624 (2010).
276 Asha, M. K. *et al.* In vitro anti-Helicobacter pylori activity of a flavonoid rich extract of Glycyrrhiza glabra and its probable mechanisms of action. *J Ethnopharmacol* **145**, 581-586, doi:10.1016/j.jep.2012.11.033 (2013).
277 Mora-Pale, M. *et al.* Antimicrobial mechanism of resveratrol-trans-dihydrodimer produced from peroxidase-catalyzed oxidation of resveratrol. *Biotechnol Bioeng* **112**, 2417-2428, doi:10.1002/bit.25686 (2015).
278 Thalji, R. K. *et al.* Structure-guided design of antibacterials that allosterically inhibit DNA gyrase. *Bioorg Med Chem Lett* **29**, 1407-1412, doi:10.1016/j.bmcl.2019.03.029 (2019).
279 Aldred, K. J. *et al.* Overcoming target-mediated quinolone resistance in topoisomerase IV by introducing metal-ion-independent drug-enzyme interactions. *ACS Chem Biol* **8**, 2660-2668, doi:10.1021/cb400592n (2013).
280 Gadelle, D., Graille, M. & Forterre, P. The HSP90 and DNA topoisomerase VI inhibitor radicicol also inhibits human type II DNA topoisomerase. *Biochemical pharmacology* **72**, 1207-1216, doi:10.1016/j.bcp.2006.07.040 (2006).
281 Park, H. *et al.* Purpurin inhibits adipocyte-derived leucine aminopeptidase and angiogenesis in a zebrafish model. *Biochem Biophys Res Commun* **450**, 561-567, doi:10.1016/j.bbrc.2014.06.017 (2014).
282 Watanabe, Y. *et al.* Adipocyte-derived leucine aminopeptidase suppresses angiogenesis in human endometrial carcinoma via renin-angiotensin system. *Clin Cancer Res* **9**, 6497-6503 (2003).
283 Kim, S. G. *et al.* 4-Hexylresorcinol inhibits transglutaminase-2 activity and has synergistic effects along with cisplatin in KB cells. *Oncol Rep* **25**, 1597-1602, doi:10.3892/or.2011.1218 (2011).
284 Kim, S. G., Lee, S. W., Park, Y. W., Jeong, J. H. & Choi, J. Y. 4-hexylresorcinol inhibits NF-kappaB phosphorylation and has a synergistic effect with cisplatin in KB cells. *Oncol Rep* **26**, 1527-1532, doi:10.3892/or.2011.1436 (2011).
285 Fesus, L. & Piacentini, M. Transglutaminase 2: an enigmatic enzyme with diverse functions. *Trends in biochemical sciences* **27**, 534-539, doi:10.1016/s0968-0004(02)02182-5 (2002).

286 Kucharczak, J., Simmons, M. J., Fan, Y. & Gelinas, C. To be, or not to be: NF-kappaB is the answer--role of Rel/NF-kappaB in the regulation of apoptosis. *Oncogene* **22**, 8961-8982, doi:10.1038/sj.onc.1207230 (2003).
287 Shapiro, A. B. & Austin, C. A. A high-throughput fluorescence anisotropy-based assay for human topoisomerase II beta-catalyzed ATP-dependent supercoiled DNA relaxation. *Anal Biochem* **448**, 23-29, doi:10.1016/j.ab.2013.11.029 (2014).
288 Crisona, N. J., Strick, T. R., Bensimon, D., Croquette, V. & Cozzarelli, N. R. Preferential relaxation of positively supercoiled DNA by E. coli topoisomerase IV in single-molecule and ensemble measurements. *Genes Dev* **14**, 2881-2892, doi:10.1101/gad.838900 (2000).
289 Spangenberg, T. *et al.* The open access malaria box: a drug discovery catalyst for neglected diseases. *PloS one* **8**, e62906, doi:10.1371/journal.pone.0062906 (2013).
290 McGovern, S. L., Caselli, E., Grigorieff, N. & Shoichet, B. K. A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *Journal of medicinal chemistry* **45**, 1712-1722, doi:10.1021/jm010533y (2002).
291 Takahashi, T., Matsuhara, S., Abe, M. & Komeda, Y. Disruption of a DNA topoisomerase I gene affects morphogenesis in Arabidopsis. *Plant Cell* **14**, 2085-2093, doi:10.1105/tpc.001925 (2002).
292 DiNardo, S., Voelkel, K. & Sternglanz, R. DNA topoisomerase II mutant of Saccharomyces cerevisiae: topoisomerase II is required for segregation of daughter molecules at the termination of DNA replication. *Proc Natl Acad Sci U S A* **81**, 2616-2620, doi:10.1073/pnas.81.9.2616 (1984).
293 Figgitt, D. P., Denyer, S. P., Dewick, P. M., Jackson, D. E. & Williams, P. Topoisomerase II: A potential target for novel antifungal agents. *Biochemical and Biophysical Research Communications* **160**, 257-262, doi:10.1016/0006-291x(89)91649-5 (1989).
294 Shen, L. L., Baranowski, J., Fostel, J., Montgomery, D. A. & Lartey, P. A. DNA topoisomerases from pathogenic fungi: targets for the discovery of antifungal drugs. *Antimicrobial agents and chemotherapy* **36**, 2778-2784, doi:10.1128/AAC.36.12.2778 (1992).
295 Agarwal, R. & Duderstadt, K. E. Multiplex flow magnetic tweezers reveal rare enzymatic events with single molecule precision. *Nat Commun* **11**, 4714, doi:10.1038/s41467-020-18456-y (2020).