The University Of Sheffield.

Access to Electronic Thesis

Author:	Io Nam Wong
Thesis title:	Bioinformatic & biochemical characterization of helicases from Bacteriophage T5
Qualification:	PhD

This thesis was embargoed until April 2017.

If this electronic thesis has been edited by the author it will be indicated as such on the title page and in the text.

Institute for Cancer Studies

Department of Oncology

The University of Sheffield

Bioinformatic and biochemical characterization of helicases from Bacteriophage T5

Io Nam Wong

The thesis submitted for the degree of

Doctor of Philosophy

April 2012

Acknowledgements

I gratefully acknowledge my supervisors, Dr. Cyril Sanders and Prof. Jon Sayers, for their helpful guidance and advice throughout my project. I also appreciate their incisive comments on my manuscripts. I thank all my colleagues for their encouragement and help. I am indebted to my parents for their continual support and love. Finally, I thank the University of Sheffield for the award of a scholarship.

Summary

Bacteriophage T5 is a bacterial virus known to have a remarkably high replication rate. It is a double-stranded DNA virus and encodes many of the proteins needed for its own replication. During replication, the viral double-stranded genomic DNA has to be separated by enzymes called helicases, which are motor proteins that utilize chemical energy from ATP to move along and unwind nucleic acid duplexes. Until now, no helicase has been characterized in bacteriophage T5. A bioinformatic analysis on the T5 replication gene cluster showed that several early gene products (D2, D6 and D10), which possess key helicase signature sequences (motifs), may be T5 helicases. This is the first report to investigate helicases of bacteriophage T5 and the study focused on bioinformatic and biochemical characterization of these three potential helicases. Here, D2 and D10 were identified to be two novel T5 helicases, showing helicase activity *in vitro* as well as having some unique properties previously uncharacterised in other helicases. However, D6 did not show ATP*ase* activity under the condition employed and a further investigation on characteristics of D6 is required.

Except for a Walker A motif, no other common conserved motifs related to helicase activity were identified in the D2 protein sequence. However, D2 was found to have a rare bipolar helicase activity giving it the ability to unwind partial duplex DNA with either a 5' or a 3' ssDNA tail (ss-dsDNA). This indicates D2 may possess some unconventional motifs relevant to its helicase activity. The extent of 5'→3' or 3'→5' unwinding activity of D2 was revealed to be dependent on 5' or 3' tail length. Interestingly, D2 displayed biased polarity preference with its 3'→5' unwinding activity being several fold greater than its 5'→3' unwinding activity when the substrates have identical tail length. Differential inhibition of the bipolar helicase activities by high NaCl concentration was also observed. The

5’→3’ unwinding activity was more sensitive to inhibition by high NaCl concentration than the 3’→5’ unwinding activity.

The D10 protein can unwind branched DNA substrates, including forks, Y-junctions and Holliday junctions, which resemble DNA replication, recombination and repair intermediates. Furthermore, D10 was shown to catalyze branch migration of the Holliday junction substrate. Intriguingly, the ability of D10 to unwind the Y-junction substrate was found to be structure-dependent and sequence-dependent. Also, the unwinding activity can be affected by the strand discontinuity of the substrate.

All the findings in this study contribute to a new insight into functional properties of helicases.

Table of Contents

Abbreviations

AAA^+	ATP*ase*s Associated with diverse cellular Activities
ADP	adenosine diphosphate
ATP	adenosine triphosphate
Chi	crossover hotspot instigator
Da	Dalton
DEPC	diethylpyrocarbonate
del	deletion
D-loop	displacement loop
DNA	deoxyribonucleic acid
ds	double-stranded
DTT	dithiothreitol
EDTA	ethylenediaminetetra-acetic acid
FST	first-step transfer
GST	glutathione S-transferase
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HSV	Herpes simplex virus
HTH	helix-turn-helix
kbp	kilobase pairs
MBP	maltose-binding protein
MCM	mini chromosome maintenance
NA	nucleic acid
NTP	nucleoside triphosphate
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polyermase chain reaction
PEI	polyethyleneimine

pI	isoelectric point
PMSF	phenylmethylsulfonyl fluoride
pnk	polynucleotide kinase
P-loop	phosphate-binding loop
RNA	ribonucleic acid
SDS	sodium dodecyl sulfate
ss	single-stranded
SSB	single-stranded DNA binding protein
SST	second-step transfer
WT	wild type

Chaper 1. Introduction

Bacteriophage T5 is a lytic enterobacteria phage that specifically infects the Gram negative bacterium *Escherichia coli*. It belongs to the T5-like viruses genus, which also includes bacteriophages BF23, 29α, BG3, PB (McCorquodale and Warner, 1988; Sayers, 2005) and the recently discovered EPS7 which can infect *Salmonella* (Hong et al., 2008). Bacteriophage T5 consists of an icosahedral head containing a linear double-stranded DNA (dsDNA) genome and a long (~250 nm) and flexible tail, but it does not have a viral envelope (Figure 1.1) (Effantin et al., 2006; Sayers, 2005). It is the most investigated member of the T5 genus and a number of unusual features make T5 an interesting and remarkable phage. T5 has the largest genome (~121 kbp, kilobase pairs) of the T-odd viruses and it carries mysterious nicks in one strand of its dsDNA genome. T5 genomic DNA also possesses large direct terminal repetitions of 10,139 bp (Wang et al., 2005) and some strong prokaryotic promoters (Gentz and Bujard, 1985; McCorquodale and Warner, 1988). The most intriguing feature of T5 is a two-step transfer of phage DNA into its host during infection. The left-end 8.3% of phage DNA, called the first-step transfer (FST) sequence, is injected into the host when the T5 phage attaches to the host cell envelope. After the expression of the FST genes, the remaining second-step transfer (SST) sequence (8.3%-100%) enters the host (Figure 1.3) (Heusterspreute et al., 1987). Although T5 has these unique and fascinating features, not much progress has been made on the area of T5 phage research in these 20 years. Recently, the entire T5 genome was sequenced and these efforts may renew interest in the T5 field.

1.1 Bacteriophage T5 genome

The T5 DNA genome is linear, double-stranded and about 121,752 bp long. It consists of two identical direct terminal repeats of 10,139 bp at both ends. 168 putative ORFs (open reading frames) have been identified in the T5 genome, but only 76 ORFs have been assigned known or hypothetical functions (Wang et al., 2005). Interestingly, there are many unexplained nicks (missing phosphodiester bonds between adjacent nucleotides of one strand) in one strand of the T5 double-stranded DNA (dsDNA) genome. The majority of nicks occur specifically at five sites (7.9%, 18.5%, 32.6%, 64.8% and 99.5% of the length of DNA from the left end of the T5 genome) and a minority of nicks occur in a variable, but non-random manner (Rhoades, 1977; Scheible et al., 1977).

The fixed five nicks were found to have the terminal consensus sequences 5'-N-A/G-OH3'^5'pG-C-G-C-N-3' (Nichols and Donelson, 1977) and McCorquodale suggested that the sequence should be 5'-G-G-OH3'^5'pG-C-G-C-G-G-3' according to the nicking sequence around the major nick at the position 64.8% (McCorquodale, 1999). Four phage-encoded endonucleases were identified to be able to produce nicks in T5 DNA, but those nicks were not coincident with the nicks of mature T5 DNA (Rogers and Rhoades, 1976), so it is still uncertain which nuclease(s) is responsible for making these enigmatic nicks.

Based on sequential gene expression, the T5 genome can be genetically classified into 5 parts (A, C, B, D and A' regions) and functionally sorted into 3 parts (pre-early, early and late gene regions) (McCorquodale and Warner, 1988; Sayers, 2005). Figure 1.2 shows the physical and genetic map of the T5 genome and three functional gene regions are discussed below.

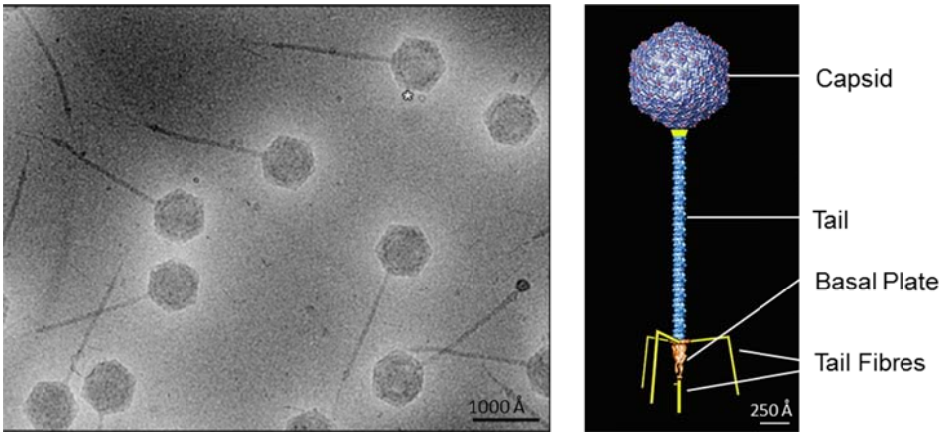


Figure 1.1 The electron micrograph (left, bar = 1000 Å) and surface model (right, bar = 250 Å) of bacteriophage T5. The right panel shows simulated surface model of bacteriophage T5, which consists of icosahedral capsid enclosing the dsDNA genome, a long non-contractile tail, basal plate and tail fibres. This figure is reproduced with permission from (Effantin et al., 2006)

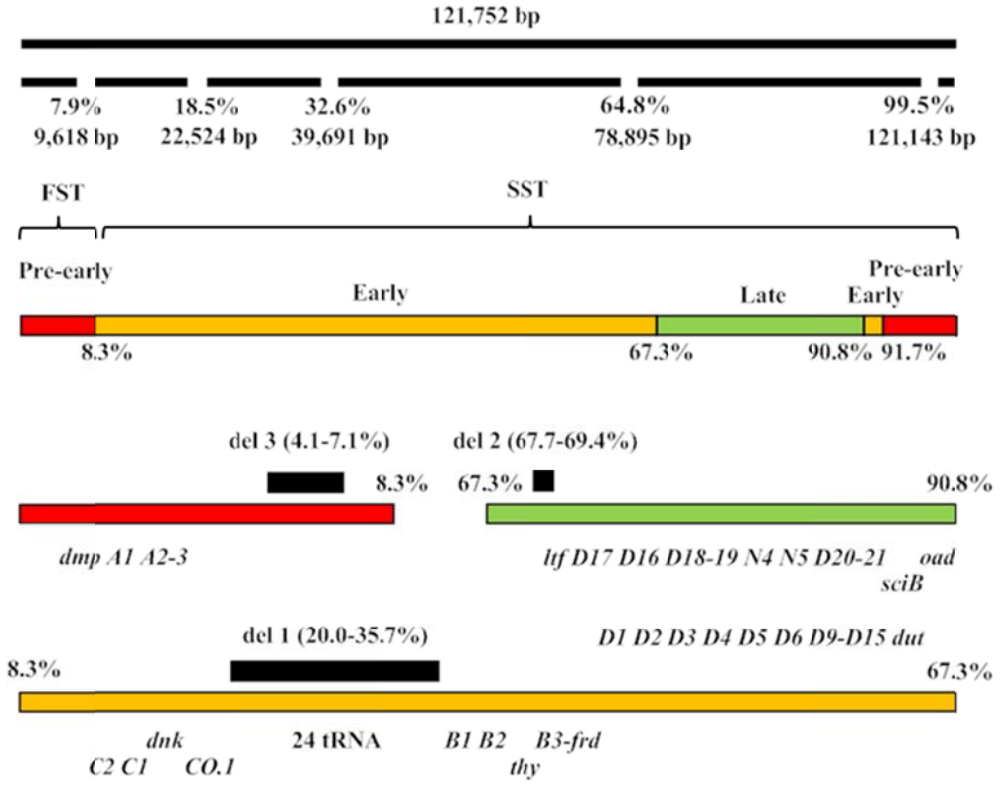


Figure 1.2 The genetic and physical map of T5 genome. The major nicks occur at five specific sites (7.9%, 18.5%, 32.6%, 64.8% and 99.5%; the equivalent bp numbers are bottomed) on one strand of the dsDNA genome. The T5 genome is injected into the host by two-step transfer mechanism. The first-step transfer (FST) sequence is the first 8.3% of phage DNA and the remaining phage DNA is the second-step transfer (SST) sequence. The genome was divided into three functional gene regions: pre-early (red), early (orange) and late (green) gene regions. Pre-early, early and late genes are located in several genetic regions (A, D, C, B, A'; A' is at the right terminal repeat). Genes encoded by plus strand are shown above (such as *D10*) and genes encoded by minus strand are shown below (such as *B3-frd*). The black bars represent three viable deletions (del-1, del-2 and del-3) in the T5 genome. The diagram is adapted from (McCorquodale and Warner, 1988; Sayers, 2005; Wang et al., 2005)

1.1.1 Pre-early gene region

The pre-early gene region is comprised of the first of two direct terminal repeats from 0% to 8.3% (the second being from 91.7% to 100%). There are 10 (5×2) putative promoters in the pre-early gene region, and several promoters were identified as having high affinity for the host's RNA polymerase (Gentz and Bujard, 1985; Wang et al., 2005). T5 bacteriophage injects its DNA into the host via a unique two-step transfer mechanism. The left terminal repeat (0%-8.3%) is the first-step transfer (FST) sequence, initially injected into the host during infection. The pre-early gene region consists of 17 ORFs and 9 pre-early proteins were detected in phage-infected cells (McCorquodale et al., 1977). Some pre-early proteins seemed not to be necessary because a mutant of the T5 related phage BF23, which carries a deletion (del-3) encoding 9 small ORFs, was viable although it failed to synthesize three or four detectable pre-early proteins (McCorquodale et al., 1977; Wang et al., 2005). Among all the ORFs, *A1*, *A2-3* (one gene), *dmp* are the best characterized pre-early genes. *A1* gene encodes a protein of 57 kDa, which is able to form homo-oligomers with a molecular weight of ~244 kDa. The A1 protein can also interact with the A2-3 protein to form hetero-oligomer complexes with a molecular weight of ~364 kDa (Beckman et al., 1971). A1 is believed to be a nuclease involved in host cell DNA

degradation since A1 mutants cannot degrade host DNA (Lanni, 1969), however, there are no reports of this gene being cloned or the protein characterized to date. A2-3 has been found to be a 15-kDa protein that is capable of binding DNA *in vitro* and changing host cell envelope structure (Snyder, 1984; Snyder and Benzinger, 1981). A2-3 has been suggested to initialize second-step transfer (SST) of phage DNA for the reason that an A2-3 mutant can degrade the host DNA, but fails to transfer SST DNA (Lanni, 1969). A 5'-deoxyribonucleotidase (~25 kDa) encoded by *dmp* is responsible for hydrolyzing degraded host DNA products (dNMPs, deoxyribonucleoside 5'-monophosphates) to deoxyribonucleosides which are finally converted to free bases and excreted by cellular enzymes (Mozer et al., 1977; Mozer and Warner, 1977; Warner et al., 1975). Except these three pre-early proteins, most pre-early proteins remain poorly characterized. Only some early studies (Davison and Brunel, 1979; Hausmann and Gold, 1966; McCorquodale and Warner, 1988; Sakaki, 1974; Warner et al., 1975; Warner et al., 1980) reported that they are involved in inhibition of host DNA, RNA and protein synthesis, host DNA degradation, inactivation of host enzymes, including *EcoRI*, *RecBC*, uracil-DNA glycosylase and DNA methylase), and activation of second-step transfer of phage DNA into the host. The inhibition of several host cell enzymes by the pre-early proteins enables the T5 phage to take over the host macromolecular synthesis machinery for its development.

1.1.2 Early gene region

The early gene region of the T5 genome consists of two sections, made up from a large section (8.3%-67.3%) and a small region (90.8%-91.7%) of the genome. It has 34 putative promoters and 111 ORFs. The early gene region includes the largest deletable section (del-1, 20.0%-35.7%) encoding 35 ORFs as well as 24 transfer RNA genes. Most phages encode few or no tRNAs because they usually use host tRNAs to translate their mRNA, but T5 encodes its own

tRNAs presumably to support the highly efficient translation/replication of this phage. The deletion of tRNA genes does not affect the viability of T5 phage, but it makes the replication rate slower (Shlyapnikov et al., 1995). Among all the ORFs, only 41 ORFs were identified to have known or probable functions. Seven ORFs seem to function as signal transducers or regulators, such as serine/threonine protein phosphatases and bacterial transcriptional regulatory protein (LuxR family). Four ORFs are associated with the phage lytic process (Wang et al., 2005) of which the best-characterized one is the *llp* gene that encodes a 7.8-kDa lipoprotein which is important to block superinfection of the host by direct interaction with the host T5 receptor FhuA (Decker et al., 1994). Eight ORFs are involved in nucleotide metabolism, including deoxynucleoside monophosphokinase (*dnk*) (Mikoulinskaia et al., 2003), thymidylate synthetase (*thy*), dihydrofolate reductase (*B3-frd*), ribonucleotide reductase (*nrdA*, *nrdB*, *nrdD*), thioredoxin (*nrdC*), and deoxyuridine triphosphatase (*dut*). A large proportion of early genes appear to reside in a gene cluster which encodes proteins related to DNA replication, transcription, recombination and repair. It has been reported that a T5 transcription-replication complex contains several phage-encoded early proteins including C2, D5, D9, and D15 (Ficht and Moyer, 1980). C2 (90 kDa) is thought to act as an alternative σ factor to change the specificity of host RNA polymerase from T5 phage early to late promoters (Szabo et al., 1975; Szabo and Moyer, 1975). D5 is a 29-kDa DNA binding protein which plays a complex role in regulating T5 gene expression and DNA replication (McCorquodale et al., 1979; Rice et al., 1979). D9 (~95 kDa, previously named D7-8-9) has been found to be a highly processive proofreading polymerase which possesses 5'-3' polymerase and 3'-5' exonuclease activities. This enzyme also has a unique strand-displacement, being able to carry out strand displacement from a nicked primer template (Chatterjee et al., 1991). D15 is the best-characterized component of the T5 replication-transcription complex. *D15* encodes a nuclease which functions as a 5'-3' exonuclease (Sayers and Eckstein,

1990) and a flap endonuclease (Ceska et al., 1996) depending on the substrate and cofactor (Feng et al., 2004). Until now, D15 is the only T5 protein whose crystal structure has been resolved (Ceska et al., 1996). The structure of D15 shares high similarity to its close homolog of T4 RNaseH which has been highlighted to have an important role in exonucleolytic processing of Okazaki fragments from DNA lagging strands during replication. Intriguingly, D9 and D15 together comprise the entire functions of *E. coli* DNA polymerase I holoenzyme, containing 5'-3' polymerase activity (D9), 3'-5' proofreading exonuclease activity (D9) and 5'-3' exonuclease activity (D15) (Sayers, 2005). In addition to the above-mentioned genes, others in the early region are also involved in replication. The *D4* gene is thought to express a NAD-dependent DNA ligase (Sayers, 2005; Wang et al., 2005). The *D10* gene product has been proposed to be a helicase (Blinov et al., 1989). The details of D10 will be discussed below.

1.1.3 Late gene region

The late gene region includes the sequence from 67.3% to 90.8%. There are 3 putative promoters and 23 ORFs in this region. Late genes mostly encode phage structural proteins for the mature virions. The *D20-21* gene product (one gene product, 32 kDa) and *N5* gene product (19 kDa) are head proteins. *N4* gene encodes a tail protein (58 kDa) and *ltf* gene encodes the L-shaped tail fibers (~148 kDa). The *ltf* gene is not essential due to the existence of a viable mutant carrying deletion (del-2 containing *ltf* gene, 67.7%-69.4%) (Heller and Krauel, 1986; Kaliman et al., 1995; Sayers, 2005; Wang et al., 2005). The *oad* (*pb5*) gene encodes the host T5 receptor (FhuA) binding protein which is responsible for irreversible specific absorption of the T5 phage to its host (Mondigler et al., 1995). Apart from structural proteins, late genes also encode other proteins, such

as *sciB* gene encoding a nicking endonuclease (Rogers et al., 1979a; Rogers et al., 1979b).

1.2 Replication cycle of bacteriophage T5

After analysing the complete sequence of the T5 genome, it has been revealed that the sequential expression of pre-early, early and late genes is consistent with the replication cycle (Wang et al., 2005). Once the T5 receptor binding protein (PB5) binds to host T5 receptor (FhuA), the phage begins first-step transfer to inject the left 8.3% of DNA into the cell. The pre-early genes are expressed rapidly to degrade host cell DNA, inhibit hostile host function, take over host macromolecular synthesis machinery, and activate second-step transfer. After the completion of DNA transfer, the expression of early genes starts at about 5 min after infection and continues for up to 20 min or longer. The early-gene products are related to many cellular processes, such as nucleotide metabolism, DNA replication, recombination and repair, and so on. Afterwards, the late proteins appear at around 10-12 min after infection and their syntheses continue until lysis. Late proteins form the head and tail structural components of the mature phage particle and introduce nicks into the phage DNA from a linear genome concatemer (McCorquodale and Warner, 1988; Sayers, 2005; Wang et al., 2005). The replication cycle of T5 phage is briefly illustrated in Figure 1.3. In comparison with T4 and T7 phages, T5 replication is poorly understood even though its infectious cycle is the fastest in all known T-odd phages. So far, only two enzymes (D9 polymerase and D15 flap endonuclease) have been certainly recognized to participate in T5 replication step. There are still a lot of unidentified enzymes, such as helicase.

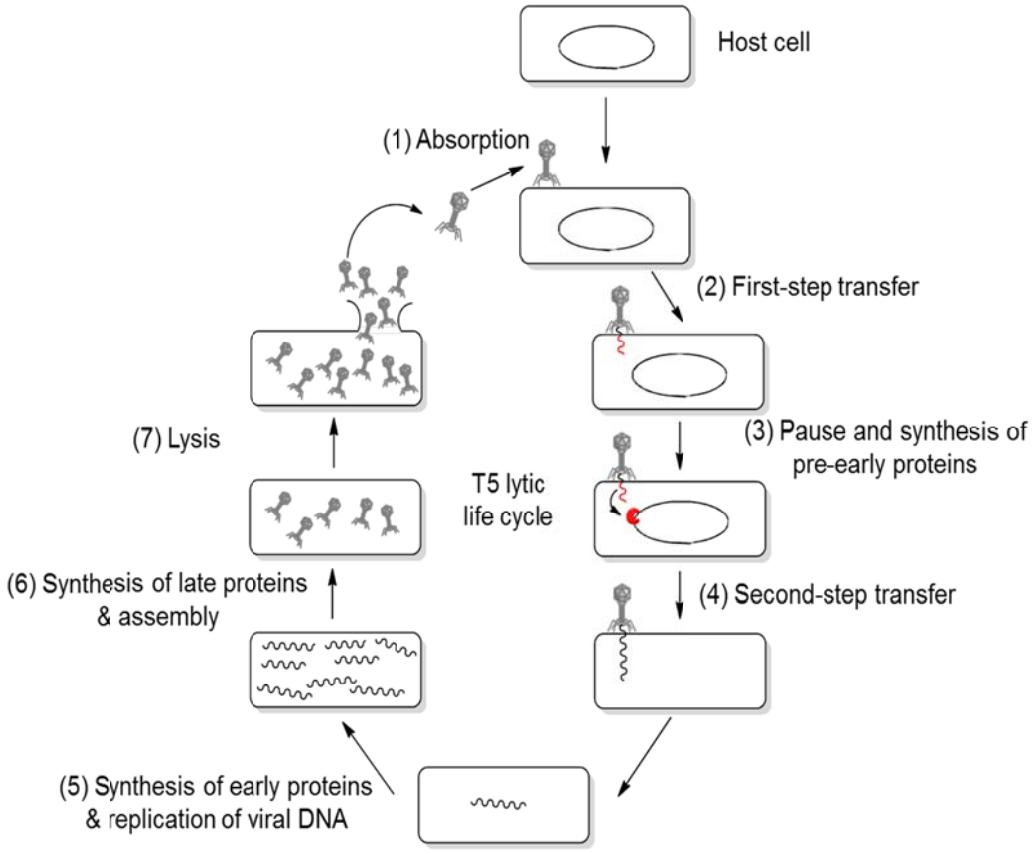


Figure 1.3 An illustration of replication cycle of bacteriophage T5. T5 phage has dsDNA genome and it undergoes a lytic life cycle to infect its host cell (*E. coli*) for its replication. (**1**) The first step is absorption when the phage attaches to the host cell envelope via the interaction of the T5 tail fibre protein PB5 with the host T5 receptor FhuA. (**2**) After the adsorption process, the phage injects the left 8.3% of DNA (first-step transfer sequence, coloured red) into the bacterial cell. (**3**) Then, the transfer of phage DNA pauses and the pre-early proteins (such as A1, A2-3, Dmp, red) are synthesized rapidly to disrupt the bacterial genome and enable the phage to take over the bacterial macromolecular synthesis machinery. (**4**) The next step is that the remaining phage DNA (second-step transfer sequence) enters the bacterial cell and then (**5**) a set of early proteins related to replication of the phage DNA (such as D9 and D15) are expressed to generate many copies of the phage genomes. (**6**) Subsequently, the late proteins are synthesized to form the head and tail components of the mature phage particle and introduce nicks on one strand of the phage DNA. Consequently, the phage genomes are packaged into the preassembled shells (head and tail) to create the mature phages. (**7**) When the phages accomplish the assembly process, the new phage particles are released by lysing the host cell wall. The figure is adapted from (McCorquodale and Warner, 1988; Sayers, 2005; Wang et al., 2005)

1.3 Helicase

Helicases are motor proteins which use chemical energy derived from NTP (nucleoside triphosphate; mostly ATP, adenosine triphosphate) hydrolysis to move along and unwind DNA, DNA-RNA or RNA duplex substrates. They are involved in many cellular processes, including DNA replication, DNA repair, recombination, transcription, translation, ribosome biogenesis, RNA splicing, RNA maturation and nuclear localization (Singleton et al., 2007). Consequently, malfunction of helicases can have disastrous effects on organisms, as exemplified by mutations in RecQ helicases WRN, BLM and RecQ that lead to Werner's, Bloom's, and Rothmund-Thomson syndromes respectively, each characterized by genomic instability and increased cancer susceptibility (Hickson, 2003; Opresko et al., 2004; Perumal et al., 2010). Therefore, it is important to understand the molecular basis of helicase action and many helicases from the prokaryotes have facilitated understanding of their eukaryotic counterparts.

1.3.1 Classification and structure of helicases

Based on a series of conserved motifs, helicases have been classified into six superfamilies (SF1-SF6) (Singleton et al., 2007). SF1 and SF2 helicases are extensively distributed from viruses to eukaryotic cells. It was originally shown that they both share similar sets of seven conserved motifs (1, 1a, 2, 3, 4, 5 and 6, Figure 1.4A) (Hall and Matson, 1999). Additionally, several studies (Korolev et al., 1998; Pause and Sonenberg, 1992; Tanner et al., 2003) have identified some new motifs in these two groups, such as TxGx, Q-motif and motif 4a, and certain motifs are only specific to SF1, SF2 or subfamilies of each superfamily. SF3 consists of helicases usually found in small DNA and RNA viruses. SF3 has five conserved motifs: A, B, B', C and R. SF4, also called DnaB-like family, is exemplified by the *E. coli* DnaB replication fork helicase. There are six motifs (H1, H1a, H2, H3, H4 and R) in this family. SF5 (Rho-like family) includes

helicases related to *E. coli* Rho transcription termination factor. This family has four motifs (1, 1a, 2 and R). SF6 (AAA^+-like family; AAA^+ means ATP*ase*s Associated with diverse cellular Activities) consists of the prokaryotic RuvB which is associated with processing Holliday junctions (an intermediate created during homologous recombination), human RuvBL1 and a number of eukaryotic MCM (mini chromosome maintenance) proteins responsible for replication initiation and elongation. There are five motifs (A, B, S1, R and S2) in SF6 (Singleton et al., 2007).

Although there are a variety of motifs across these six superfamilies of enzymes, three universal motifs are identified in all of them. The three universal motifs include Walker A motif (phosphate-binding loop, P-loop), Walker B motif (Mg^{2+}-binding aspartic acid motif) and an arginine finger (Figure 1.4A). These three motifs are conserved in many NTP*ase* proteins and involved in NTP binding, hydrolysis and the coupling of these activities to the conformational changes of the proteins (Caruthers and McKay, 2002; Hall and Matson, 1999; Hanson and Whiteheart, 2005; Singleton et al., 2007). Their functions and consensus sequences are summarized in Table 1.1.

After a number of crystal structures of helicases from different superfamilies were resolved, it has been revealed that the functional state of SF1 and SF2 helicases are monomeric or dimeric (two subunits), and the active form of SF3-6 helicases are typically hexameric (six subunits) (Figure 1.4B) (Singleton et al., 2007). In addition, it has also been shown that the tandem repeat of a RecA-like fold (domain) commonly exists in the structure of SF1 and SF2 helicases (Figure 1.4C), and six individual RecA- or AAA^+-like folds are arranged in the six subunits of SF3-6 helicases (Figure 1.4D). The above-mentioned three universal motifs generally reside in two adjacent RecA/AAA^+-like folds. This explains why the nucleotide (*e.g.* ATP, ADP and dTTP) binding site, which is related to the three universal motifs, is usually located at the interface of the two adjacent

RecA/AAA^{+}-like folds either within the same polypeptide chain of monomeric helicases (Figure 1.4C) (Sengoku et al., 2006; Subramanya et al., 1996) or two neighboring subunits of multimeric helicases (Figure 1.4D) (Enemark and Joshua-Tor, 2006; Singleton et al., 2000).

Universal motifs	**Consensus sequence**	**Function**	**References**
Walker A	GxxxxG<u>K</u>(T/S)	NTP binding	(Hall and Matson, 1999; Hanson and Whiteheart, 2005; Walker et al., 1982)
Walker B	hhhh<u>DE</u>	Mg^{2+} binding and NTP hydrolysis	(Hanson and Whiteheart, 2005)
Arginine finger	SF1: V(A/G)h(T/S)<u>R</u>xoo SF2: QxxG<u>R</u>xx<u>R</u>	NTP-derived energy coupling to the conformational changes of the protein	(Hall and Matson, 1999)

Table 1.1 The consensus sequences and functions of three universal motifs of helicases. The sequences are represented in single-letter-code format and the key residues are underlined. x, h and o denote any amino acid, a hydrophobic amino acid and a hydrophilic amino acid respectively. Notes: The arginine finger in SF3-6 shares structural similarity, but no consensus primary sequence has been identified.

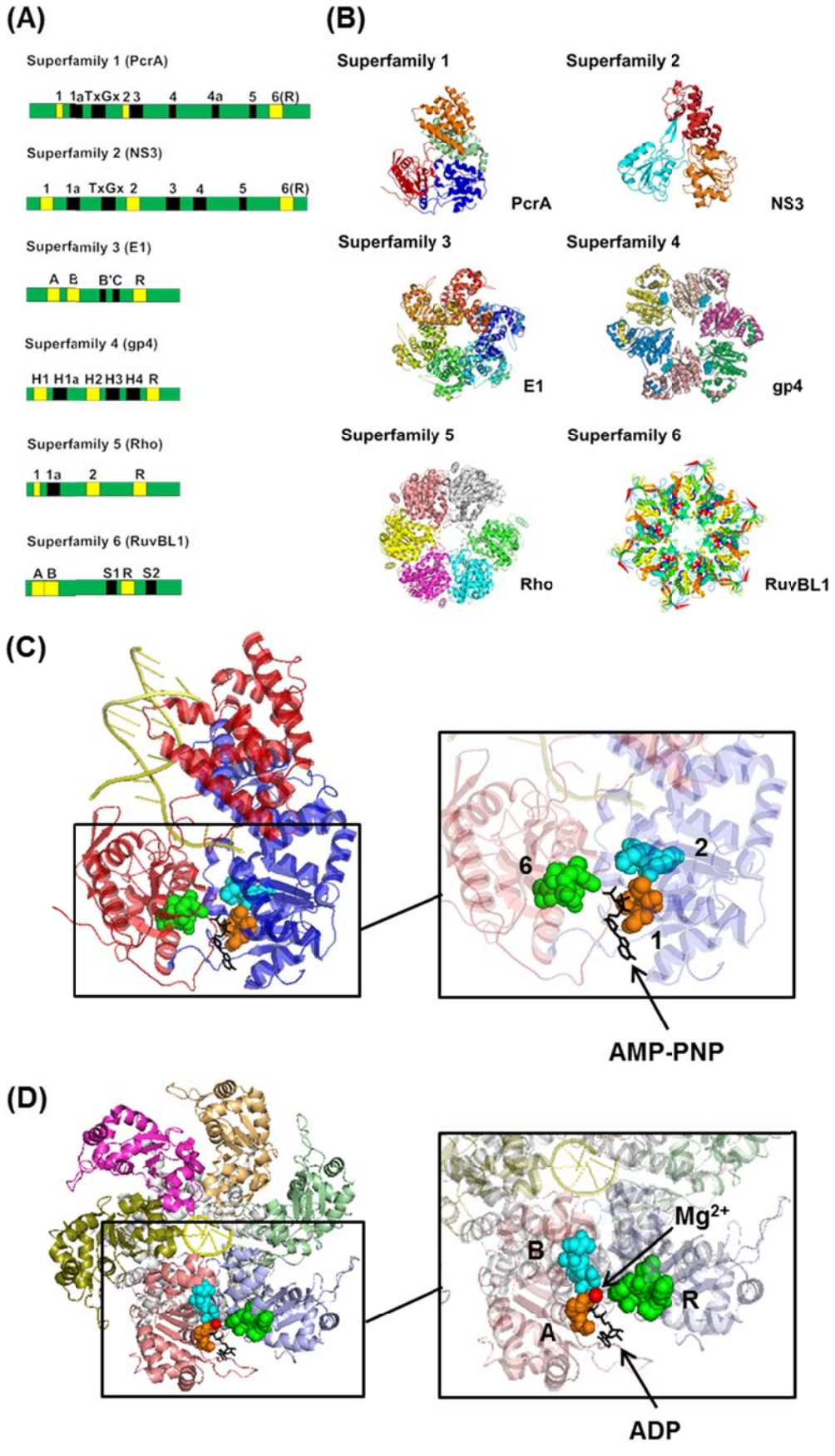

(A)
Superfamily 1 (PcrA)
1 1aTxGx 2 3 4 4a 5 6(R)
Superfamily 2 (NS3)
1 1a TxGx 2 3 4 5 6(R)
Superfamily 3 (E1)
A B B'C R
Superfamily 4 (gp4)
H1 H1a H2 H3 H4 R
Superfamily 5 (Rho)
1 1a 2 R
Superfamily 6 (RuvBL1)
A B S1 R S2
(B)
Superfamily 1
PcrA
Superfamily 2
NS3
Superfamily 3
E1
Superfamily 4
gp4
Superfamily 5
Rho
Superfamily 6
RuvBL1
(C)
6
2
1
AMP-PNP
(D)
Mg2+
B
R
A
ADP

Figure 1.4 Schematic diagram representing conserved motifs and structures of helicases from the six superfamilies. (**A**) The conserved motifs of helicases of SF1-6. Three universal motifs, Walker A (1 or A), Walker B (2 or B) and an arginine finger (6 or R), are shown in yellow. (**B**) Cartoon representations of structures of helicases from six superfamilies. The SF1 and SF2 helicases are monomeric. The helicases from SF3-6 are hexameric. The representative helicases are: SF1, *G. stearothermophilus* PcrA, PDB code: 3PJR; SF2, Hepatitis C virus NS3, PDB code: 1A1V; SF3, Bovine papillomavirus type 1 E1, PDB code: 2GXA; SF4, T7 phage gp4, PDB code: 1E0J; SF5, *E. coli* Rho, PDB code: 1PV4; SF6, human RuvBL1, PDB code: 2C9O. (**C**) Structure of a SF1 helicase, PcrA with a DNA substrate and a bound nucleotide. The DNA substrate (a dsDNA molecule with a 3' ssDNA tail) is coloured in yellow and the bound nucleotide (AMP-PNP, a nonhydrolyzable ATP analogue) is coloured in black. The monomeric PcrA contains tandem RecA-like folds (N-terminal, blue; C-terminal, red). The ATP analogue is bound at the interface of the two RecA-like domains, and surrounded by the three universal motifs. The Walker A (1, orange sphere) and B (2, cyan sphere) motifs are located on the N-terminal RecA-like domain. The arginine finger (6, green sphere) resides in the C-terminal RecA-like domain. (**D**) Structure of a typical hexameric helicase, BPV E1 in complex with ssDNA, ADP and Mg^{2+}. The ssDNA substrate, represented in yellow circular cartoon, is bound within the central hexamer channel. The bound ADP and the divalent Mg^{2+} cation are indicated in black stick and red sphere modes, respectively. The E1 helicase consists of six AAA^{+}-like folds, arranged in six subunits and shown in a different colour (pink, pale orange, pale green, pale blue, pale red and pale yellow). The E1 hexamer has six nucleotide-binding sites, each located at the interface of two adjacent AAA^{+}-like folds. Like the PcrA helicase, a nucleotide-binding site is surrounded by the three universal motifs (A, B and R, indicated in the same way as in the panel **C**). For simplicity, only one nucleotide binding site and a set of universal motifs are shown. The location of Walker A and B motifs which are contributed by one AAA^{+}-like domain is opposite to the location of an arginine finger which is contributed by another neighbouring one. The diagram is adapted from (Enemark and Joshua-Tor, 2006; Hall and Matson, 1999; Hickman and Dyda, 2005; Singleton et al., 2007; Whelan et al., 2011)

1.3.2 Mechanisms of helicase action

Helicases are enzymes coupling NTP binding and hydrolysis to conformational changes which result in translocation and unwinding along nucleic acid duplexes. Until now, not a unified mechanism can be used to explain helicase action. Whereas, there are a wide range of proposed mechanisms for helicase translocation and unwinding movements based on a number of well-characterized helicases. The mechanisms are briefly discussed below in four key aspects including functional state of helicase, helicase translocation

mechanisms, mechanisms for nucleic acid unwinding and coupling of NTP*ase* to translocation and unwinding.

1.3.2.1 Functional state of helicase

As stated above, the helicases from the six superfamilies generally assemble in different oligomeric states to become fully active. SF3-6 helicases (e.g., BPV E1, T7 gp4, *B. stearothermophilus* DnaB and *E. coli* Rho) tend to form a hexamer (Figure 1.5A). Six protein subunits encircle the nucleic acid to form a ring-shaped central channel (Figure 1.4D), which increases the efficiency and processivity of these types of helicases during translocation and unwinding. The subunits of these hexameric helicases usually display cooperativity in NTP*ase* or unwinding activities (Bailey et al., 2007; Donmez and Patel, 2006; Enemark and Joshua-Tor, 2006; Singleton et al., 2000; Skordalakes and Berger, 2003).

Regarding the non-ring-shaped SF-1 and SF-2 helicases, although many helicases in these two superfamilies function as monomers (T4 Dda, HCV NS3, *E. coli* RecQ, Figure 1.5B) (Levin et al., 2004; Nanduri et al., 2002; Xu et al., 2003), there is considerable evidence that the formation of dimer or oligomer is an essential or promoting factor for the activity of some SF1 and SF2 helicases (Donmez and Patel, 2006). For instance, homodimeric formation of UvrD (Figure 1.5C) and heterotrimeric formation of RecBCD (Figure 1.5D) can enhance the helicase processivity and increase the unwinding activity significantly. (Dillingham et al., 2005; Maluf et al., 2003). In some cases, unlike ring-shaped helicases, the multiple molecules of this class of helicases do not assemble to form a stable oligomer and do not show cooperativity in NTP*ase* or unwinding activities. Instead, they individually load onto a tracking strand (Figure 1.5E), which either prevents backward slippage of helicase or simply supplements the molecule of helicase dissociating from the tracking strand to enhance the overall efficiency and activity of the helicase. Therefore, multiple molecules of helicase

are generally more active and efficient than the monomeric protein, especially with the substrate containing longer ssDNA overhangs flanking the duplex region because more monomeric proteins are able to load onto a tracking strand. This observation is exemplified by T4 Dda (Byrd and Raney, 2004, 2005), HCV NS3 (Tackett et al., 2005), and *B. stearothermophilus* PcrA (Dillingham et al., 2000). Another explanation is that the additional helicase molecules binding to the tracking strand may act as single-stranded DNA binding (SSB) proteins to prevent the separated strands from reannealing, which facilitates the helicase-mediated unwinding.

Generally speaking, hexameric helicases display higher processivity than monomeric/dimeric helicases since the nucleic acid is tightly bound within the central channel formed by the hexamer (Bujalowski et al., 1994; Dong et al., 1995; Patel and Hingorani, 1993). The high processivity of the helicase is required for unwinding lengthy chromosomes without dissociation and that is why many hexameric helicases (e.g., T7 gp4, T4 gp41 and BPV E1) are associated with replication forks directly (Dong et al., 1996; Kim et al., 2002; Yao et al., 2009). In contrast, monomeric/dimeric helicase with relatively low processivity are mostly dedicated to the repair and recombination processes, which do not need extended association of helicase with the nucleic acid (Gyimesi et al., 2010; McGlynn and Lloyd, 2001; Nelson et al., 2009).

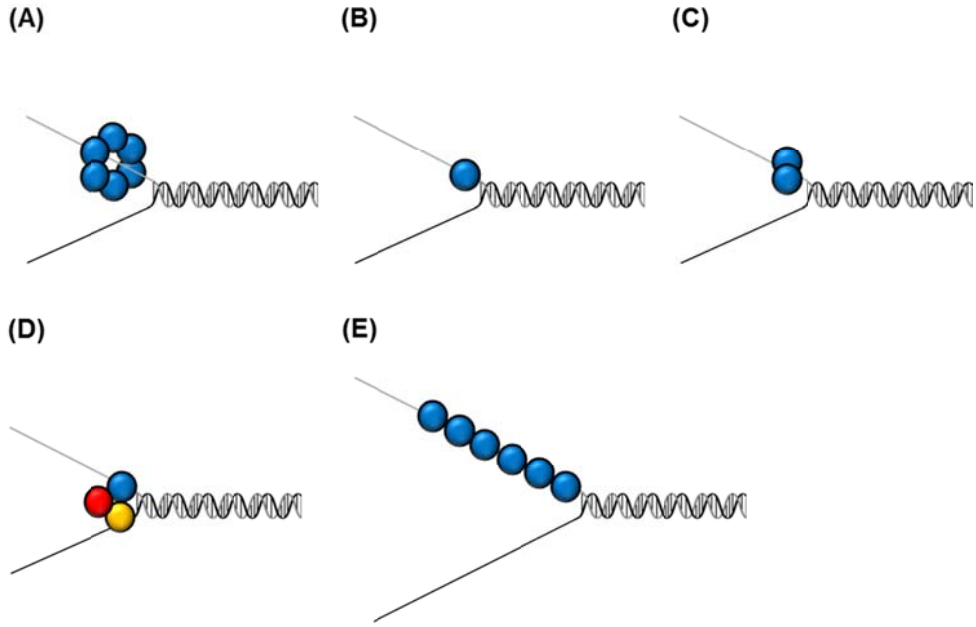


Figure 1.5 Functional state of helicase. (**A**) Ring-shaped helicase hexamer. (**B**) Helicase monomer. (**C**) Helicase homodimer. (**D**) Helicase heterotrimer. (**E**) Helicase oligomer. The figure is adapted from (Donmez and Patel, 2006)

1.3.2.2 Helicase translocation mechanisms

A helicase is typically required to unwind nucleic acid stretches longer than its own binding site during replication, recombination and repair processes, so it has the ability to translocate along either ssNA (single-stranded nucleic acid, i.e., ssDNA or ssRNA) or dsNA (double-stranded nucleic acid). The helicase which moves along ssNA usually demonstrates a biased directionality (either a 5'→3'or a 3'→5') (Bochman and Schwacha, 2009; Patel and Donmez, 2006; Singleton et al., 2007). In contrast, it is difficult to define the directionality of the helicase that transverses dsNA, because dsNA is a bipolar molecule with two nucleic acid strands running in opposite directions. In this case, to monitor how the enzyme is initially loaded onto the substrate is important for determining its directionality (Firman and Szczelkun, 2000; Singleton et al., 2007).

Until now, the mechanisms of helicase translocation are not fully understood,

but it is believed that helicase translocation is driven by NTP binding and hydrolysis. Based on a variety of biochemical, biophysical and structural data, different models have been proposed for elucidating this NTP*ase*-coupled translocation of different groups of helicases. Three major models for broad types of helicases are discussed below. All of these models involve helicase conformation and nucleic acid affinity changes, which are associated with NTP binding and hydrolysis (Figure 1.6) (Levin et al., 2005; Soultanas and Wigley, 2000, 2001).

Stepping model — many experimental studies supported stepping model is applicable to translocation of many monomeric and dimeric helicases, including *B. stearothermophilus* PcrA (Velankar et al., 1999), T4 Dda (Spurling et al., 2006), *E.coli* UvrD (Fischer et al., 2004) and Rep (Ha et al., 2002). This model has been generally described as an inchworm mechanism for monomeric helicases (Soultanas and Wigley, 2001; Velankar et al., 1999; Yu et al., 2006) as well as dimeric inchworm and rolling mechanisms for dimeric helicases (Lee and Yang, 2006; Lohman and Bjornson, 1996; Wong and Lohman, 1992).

In the inchworm mechanisms, the helicase has two nucleic acid binding sites, contributed by two RecA-like domains of a monomeric helicase or two subunits of a dimeric helicase, with weak or tight affinity modulated by NTP binding, hydrolysis and release (Lee and Yang, 2006; Maluf et al., 2003). Figure 1.6A presents how a helicase adopts the inchworm mechanism to translocate along an ssNA strand in a stepwise manner. The event starts with the helicase having one nucleic acid binding site bound tightly to the NA (closed hand) and the other site bound weakly to the NA (open hand). The weak site easily dissociates from the NA and moves forward to bind a position ahead by a power stroke motion. While the weak site transits into a tightly bound state, the original tight site becomes a weak site to dissociate from the DNA and move forward to get close to the current tight site by a power stroke motion. After several additional affinity

changes of the binding sites, all the binding sites resume the beginning state and the helicase has already stepped forward in one direction. There are six conformational changes in the helicase in a full cycle of the stepping inchworm mechanism (Figure 1.6A) (Patel and Donmez, 2006; Soultanas and Wigley, 2000, 2001). The rolling mechanism (a variant of inchworm mechanism) was derived from the biochemical studies on Rep helicase dimer (Lohman and Bjornson, 1996; Wong and Lohman, 1992). By contrast, each subunit of the helicase takes a turn as the leading subunit in the rolling mechanism (Figure 1.9C) instead of the fixed leading subunit in the inchworm mechanism.

Brownian ratchet model — Levin *et.al* have proposed a Brownian ratchet model as an alternative to the stepping model (Levin et al., 2005; Levin et al., 2003). This model requires only one nucleic acid binding site and invokes two conformational states of the helicase with weak and tight ssNA binding affinities. The transition between these two states is regulated by NTP binding and hydrolysis. The Brownian ratchet model predicts that the helicase has limited processivity of translocation, which is different from normally high processivity of translocation in the stepping model (Dillingham et al., 2000; Tomko et al., 2007). HCV (Hepatitis C virus) NS3 helicase is regarded as the representative for this model (Dumont et al., 2006; Levin et al., 2005). For NS3 helicase, ATP binding and hydrolysis switch the helicase between two conformational states (tight and weak DNA-binding states). In the nucleotide-free state, NS3 binds to ssDNA tightly and it is not possible for the helicase to translocate along the nucleic acid (Dumont et al., 2006). In the ATP-bound state, the interaction of the helicase with ssDNA become weaker, so the enzyme can move randomly in either direction due to Brownian motion or possibly dissociate from the nucleic acid on account of the observed low processivity of NS3 (Dumont et al., 2006). Because this weak state has a short lifetime (ATP is hydrolyzed at a very fast rate), the random movement of the helicase lasts very short so that the position of the helicase is close to its starting position. After ATP is hydrolyzed and the product

(ADP) is released, the helicase makes a forward movement by a power stroke motion and rebinds to the nucleic acid tightly. If Brownian motion drives the helicase in the backward direction, the helicase will end up in its original position, but if Brownian motion moves the helicase forward, the helicase will end up in a position ahead from the original position. Repetition of these steps results in the net movement of the helicase along the nucleic acid at one direction (Figure 1.6B) (Levin et al., 2005; Patel and Donmez, 2006).

Spiral staircase model — a spiral staircase model, which is exemplified by BPV E1 and *E. coli* Rho, is specifically proposed for translocation of the hexameric helicase along ssNA (Enemark and Joshua-Tor, 2006; Thomsen and Berger, 2009). In this model, the ring-shaped hexameric helicase encircles and binds a ssNA strand within its central channel, which is like a bolt (the ssNA strand) through a nut (the helicase hexamer) (Enemark and Joshua-Tor, 2006; Patel, 2009; Pyle, 2009; Thomsen and Berger, 2009). On the basis of the structures of E1 and Rho in complex with ssNA and bound nucleotides (ATP or ADP), it has revealed that six subunits of the helicases do not lie on the same plane and they organize into a spiral staircase to track the ssNA backbone in a specific direction sequentially (E1: 3'→5' on DNA; Rho: 5'→3' on RNA) (Figure 1.6C) (Enemark and Joshua-Tor, 2006; Thomsen and Berger, 2009). In the case of Rho, each subunit keeps continuous interaction with one unique nucleotide of the ssRNA in the interior channel. Rho has six ATP-binding sites located at the subunit interfaces. Each of the sites exists in one of four states including nucleotide-free (E, empty), ADP-bound (D), ATP-bound (T) and ATP hydrolysis-competent (T^*). The T^* subunits possess high affinity for RNA. The T and E subunits interact with RNA weakly, and the D subunits have limited RNA-binding activity. In a translocation cycle, Thomsen *et al.* suggested that Rho hydrolyzes six ATP molecules at the subunit interfaces in a sequential manner so that the directional movements of the subunits driven by ATP hydrolysis are also in an orderly fashion (one by one, step by step) (Figure 1.7). In every

ATP-hydrolysis cycle, a T^* subunit releases one nucleotide from 5' end and a D subunit grabs one nucleotide from 3' end, resulting in Rho to translocate along RNA in the 5'→3' direction with a step of one nucleotide per ATP hydrolyzed. For a full cycle, the overall helicase hexamer hydrolyzes six ATP molecules, releases six ADP molecules and translocate six nucleotides (each subunit translocates one nucleotide) (Figure 1.7) (Patel, 2009; Thomsen and Berger, 2009). Recently, Schwartz and colleagues reported Rho is likely to translocate more than one nucleotide per ATP hydrolyzed using a chemical interference method (Schwartz et al., 2009).

All of the above models are mainly proposed for the ssNA translocation of helicases. Compared to the ssNA translocation, little is known about the dsNA translocation as many aspects of dsNA translocation are still experimental intractable (the dsNA unwinding could interfere with the dsNA translocation to confound the results). Even though a triplex displacement assay has been developed to measure the kinetic properties (e.g., the translocation rate, the step size) of helicases for dsDNA translocation (Firman and Szczelkun, 2000; Seidel et al., 2008), there is no mechanism available for explaining this activity due to lack of structural and biochemical data. However, the crystal structure of Rad54 (a SWI2/SNF2 chromatin-remodeling enzyme which shares high structure similarity with SF2 helicases) bound to dsDNA suggested that Rad54 adopts an inchworm mechanism similar to that described for PcrA to translocate on dsDNA (Singleton et al., 2007; Thoma et al., 2005).

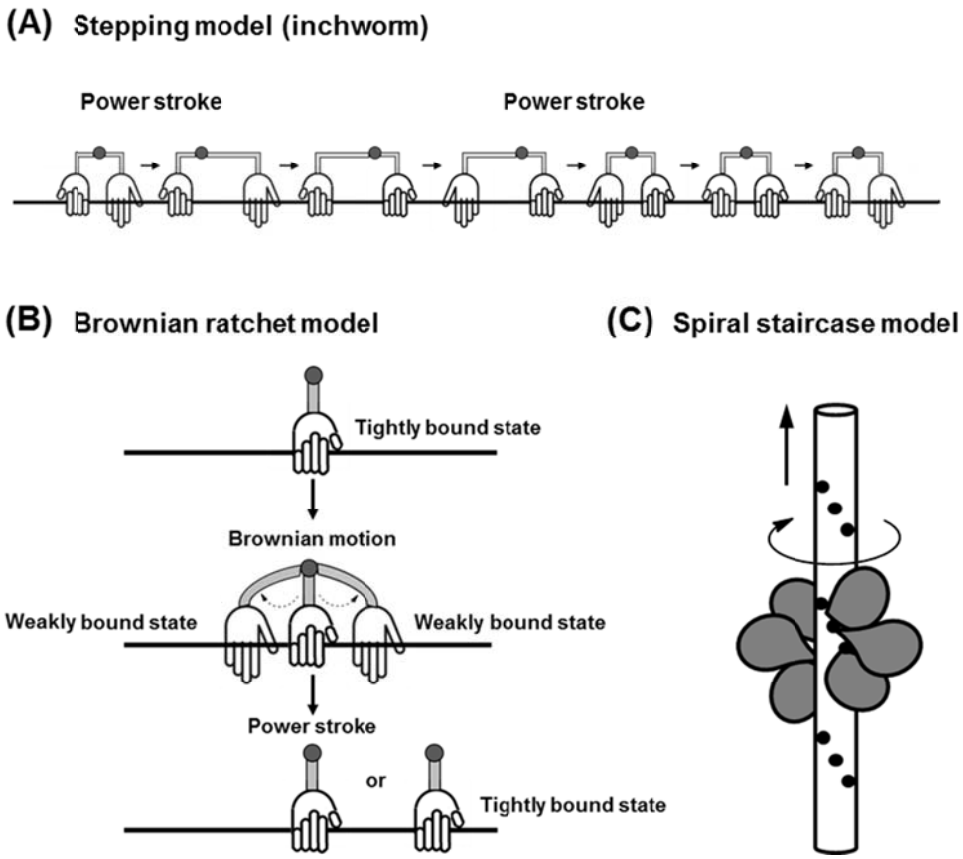


Figure 1.6 Mechanistic models for helicase translocation. (**A**) Stepping model (inchworm). The helicase initially possesses a tight (closed hand) and a weak (open hand) nucleic acid binding site. The nucleic-acid affinities of these two sites are modulated by NTP binding, hydrolysis and release. The affinity changes result in the conformational changes required for translocation of the helicase along the ssNA (power stroke) in one direction. For one cycle of the inchworm mechanism, the helicase needs to undergo six conformational changes. (**B**) Brownian ratchet model. The helicase has one nucleic acid binding site in either tightly or weakly bound state, determined by the state of the NTP-binding site. For NS3, in the nucleotide-free state (tight), the helicase binds to the nucleic acid tightly so that it cannot move along the nucleic acid. In the ATP-bound state (weak), the helicase fluctuates to move in either direction by Brownian motion. After ATP is hydrolyzed and the product is released (weak to tight), the helicase makes a step forward by a power stroke motion and rebind to the nucleic acid tightly. Depending on which direction Brownian motion drives the helicase in, the helicase molecule moves forward or goes back to the original position. (**C**) Spiral staircase model. The helicase hexamer encircles and binds the nucleic acid in the central channel. Six nucleotides (black circles) are arranged in the spiral staircase-like structure formed by six subunits (grey). Since the hexameric helicases (Rho and E1) couple ATP hydrolysis to their translocation actions, the sequential ATP hydrolysis in the subunits leads to the ordered movement of helicase hexamer along the nucleic acid in one direction. The diagrams are adapted from (Patel and Donmez, 2006; Pyle, 2009)

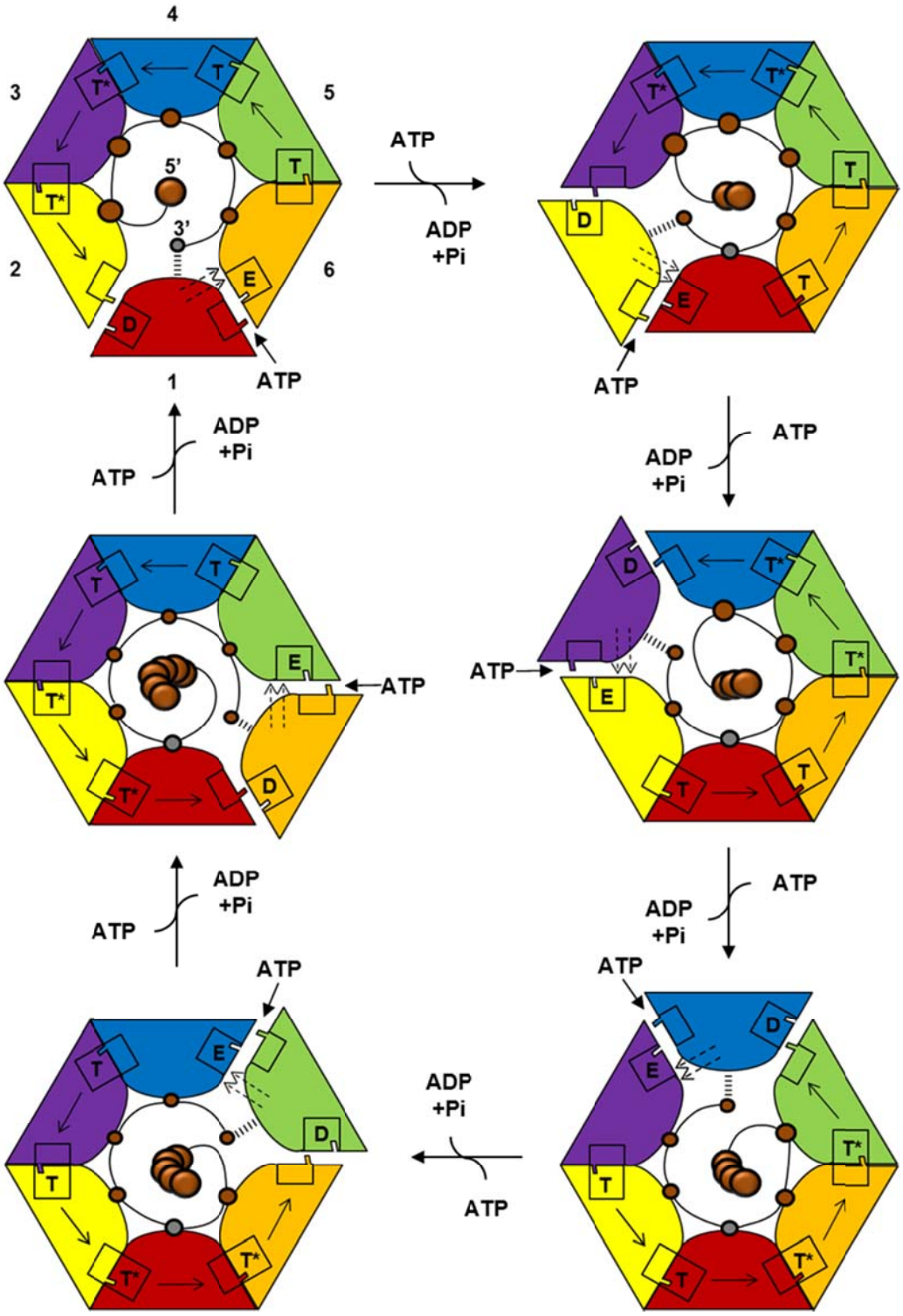


Figure 1.7 Schematic diagram for a Rho translocation cycle. The helicase coordinates the movements of subunits (coloured trilateral figures) along RNA with the sequential ATP hydrolysis cycles. In each ATP hydrolysis cycle, a T^* subunit releases one nucleotide from 5' end and a mobile D subunit makes a contact with a nucleotide from 3' end, resulting in Rho to step forward a nucleotide in the 5'→3' direction. Six nucleotides contacted by the subunits are indicated by brown circles and the grey one is used as a reference point. The free 5' end

nucleotides (released by T^* subunit) spiralling out of the plane of the paper are illustrated as brown spheres. The incoming 3' end nucleotides are connected by dashed lines. The solid arrows show the progression toward subsequent steps in the ATP hydrolysis cycle. The dotted arrows indicate the movements of the D subunit to its partner subunit after ATP binding. This diagram is adapted from (Thomsen and Berger, 2009)

1.3.2.3 Mechanisms for nucleic acid unwinding

In many cases, the helicase couples its unidirectional translocation along the nucleic acid to local base pair separation as described below in passive and active unwinding mechanisms. However, the relationships between translocation and the nucleic acid unwinding have not been established for most helicases and there are still many helicases that are able to translocate along the nucleic acids uncoupled from unwinding, so the proposed mechanisms for translocation and nucleic acid unwinding are introduced separately here. In order to explain how the helicase unwinds the nucleic acids, the mechanisms for the nucleic acid unwinding are discussed in two aspects: active vs passive and base pair separation models.

"Active" vs "Passive" — Lots of helicases were found to be able to translocate along ssNA strand with defined polarity in either a 5'→3' or a 3'→5' direction via one of the above-mentioned mechanisms and unwind dsNA at the ss-dsNA junction. Von Hippel and Lohman groups proposed the helicase separates the dsNA using either a "passive" or an "active" mechanism (Amaratunga and Lohman, 1993; Geiselmann et al., 1993). In a passive mechanism, the helicase waits for the terminal base pairs at the junction to open by thermal fluctuations. When the terminal base pairs open, the helicase moves along the tracking strand and binds the newly opened bases to physically stop them from reannealing (Figure 1.8A). This mechanism is simple, but it is supposed to be efficient because the terminal base pairs open and close at a relatively rapid rate (Gueron and Leroy, 1995; Nonin et al., 1995). Since the base-pair opening rate is often lower than the helicase translocation rate, the

helicase unwinding rate could not reach the maximum value (the helicase ssNA translocation rate) according to the passive mechanism. However, the unwinding rates of many helicases, such as *E. coli* UvrD and RecQ, were found to be similar to their translocation rates, so these helicases were suggested to use a more efficient mechanism (an active mechanism) to unwind their substrates (Fischer et al., 2004; Manosas et al., 2010; Sun et al., 2008). In an active mechanism, the helicase actively participates in destabilization of the dsNA near the junction to stimulate its unwinding rate (Figure 1.6B), but the mechanism for the destabilization remains elusive (Amaratunga and Lohman, 1993; Pyle, 2008; Soultanas et al., 2000; Velankar et al., 1999).

Recently, several theoretical and experimental studies have attempted to differentiate between active and passive mechanisms by using the degree of activeness of helicases (the extent of the helicase to destabilize the dsNA region near the junction) (Betterton and Julicher, 2005; Johnson et al., 2007; Manosas et al., 2010). The degree of activeness is represented by the destabilization energy (ΔG_d, the free energy is used to indicate the probability of helicase-mediated dsNA destabilization at the junction) (Johnson et al., 2007) or the ratio of the helicase dsNA unwinding rate to the helicase ssNA translocation rate (V_{un}/V_{trans}) (Manosas et al., 2010). Wang *et al.* and Croquette *et al.* developed two scales respectively for defining a purely passive helicase ($\Delta G_d \sim 0\ k_BT$, where k_B is the Boltzmann's constant and T is the temperature; $V_{un}/V_{trans} \sim 0.1$) and an optimally active helicase ($\Delta G_d \sim 3.4\ k_BT$; $V_{un}/V_{trans} \sim 0.9$) (Figure 1.6C) (Johnson et al., 2007; Manosas et al., 2010). No matter which scale is used, many helicases (such as T7 gp4 and NS3) fall in the middle of scale between passive and active mechanisms for unwinding. Because these two mechanisms are not contradictory, it is likely that the helicases combine them together for unwinding the nucleic acids (Manosas et al., 2010; Pyle, 2008).

In addition, it has been suggested that the existence of passive and active mechanisms is related to the function of helicases (Manosas et al., 2010). Some

remarkable studies on the replicative helicases (e.g., T7 gp4, T4 gp41 and *E. coli* DnaB) revealed that these helicases alone are mainly passive (Johnson et al., 2007; Kim et al., 1996; Lionnet et al., 2007), but their unwinding rates are greatly stimulated and almost approach the ssNA translocation rates in the presence of other replisomal partner proteins (such as T7 polymerase) (Delagoutte and von Hippel, 2001; Johnson et al., 2007; Stano et al., 2005; von Hippel and Delagoutte, 2001). In comparison with the replicative helicases, the repair helicases (such as UvrD and RecQ) tend to be mainly active and work alone to unwind dsNA at their maximal rates (Manosas et al., 2010; Sun et al., 2008; von Hippel and Delagoutte, 2001).

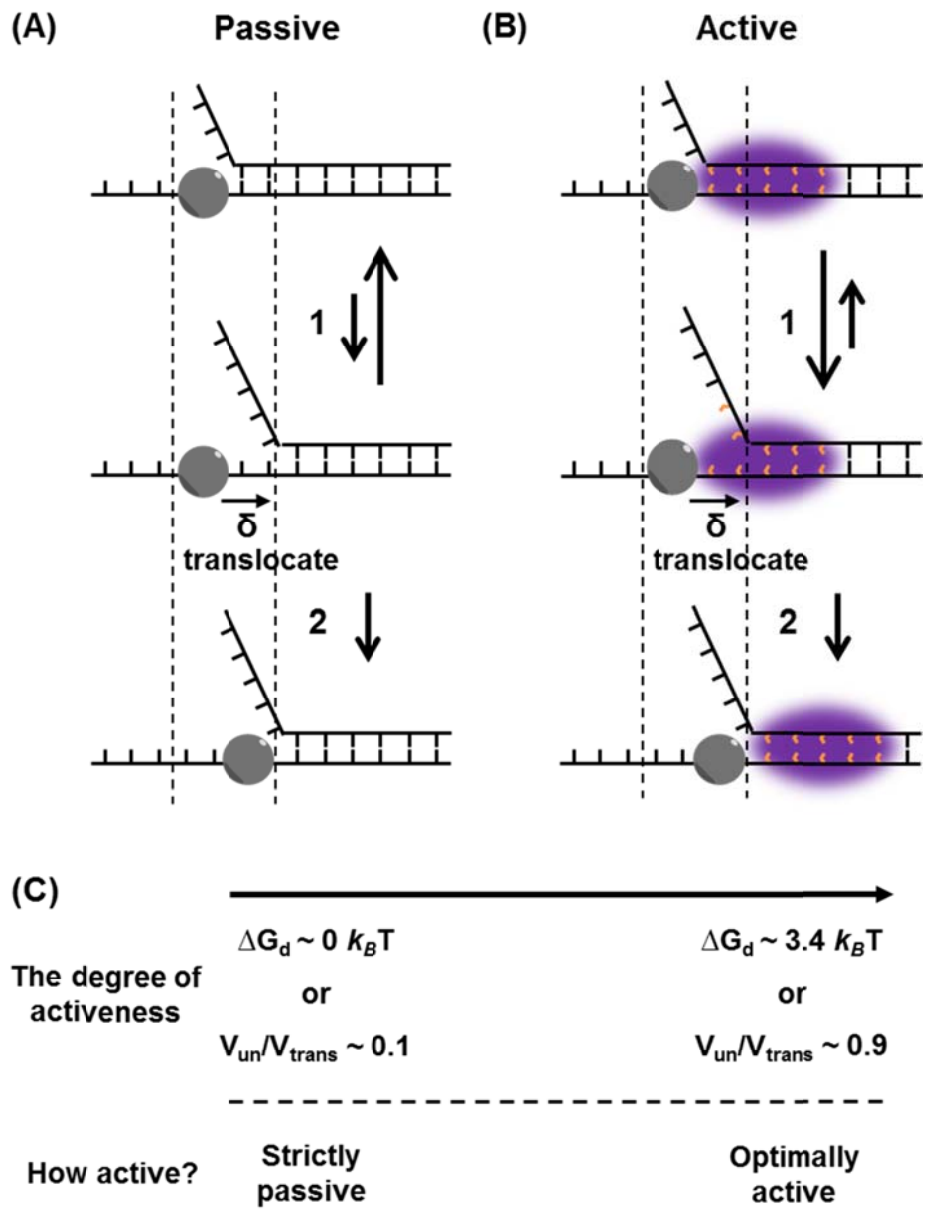


Figure 1.8 Illustration of passive and active mechanisms. (**A**) A cartoon of the passive mechanism. In this model, the helicase waits for the terminal base pairs to open by thermal fluctuations (step 1). When the number of ssNA bases between the helicase and the junction is equal to or more than the helicase step size (δ, bases are translocated per catalytic cycle; here is two bases translocated per catalytic cycle), the helicase is able to translocate and bind to the newly opened bases in order to stop them from reannealing (step 2). (**B**) A cartoon of the active mechanism. In this model, the helicase actively destabilizes the base pairs near the junction (orange pairings and purple shading) (step 1). This makes the base-pair opening more frequent so that it is more likely for the helicase to step forward (step 2). (**C**) Two different scales (ΔG_d or V_{un}/V_{trans}) for the degree of activeness of helicases. When $\Delta G_d \sim 0\ k_BT$ or $V_{un}/V_{trans} \sim 0.1$, the helicase may be considered strictly passive, and when $\Delta G_d \sim 3.4\ k_BT$ or $V_{un}/V_{trans} \sim 0.9$, the helicase may be considered optimally active. The figure is adapted from (Johnson et al., 2007; Manosas et al., 2010)

Base pair separation models — As the translocation mechanisms, there are diverse proposed models for explaining base-pair separation activities of helicases. These models reflect various biochemical properties of helicases, such as functional state, interaction with nucleic acid substrate, and the effect of NTP binding and hydrolysis on the mode of unwinding. The common feature of these models is that the base pair separation occurs at the junction of single-stranded and double-stranded regions. According to how the helicase unwinds nucleic acids, the models can be divided into two broad types— "steric" and "pump" models.

(a) Steric model. It is a simple mechanistic model that the helicase translocates along one strand of the nucleic acid while physically displacing the complementary strand at the ss-dsNA junction via steric exclusion, like a wedge or a wire stripper. Many ring-shaped or non-ring-shaped helicases (such as T4 Dda, T7 gp4 and *E. coli* DnaB) have been postulated to adopt this model to unwind nucleic acids (Figure 1.9A and D). (Ahnert and Patel, 1997; Jezewska et al., 1998; Kaplan, 2000; Kaplan et al., 2003; McGeoch et al., 2005). Furthermore, some helicases may employ additional ways to destabilize the double-stranded region near the junction before separating the strands by directional translocation (Figure 1.9B and C), but the mechanism for this destabilization is still unknown (Cheng et al., 2002; Lohman and Bjornson, 1996; Velankar et al., 1999).

(b) Pump model. Expect the steric model, several pump models have been brought forward to account for the observations that some ring-shaped helicases (i.e., MCM and SV40 LTag) can translocate along and unwind dsDNA to induce initiation of DNA replication or homologous DNA recombination at a Holliday junction. The hexameric helicases are thought to unwind dsDNA possibly by either twisting it apart (rotary pump) or extruding it though channels within the helicase (dsDNA pump) (Bochman and Schwacha, 2009; Takahashi et al., 2005). These models require that the central channels of these ring-shaped helicases are

wide enough to bind dsDNA, which was supported by the structure of SV40 LTag hexamer (Gai et al., 2004; Li et al., 2003). Except the big dsDNA central channel, SV40 LTag also has six narrower side channels vertical to the central channel. Thus, it was proposed the hexameric LTag pumps dsDNA into the central channel and extrudes ssDNA out through its side channels (Figure 1.9G) (Gai et al., 2004; Li et al., 2003). In addition, LTag was identified to function as double hexamers (a head-to-head complex) (Alexandrov et al., 2002; Valle et al., 2006), so LTag was suggested to operate as two associated dsDNA pumps (a hexamer equals to a pump) (Takahashi et al., 2005). Due to the double hexamer formation of LTag, the separated ssDNA strands should form two loops of ssDNA emanate (bunny ear) (Figure 1.9G), which were confirmed in electron microscopic images (Wessel et al., 1992). A variant of the dsDNA pump model, called ploughshare model, was postulated for MCM2-7 helicase-mediated DNA unwinding. The variations in this model are that the helicase functions as a single hexamer and physically separates the dsDNA strands by passing them over a rigid "wedge" structure (ploughshare) (Figure 1.9F) (Takahashi et al., 2005). Additionally, a rotary pump model was proposed as an alternative model for MCM2-7 DNA unwinding. This model was motivated by the findings that F_1-ATP*ase* couples ATP binding and hydrolysis to the rotational movements of protein components within its central channel (Hingorani et al., 1997; Zhou et al., 1997) and MCM complexes are abundant in the pre-replication complex (multiple helicase hexamers per replication fork) (Bowers et al., 2004; Edwards et al., 2002; Randell et al., 2006). The rotary pump model suggested that two groups (at least two) of helicase hexamers load onto the replication origin, translocate away from one another and ultimately anchor at the sites distant from the origin. Then, two groups of helicase hexamers coordinately rotate dsDNA in opposite directions, causing the torsional stress to unwind the intervening dsDNA region (Figure 1.9E) (Laskey and Madine, 2003; Takahashi et al., 2005). Even though the pump models were initially proposed for the ring-shaped helicase (particularly MCM),

some of them were partially based on the data from non-ring-shaped helicases. For instance, the ploughshare structure was inspired by the crystal structure of RecBCD (Bochman and Schwacha, 2009; Singleton et al., 2004; Takahashi et al., 2005). Hence, it is possible that the pump models are also applicable to some non-ring-shaped helicases, especially those capable of translocating along and unwinding dsDNA.

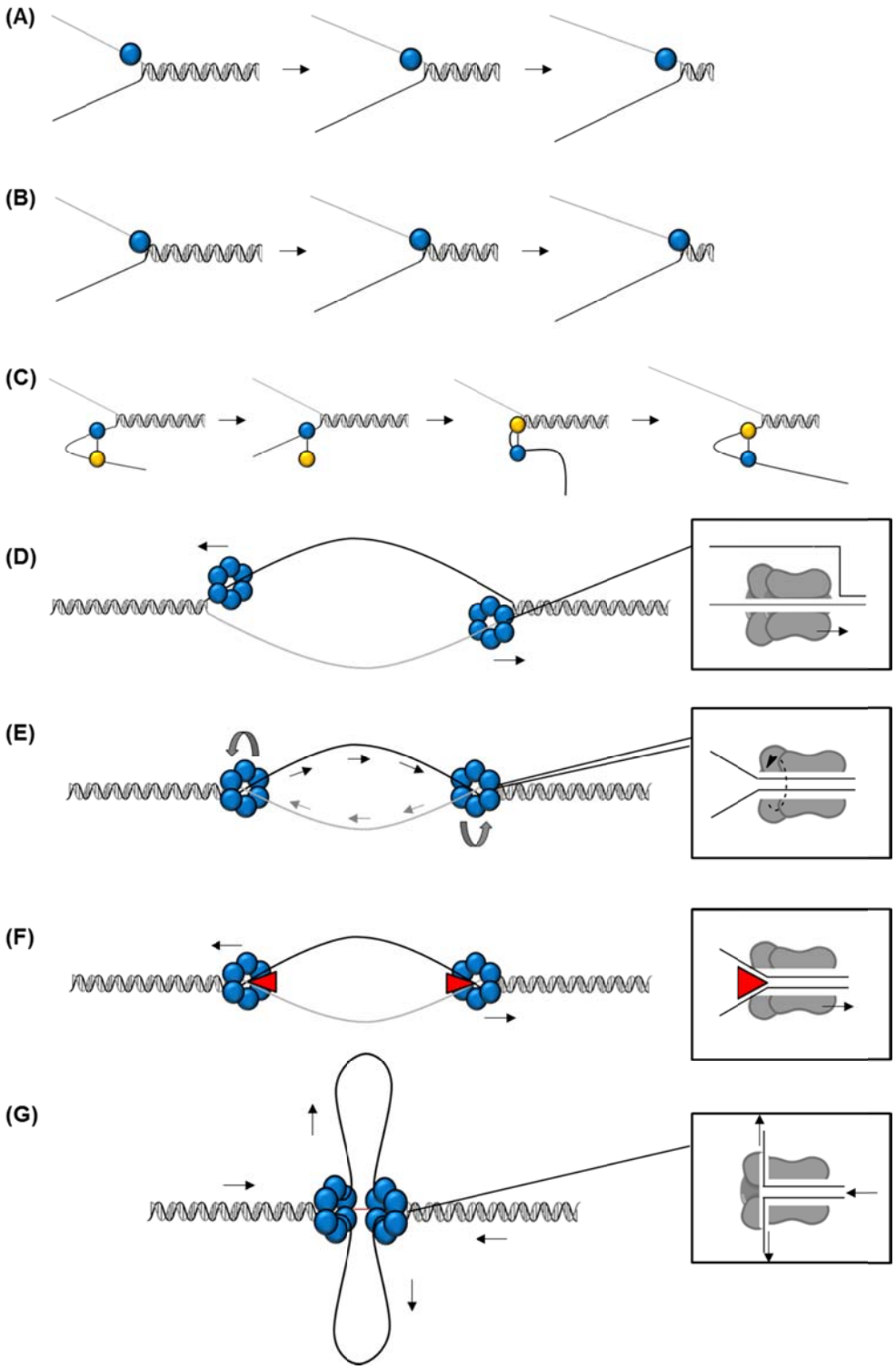


Figure 1.9 Proposed base separation models. (**A**) and (**D**) are steric models. The helicase (**A**, monomer; **D**, ring-shaped hexamer) translocates along one strand while displacing the complementary strand at the ss-dsNA junction by steric exclusion. (**B**) and (**C**) show that the helicase (**B**, monomer; **C**, dimer) destabilizes the duplex region close to the junction before fully separation of the double strands by directional translocation. (**E-G**) Pump models. The right rectangles highlight the interior structures of the helicase hexamer:DNA complex and

reveal how the helicases unwind the duplex DNA. (**E**) The rotary pump model. There are two immobilized hexameric helicase rings at the sites distant from the replication origin. They coordinately rotate dsDNA in opposite directions, generating torsional stress to unwind duplex DNA in the intervening region. (**F**) The ploughshare model. The helicase ring encircles and translocates on dsDNA. It drags a rigid wedge structure (like a ploughshare, red triangle) to separate dsDNA strands. (**G**) The dsDNA pump model. Two physically connected (red line) helicase hexamers pump dsDNA into their central channels and extrude ssDNA strands out through their side channels. The figure is adapted from (Lohman and Bjornson, 1996; Patel and Donmez, 2006; Takahashi et al., 2005)

1.3.2.4 Coupling of NTP*ase* to translocation and unwinding

As described above, NTP binding and hydrolysis are coupled to translocation and unwinding activities of helicases. All helicases possess one or multiple NTP-binding sites, which exist in one of three interconverting states (T, NTP bound; D, NDP bound; E, empty). The transition of these three states causes changes in nucleic acid binding affinity and configuration of the helicase, resulting in translocation and/or unwinding. As mentioned above, a NTP-binding site is commonly located at the interface between two RecA/AAA^{+}-like domains (Figure 1.4C and D). In monomeric helicases, a single NTP binding site regulates the conformational changes of the helicases. Whereas, there are multiple NTP binding sites in multimeric helicases and these sites can potentially work coordinately to modulate the action of helicases. For example, when dimeric Rep helicase translocates along ssDNA, two ATP-binding sites appear in two different states (the site on the DNA-bound subunit are occupied by ATP and another site on the DNA-free subunit are bound by ADP) (Hsieh et al., 1999). Hexameric helicases are capable of binding and hydrolyzing six NTPs, so their coordinated NTP binding and hydrolysis are much complicated (Patel and Donmez, 2006; Singleton et al., 2007). Until now, several models were proposed to elucidate the cooperativity of NTP binding and hydrolysis between the subunits of the hexameric helicases, including three-site sequential (Hingorani et al., 1997; Stitt and Xu, 1998), all sites sequential (Enemark and Joshua-Tor, 2006; Singleton et

al., 2000; Thomsen and Berger, 2009), stochastic (Crampton et al., 2006; Liao et al., 2005; Martin et al., 2005) and concerted (Gai et al., 2004) models (Figure 1.10).

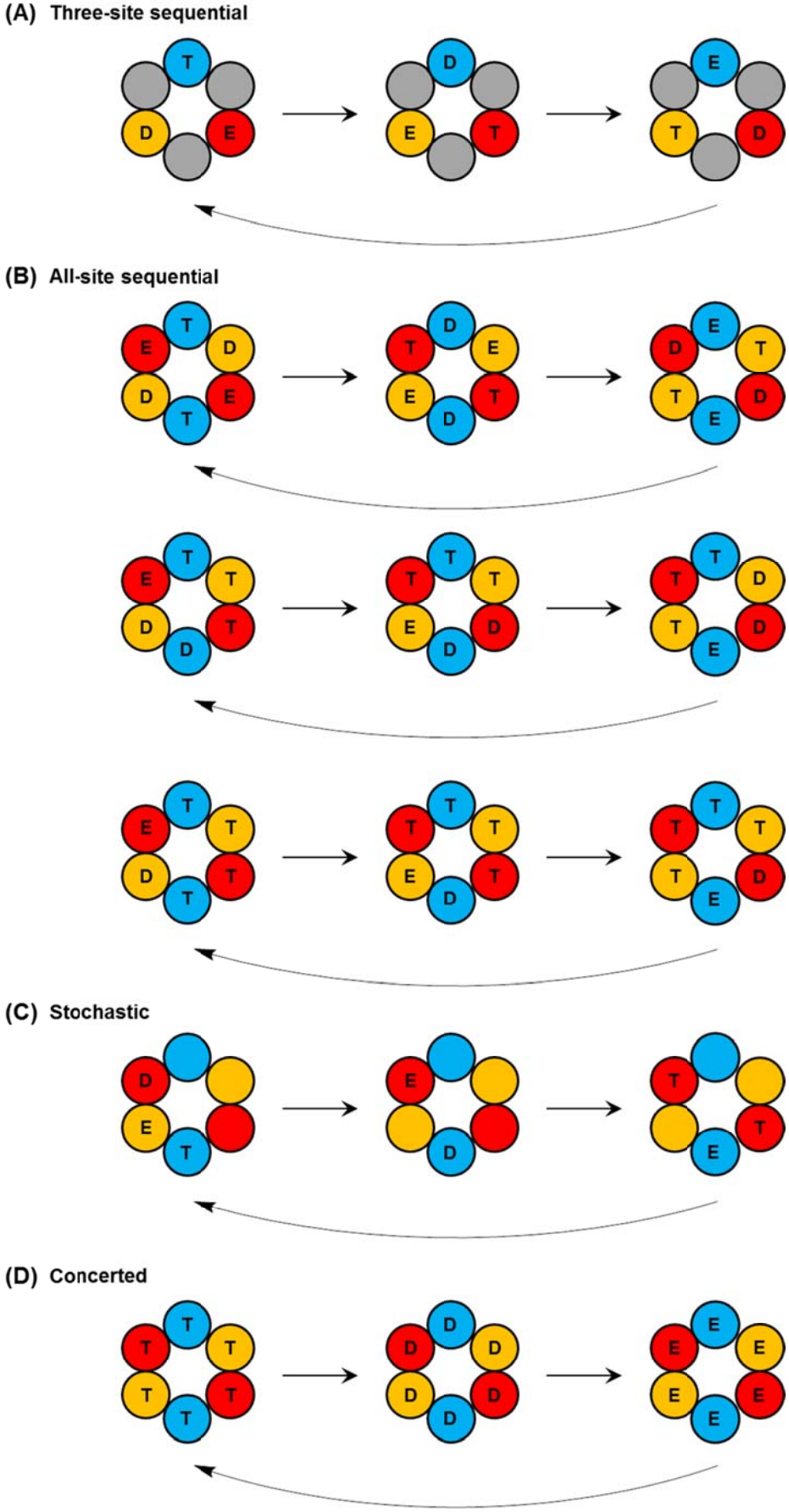

(A) Three-site sequential
(B) All-site sequential
(C) Stochastic
(D) Concerted

Figure 1.10 Proposed models for NTP binding and hydrolysis of hexameric helicases. One circle represents a NTP binding site and each site exists in one of three states: NTP bound (T), NDP bound (D) or apo (empty, E). (**A**) Three-site sequential model. In six NTP-binding sites, three or four are active (coloured) and the remainder are not active (grey). Only the active sites are catalytic and they cycle through three states coordinately and sequentially. (**B**) All-site sequential model. All sites are active and catalyze NTP hydrolysis cycles sequentially. NTP hydrolysis is coordinated between the neighbouring NTP-binding sites in different states (T7 gp4) or multiple same states (BPV E1 and *E. coli* Rho). (**C**) Stochastic model. All sites are catalytically active, but unlike sequential model, they catalyze NTP hydrolysis cycles independently and randomly. (**D**) Concerted model. All six sites are in the same state and transit between T, D, and E states simultaneously. This diagram is adapted from (Singleton et al., 2007; Thomsen and Berger, 2009)

1.4 Helicases — feasible drug targets for anti-cancer and anti-pathogenic therapies

Since different helicases in multiple organisms have been identified to be essential for many aspects of nucleic acid metabolism, the helicases have been assumed to be potential drug targets for anti-cancer and anti-pathogenic therapies (Aggarwal and Brosh, 2009; Gupta and Brosh, 2008; Kwong et al., 2005; Tuteja, 2007). Hence, it is of great interest to find a viable approach to modulate the expression or function of the helicase related to tumourigenesis or pathogenesis. Several advanced techniques (including RNA interference, small molecule and peptide-based approaches) were found to be able to do so, and the small molecule approach has been widely used in many preclinical studies (Aggarwal et al., 2011; Belon and Frick, 2009; Grandori et al., 2004; Gupta and Brosh, 2008; Lehoux et al., 2012). It has proposed that the inhibition of the activities of some specific helicases (such as human WRN, HCV NS3 and *P. falciparum* PfDH60) may suppress the progression of carcinogenesis or pathogen infections (Chen et al., 2009; Grandori et al., 2004; Pradhan and Tuteja, 2006). Therefore, to screen for helicase inhibitors has been a growing field of interest in recent years because it provides the opportunity to find potential drugs that selectively kill tumour cells or pathogens. Until now, although no helicase inhibitor derived from these efforts

has been used in the clinical treatment, several helicases (Table 1.2) have been clearly confirmed to be potent drug targets in many *in vitro* and *in vivo* studies (Aggarwal et al., 2011; Aiello et al., 2009; Chen et al., 2009; Pradhan and Tuteja, 2006). The potential of helicases as anti-cancer and anti-pathogenic drug targets is discussed in more detail below.

(a) Helicases as anti-cancer drug targets: the Achilles' heel of conventional anti-cancer therapies relying on ionizing radiation or DNA damaging chemotherapy drugs is the cytotoxic effects of the therapies on normal cells and the resistance of tumour cells to such treatments. Cancer cells tend to confer resistance to ionizing radiation or chemotherapeutic agents via up-regulation of DNA damage response and repair pathways (Aggarwal and Brosh, 2009; Gupta and Brosh, 2008). In addition, compared with normal cells, tumour cells generally have deficiencies in certain DNA damage repair pathways due to the mutagenic nature of tumours (Aggarwal and Brosh, 2009). Because helicases play critical roles in DNA damage repair pathways, inhibition of the helicases upregulated in tumour cells may prove to be an effective strategy for combating cancers. Among all possible target helicases, WRN helicase (a RecQ family helicase), which is involved in DNA damage response and repair, as well as telomere maintenance, is the most extensively studied (Aggarwal and Brosh, 2009; Gupta and Brosh, 2008; Turley et al., 2001). The up-regulation of WRN has been observed in many transformed cells (resembling *in vitro* tumour cells) (Kawabe et al., 2000; Turley et al., 2001) and the oncogene *c-myc* was found to promote its expression by stimulating transcription of the *WRN* gene (Grandori et al., 2004; Grandori et al., 2003). Knockdown of WRN expression by RNA interference caused slow proliferation and increased apoptosis in c-myc transformed lymphomas (Grandori et al., 2004), so WRN helicase has been thought to be a useful target for anti-cancer therapy. Recently, a small molecule compound (NSC19360, 1-(propoxymethyl)-maleimide) which specifically inhibits WRN activity was identified to act synergistically with the chemotherapy

drug (Topotecan) to suppress cell proliferation and induce DNA damage in HeLa cells (Aggarwal et al., 2011). Except WRN helicase, other helicases (i.e., BLM and PIF1) related to DNA repair and telomere metabolism are being exploited as anti-cancer drug targets (Gupta and Brosh, 2008).

(b) Helicases as anti-pathogenic drug targets: many pathogens encode their own helicases required for genome replication and their helicases are generally different from the host cell helicases, so the helicases of the pathogens are attractive targets for anti-pathogenic drug discovery. A number of laboratories have successfully identified helicase inhibitors (Table 1.2) for suppressing pathogen replication using high-throughput screening. For example, HSV (herpes simplex virus) helicase-primase inhibitors (e.g., ASP2151, an oxadiazolephenyl derivative; BAY 57-1293, an thiazole urea derivative; BILS 179 BS, an amino-thiazolylphenyl-containing compound) (Chono et al., 2010; Crute et al., 2002; Kleymann et al., 2002), HCV NS3 helicase inhibitors (e.g., halogenated benzimidazoles; acridone derivatives; QU663, a quinolone-based compound) (Borowski et al., 2003; Maga et al., 2005; Stankiewicz-Drogoń et al., 2010), SARS NSP13 helicase inhibitor (bananin, an adamantane derivative) (Tanner et al., 2005), *B. anthracis* DnaB and DnaC inhibitor (compound 2, a coumarin-type compound) (Aiello et al., 2009), *P. falciparum* PfDH60 helicase inhibitors (e.g., actinomycin, daunorubicin, ethidium bromide) (Pradhan and Tuteja, 2006). Except directly targeting helicases of pathogens, there is another strategy to target host helicases that are required for pathogen replication, but this strategy could cause greater toxicity to the host. An interesting example for using this strategy is to inhibit HIV replication by targeting human DDX3 RNA helicase. DDX3 was found to be essential for HIV mRNA export from nucleus to cytoplasm (Yedavalli et al., 2004) and knockdown of DDX3 expression did not affect the viability of normal cells (Ishaq et al., 2008), so it has become a potential target for HIV therapy. Recently, two ring expanded nucleoside analogues were identified to suppress HIV replication in T cells and macrophages and they did not display any

significant toxicity to cells and mice (Yedavalli et al., 2008). Furthermore, it is suggested that DDX3 may also be a good target for limiting HBV (Hepatitis B virus) and HCV infections because these two viruses encode proteins which interact with DDX3 and modulate its activity (Ariumi et al., 2007; Wang et al., 2009).

Due to the increasing knowledge about the structures of helicases and their inhibitors, some inhibitors were found to suppress helicase activities by occupying their NTP-binding and DNA/RNA-binding sites (Belon and Frick, 2009; Gupta and Brosh, 2008), such as halogenated benzimidazoles compete with ATP to bind in the nucleotide-binding site of NS3 helicase (Borowski et al., 2003). However, the mechanisms of inhibitors binding to unknown sites remain elusive (Belon and Frick, 2009; Gupta and Brosh, 2008). In addition to targeting helicases themselves, the recent advances in understanding the interactions of helicases with other partner proteins (primase, single-stranded DNA binding protein, nuclease, etc.) have encouraged more attempts to find peptides or small molecule compounds to interfere the helicase interactions for therapeutic use (Brosh et al., 2001; Chono et al., 2010; Doherty et al., 2005; Lehoux et al., 2012; Sharma et al., 2005). The continued work for screening potential helicase-targeted drugs and recent advanced understanding in this field will hopefully translate basic research findings of helicases into clinical applications.

Helicase	Organisms	Superfamily	(Proposed) Function	Potent small molecule inhibitor(s)	Reference
Anti-cancer					
WRN	Human	SF2	DNA repair Telomere maintenance	1-(propoxymethyl) -maleimide	(Aggarwal et al., 2011)
Anti-viral					
DDX3	Human	SF2	mRNA translation	ring expanded nucleoside analogues	(Yedavalli et al., 2008)
NS3	Hepatitis C virus	SF2	RNA genome replication and packaging	halogenated benzimidazoles, QU663, acridone derivatives	(Borowski et al., 2003; Maga et al., 2005; Stankiewicz-Drogoń et al., 2010)
UL5	Herpes simplex virus	SF1	DNA genome replication	ASP2151, BAY 57-1293, BILS 179 BS	(Chono et al., 2010; Crute et al., 2002; Kleymann et al., 2002)
NSP13	SARS	SF1	/	bananin	(Tanner et al., 2005)
Anti-bacterial					
DnaB	*B. anthracis*	SF4	DNA genome replication	compound 2, a coumarin-type compound	(Aiello et al., 2009)
DnaC	*S.aureus*	SF4	DNA genome replication	compound 2, a coumarin-type compound	(Aiello et al., 2009)

Anti-malarial					
PfDH60	*P. falciparum*	SF2	/	actinomycin, daunorubicin, ethidium bromide, netropsin, nogalamycin	(Pradhan and Tuteja, 2006)

Table 1.2 Examples of helicases as potential anti-cancer and anti-pathogenic drug targets. "/" represents unknown function.

1.5 Aims

Given the critical roles of helicases in many cellular processes and their huge potential in the treatment of many diseases (i.e., cancer, hepatitis, HIV), it is vital to understand the underlying mechanisms of helicase action. In recent years, many research studies on helicases from bacteriophage T4 and T7 have provided lots of useful details about helicase mechanisms. In contrast, there has been no progress on helicases from bacteriophage T5 since the first bioinformatic analysis on the putative helicase gene *D10* (Blinov et al., 1989). According to the T5 genome analysis, *D10* gene resides in a gene cluster which encodes proteins related to replication, recombination and repair of the phage genome (Sayers, 2005). Additionally, the bioinformatic analyses on this gene cluster identified two more potential helicase genes, *D2* and *D6*. The high efficiency of T5 viral replication suggests that its helicases may have developed an intricate unwinding mechanism. The aims of this study were to verify the helicase activities of these three proteins using several biochemical assays and to investigate how they act on their substrates. This study may explore some new T5 helicases, their roles in efficient T5 replication and some novel characteristics and underlying mechanisms of helicases.

Chapter 2. Materials and Methods

2.1 Cloning

The cloning of three bacteriophage T5 potential helicase genes (*D2*, *D6* and *D10*) were performed using standard procedures (including PCR, digestion, ligation and restriction screening) as described below.

2.1.1 Polyermase chain reaction (PCR)

The Bacteriophage T5 D2, D6 and D10 open reading frames (ORF, see GenBank accession numbers in Table 2.1) were amplified from T5 phage genomic DNA which was isolated as described (Sayers and Eckstein, 1990) by PCR with *Pfu* DNA polyermase (Promega) using their corresponding sense and antisense primers, shown in Table 2.1. The PCR reactions were performed with pre-denaturation at 95°C for 4 min, followed by 30 cycles of denaturation at 95°C for 50 s, annealing at 50°C for 50 s and extension at 72°C for 4 or 6 min (6 min for D2, 4 min for D6 and D10), then followed by a final extension at 72°C for 10 min and stored at 4°C.

The PCR products of D2 Walker A mutant (K405E, lysine to glutamic acid) and D10 arginine finger mutant (R389N, arginine to asparagine) were generated by overlap extension PCR (Ho et al., 1989) using the sense primer, anti-sense primer, and two internal primers carrying the desired point mutation (Table 2.1). All the PCR products were analysed on 0.8% w/v agarose gels in 1× TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA) and purified using QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's protocol.

Oligonucleotide	Primer	Sequence (5'-3')	Restriction enzyme sites or desired point mutation
D2 gene (GenBank accession no. NC_005859)			
P1	sense	ATGCGGTCTCTTATGGTGTTTTCTATCCTCCAAGG	*BsaI*
P2	antisense	TCGCGGATCCTCATGCATCTTCATTAGTTG	*BamHI*
P3	internal	GAAACTACGGCAGTAACTAAATGG	K405E
P4		ACTGCCGTAGTTTCACCAGTACCAAG	
D6 gene (GenBank accession no. AY692264.1)			
P5	sense	TTTTGGATCCCTCCTTCCCCTTTTTGCA	*BamHI*
P6	antisense	TTTTGTCGACTCATATATCGCTGGCCTG	*SalI*
D10 gene (GenBank accession no. YP_006952)			
P7	sense	AATTGAATTCTTAAGGTTGTTATATCTAATAAAG	*EcoRI*
P8	antisense	ATTTAAGCTTTTATGAGCTGTTGCCAAATGCA	*HindIII*
P9	internal	AATGTTCAACGTATTGTC	R389N
P10		ATACGTTGAACATTGCCTGCAAGCTGTTC	

Table 2.1 Primers used for DNA cloning. They introduce the restriction enzyme sites or the desired mutations (underlined) for ligation or site-directed mutagenesis. The start and stop codons are shaded in grey.

2.1.2 Digestion and ligation

About 2 μg of D2, D6 and D10 PCR products (with or without point mutation) were double digested with two restriction enzymes (D2: *BsaI* and *BamHI* in 1× NEBuffer 4; D6: *BamHI* and *SalI* in 1× NEBuffer 3; D10: *EcoRI* and *HindIII* in 1× MULTI-CORE™ buffer; all the enzymes and buffers are supplied by New England BioLabs or Promega) respectively at 37°C for 3 hours. After double digestion, the reaction products were analyzed on 0.8% w/v agarose gels in 1× TAE buffer and gel-purified by QIAquick Gel Extraction Kit (Qiagen) following the manufacturer's protocol.

The digested products (about 0.12 pmol) were ligated into about 0.04 pmol of pGEX-KG (Guan and Dixon, 1991) or pHisMAL (modified pET15b vector carrying an N-terminal His and MBP tag sequence, provided by Dr. Cyril Sanders) with 1 μl of T4 DNA ligase (400 units/μl, New England BioLabs) in 1× T4 ligase buffer at 25°C for 1 hour in order to construct the plasmids, pHisMAL-D2, pGEX-KG-D6 and pGEX-KG-D10 (Figure 2.1).

2.1.3 Restriction-enzyme screen

The ligation products were transformed into competent *E. coli* DH5α cells (Table 2.2), prepared by the $CaCl_2$ method described by Hanahan (Hanahan, 1983). After transformation, the positive clones were selected by restriction-enzyme screen. First, antibiotic-resistant colonies were individually picked to inoculate into 5 ml LB broth media with appropriate antibiotics and supplements, followed by incubation at 37°C overnight. Afterwards, plasmids were extracted from the cultures using QIAprep Spin Miniprep Kit (Qiagen), according to the manufacturers' instructions. Once isolated, each plasmid was digested, as described above, with restriction enzymes to verify the presence of desired DNA insert.

2.1.4 DNA sequencing

All the plasmids were sequenced to confirm sequence integrity by BigDye Terminator v3.1 Cycle Sequencing Kit and ABI PRISM™ 3730 DNA Analyzer (Applied Biosystems) at the Genetics Core Facility, The University of Sheffield.

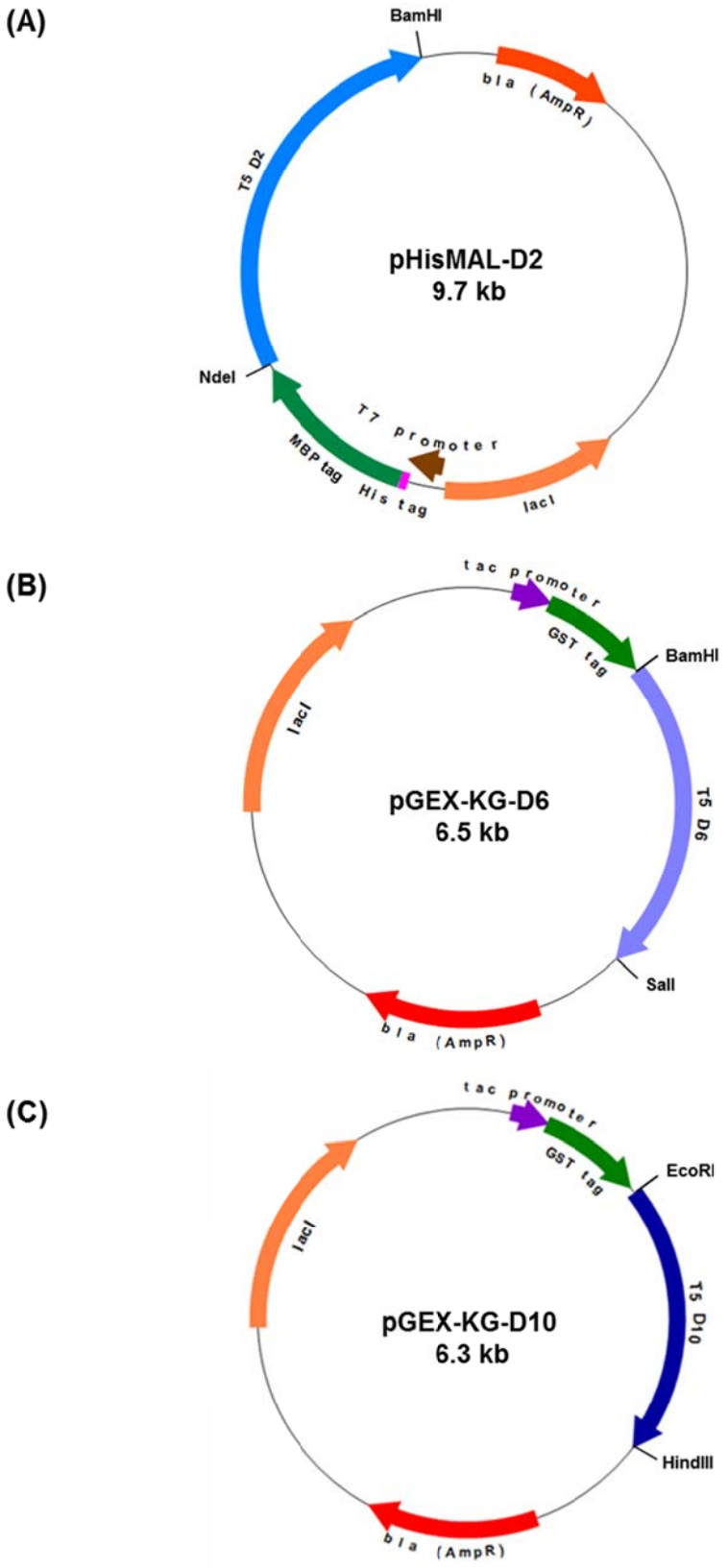


Figure 2.1 The plasmid maps of pHisMAL-D2 (**A**), pGEX-KG-D6 (**B**) and pGEX-KG-D10 (**C**).

2.2 Protein purification and analysis

The plasmids pHisMAL-D2, pGEX-KG-D6 and pGEX-KG-D10 (Figure 2.1) were transformed into the $CaCl_2$ competent *E. coli* XL1-blue or BL21 (DE3) cells to express His-MBP-tagged D2, GST-tagged D6 and GST-tagged D10 proteins (Table 2.2) for facilitating protein purification and analysis.

Strain or plasmid	Genotype or relevant characteristics	Source or Reference
***E. coli* strains**		
DH5α	F^-, φ80d*lacZ*ΔM15, Δ(*lacZYA-argF*)U169, *deoR*, *endA*1, *gyrA*96, *hsdR*17 (r_k^-, m_k^+), *phoA*, *recA*1, *relA*1, *supE*44, *thi*-1, λ^-	Invitrogen
BL21(DE3)	F^-, *dcm*, *gal*, *hsdS*$_B$ (r_B^- m_B^-), *ompT*, λ (DE3)	Novagen
XL1-Blue	*recA1*, *endA1*, *gyrA96*, *thi-1*, *hsdR17*, *supE44*, *relA1*, *lac* [F′ *proAB lacI*qZΔ*M15* Tn*10* (Tet^r)]	Stratagene
Plasmids		
pHisMAL-D2	pHisMAL vector containing T5 *D2* gene to express N-terminal His-MBP-tagged D2 protein, Ampicillin resistant	This study
pGEX-KG-D6	pGEX-KG vector containing T5 *D6* gene to express N-terminal GST-tagged D6 protein, Ampicillin resistant	This study
pGEX-KG-D10	pGEX-KG vector containing T5 *D10* gene to express N-terminal GST-tagged D10 protein, Ampicillin resistant	This study

Table 2.2 *E. coli* strains and vectors used in this study.

2.2.1 Protein purification

2.2.1.1 His-MBP-tagged D2

A day culture (5 ml) derived from a single colony of pHisMAL-D2 (WT or

K405E) transformed *E. coli* BL21(DE3) was used to inoculate 1 litre of 2×YT medium (20 g/L tryptone, 10 g/L yeast extract and 5 g/L NaCl, supplemented with 100 μg/ml ampicillin and 1% w/v glucose). The culture was grown at 37°C until an OD_{600} of 1.0 was reached and IPTG was added to a final concentration of 0.5 mM to induce His-MBP-tagged D2 (WT or K405E) protein expression at 25°C for 8 h. After expression, cells were harvested by centrifugation at 3000 × *g* for 20 min (Beckman Coulter J2-21, JA-14 rotor) and the cell pellets were washed with bacterial resuspension buffer (20 mM Tris-HCl pH 7.5, 0.2 M NaCl). The cells were then centrifuged at 3000 × *g* for 20 min, collected and stored at -80°C.

All the following purification steps for wild type and K405E proteins were identical and performed at 4°C. The frozen cell pellets were thawed on ice and resuspended in lysis buffer (25 mM Tris-HCl pH 7.5, 1 M NaCl, 1 mM EDTA, 1 mM PMSF, 5 mM DTT, 10% v/v glycerol). The bacterial cell wall was degraded by adding lysozyme (1 mg/ml, stirring at 4°C for 30 min). Then, the cells were lysed by sonication and cleared at 40000 × *g* for 30 min (Beckman Coulter J2-21, JA-20 rotor). Polyethylenimine (5% w/v, pH 8.0, Sigma-Aldrich) was added to the supernatant to a final concentration of 0.5% w/v. The mixture was centrifuged at 40000 × g for 5 min to remove precipitated nucleic acid. The supernatant was then incubated with amylose agarose beads (1 ml per 20 grams of *E. coli* cells, New England BioLabs) for over 12 h. The beads were washed in batch (50 bead volumes of lysis buffer), loaded onto a column, and washed with 50 bead volumes of lysis buffer with 0.5 M NaCl, no PMSF. The protein was eluted from the beads with elution buffer (25 mM Tris-HCl pH 7.5, 0.5 M NaCl, 1 mM DTT, 20 mM maltose, 10% v/v glycerol). Afterwards, the protein extract was 5× diluted with His buffer (25 mM Tris-HCl pH 7.5, 0.5 M NaCl, 10% v/v glycerol) and applied to HisTrap™ HP affinity column (high-performance nickel affinity column, GE Healthcare). Protein was eluted with a linear gradient of 0.01-0.25 M imidazole in His buffer. Finally, D2 was further purified by anion exchange

chromatography (Source Q, GE Healthcare, 25 mM Tris-HCl pH 8.5, 1 mM EDTA, 5 mM DTT, 10% v/v glycerol, 0.1-0.4 M NaCl gradient). Peak fractions were pooled, concentrated, aliquoted and stored at -80°C. The protein concentration of His-MBP-tagged D2 WT and K405E was determined by Bio-Rad assays (Bradford, 1976), according to the manufacturers' instructions, using BSA (bovine serum albumin) as a standard.

2.2.1.2 GST-tagged D6

The GST-tagged D6 was expressed in *E. coli* XL1-blue at 25°C for 6 h by addition of 0.5 mM IPTG when the OD_{600} of the cell cultures reached 1.0. The cells were harvested and frozen as described above. The cell pellets were lysed as described above. Polyethylenimine (5% w/v, pH 8.0) was added to the cleared lysate to a final concentration of 0.25% w/v. The solution was cleared at 25000 × *g* for 5 min and the protein was precipitated by addition of ammonium sulphate to 50% saturation (0.313 g of $(NH_4)_2SO_4$ per ml of supernatant) and centrifugation at 25000 × g for 30 min. The precipitate was dissolved in lysis buffer (1 ml per 10 grams of cells) and incubated with glutathione sepharose beads (GE Healthcare; 1 ml of beads per 20 grams of cells) for over 12 h. The beads were washed sequentially with 50 bead volumes of lysis buffer and lysis buffer with 0.2M NaCl and no PMSF. The GST-D6 protein was eluted with GST elution buffer (25 mM Tris-HCl pH 7.5, 0.2 M NaCl, 1 mM EDTA, 5 mM DTT, 20 mM reduced glutathione, 10% v/v glycerol). GST-D6 protein concentration was measured by Bradford assay as describe above.

2.2.1.3 D10

The GST-tagged D10 WT and R383N proteins were expressed in *E. coli* XL1-blue at 25°C for 8 hours by adding 0.5 mM IPTG after the cultures reached

an OD_{600} of 1.0. Cell pellets were harvested and stored at -80°C as described above and all the purification steps for wild type and R389N proteins were identical and performed at 4°C. Cell pellets were resuspended in lysis buffer (25 mM Tris-HCl pH 7.5, 1 M NaCl, 1 mM EDTA, 1 mM PMSF, 5 mM DTT, 10% v/v glycerol), followed by the addition of lysozyme, sonication and centrifugation (40000 × *g* for 30 min). Polyethylenimine was added to the supernatant to 0.65% w/v and the solution was cleared (25000 × *g* for 5 min). GST-D10 protein was precipitated by addition of ammonium sulphate to 40% saturation (0.243 g $(NH_4)_2SO_4$/ml) and centrifugation at 25000 × *g* for 30 min. The precipitate was dissolved in lysis buffer (1 ml per 10 grams of cells) and incubated with glutathione sepharose beads (1 ml of beads per 20 grams of cells) for ~18 h. Beads were washed sequentially with 50 bead volumes of lysis buffer, and 50 bead volumes of lysis buffer with 0.3 M NaCl, no PMSF. The GST-D10 protein was eluted with GST elution buffer (25 mM Tris-HCl pH 7.5, 0.3 M NaCl, 1 mM EDTA, 5 mM DTT, 20 mM reduced glutathione, 10% v/v glycerol) and digested with thrombin (~10 units per 10 grams of cells) for ~18 h to cleave off the GST fusion partner. After digestion, the protein extract was diluted with 1.5 volumes of S buffer (10 mM sodium phosphate pH 6.5, 1 mM EDTA, 5 mM DTT, 10% v/v glycerol) and applied to a 1 ml Source-S column (cation exchange column, GE Healthcare). Protein was eluted with a linear gradient of 0.2-0.7 M NaCl in S buffer. D10 was further purified by gel filtration chromatography (Superdex 75, GE Healthcare; 20 mM Tris-HCl pH 7.5, 0.3 M NaCl, 1 mM EDTA, 1 mM PMSF, 5 mM DTT, 10% v/v glycerol). Peak fractions were pooled, concentrated, aliquoted and stored at -80°C. D10 protein concentration was determined in the presence of 7 M guanidine hydrochloride by UV spectrometry using a molar extinction coefficient of 58330 M^{-1} cm^{-1}.

2.2.2 Protein analysis

2.2.2.1 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

The purity and molecular weight of each protein sample was determined by SDS-PAGE. A standard SDS-PAGE gel consisted of a 5% (29:1) acrylamide stacking gel (125 mM Tris-Cl pH 6.8, 0.1% w/v SDS) and a 10% or 12% (29:1) acrylamide resolving gel (375 mM Tris-Cl pH 8.8, 0.1% w/v SDS). The protein sample was mixed with 4× Laemmli sample buffer (250 mM Tris-Cl, pH 6.8, 8% w/v SDS, 10% v/v 2-mercaptoethanol, 0.1% w/v bromophenol blue, and 40% v/v glycerol), boiled at 95°C for 5 min and loaded onto the gel. The gel was run at a constant voltage of 215 V in 1× Tris-Glycine SDS running buffer (25 mM Tris, 0.1% w/v SDS and 192 mM glycine) until the bromophenol blue dye reached at the bottom of the gel. Then, the gel was transferred to Coomassie brilliant blue R-250 staining solution (40% v/v methanol, 10% v/v acetic acid, and 1.26 g/L Coomassie brilliant blue R-250) for 2 min and de-stained with destaining solution (9.4% v/v methanol and 9.4% v/v acetic acid) for 10 min.

2.2.2.2 ATP*ase* assays

ATP*ase* assays were performed in 20 mM HEPES-NaOH pH 7.5, 20 mM (for D2 reactions) or 100 mM NaCl (for D6 and D10 reactions), 0.1% v/v NP40 alternative (Calbiochem), 0.1 mg/ml BSA, 2 mM DTT, 0.0125 μM [γ-^{32}P]ATP (6000 Ci/ mmol), 5 mM $MgCl_2$, 5 mM ATP at 37°C for 15 min (D2) or at 22°C for 10 min (D6 and D10), unless stated otherwise. The release of radioactive phosphate was determined using the charcoal-binding assay of Iggo and Lane (Iggo and Lane, 1989).

2.2.2.3 Helicase and strand annealing assays

For each helicase substrate, one or multiple of the component oligonucleotides (5 μM, Sigma-Aldrich), as indicated, was 5' end-labelled with γ-^{32}P using T4 polynucleotide kinase (T4 pnk, New England BioLabs) and [γ-^{32}P]ATP (7.5 μM, 6000 Ci/ mmol) in 1× T4 pnk buffer at 37°C for 1 h. The kinase was inactivated at 95°C for 5 min. To generate the test DNA substrate, illustrated in Appendix A, the labelled oligonucleotide and the appropriate complementary oligonucleotides (labelled or not) were annealed by heating them to 100°C and allowing them to cool slowly in the annealing buffer (1 mM Tris-Cl 8.0, 100 mM NaCl, 0.1 mM EDTA). The end-labelled substrates were resolved and gel purified from 8% (19:1) polyacrylamide gels (1× TBE buffer; 89 mM Tris, 89 mM boric acid, 2 mM EDTA). The labelled substrates and their labelled component oligonucleotides (50× dilution of the original labelling reactions, containing the final concentration of 100 nM labelled oligonucleotides as standards) were spotted onto a Whatman DE-80 ion-exchange paper. The concentration of the labelled substrates was determined on the basis of the specific activity of the labelled component oligonucleotide standards.

Helicase or strand annealing reactions (0.1 nM substrates) were performed in 20 mM HEPES-NaOH pH 7.5, various [NaCl] as indicated in the legend of the figure, 0.1% v/v NP40 alternative, 0.1 mg/ml BSA, 2 mM DTT, 5 mM $MgCl_2$, 5 mM ATP at 37°C for 15 min (for D2 reactions) or 20 min (for D10 reactions) and stopped by the addition of 6× termination buffer (120 mM EDTA, 0.6% w/v SDS, 1% w/v bromophenol blue, 60% v/v glycerol, with 2 mg/ml proteinase K for D2 reactions or without proteinase K for D10 reactions). Reaction products were resolved on 8% (19:1) polyacrylamide gels containing 0.05% w/v SDS (1× TBE/0.05% w/v SDS running buffer), visualized and quantified by phosphorimaging.

2.2.2.4 Gel-shift assays

The substrates for DNA binding reactions were end-labelled with γ-^{32}P and purified as described above. The reactions (0.05 nM substrates for D2 reactions or 0.1 nM substrates for D10 reactions) were performed in 20 mM HEPES-NaOH pH 7.5, 100 mM (D2) or 135 mM NaCl (D10), 0.1% v/v NP40 alternative, 0.1 mg/ml BSA, 2 mM DTT, 1 mM EDTA, 10% v/v glycerol at 22°C for 15 min (D2) or 20 min (D10) in the absence of 5 mM ATP/Mg^{2+}, as indicated in the legend. The binding reactions were separated on 5% (80:1) or 6% (30:1) polyacrylamide gels in 0.25× TBE buffer, visualized and quantified as above.

2.2.2.5 Hydroxyl radical footprinting

Hydroxyl radical footprinting of D10 in solution was carried out on the basis of the general guidelines of Dixon, *et al*. (Dixon et al., 1991). The binding reactions (50 µl, containing 10 nM of Y20 substrate with one of its three strands ^{32}P end-labelled) were set up as described for the gel-shift assays, except that glycerol was omitted. Following a 20-minute incubation at 22°C, the hydroxyl radical was generated by addition of 0.375 mM sodium ascorbate, 0.0275% w/v H_2O_2, 1 mM $(NH_4)_2FeSO_4 \cdot 6H_2O$ and 2 mM EDTA for 2 min and quenched with 0.25 volume of 200 mM thiourea. Cleaved products were extracted twice with phenol/chloroform and analyzed on 15% (19:1) polyacrylamide gels containing 8M Urea in 1× TBE buffer.

2.2.2.6 DEPC interference

The Y20 substrates for DEPC interference were prepared by 5' end-labelling one strand with ^{32}P and modifying it with diethylpyrocarbonate (DEPC, Appendix B). About 60 ng of end-labelled DNA was carbethoxylated by adding 4

µl of DEPC (diethyl pyrocarbonate, 99%, Sigma-Aldrich) in 200 µl of cacodylate buffer (50 mM sodium cacodylate pH 7.0, 1 mM EDTA). After incubation at 37°C for 20 min, the DNA was ethanol precipitated twice and annealed with its complementary oligonucleotides, before gel-purification and quantification as described above. For helicase reactions, 5 nM substrate and 0.005 nM D10 were used. After electrophoresis, the desired helicase reaction products were detected by autoradiography and the appropriate bands excised from the gel. The DNA was soak-eluted in TE buffer (1 mM Tris-HCl pH 8.0, 100 mM NaCl, 0.1 mM EDTA) at 4°C overnight and recovered by phenol/chloroform extraction, ethanol precipitation and CTAB (cetyl trimethyl ammonium bromide) precipitation. The DNA was then cleaved with piperidine, extracted with butanol and ethanol precipitated. Finally, the products were analyzed on urea-polyacrylamide sequencing gels, as described above.

Chapter 3. Bacteriophage T5 D2 helicase

3.1 Results

3.1.1 Bioinformatics

For this study, the protein sequence of the D2 protein (accession no. AAS77154) was analysed for putative helicase motifs, domains and structures using a number of web-based bioinformatic tools.

3.1.1.1 The analysis of amino acid sequence

The D2 protein is predicted to be a 105-kDa polypeptide with an isoelectric point (pI) of 6.18 using ProtParam tool (http://web.expasy.org/protparam/). Analysis with InterProScan (http://npsa-pbil.ibcp.fr/iprscan/iprscan) identified that D2 may possess one conserved domain (PF02399, Pfam database; at amino acids 392-532), which is also found in herpesvirus origin-binding proteins and probably involved in DNA origin-dependent replication. In addition, PROSCAN (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_proscan.html) found one significant classical ATP-binding protein motif: Walker A motif (motif I, identity) at amino acids 399-406 within the putative conserved domain (Figure 3.1), suggesting that D2 may have the ability to bind and hydrolyze ATP for its potential replication activity. In addition, BLASTP analysis for the D2 sequence revealed there are several weak hits to some DNA replication origin-binding helicases, including human herpesvirus UL9.

3.1.1.2 Secondary structural analysis of the D2 protein

The secondary structure of D2 was predicted using GOR IV method (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html), showing it contains 34 alpha helices, 28 extended strands and 61 random coils

(Figure 3.1). Furthermore, helix-turn-helix (HTH) prediction software (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_hth.html) detected that there is a HTH DNA-binding motif located at amino acids 791-812 (Figure 3.1), postulating D2 is a DNA-binding protein.

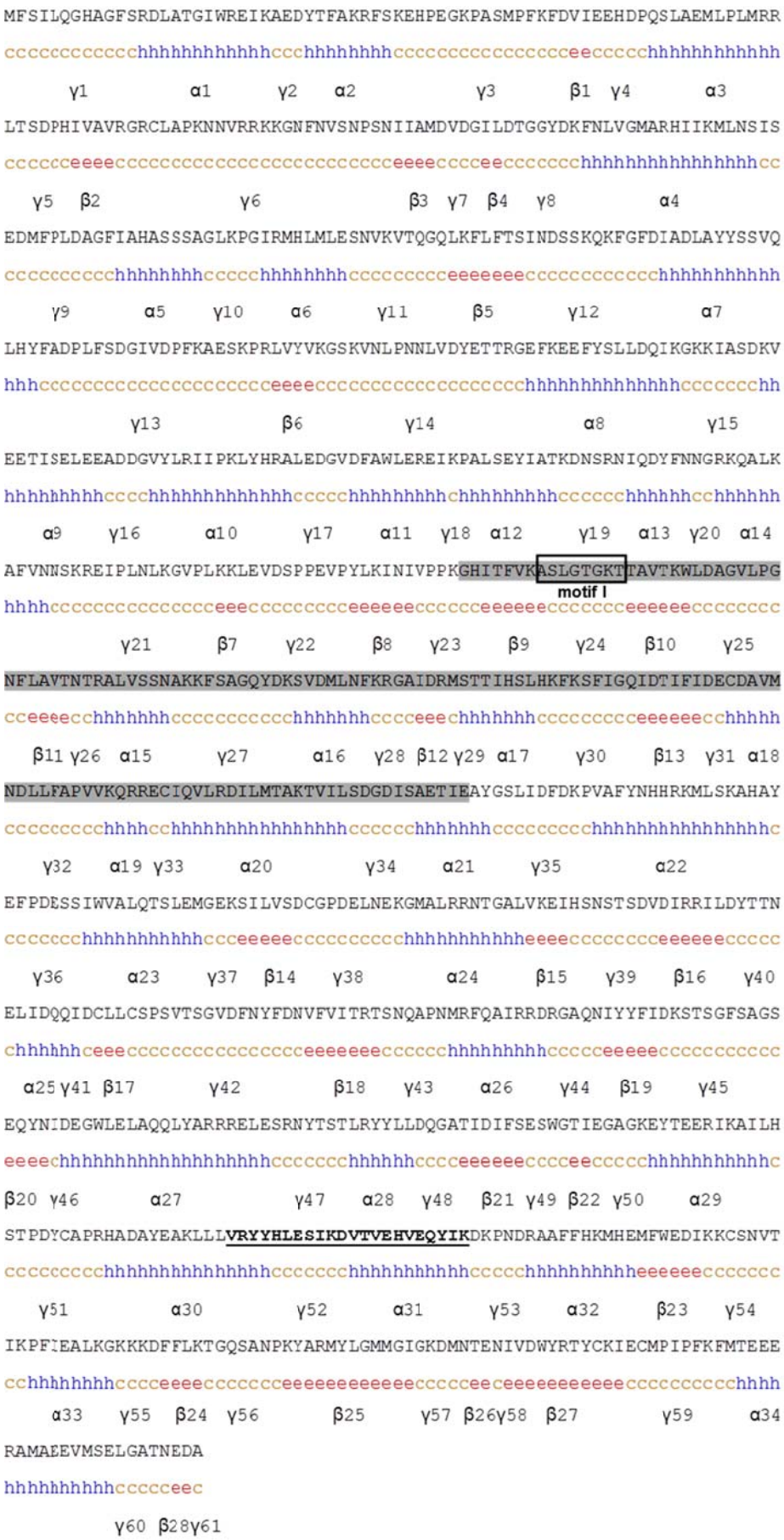

MFSILQGHAGFSRDLATGIWREIKAEDYTFAKRFSKEHPEGKPASMPFKFDVIEEHDPQSLAEMLPLMRR
cccccccccccchhhhhhhhhhhhcccchhhhhhhhccccccccccccccccceeccccchhhhhhhhhhhh
γ1 α1 γ2 α2 γ3 β1 γ4 α3
LTSDPHIVAVRGRCLAPKNNVRRKKGNFNVSNPSNIIAMDVDGILDTGGYDKFNLVGMARHIIKMLNSIS
ccccceeeeccccccccccccccccccccccccceeeeccccceeccccccchhhhhhhhhhhhhhhhcc
γ5 β2 γ6 β3 γ7 β4 γ8 α4
EDMFPLDAGFIAHASSSAGLKPGIRMHLMLESNVKVTQGQLKFLFTSINDSSKQKFGFDIADLAYYSSVQ
ccccccccccchhhhhhhhccccchhhhhhhhccccccccceeeeeeeccccccccccchhhhhhhhhhh
γ9 α5 γ10 α6 γ11 β5 γ12 α7
LHYFADPLFSDGIVDPFKAESKPRLVYVKGSKVNLPNNLVDYETTRGEFKEEFYSLLDQIKGKKIASDKV
hhhccccccccccccccccccccceeeeccccccccccccccccchhhhhhhhhhhhhhccccccchh
γ13 β6 γ14 α8 γ15
EETISELEEADDGVYLRIIPKLYHRALEDGVDFAWLEREIKPALSEYIATKDNSRNIQDYFNNGRKQALK
hhhhhhhhhccccchhhhhhhhhhhhhcccccchhhhhhhhhchhhhhhhhhccccccchhhhhhccchhhhhh
α9 γ16 α10 γ17 α11 γ18 α12 γ19 α13 γ20 α14
AFVNNSKREIPLNLKGVPLKKLEVDSPPEVPYLKINIVPPKGHITFVKASLGTGKTTAVTKWLDAGVLPG
motif I
hhhhccccccccccccccceeeccccccccceeeeeeccccccceeeeeeccccccceeeeeeccccccccc
γ21 β7 γ22 β8 γ23 β9 γ24 β10 γ25
NFLAVTNTRALVSSNAKKFSAGQYDKSVDMLNFKRGAIDRMSTTIHSLHKFKSFIGQIDTIFIDECDAVM
cceeeeccchhhhhhhccccccccccchhhhhhhccccceeechhhhhhhccccccccceeeeeeccchhhhh
β11 γ26 α15 γ27 α16 γ28 β12 γ29 α17 γ30 β13 γ31 α18
NDLLFAPVVKQRRECIQVLRDILMTAKTVILSDGDISAETIEAYGSLIDFDKPVAFYNHHRKMLSKAHAY
ccccccccchhhhccchhhhhhhhhhhhhhhhcccccchhhhhhhccccccccchhhhhhhhhhhhhhhhc
γ32 α19 γ33 α20 γ34 α21 γ35 α22
EFPDESSIWVALQTSLEMGEKSILVSDCGPDELNEKGMALRRNTGALVKEIHSNSTSDVDIRRILDYTTN
ccccccchhhhhhhhhhhccceeeeeccccccccccchhhhhhhhhhheeeeccccccccceeeeeeccccc
γ36 α23 γ37 β14 γ38 α24 β15 γ39 β16 γ40
ELIDQQIDCLLCSPSVTSGVDFNYFDNVFVITRTSNQAPNMRFQAIRRDRGAQNIYYFIDKSTSGFSAGS
chhhhhhceeeccccccccccccccceeeeeeeccccccchhhhhhhhhccccceeeeeccccccccccc
α25 γ41 β17 γ42 β18 γ43 α26 γ44 β19 γ45
EQYNIDEGWLELAQQLYARRRELESRNYTSTLRYYLLDQGATIDIFSESWGTIEGAGKEYTEERIKAILH
eeeechhhhhhhhhhhhhhhhhhcccccccchhhhhcccceeeeeeccccceeccccchhhhhhhhhhhc
β20 γ46 α27 γ47 α28 γ48 β21 γ49 β22 γ50 α29
STPDYCAPRHADAYEAKLLLVRYYHLESIKDVTVEHVEQYIKDKPNDRAAFFHKMHEMFWEDIKKCSNVT
cccccccccchhhhhhhhhhhhhhhcccccccchhhhhhhhhhhcccccchhhhhhhhhheeeeeeccccccc
γ51 α30 γ52 α31 γ53 α32 β23 γ54
IKPFIEALKGKKKDFFLKTGQSANPKYARMYLGMMGIGKDMNTENIVDWYRTYCKIECMPIPFKFMTEEE
cchhhhhhhhcccceeeeccccccceeeeeeeeeeeeccccceeceeeeeeeeeeccccccccccchhhh
α33 γ55 β24 γ56 β25 γ57 β26 γ58 β27 γ59 α34
RAMAEEVMSELGATNEDA
hhhhhhhhhhccccceec
γ60 β28 γ61

Figure 3.1 The predicted secondary structure of D2. The amino acid sequence of D2 is shown in one-letter-code format. The secondary structure of each amino acid is indicated by h (alpha helix, blue), e (extended strand, red) and c (random coil, orange). The secondary structure elements are listed in order, such as α1 (the 1st alpha helix), β5 (the 5th extended strand), γ6 (the 6th random coil). The conserved domain probably involved in DNA replication is shaded in grey. The putative Walker A motif (motif I) is boxed and the helix-turn-helix DNA-binding motif is highlighted in underlined bold script.

3.1.2 Expression, purification and characterization of D2

The D2 ORF was cloned into pHisMAL (Figure 2.1) to express the recombinant His-MBP-tagged D2 in *E. coli* BL21(DE3). As mentioned above, the bioinformatic analysis showed that one classical ATP-binding motif (Walker A, which is also one of three universal motifs of helicases) was identified in D2 (Figure 3.1). Therefore, we generated a D2 Walker A mutant construct (K405E, replacing the conserved lysine with glutamic acid), which is expected to lose ATP*ase* activity, for use as a control in comparison to wild-type protein.

Figure 3.2A shows an SDS-PAGE analysis of the protein samples of His-MBP-tagged D2 WT and K405E at each purification step from cell lysate of induced *E. coli* (S, supernatant from bacterial lysate after sonication; M, eluate from amylose column; H, eluate from HisTrap™ HP column; Q, eluate from Source Q column). The molecular weight of the major band of His-MBP-tagged D2 WT and K405E protein samples from the final Source Q column is approximately 150 kDa, which is nearly equivalent to that predicted for His-MBP-tagged D2 (Figure 3.2, lanes 4 and 8). The purity of the protein samples at each step was estimated by densitometric analysis of the SDS-PAGE gel in Figure 3.2A. His-MBP-tagged D2 WT and K405E were finally obtained to at least 98% pure (Table 3.1). For simplicity, His-MBP-tagged D2 is referred to as D2 below.

Furthermore, the ATP*ase* activity of the D2 WT and K405E protein samples from each purification step and His-MBP tag control (fixed total protein amount,

74.5 ng) was tested in the presence of ssDNA substrate (T55, 55-base poly T oligonucleotide). For D2 WT protein samples, the protein samples contained ATP*ase* activity and the enzymatic activity increased (Figure 3.2) as D2 WT became more pure after each purification step (the purity and specific activity of D2 after each step are shown in Table 3.1). For D2 K405E protein samples, the ATP*ase* activity gradually decreased as the purification of D2 K405E progressed and the final purified D2 K405E protein sample lacked any significant ATP*ase* activity (Figure 3.2 and Table 3.1). Excluding the possibility that the ATP*ase* activity is stimulated by the His-MBP tag partner (HM, Figure 3.2 and Table 3.1), all these data suggested the majority of observed ATP*ase* activity is contributed by D2 WT protein and a mutation (K405E) almost completely abolished the ATP*ase* activity as expected. However, the mutation did not disrupt D2 ssDNA-binding activity assessed by gel-shift assays with a ^{32}P-labelled T25 (without ATP/Mg^{2+}), and the substrate was bound by WT (lanes 1-4) and K405E (lanes 5-8) to form one protein-DNA complex (Figure 3.3A).

In addition, it was shown that the stimulation of D2 ATP*ase* activity occurred in a DNA-dependent (Figure 3.3B) and DNA-structure-dependent manner (Figure 3.3C). ATP*ase* assays were performed in the presence of various linear DNA substrates, including ssDNA (T12, 12-base dT), blunt-ended dsDNA (ds20, 20 bp), and partially double-stranded DNA molecules with either a 5' or 3' ssDNA overhang (5'T5-ds20 and ds20-3'T5, 20 bp and 5-base dT). The ATP*ase* activity of D2 could be stimulated by all the test substrates and there is a correlation between the extent of the ATP*ase* activity and the structure of the substrate (D2 ATP*ase* stimulation: T12 < ds20 < 5'T5-ds20 < ds20-3'T5, Figure 3.3C).

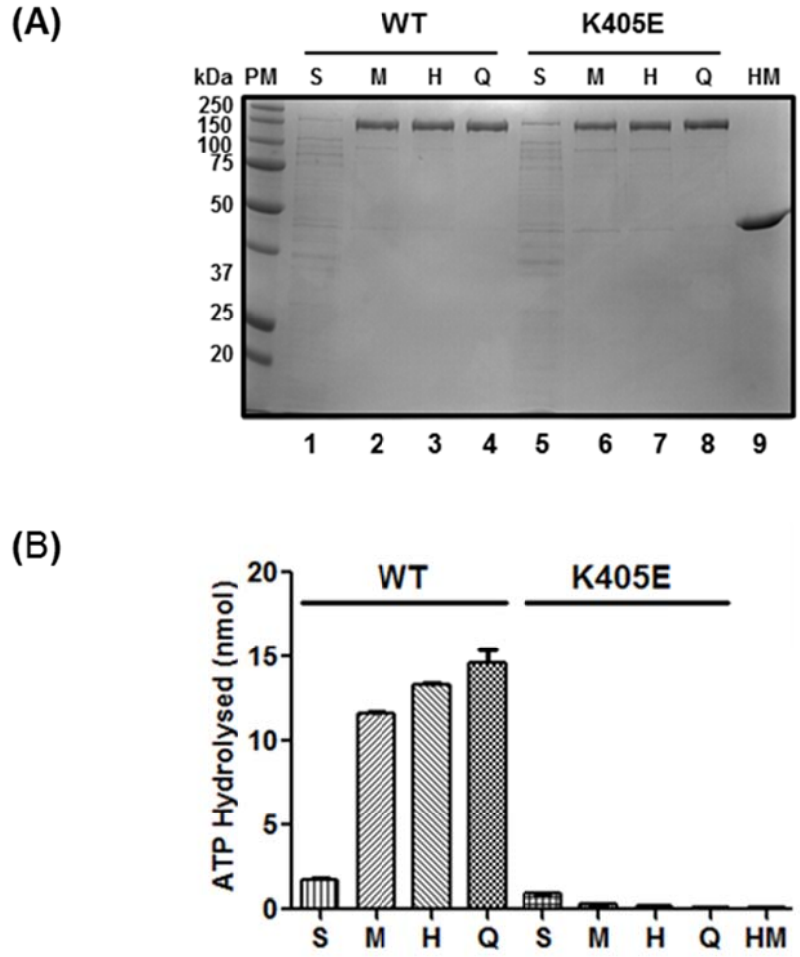


Figure 3.2 Purification and ATP*ase* activity of bacteriophage T5 D2. (**A**) 12% SDS-PAGE analysis of the protein samples of His-MBP-D2 wild-type (WT, lanes 1-4) and Walker A mutant (K405E, lanes 5-8) at each purification step — supernatant from bacterial lysate after sonication (S), eluate from amylose column (M), eluate from HisTrap^{TM} HP column (H), eluate from Source Q column (Q); His-MBP tag (HM) for use as control; 1 µg total protein. The molecular weight of the purified proteins corresponded to that predicted for His-MBP-D2 (~150 kDa). (**B**) The ATP*ase* activity of His-MBP-D2 WT and K405E protein samples from each purification step and His-MBP tag (74.5 ng total protein) was determined in the presence of T55, a single-stranded 55-base oligo-dT substrate. For D2 WT protein samples, the ATP*ase* activity increased as D2 WT became more pure after each purification step. For D2 K405E protein samples, the ATP*ase* activity decreased as the purification of D2 K405E progressed, indicating the ATP*ase* activity was almost completely abolished by the point mutation K405E in the putative Walker A motif. His-MBP tag had no ATP*ase* activity detected. The experiment was repeated three times (n=3, mean and SEM).

Sample	Total protein (ng)	Purity (%)	Specific activity ($\mu mol \cdot min^{-1} \cdot mg^{-1}$)
D2 (WT)			
S	74.5	~20%	1.56
M		~96%	10.36
H		~96%	11.91
Q		~99%	13.07
D2 (K405E)			
S	74.5	~11%	0.79
M		~86%	0.24
H		~93%	0.16
Q		~98%	0.11
His-MBP tag	74.5	~99%	0.08

Table 3.1 A summary of purification of T5 D2 WT and K405E from cell lysate of induced *E. coli*. The indication for each purification step (S, M, H, Q) is described as above. The purity of D2 protein was estimated by densitometry of the SDS-PAGE gel in Figure 3.2**A** using GelQuant.NET software provided by biochemlabsolutions.com. The specific activity gives a measurement of the ATP*ase* activity of D2 per milligram of total protein (expressed in $\mu mol \cdot min^{-1} \cdot mg^{-1}$). The specific activity is calculated on the basis of the data from Figure 3.2**B**.

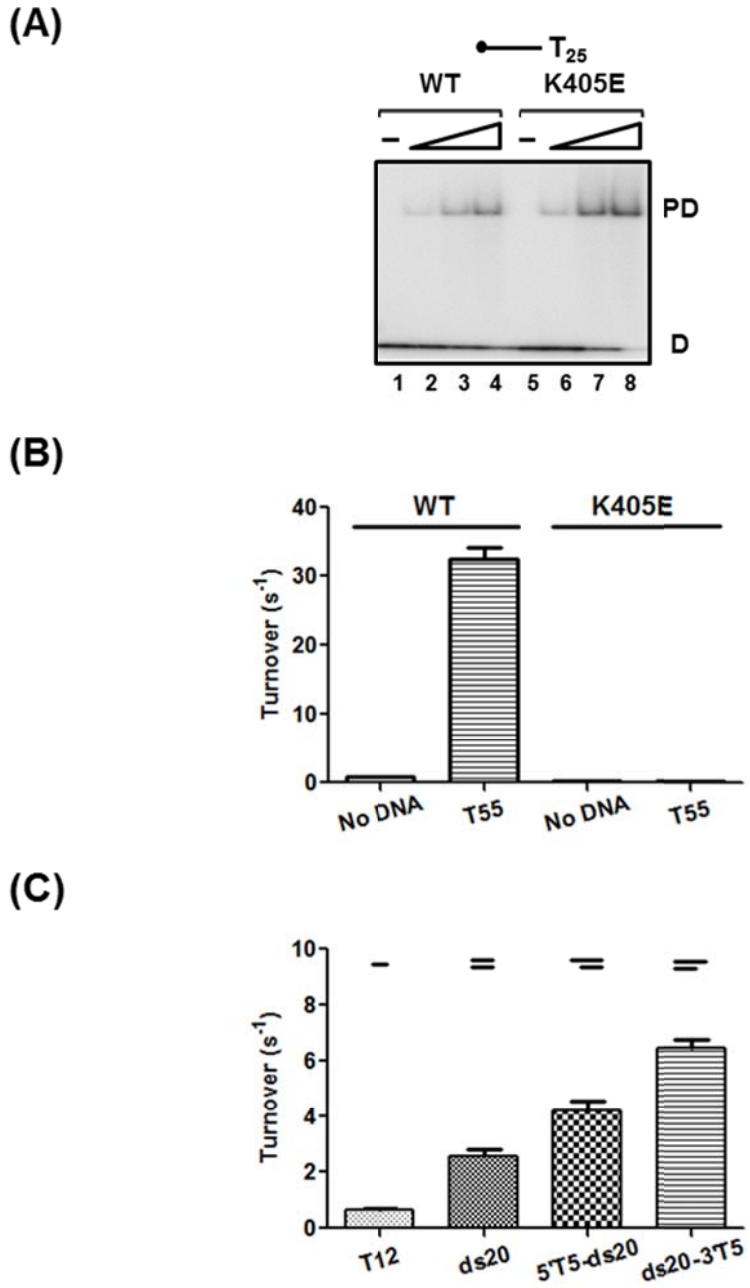


Figure 3.3 ssDNA-binding and ATP*ase* activities of T5 D2. (**A**) Binding of D2 WT and K405E to ssDNA (T25). Oligonucleotide T25 was ^{32}P-end-labelled at its 5' end as indicated by a black circle, which is also used for representing the labelling of the substrates in all subsequent figures. The binding reactions (0.05 nM substrate; 0.01, 0.05 and 0.1 nM D2 WT and K405E) were incubated at 22°C for 15 min, resolved on a 5% (80:1) acrylamide gel and visualized by phosphorimaging. Lanes 1 and 5, no protein control; lanes 2-4, D2 WT (0.01, 0.05 and 0.1 nM); lanes 6-8, D2 K405E (0.01, 0.05 and 0.1 nM); PD, protein-DNA complex; D, free DNA probe. (**B**) ATP*ase* activities of D2 WT and K405E (25 nM) were determined in the absence or presence of T55 (25 nM). (**C**) ATP*ase* activity of D2 WT (25 nM) was measured in the presence of various DNA substrates (25 nM) – ssDNA, T12; blunt-ended dsDNA, ds20; partial duplexes with a 20 bp and a 5' or 3' 5-base oligo-dT overhang, 5'T5-ds20 and ds20-3'T5.

3.1.3 D2 is a bipolar helicase

Given that the ATP*ase* activity of D2 was stimulated by blunt-ended dsDNA and partial duplex DNA with a 5' or 3' ssDNA overhang, the ability of D2 to unwind such substrates was next studied using helicase assays. D2 did not significantly unwind blunt-ended dsDNA substrates (ds20 and ds40, Figure 3.4A, lanes 1-10), but it could unwind dsDNA with a 5' or 3' ssDNA overhang (substrates 5'T55-ds20 and ds20-3'T55, 20 bp with a 5' or 3' 55-base oligo-dT overhang, lanes 11-20), suggesting the ssDNA overhang is necessary for DNA unwinding by D2. In general, DNA helicases show specific directionality (in a 5'→3' or 3'→5' direction) for unwinding the duplex substrates with either a 5' or 3' ssDNA overhang. However, our data revealed D2 is capable of unwinding both 5' and 3' oligo-dT overhang substrates, implying that D2 is a bipolar helicase. Moreover, the 3'→5' unwinding activity was greater than the 5'→3' unwinding activity (>3 folds, at the intermediate concentration of 1 nM protein, lane 18 compared to lane 13 and the graph on the right).

To complement the helicase assays, the binding of D2 to the helicase substrates was also examined using gel-shift assays in the absence of ATP/Mg^{2+}. No significant binding of D2 to ds20 and ds40 was observed (Figure 3.4B, lanes 5-12). However, D2 could bind to 5'T25-ds20 and ds20-3'T25 to form a protein-DNA complex (lanes 13-20) and it showed a higher binding affinity for ds20-3'T25 in comparison with 5'T25-ds20 (>8 folds, at 0.05 nM D2, lane 19 compared to lane 15).

The experimental binding and helicase results suggested that D2 displayed stronger helicase activity with the substrate for which it has higher affinity. Next, all activities of D2 K405E were compared with those of wild type D2 and the results indicated that the mutant totally lost bipolar helicase and ATP*ase* activities, but retained full DNA-binding activity (Figure 3.5), confirming D2 is

responsible for the observed unwinding and ATP*ase* activities rather than other contaminant proteins.

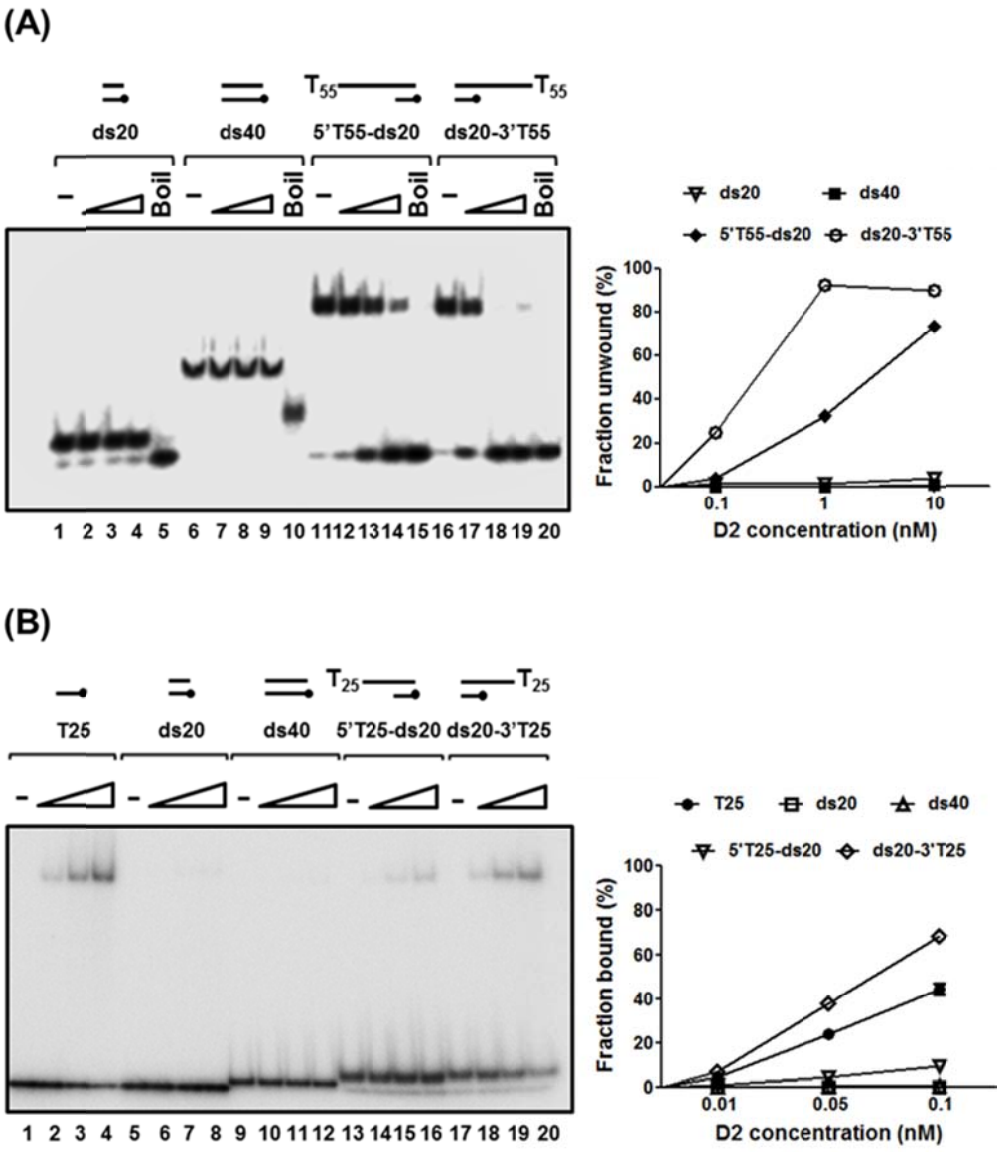


Figure 3.4 D2 binding and unwinding activities on various DNA substrates, including ssDNA, T25; blunt-ended dsDNA, ds20 and ds40; partial duplexes with a 5' or 3' dT overhang, 5'T25-ds20, 5'T55-ds20, ds20-3'T25 and ds20-3'T55. (n=3) (**A**) The helicase reactions (0.1 nM substrate; 0.1, 1, 10 nM D2) were incubated at 37°C for 15 min in the reaction buffer with 20 mM NaCl. The reaction products were resolved on 8% (19:1) acrylamide gels, visualized and quantified by phosphoimaging. The quantitative data (n=3, mean and SEM) for the extents of unwinding are shown in the graph on the right. Lanes 1, 6, 11 and 16, no protein control (-); lanes 5, 10, 15 and 20, heat-denatured substrate control (Boil); lanes 2-4, Lanes 7-9, lanes 12-14 and lanes 17-19, D2 protein titrations (0.1, 1, 10 nM) and indicated substrates. (**B**) The binding reactions (0.05 nM substrate; 0.01, 0.05 and 0.1 nM D2) were assessed by gel-shift assay in the absence of ATP/Mg^{2+}. Lanes 1, 5, 9, 13 and 17, no protein control (-); lanes 2-4, lanes 6-8, lanes 10-12, lanes 14-16 and lanes 18-20, with D2 protein titrations (0.01, 0.05 and 0.1 nM) and indicated substrates. Statistical data (fraction bound) are shown in the graph on the right. The gels presented in this figure are the representative gels of triplicate experiments, which applies to all subsequent figures unless otherwise stated.

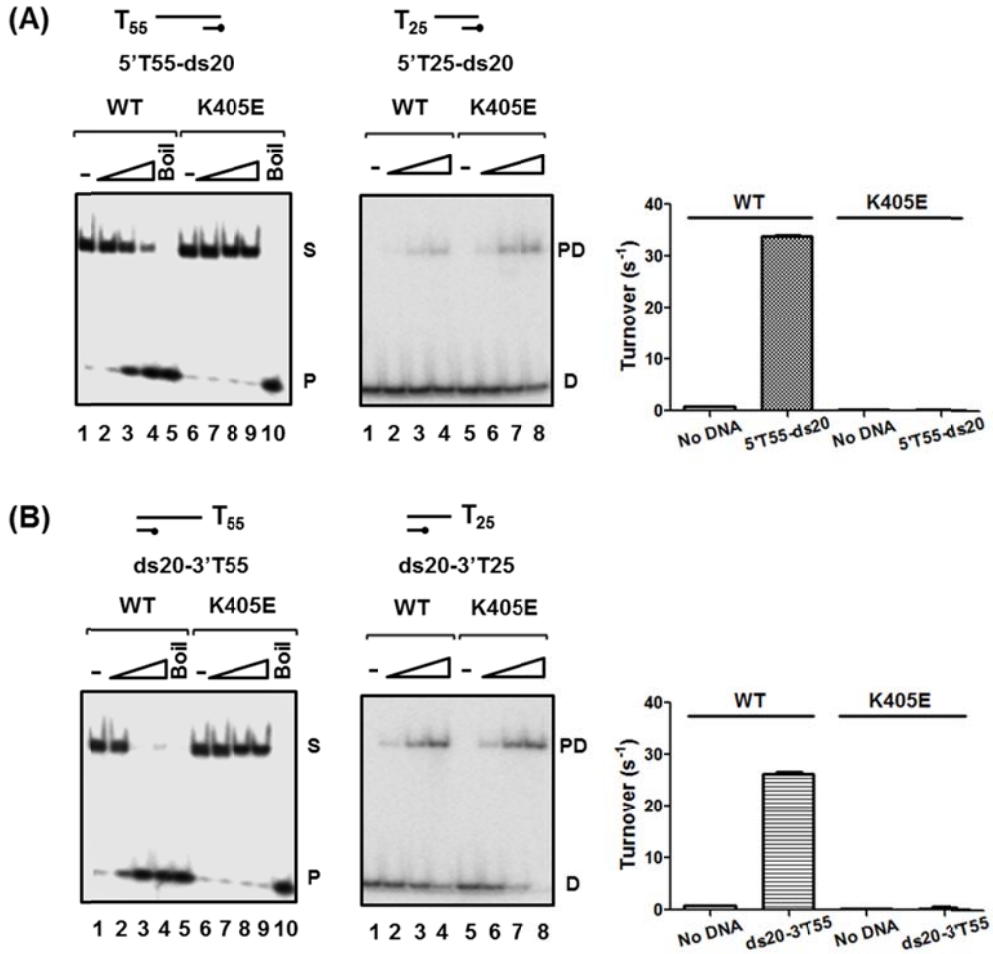


Figure 3.5 Helicase, DNA-binding and ATP*ase* activities of D2 WT and K405E on partial duplex DNA with a 20 bp and a 5' dT overhang (**A**) or 3' dT overhang (**B**). Helicase assays (0.1 nM 5'T55-ds20 or ds20-3'T55; 20 mM NaCl). Lanes 1 and 6, no protein control (-); lanes 5 and 10, heat-denatured substrate control (Boil); lanes 2-4, D2 WT (0.1, 1, 10 nM); lanes 7-9, D2 K405E (0.1, 1, 10 nM); S, native substrate; P, ssDNA product as indicated by boiling the substrate. Gel-shift assays (0.05 nM 5'T25-ds20 or ds20-3'T25), Lanes 1 and 5, no protein control (-); lanes 2-4, D2 WT (0.01, 0.05 and 0.1 nM); lanes 6-8, D2 K405E (0.01, 0.05 and 0.1 nM); PD, protein-DNA complex; D, free DNA probe. ATP*ase* activities of D2 WT and K405E (25 nM) were determined in the absence or presence of the DNA substrates (25 nM). In comparison with D2 WT, the mutant (K405E) retained DNA-binding activity, but totally lost helicase and ATP*ase* activities. (n=3)

3.1.4 Binding of D2 to ssDNA

Next, the binding of D2 to ssDNA was analysed using gel-shift assays (no ATP/Mg^{2+}) with radiolabeled poly T oligonucleotides of increasing length (from 12 to 55 bases; T12, T25, T35, T45, T55; Figure 3.6A). D2 protein at 0.01-0.1 nM was unable to bind oligo-dT substrate of 12 bases (lanes 1-4) or less (not shown). Furthermore, the detectable D2-DNA complexes were observed with oligo-dT substrate of 25 bases or more and the ssDNA-binding affinity of D2 increased when the length of the substrate was extended (lanes 5-20 and graphed data). One stable complex (C1) formed on the substrates T25, T35, T45, T55 and a slower migrating complex (C2) started to appear at high D2 concentrations (0.05 and 0.1 nM) with substrates T45 (lanes 15-16) and T55 (lanes 19-20), suggesting complex 2 is likely to involve multimers of D2. Because the stoichiometry for D2 ssDNA binding has not been determined, the composition of each D2-DNA complex is currently unknown.

Also, D2 ATP*ase* activity stimulated by these ssDNA substrates was assessed using ATP*ase* assays. Very little stimulation of D2 ATP*ase* activity occurred in the presence of the substrate T12. There was a gradual increase in the stimulation of D2 ATP*ase* activity by the substrates of increasing length (T25, T35, T45) and the extents of the activity in the presence of T45 or T55 reached a stable plateau (Figure 3.6B).

Based on these results, the binding of D2 to ssDNA was shown to be length-dependent and the similar dependence of D2 ATP*ase* activity on ssDNA length was also observed (Figure 3.6). In addition, 25 bases may be the minimal substrate length required for D2 binding to ssDNA. This is the basis for the substrates chosen for the binding analysis above.

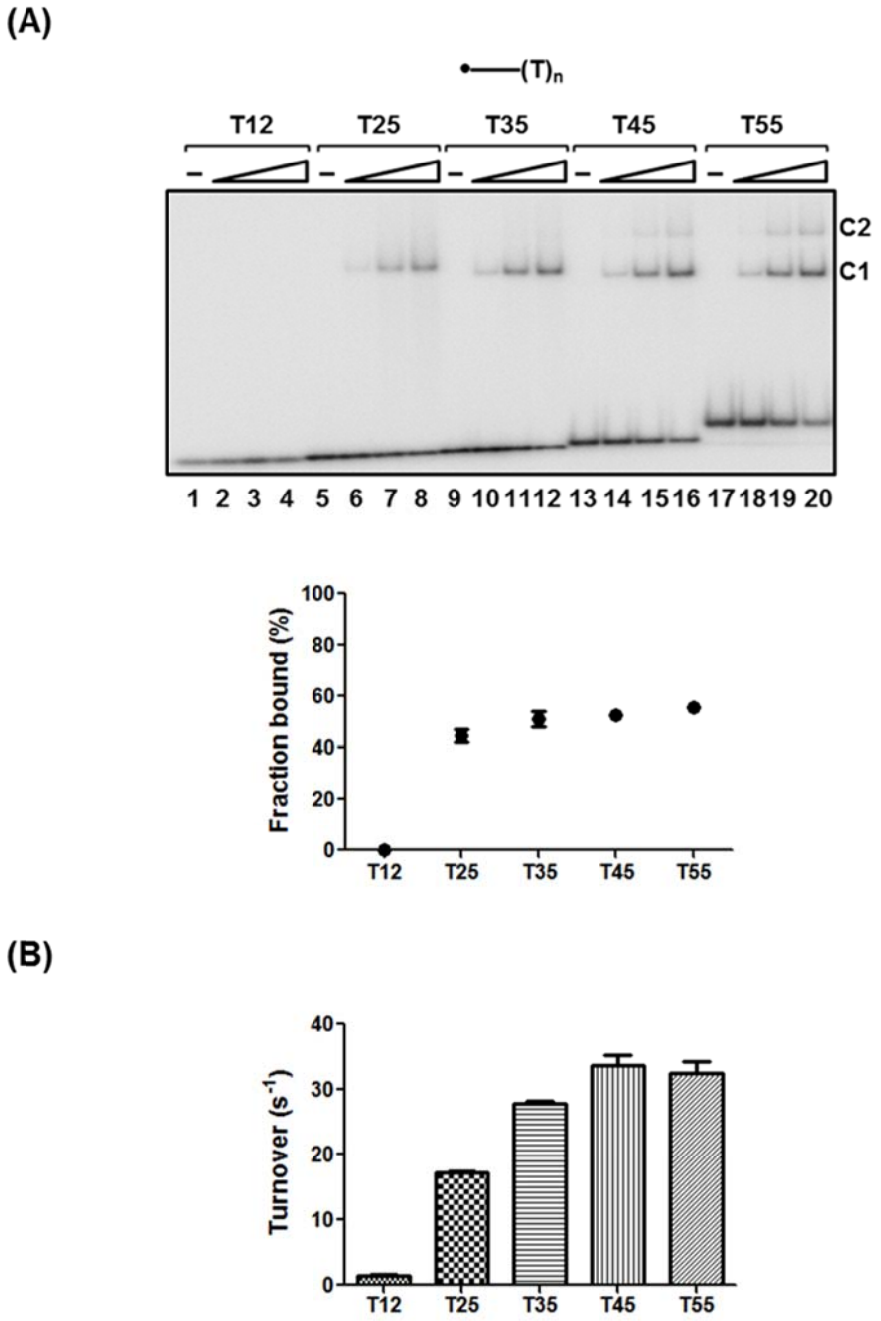


Figure 3.6 D2 interactions with ssDNA. (**A**) The oligo-dT substrates of increasing length (from 12 to 55 bases, 0.05 nM) were radiolabelled at their 5' ends. Reactions (0.01, 0.05, 0.1 nM D2) were incubated at 22°C for 15 min and assessed by gel-shift assay in the absence of ATP/Mg^{2+}. Lanes 1, 5, 9, 13 and 17, no protein control (-); lanes 2-4, lanes 6-8, lanes 10-12, lanes 14-16 and lanes 18-20, D2 protein titrations (0.01, 0.05, 0.1 nM). Significant binding was observed with probes of 25 bases or longer (lanes 5-20) and two protein-DNA complexes (C1 and C2) formed on probes of 45 bases or longer. The bottom graph showed that the binding affinity of D2 for ssDNA increases while the length of probes is extended from 12 to 55 bases (data for 0.1 nM D2, n=3, mean and SEM). **(B)** Stimulation of D2 ATP*ase* activity by oligo-dT substrates of increasing length. D2 (25 nM) was incubated with T12, T25, T35, T45 and T55 (25 nM) at 37°C for 15 min and the activity correlated with the length of the substrates. (n=3)

3.1.5 ssDNA tail length dependence of D2 unwinding

As described above, D2 unwound partial duplex DNA with a 5' or 3' ssDNA overhang (tail) and the ssDNA tail is essential for initiation of D2 helicase activity. Next, the dependence of D2 helicase activity on the length of 5' or 3' single-stranded tail was investigated using substrates with a 20-bp of dsDNA and a dT tail of increasing length (from 0 to 55 bases). Blunt-ended dsDNA (ds20, without tail) was not significantly unwound (Figure 3.7A and B, lanes 1-5). The 3' tail dependence data in Figure 3.7B (3' tail substrates) indicated that D2 requires a 5-base 3' tail for efficient DNA unwinding (~76% of ds20-3'T5 unwound, with the highest D2 concentration, 10 nM, the graphed data). When the 3' tail length was extended from 5 to 55 bases (10-base increments), the extent of unwinding activity increased slightly and reached a plateau at the 3' tail of 15 bases (~85% of ds20-3'T15 unwound with 10 nM of D2).

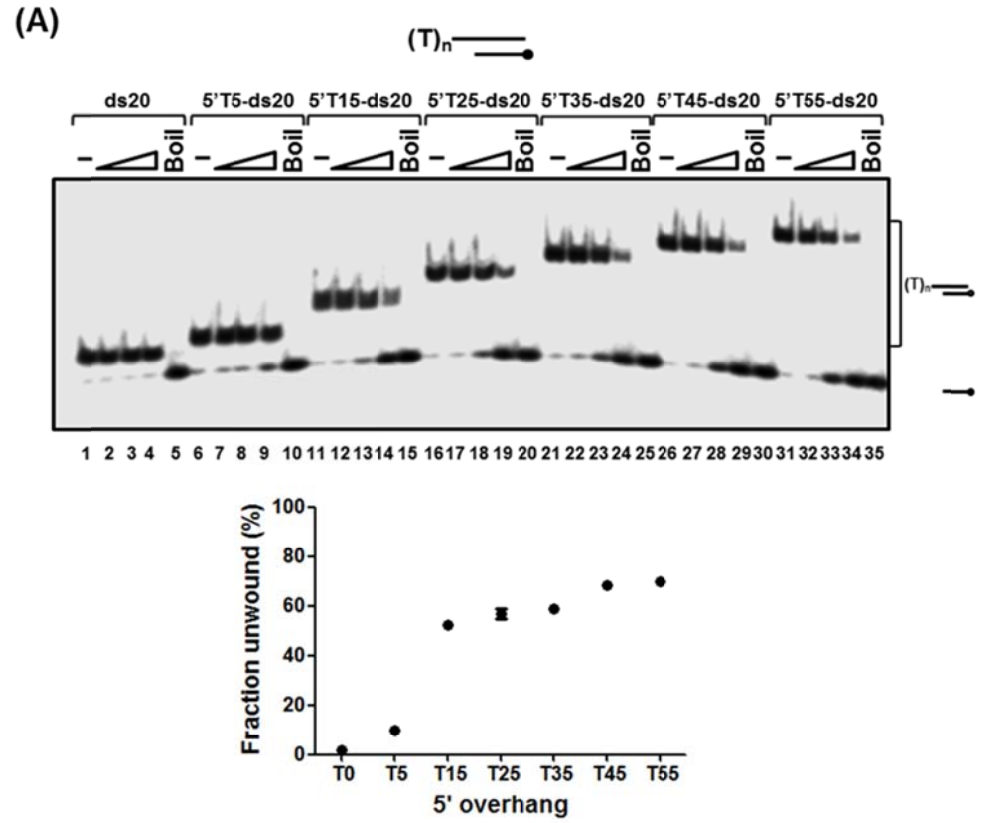
The 5' tail dependence data in Figure 3.7A (5' tail substrates) showed that 5' T5-ds20 with a 5-base 5' tail, was a poorer substrate for D2 (~10% of 5'T5-ds20 unwound with 10 nM of D2, the statistical graph) compared to ds20-3'T5 (described above). The extent of D2 unwinding activity increased sharply as the 5' tail length was extended from 5 to 15 bases (~50% of 5'T15-ds20 unwound) and rose slightly thereafter when the 5' tail length was increased to 35 bases. A further increase in the efficiency of initiation of unwinding was observed when the 5' tail length was increased to 45 bases (~70% of 5'T45-ds20 unwound), but no further increase in the extent of unwinding was seen when the 5' tail length was extended up to 55 bases. In summary, compared to the 5' tail substrates, substrates with shorter 3' tails were better substrates for the initiation of D2 unwinding.

To complement the helicase assays, the binding of D2 to the tail substrates was also analyzed using the gel-shift assays (without ATP/Mg^{2+}). The binding data in Figure 3.8B (3' tail substrates) indicated that D2 did not bind to the

substrate with a 3' tail of 5 bases or less (ds20 and ds20-3'T5, lanes 1-8 and the statistical graph). The extent of binding increased sharply as the 3' tail was extended from 15 to 35 bases (from ~20% of ds20-3'T15 bound to ~75% of ds20-3'T35 bound, 0.1 nM D2) and one predominant protein-DNA complex (C1) was observed (lanes 9-20). Further increase in the tail length resulted in a second complex (C2) formed on ds20-3'T45 and ds20-3'T55 at higher D2 concentrations.

The binding data in Figure 3.8A (5' tail substrates) showed there was no detectable D2 binding to the substrate with a 5' tail of 15 bases or less (ds20, 5'T5-ds20 and 5'T15-ds20, lanes 1-12 and the graphed data). The substrates with a 5' tail from 25 to 55 bases (5'T25 to T45-ds20) were bound by D2 with gradually increasing affinity (from ~10% of 5'T25-ds20 bound to ~42% of 5'T55-ds20 bound, 0.1 nM D2) and there was one main complex observed (C1). A second complex (C2) formed on 5'T55-ds20 at higher D2 concentrations. All the binding data indicated that the binding affinity of D2 for the tail substrates and its ability to assemble higher-order complexes on the substrates are tail-length-dependent. 3' tail substrates bind D2 with higher affinity than 5' tail substrates when the oligo dT stretch is short.

Furthermore, the stimulation of D2 ATP*ase* activity by the tail substrate is also in a tail-length-dependent fashion. With the concentration of D2 fixed at 25 nM, the ATP*ase* activity was measured in the presence of the tail substrate with increasing tail length. The ATP*ase* activity progressively increased when the tail length was extended (5' tail length from 0 to 55 bases, Figure 3.9A; 3' tail length from 0 to 55 bases, Figure 3.9B). Compared with 5' tail substrates, shorter tail length is required for 3' tail substrates to stimulate D2 ATP*ase* activity to similar extents (such as 5'T25-ds20 vs ds20-3'T15). This observation is consistent with results from the above helicase and gel-shift analyses.


(A)
(T)n
ds20
5'T5-ds20
5'T15-ds20
5'T25-ds20
5'T35-ds20
5'T45-ds20
5'T55-ds20
–
Boil
(T)n
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35
Fraction unwound (%)
100
80
60
40
20
0
T0 T5 T15 T25 T35 T45 T55
5' overhang

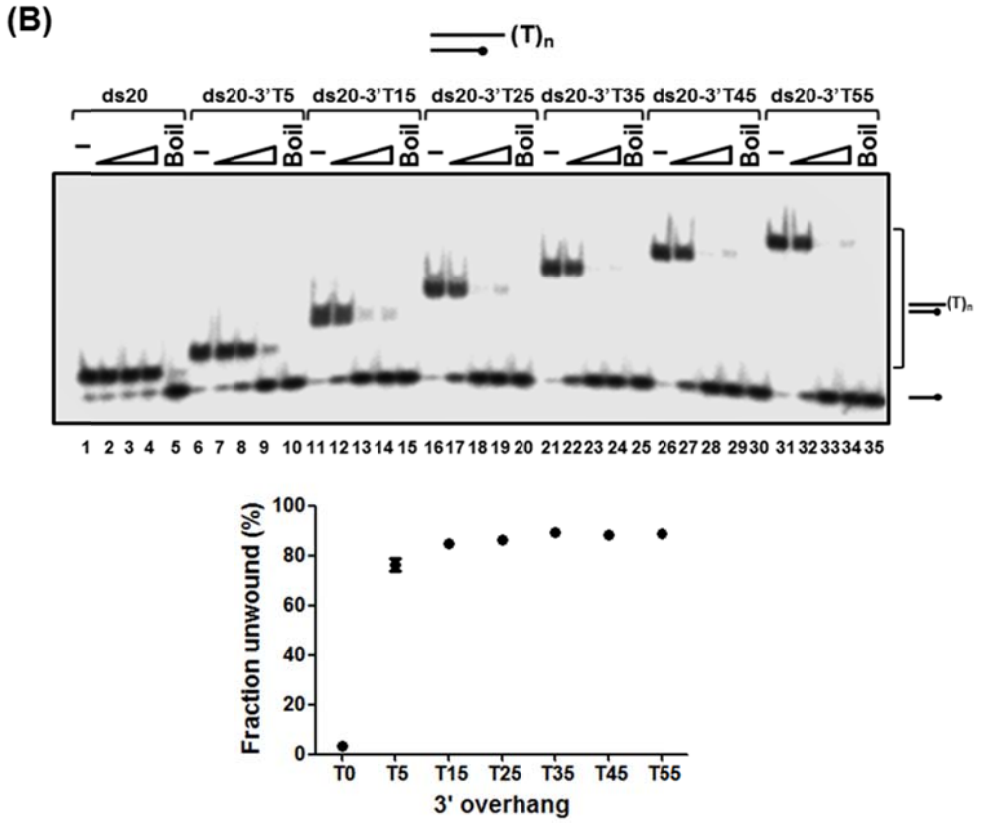
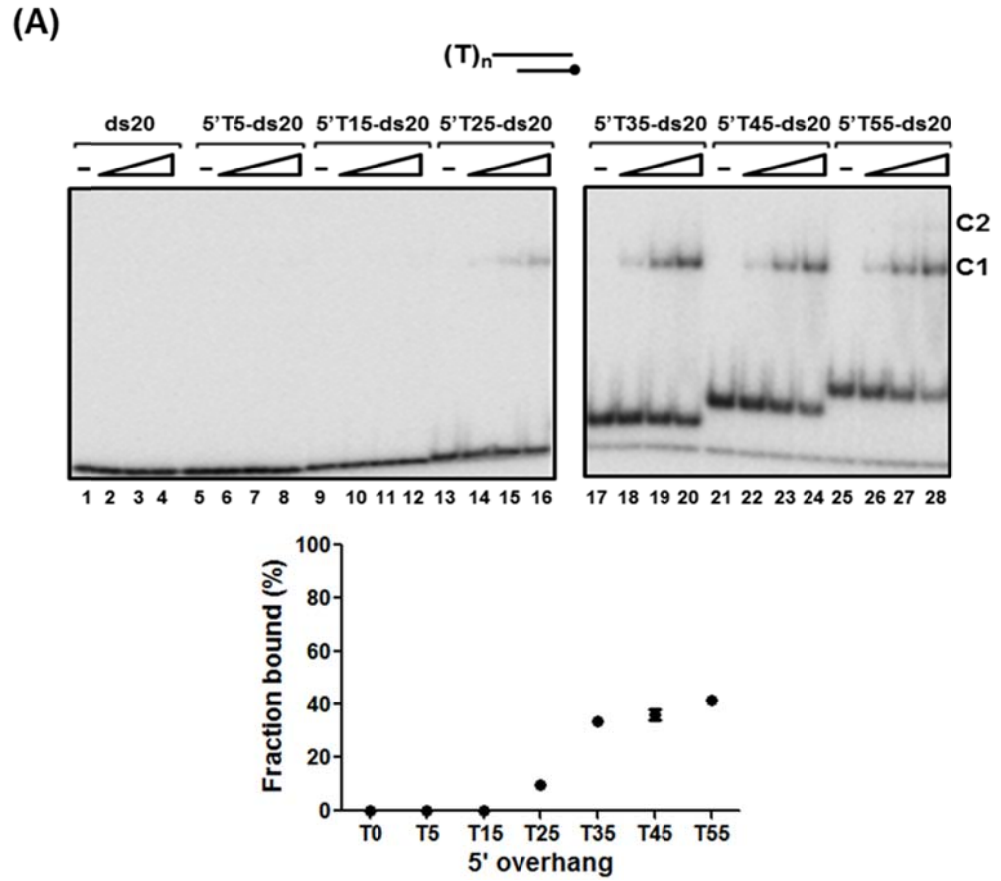

(B)
(T)n
ds20
ds20-3'T5
ds20-3'T15
ds20-3'T25
ds20-3'T35
ds20-3'T45
ds20-3'T55
–
Boil
(T)n
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35
Fraction unwound (%)
100
80
60
40
20
0
T0 T5 T15 T25 T35 T45 T55
3' overhang

Figure 3.7 Helicase activity of D2 on partial duplex substrates with a 5' (**A**) or 3' (**B**) ssDNA tail of increasing length from 0 to 55 bases. The ^{32}P-end-labelled substrates were constructed to contain a 20-bp duplex and a 5' or 3' oligo-dT tail of increasing length (from 0 to 55 bases). Helicase reactions (0.1 nM substrate; 0.1, 1, 10 nM D2; 20 mM NaCl) were incubated at 37°C for 15 min and products were resolved on 8% (19:1) polyacrylamide gels, visualized and quantified by phosphorimaging. Lanes 1, 6, 11, 16, 21, 26 and 31, no protein control (-); lanes 5, 10, 15, 20, 25, 30 and 35, heat-denatured substrate control (Boil); lanes 2-4, lanes 7-9, lanes 12-14, lanes 17-19, lanes 22-24, lanes 27-29 and lanes 32-34, D2 protein titrations (0.1, 1, 10 nM) and indicated substrates. The graphs at the bottom of the acrylamide gels showed statistical data for the extents of unwinding (data for 10 nM D2, three repeats, mean and SEM).


(A)
(T)n
ds20
5'T5-ds20
5'T15-ds20
5'T25-ds20
5'T35-ds20
5'T45-ds20
5'T55-ds20
C2
C1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
17 18 19 20 21 22 23 24 25 26 27 28
Fraction bound (%)
100
80
60
40
20
0
T0 T5 T15 T25 T35 T45 T55
5' overhang

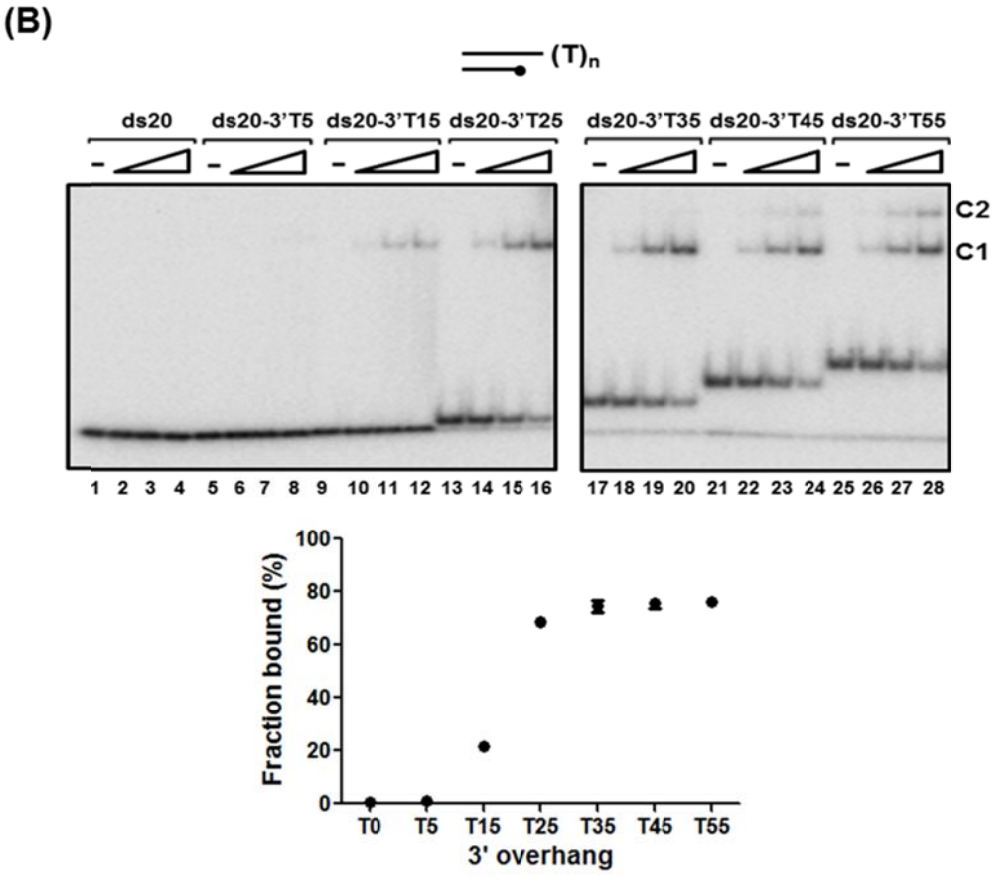

(B)
(T)n
ds20
ds20-3'T5
ds20-3'T15
ds20-3'T25
ds20-3'T35
ds20-3'T45
ds20-3'T55
C2
C1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
17 18 19 20 21 22 23 24 25 26 27 28
Fraction bound (%)
100
80
60
40
20
0
T0 T5 T15 T25 T35 T45 T55
3' overhang

Figure 3.8 Binding of D2 to partial duplex substrates with a 5' (**A**) or 3' (**B**) ssDNA tail of increasing length from 0 to 55 bases. Binding reactions (0.05 nM probe; 0.01, 0.05, 0.1 nM D2) were performed at 22°C for 15 min in the absence of ATP/Mg^{2+} and separated on 5% (80:1) native polyacrylamide gels. Lanes 1, 5, 9, 13, 17, 21 and 25, no protein control (-); lanes 2-4, lanes 6-8, lanes 10-12, lanes 14-16, lanes 18-20, lanes 22-24 and lanes 26-28, D2 protein titrations (0.01, 0.05, 0.1 nM). With 5' tail substrates, the stable protein-DNA complex started to appear when the 5' tail length increased to 25 bases or more and two complexes (C1 and C2) were detected with 5'T55-ds20. With 3' tail substrates, the protein-DNA complex started to form on the substrates with a 3' tail of 15 bases or more and two complexes (C1 and C2) were observed with ds20-3'T45 and ds20-3'T55. Two bottom graphs showed the statistical data (0.1 nM D2, n=3) for binding of D2 to such substrates.

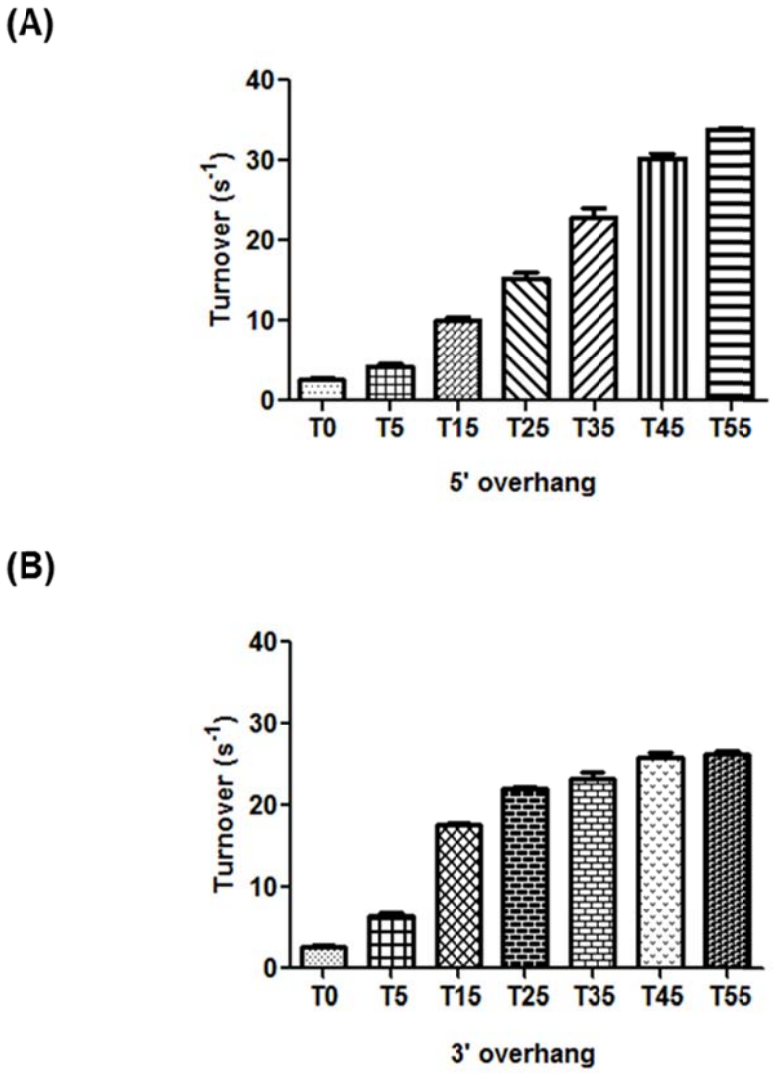


Figure 3.9 Stimulation of D2 ATP*ase* activity by partial duplex substrates with a 5' (**A**) or 3' (**B**) ssDNA tail of increasing length from 0 to 55 bases. D2 (25 nM) was incubated with partial duplexes with a 5' dT tail (25 nM; ds20, 5'T5-ds20, 5'T15-ds20, 5'T25-ds20, 5'T35-ds20, 5'T45-ds20 and 5'T55-ds20) or partial duplexes with a 3' dT tail (25 nM; ds20, ds20-3'T5, ds20-3'T15, ds20-3'T25, ds20-3'T35, ds20-3'T45 and ds20-3'T55) at 37°C for 15 min and the activity correlated with the tail length of the substrates. (n=3)

3.1.6 The effect of NaCl concentration on D2 helicase activity

Generally, the unwinding activity of many helicases is significantly affected by NaCl concentration, so the condition of NaCl concentration was routinely optimized for helicase reactions. In this study, the optimal NaCl concentration for 5'→3' or 3'→5' unwinding activity of D2 was determined using helicase assays with substrate 5'T55-ds20 or ds20-3'T55 in the buffer with increasing NaCl concentration from 20 mM to 140 mM. When the NaCl concentration was increased, there was a significant decrease in the 5'→3' unwinding activity (Figure 3.10A), but there was only minor decrease in the 3'→5' unwinding activity (Figure 3.10B). In order to simplify these results, Figure 3.10C only showed D2 bidirectional unwinding activities in the buffer with the extreme NaCl concentration (20 mM and 140 mM). It clearly revealed the differential inhibition of the bipolar helicase activities by the high NaCl concentration. The extent of 5'→3' unwinding activity considerably dropped (from ~73.4% to ~6.8% of 5'T55-ds20 unwound by 10 nM D2, 20 and 140 mM NaCl respectively), but the extent of 3'→5' unwinding activity was appreciably inhibited (from ~89.6% to 89.4% of ds20-3'T55 unwound by 10 nM D2, 20 and 140 mM NaCl respectively).

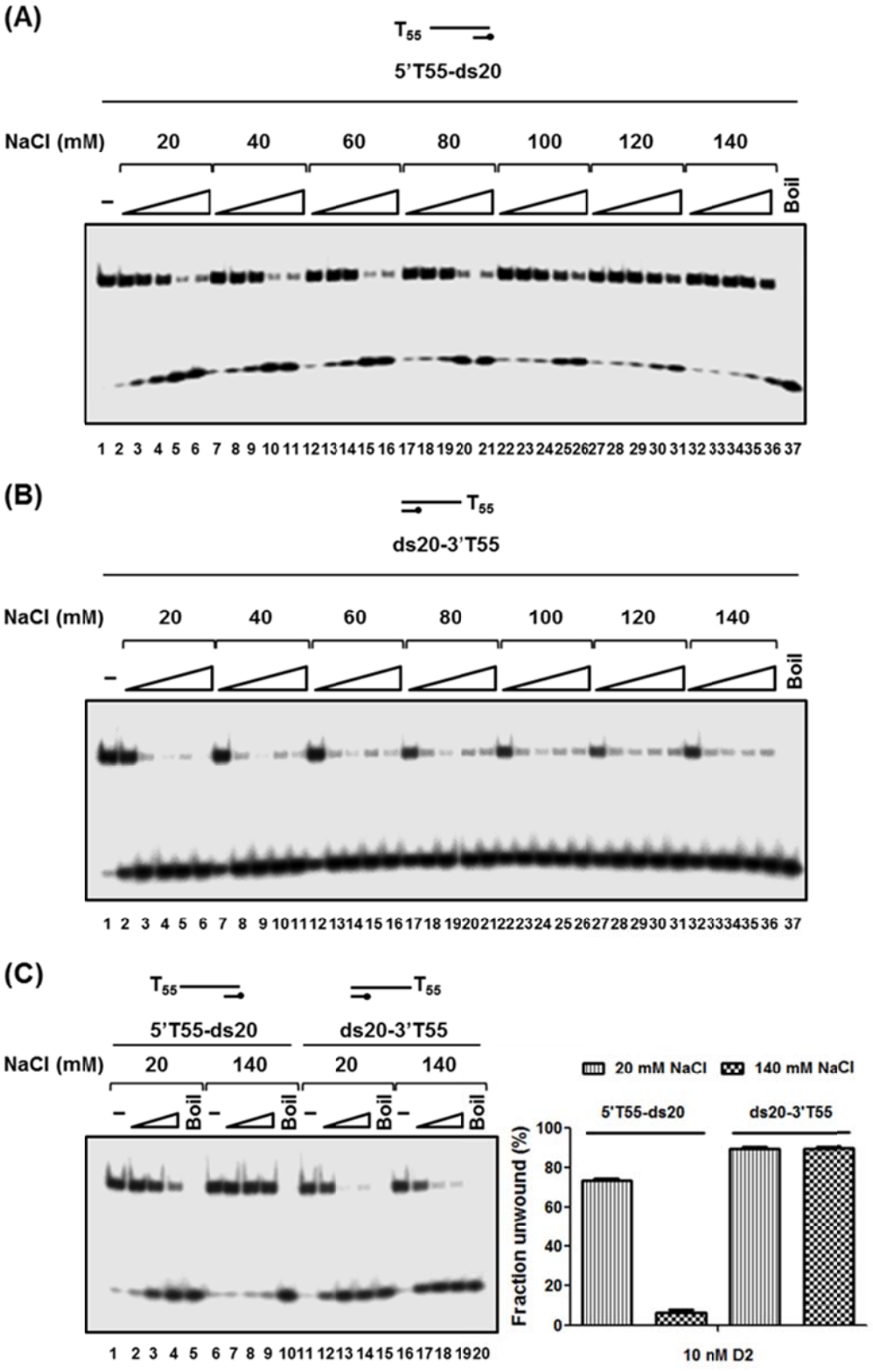


Figure 3.10 Effect of NaCl concentration on the bipolar DNA unwinding activity of D2. The unwinding of 5'T55-ds20 (**A**) and ds20-3'T55 (**B**) by D2 under various [NaCl] conditions (20-140 mM). Lane 1, no protein control (-); lane 37, heat-denatured substrate control (Boil); lanes 2-6, lanes 7-11, lanes 12-16, lanes 17-21, lanes 22-26, lanes 27-31 and lanes 32-36, D2 protein titrations (0.1, 0.5, 1, 10, 100 nM) at the indicated [NaCl]. (**C**) The 5'→3' unwinding

activity is more sensitive to increasing NaCl concentration than the 3'→5' unwinding activity. At 20 mM NaCl, D2 showed bipolar unwinding activity (lanes 1-5 and lanes 11-15), while at 140 mM NaCl, the 5'→3' unwinding activity was significantly inhibited (lanes 6-10), but not the 3'→5' unwinding activity (lanes 16-20). (n=3)

3.2 Discussion

This is the first study to report the characteristics of bacteriophage T5 D2 protein. Initially, the D2 protein sequence was analyzed using a number of online bioinformatic programmes. D2 was predicted to probably consist of a HTH DNA-binding motif and a conserved Walker A motif (Figure 3.1), suggesting D2 may be a DNA-binding protein and may have ATP*ase* activity. Moreover, D2 was shown to possess a conserved domain which was identified in the herpesvirus origin-binding proteins/helicases (Figure 3.1). The exact function of this conserved domain is unknown, but it was assumed to be responsible for DNA origin-dependent replication (Boehmer et al., 1994; Nicholas, 1994). This suggests D2 may play a role in T5 origin-dependent replication.

Expression constructs for MBP-His-tagged D2 WT (wild-type) and K405E (with a mutation in the Walker A motif) were generated. The D2 WT and mutant recombinant proteins were successfully purified for further biochemical analyses. D2 was demonstrated to possess a DNA-dependent ATP*ase* activity stimulated by a variety of linear DNA substrates, including ssDNA, blunt-ended dsDNA and dsDNA with a 5' tail or a 3' tail (Figure 3.3 and 3.5). These results fit with the prediction of a conserved Walker A motif within D2 protein sequence.

Next, the ability of D2 to unwind dsDNA substrates (with or without an ssDNA tail) was examined. D2 was finally determined to be a rare bipolar helicase since it could unwind partial dsDNA with a 5' or 3' ssDNA tail (Figure 3.4A). In addition, D2 helicase was shown to require a free 5' or 3' ssDNA tail for initiation of DNA unwinding because it was unable to unwind blunt-ended dsDNA (Figure 3.4A). Interestingly, the degree of each unidirectional unwinding

by D2 is not equivalent. D2 showed greater 3'→5' unwinding activity than 5'→3' unwinding activity (Figure 3.4A). Complementing this, the binding of D2 to the linear DNA substrates (ssDNA, dsDNA, 5' tail dsDNA and 3' tail dsDNA) revealed that D2 has higher binding affinity for the dsDNA substrates that it prefers to unwind (Figure 3.4B, the D2 binding affinity for the substrate: 3' tail dsDNA> 5' tail dsDNA> blunt-ended dsDNA), suggesting there is a positive correlation between the binding affinity of D2 for the substrate and the extent of D2 unwinding activity. Furthermore, the binding data also implicated that D2 binds preferentially to ssDNA and/or ss-dsDNA junction of the tail substrate since its interaction with blunt-ended dsDNA is very weak (Figure 3.4B).

A series of 20-bp duplex substrates with either a 5' or 3' ssDNA tail ranging from 0 to 55 bases were used for a systematic investigation of D2 DNA unwinding and binding as a function of 5' or 3' tail length. The 5'→3' or 3'→5' unwinding activities of D2 was found to be dependent on 5' tail length or 3' tail length. When the tail length was extended, the extents of unwinding activities generally increased before they reached the maximum (Figure 3.7). Furthermore, the binding of D2 to the tail substrates was also found to be tail-length-dependent. The binding affinity for the substrate increased progressively with tail length (Figure 3.8). D2 bound to the substrates (5'T25 to T55-ds20, ds20-3' T15 to T55), forming one predominant protein-DNA complex (C1). A second slower migrating complex (C2) was observed with the tail substrates having longer ssDNA tails (5'T55-ds20, ds20-3'T45 and ds20-3'T55), which were more efficiently unwound (Figure 3.8). Even though the composition of the D2 unwinding complex is not known, the multiple binding of proteins to the tail substrates with longer ssDNA tails suggested that these substrates are able to recruit multiple D2 proteins and they may function cooperatively to enhance D2 helicase action.

It was also noted that the binding affinity of D2 and the ability to form higher-order complexes on the substrates with tails of increasing length correlated with the relative ability to unwind the substrate (Figure 3.7 and 3.8). However, there are minor inconsistencies between the helicase and DNA-binding activities of D2. No detectable D2 binding to substrates 5'T15-ds20 and ds20-3'T5 was seen (Figure 3.8), but D2 could efficiently unwind these two substrates (Figure 3.7). This may be due to the fact that the DNA-binding and helicase experiments were performed under different conditions (such as buffer composition, protein and substrate concentration) or weak/transient interaction between D2 and DNA cannot be detected by the gel-shift assay. These reasons may also explain why no higher-order D2-DNA complex (C2) formed on 5'T45-ds20, but the enhanced helicase activity of D2 (possibly caused by binding of multiple proteins to DNA) was observed (additional ~20% of the DNA molecules of 5'T45-ds20 and 5'T55-ds20 are unwound compared with 5'T15-ds20, 5'T25-ds20 and 5'T35-ds20; Figure 3.7A, graphed data).

The bioinformatic analysis of D2 showed that it does not belong to any known helicase family. Also, only one of three universal motifs of SF1-SF6 helicases (Walker A motif, ATP-binding motif) was identified in D2 (Figure 3.1), implying D2 may possess some unconventional motifs or domains relevant to its helicase and ATP*ase* activities. A single amino acid substitution (K405E) in the Walker A motif abolished D2 ATP*ase* and bipolar helicase activities, but retained DNA-binding activity (Figure 3.3 and 3.5). This suggests both 5'→3' and 3'→5' motor units of D2 share the same catalytic motif (at least Walker A motif). Moreover, differential inhibition of D2 bidirectional unwinding activities by high NaCl concentration (Figure 3.10) suggests that the unknown functional motifs or domains, specifically related to 5'→3' unwinding activity, are more sensitive to monovalent sodium cations compared with those specifically related to 3'→5' unwinding activity.

In general, most helicases demonstrated unidirectional helicase activity and bipolar helicase activity is rare among known helicases. A literature search reveals that only ten bipolar helicases have been discovered in all organisms, including *E. coli* RecBCD (Dillingham et al., 2003; Taylor and Smith, 2003), *Staphylococcus aureus* and *Bacillus anthracis* PcrA (Anand and Khan, 2004; Naqvi et al., 2003), *Sulfolobus acidocaldarius* HerA (Constantinesco et al., 2004), *Plasmodium falciparum* PfDH45 and PfDH60 (Pradhan and Tuteja, 2006, 2007), pea PDH47 (Vashisht et al., 2005), rabbit eIF4A (Rozen et al., 1990), human p68 (Huang and Liu, 2002) and hChlR1 (Hirota and Lahti, 2000). Hence, D2 may be the first identified bipolar helicase encoded by a virus.

D2 displayed a biased polarity preference with its 3'→5' unwinding activity being greater than its 5'→3' unwinding activity when the 5' and 3' tail test substrates have identical tail length. Different levels of unwinding activity in the 5' or 3' direction are common to most of the DNA bipolar helicases characterized to date (Anand and Khan, 2004; Dillingham et al., 2005; Hirota and Lahti, 2000; Naqvi et al., 2003; Pradhan and Tuteja, 2007; Vashisht et al., 2005). This common feature may have important biological implications. A major barrier to understand the exact role of this feature is that there is very little information available about the biological function of most bipolar helicases.

Recently, several studies on *E. coli* RecBCD bipolar helicase complex have proposed a model to show how the biased polarity preference for unwinding relates to the function of RecBCD. RecBCD, which is involved in the major pathway of recombinational repair, possesses two independent helicase subunits with opposite polarity (RecB: 3'→5' helicase, RecD: 5'→3' helicase) to translocate along both ssDNA strands of anti-parallel DNA duplex (Dillingham and Kowalczykowski, 2008; Dillingham et al., 2005). The bidirectional unwinding biased toward one direction makes two helicase subunits move on opposite strands at different speed to unwind the DNA. The faster motor is a

leading motor to act as a genuine helicase, while the slower motor is a trailing motor to act as an ssDNA translocase. Two motors with different speeds were reported to adjust the overall speed of RecBCD. Before Chi (crossover hotspot instigator) sequence recognition, RecD is the faster motor and an ssDNA loop forms ahead of the slower RecB motor. When the Chi sequence is recognized and tightly bound by RecC, RecB becomes the leading motor and the speed of RecD is reduced to be below that of RecB. Consequently, the existing ssDNA loop shortens and a new ssDNA loop grows and accumulates between RecB and RecC. This ssDNA loop is loaded by multiple RecA proteins and cut by a nuclease upstream the Chi sequence. The resulting RecA/Chi-containing ssDNA undergoes the strand invasion with homologous duplex to produce D-loop (Displacement loop) to initiate recombination for DNA repair (Figure 3.11) (Amundsen and Smith, 2007; Dillingham and Kowalczykowski, 2008; Spies et al., 2007).

Our data presented in this study are not enough to reveal D2 biological function(s) and the functional role of its biased polarity preference. However, several possibilities may be envisioned. (a) One possible role of D2 may involve the initiation of origin-dependent replication since it shares some sequence similarity with many DNA replication origin-binding proteins/helicases, observed in the bioinformatic analysis. (b) Concerning the potential replicative helicase activity of D2, D2 could unwind DNA at the replication fork. To refer to the above RecBCD model, the active D2 (monomer, multimer?) may contain two motor units of opposite polarity (sharing the same catalytic motif, different from RecBCD). Assuming that the 3'→5' motor unit may bind to the 3' tail of the replication fork and the 5'→3' motor unit may bind to the 5' tail, the polarity preference of D2 for unwinding duplex DNA in 3'→5' direction greater efficiently than in 5'→3' direction may lead to two motor units with different speeds and formation of an ssDNA loop, as described in the RecBCD model. Two helicase motors with different speeds may act as a molecular "gearbox" in

DNA replication. The switch of the acting leading motor between the two motors may adjust the overall speed of the replication machinery to coordinate continuous leading-strand and discontinuous lagging-strand DNA syntheses in the replication fork. The possible ssDNA loop formation and release may play a role in preventing leading-strand synthesis from outpacing lagging-strand synthesis (Figure 3.12). All of these estimations need confirming by further genetic and biochemical analyses.

In conclusion, D2 is the first identified viral bipolar helicase. It has some interesting characteristics, such as biased bidirectional unwinding activity with tail length dependence and different salt sensitivity as well as lack of conventional motifs of SF1-SF6 helicases, so it is of great interest to further understand D2 helicase for finding some unknown characteristics and mechanisms of bipolar helicases.

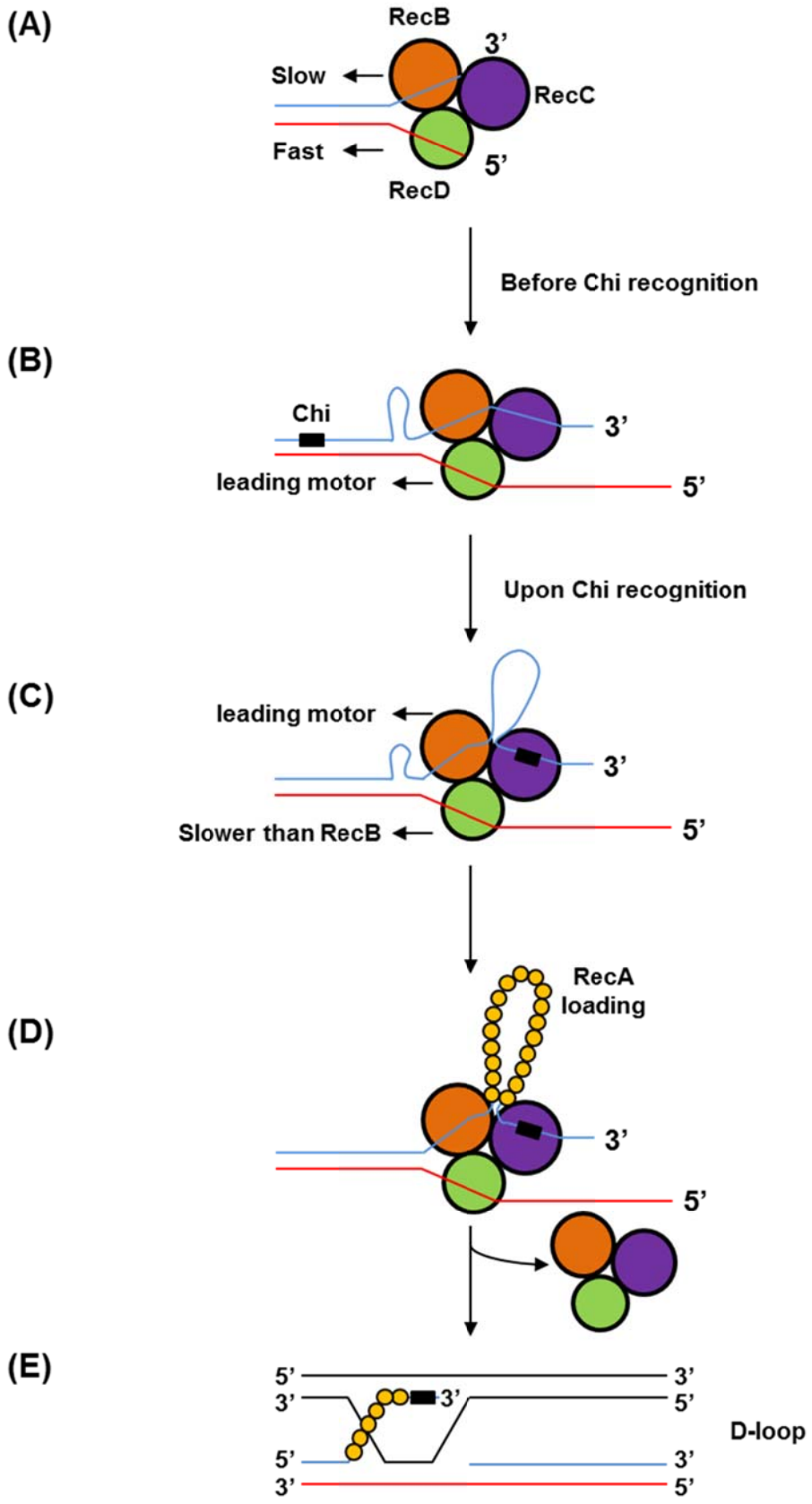


Figure 3.11 The illustration of RecBCD-catalyzed recombination by uncoupled translocation of two helicase subunits of RecBCD. (**A**) RecBCD initiates unwinding duplex DNA by two

helicase motors with opposite polarity and different speed (RecB: 3'→5' slow motor, RecD: 5'→3' fast motor). The three subunits of RecBCD are coloured (RecB: orange; RecC: purple; RecD: green). (**B**) Before Chi (crossover hotspot instigator) sequence recognition, RecD is the leading motor and an ssDNA loop forms ahead of RecB motor. (**C**) Upon Chi recognition by RecC, RecB becomes the leading motor and the speed of RecD is reduced to be slower than that of RecB. The existing ssDNA loop shortens and a new ssDNA loop grows and accumulates between RecB and RecC. (**D**) This ssDNA loop is loaded by multiple RecA proteins and RecBCD is released from the DNA. (**E**) The ssDNA loop is cut with a nuclease upstream the Chi sequence and the resulting RecA/Chi-containing ssDNA undergoes the strand invasion with homologous duplex to produce a D-loop (Displacement loop) for initiation of recombinational repair. This figure is adapted from (Dillingham and Kowalczykowski, 2008; Spies et al., 2007)

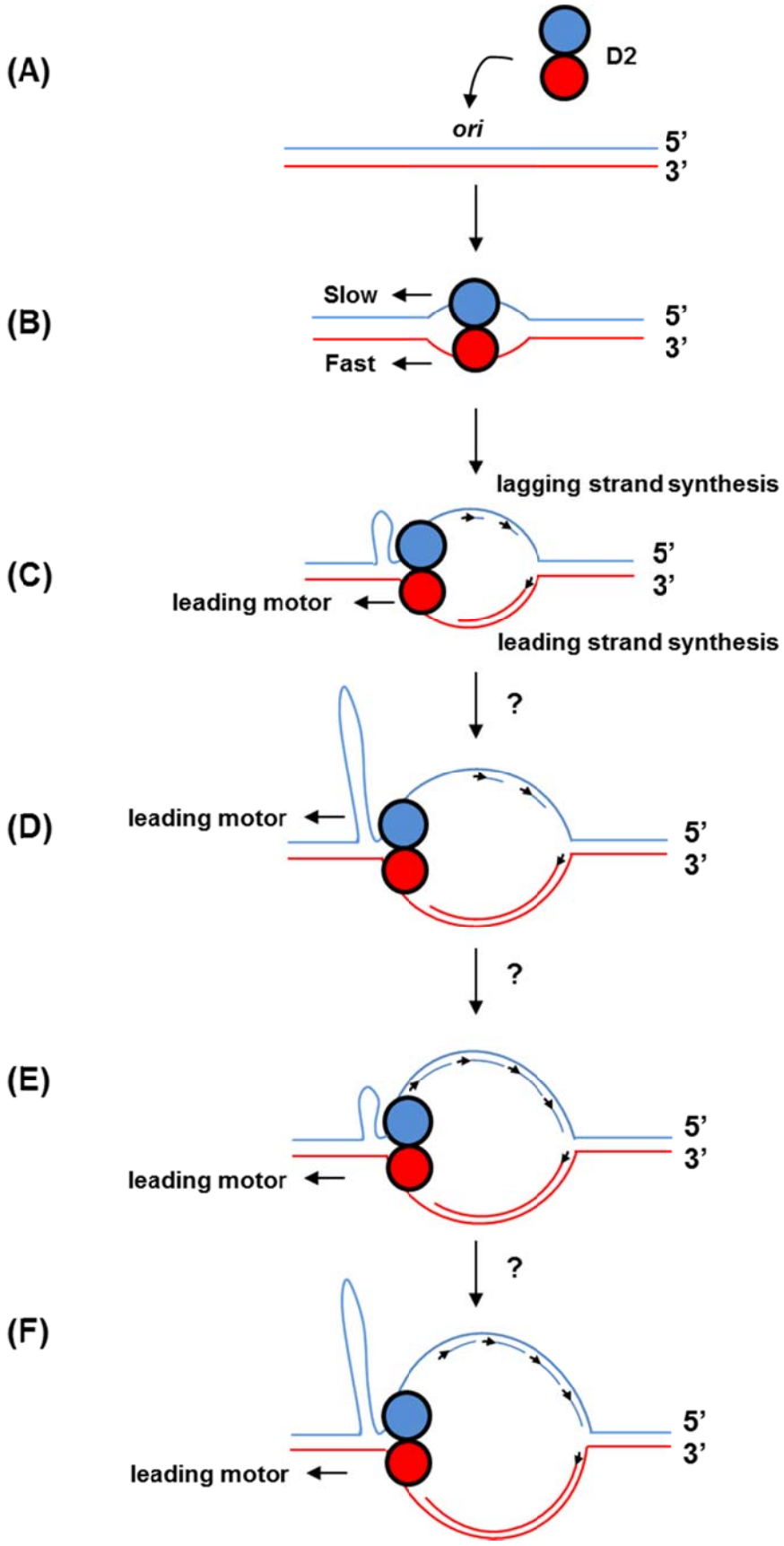

(A)
D2
ori
5'
3'
(B)
Slow
Fast
5'
3'
lagging strand synthesis
(C)
leading motor
5'
3'
leading strand synthesis
?
(D)
leading motor
5'
3'
?
(E)
leading motor
5'
3'
?
(F)
leading motor
5'
3'

Figure 3.12 Proposed function of D2 protein and possible functional role of its biased polarity preference. This model assumes that D2 (monomer, multimer?) may contain two motor units with opposite polarity and the two motor units may bind to opposite ssDNA strands of a replication fork and move in the same overall direction to unwind the duplex DNA. (**A**) D2 probably recognizes and binds to the replication origin (*ori*), and then begins to unwind the duplex DNA. (**B**) Two helicase motors with different speeds may act as a molecular gearbox in DNA replication. (**C**-**E**) The switch in the leading motor between two motors (determined by unknown factors) may adjust the overall speed of replication machinery to coordinate continuous leading-strand and discontinuous lagging-strand DNA syntheses in the replication fork. The ssDNA loop formation and release may play a role in preventing leading-strand synthesis from outpacing lagging-strand synthesis.

Chaper 4. Bacteriophage T5 D6

4.1 Results

4.1.1 Bioinformatics

The potential helicase motifs and domains within the D6 protein sequence (Genbank accession no. AAU05256.1) were identified using the methods described in Section 3.1.1.

4.1.1.1 The analysis of amino acid sequence

The ProtParam tool predicted that D6 is a 57.5-kDa protein with an isoelectric point of 5.37. InterProScan analysis tool identified that D6 may contain one conserved DnaB-like helicase C-terminal domain (PF03796, Pfam database; at amino acids 194-279), which consists of an ATP-binding site involved in ATP hydrolysis. Moreover, the PROSCAN tool predicted that D6 may possess two conserved motifs: Walker A motif (motif I, 100% identity) at amino acids 207-214 and Walker B motif (motif II, 69% similarity) at amino acids 410-418, which constitute an ATP*ase* active site responsible for binding and hydrolyzing ATP. Also, BLASTP analysis for the D6 sequence showed there are several weak hits to replicative DNA helicases from bacteria and viruses. These all proposed that D6 might have the ability to hydrolyze ATP for its potential helicase activity.

4.1.1.2 Secondary structural analysis of the D6 protein

The GOR IV tool predicted that D6 probably contains 15 alpha helices, 14 extended strands and 27 random coils. In addition, a helix-turn-helix DNA-binding motif was identified at the amino acids 243-264 of D6 using the HTH prediction software, suggesting D6 could bind DNA (Figure 4.1).

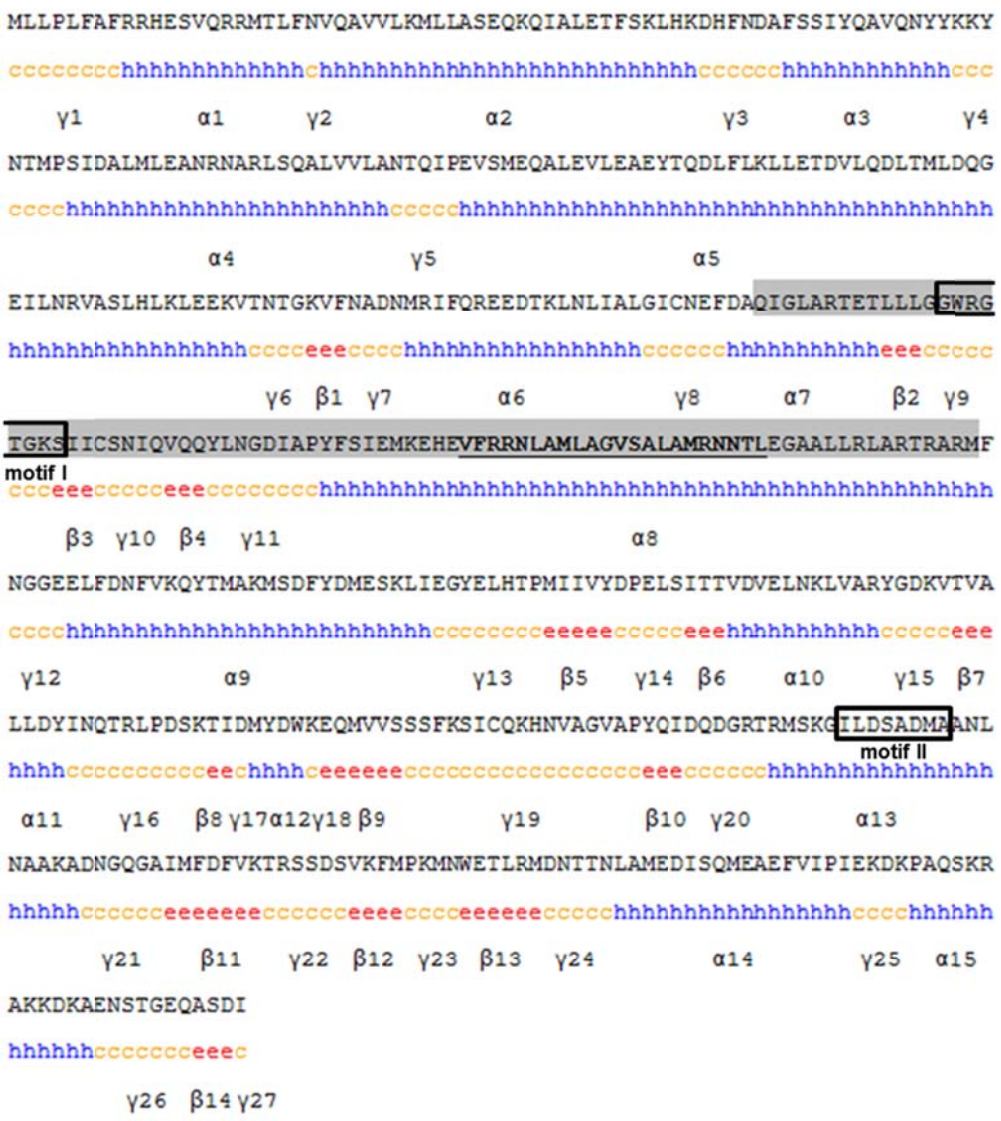


Figure 4.1 The predicted secondary structure of D6. The amino acid sequence of D6 is represented in one-letter-code format. The secondary structure of each amino acid is shown by h (alpha helix, blue), e (extended strand, red) and c (random coil, orange). The secondary structure elements (α, alpha helix; β, extended strand; γ, random coil) are listed in order. The conserved helicase C-terminal domain probably responsible for ATP binding and hydrolysis is shaded in grey. The putative Walker A motif (motif I) and Walker B motif (motif II) are boxed and the helix-turn-helix DNA-binding motif is highlighted in underlined bold script.

4.1.2 Expression, purification and characterization of D6

The D6 ORF was cloned into pGEX-KG to express GST-tagged D6, containing a thrombin cleavage site between GST tag and D6 protein, in *E. coli*. The GST-tagged D6 was purified and analyzed by SDS-PAGE (Figure 4.2A, lane

1). The molecular weight of the major purified protein is approximately 75 kDa, slightly smaller than that predicted for the GST-D6 (~83 kDa). Since the NaCl concentration was found to affect the removal of GST tag from the GST-fusion protein by thrombin cleavage and the solubility of the thrombin cleavage product in Dr. Cyril Sanders' lab, the NaCl concentration was optimized for thrombin to completely remove the GST tag from the GST-D6 protein. After the digestion, there are two major cleaved products (~55 kDa corresponding to the molecular weight predicted for D6 and ~25 kDa corresponding to that of GST tag, lanes 2-9). However, GST-D6 was not fully digested (<28% digested) by thrombin (15 units of thrombin per mg of GST-D6 at 4°C for >18 h) over the NaCl concentration range tested (from 200 to 800 mM, lanes 3-9). One possible reason for poor cleavage is that D6 protein which is free from GST could mask the thrombin cleavage site or cause steric hindrance between thrombin and the cleavage site. Furthermore, the stimulation of GST-D6 ATP*ase* activity was examined in the presence of various linear DNA substrates, including ssDNA (T55, 55-base dT), blunt-ended dsDNA (ds40, 40 bp), and partially double-stranded DNA molecules with either a 5' or 3' ssDNA overhang (5'T55-ds20 and ds20-3'T55, 20 bp and a 55-base dT overhang). In contrast to without DNA control, almost no ATP*ase* activity was observed in the presence of any tested substrates (Figure 4.2B).

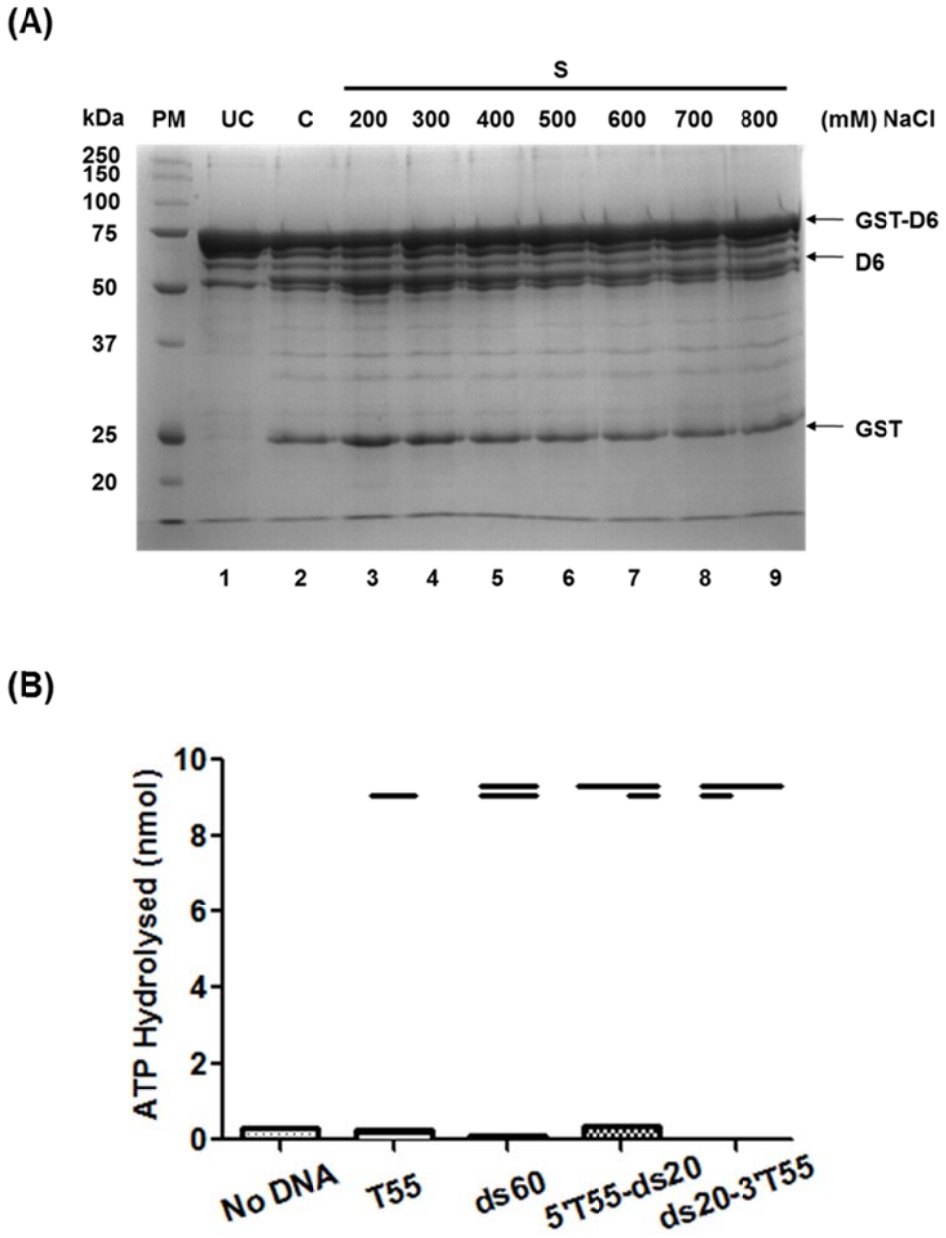


Figure 4.2 Purification and ATP*ase* activity of GST-tagged D6. (**A**) 10% SDS-PAGE analysis of purification of GST-D6 (lane 1) and thrombin cleavage of GST-D6 under the conditions of different NaCl (lanes 2-9). The thrombin cleavage reactions (15 cleavage units of thrombin per mg of GST-D6) were incubated at 4°C for >18 h in the GST elution buffer with increasing [NaCl] (from 200 mM to 800 mM). The samples representing the total cleaved products (200 mM NaCl, lane 2, C) and the soluble fractions of the cleaved products under various [NaCl] as indicated (from 200 mM to 800 mM, lanes 3-9, S) were analysed by SDS-PAGE to estimate the yield, purity and extent of thrombin digestion. The molecular weight of the major purified protein (~75 kDa) is slightly smaller than that predicted for GST-D6 (~83 kDa), but the molecular weight of the cleaved product (free from the GST tag, 55 kDa) corresponded to that predicted for D6 (~57 kDa). (**B**) ATP*ase* activity of GST-D6 (40 nM) was measured in the absence of DNA substrate or presence of various DNA substrates (40 nM) – ssDNA, T55;

blunt-ended dsDNA, ds60; partial duplexes with a 20 bp and a 5' or 3' 55-base oligo-dT overhang, 5'T55-ds20 and ds20-3'T55.

4.2 Discussion

Here, the D6 protein sequence was analyzed using above-mentioned bioinformatic tools. D6 was shown to possibly contain two conserved motifs, Walker A and B, which consitutue a classical ATP*ase* active site responsible for ATP binding and hydrolysis. In addition, a hypothetical helix-turn-helix DNA-binding motif was identified in D6. All these data suggested D6 may have the ability to bind DNA and hydrolyze ATP. However, our preliminary experimental results showed that GST-tagged D6 has no significant ATP*ase* activity stimulated by linear DNA substrates (including ssDNA, blunt-ended dsDNA and partially DNA duplex with either a 5' or 3' ssDNA overhang) under the condition tested (Figure 4.2B). If D6 did not have ATP*ase* activity, it would be less likely to be a helicase since all known helicases are able to catalyze ATP hydrolysis (Tuteja and Tuteja, 2004).

The observation that GST-D6 is unable to hydrolyze ATP is possibly due to the following reasons: (a) the reaction conditions (such as buffer pH, temperature, time duration and DNA substrate) are not optimal. (b) GST-D6 protein is not pure enough for ATP*ase* assays, and other contaminant proteins (at least three, Figure 4.2A, lane 1) may hinder the reactions. (c) The additional GST moiety of GST-D6 may inhibit D6 ATP*ase* activity by changing the conformation of the motifs related to ATP hydrolysis.

Although the D6 protein sequence (mainly the putative DnaB-like helicase C-terminal domain) was noted to share sequence similarity to some replicative DNA helicases of bacteria and viruses, there is still no experimental evidence to show whether D6 is a helicase or not.

The progress on D6 in this study is therefore limited. In the future, the

protein purification for the active D6 protein and the conditions for the detailed *in vitro* functional analyses on D6 could be established and optimized. These all would facilitate determining whether D6 is a bacteriophage T5 helicase.

Chaper 5. Bacteriophage T5 D10 helicase

5.1 Results

5.1.1 Bioinformatics

The D10 protein sequence was analyzed to identify possible helicase motifs, domains and structures using the above-mentioned bioinformatic tools as well as the protein structure modeling and visualization programs, PHYRE (http://www.sbg.bio.ic.ac.uk/phyre2/) and PyMOL (http://www.pymol.org/).

5.1.1.1 The analysis of amino acid sequence

The D10 protein was predicted to be a 50-kDa protein with an isoelectric point (pI) of 8.69 using the ProtParam tool and to possess two conserved helicase domains: SF1 and SF2 helicase ATP-binding type-1 domain (accession no. PS51192, PROSITE database) at amino acids 95-240 as well as SF1 and SF2 helicase C-terminal domain (accession no. PS51194, PROSITE database) at amino acids 289-439 by the InterProScan tool. Within D10, PROSCAN found there are two conserved motifs, a Walker A motif (motif I, 100% identity) at amino acids 108-115 and a Walker B (motif II, 90% similarity) at amino acids 188-197, which constitute an ATP*ase* active site responsible for ATP binding and hydrolysis. A search for homologues of the D10 sequence using BLASTP revealed several homologous helicases from bacteria and viruses (Figure 5.1), providing some support to the hypothesis that D10 is a T5 helicase. In addition, ClustalW sequence alignment of D10 along with its closest homologue UvsW helicase from bacteriophage T4 (Figure 5.2) has shown that D10 may have five SF2 conserved helicase motifs, including three above-mentioned universal motifs (motif I, Walker A; motif II, Walker B; motif VI, an arginine finger) of SF1-SF6 family helicases.

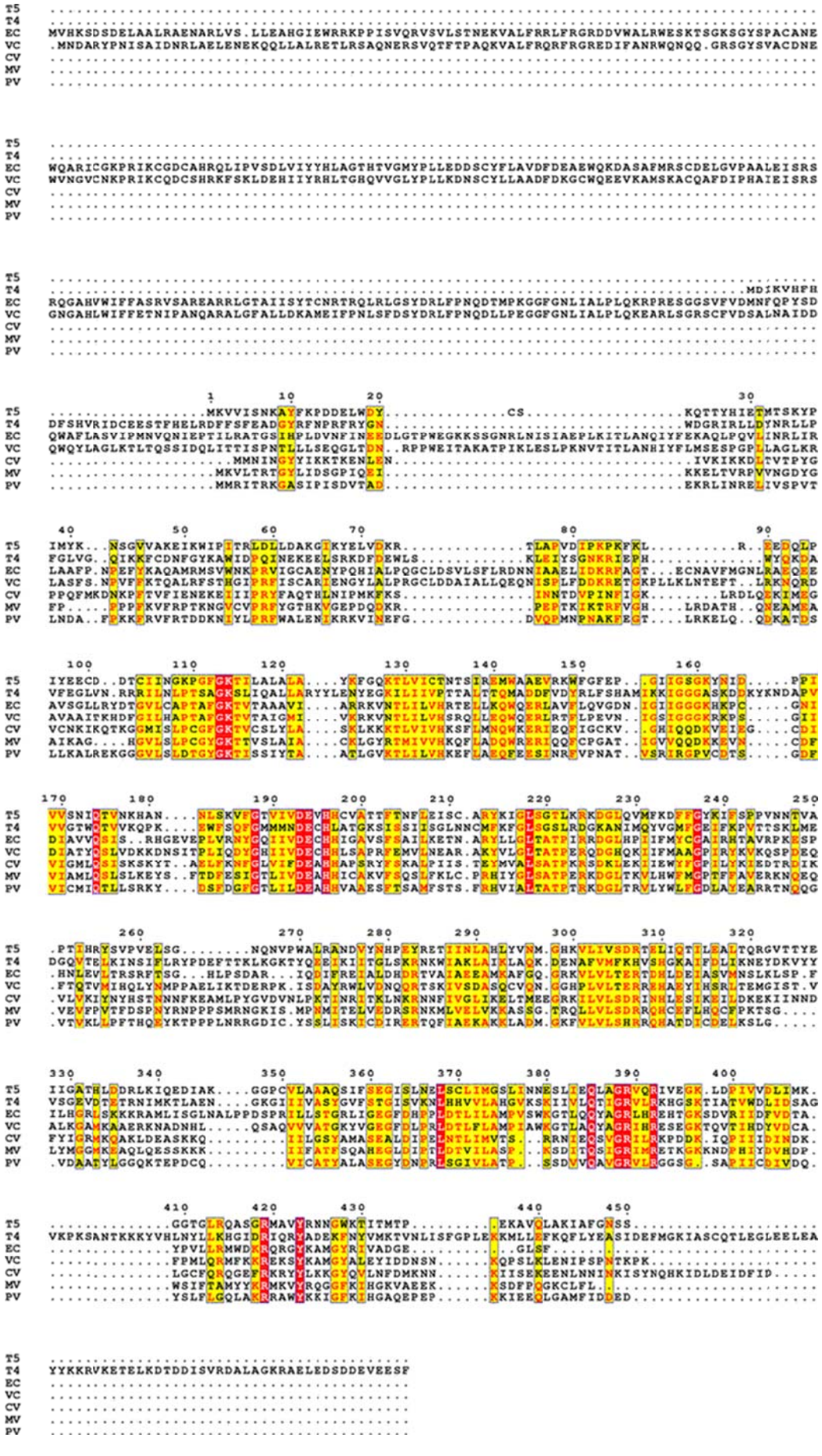


Figure 5.1 The multiple sequence alignment of D10 with other six homologous helicases. These seven proteins are: T5, T5 phage D10, YP_006952; T4, T4 phage UvsW, NP_049796.1; EC, *Escherichia coli* ED1a, YP_002396352.1; VC, *Vibrio cholerae* RC385, ZP_06943579.1; CV, *Cafeteria roenbergensis virus* BV-PW1, YP_003969949.1; MV, *Micromonas sp.* RCC1109 *virus* MpV1, YP_004061947.1; PV, *Paramecium bursaria Chlorella virus* FR483, YP_001425838.1. The identical residues are shaded in red and highly conserved residues are in red script.

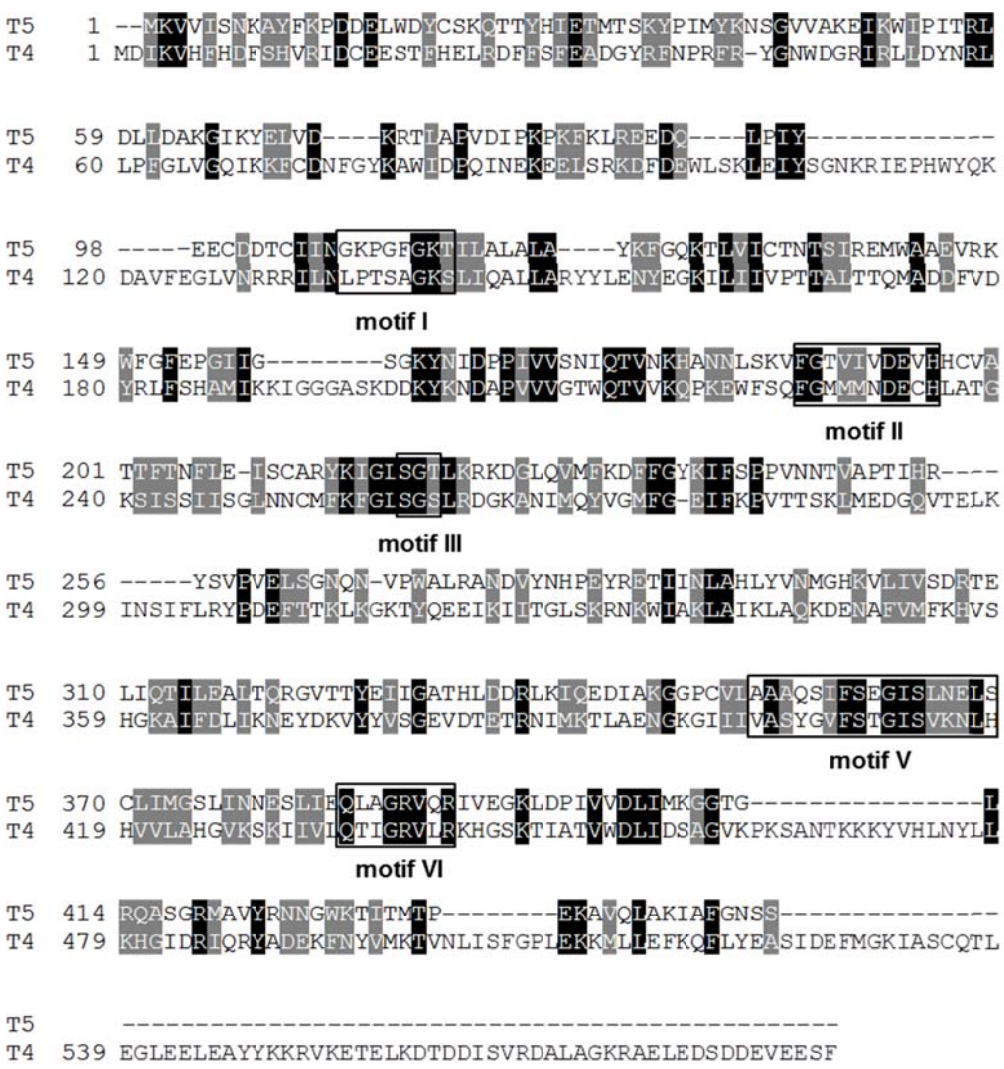


Figure 5.2 ClustalW alignment of T5 D10 (YP_006952) against T4 UvsW (NP_049796.1). The sequences display about 20% identity and 46% similarity, with five conserved SF2 motifs. Identical and conserved amino acid residues are shaded in black and grey, respectively. The five conserved helicase motifs are boxed. The locations of the conserved motifs are based on the previous study on T4 UvsW (26).

5.1.1.2 Structural analysis of the D10 protein

The theoretical secondary structure analysis of D10 predicted that it potentially has 14 alpha helices, 23 extended strands, 36 random coils, and a helix-turn-helix DNA-binding motif at amino acids 114-125, analyzed by GOR IV and HTH programs (Figure 5.3). This suggested D10 may be a DNA-binding protein. Because the crystal structure of D10 has not been determined, the predicted 3D structure of D10 was modeled via PHYRE using its closest homologue T4 UvsW as the template. The predicted structure of D10 was visualized and analyzed using PyMOL software and the conserved helicase domains as well as motifs are highlighted (Figure 5.4). In both of Figure 5.3 and 5.4, Walker A, Walker B and motif III reside in the SF1 and SF2 helicase ATP-binding type-1 domain as well as an arginine finger and motif V are located in the SF1 and SF2 helicase C-terminal domain. Based on the description about the nucleotide (ATP) binding site in Section 1.3.1, the D10 nucleotide binding site (associated with these three possible universal motifs, Walker A, B and an arginine finger) is possibly located at the interface of two putative domains.

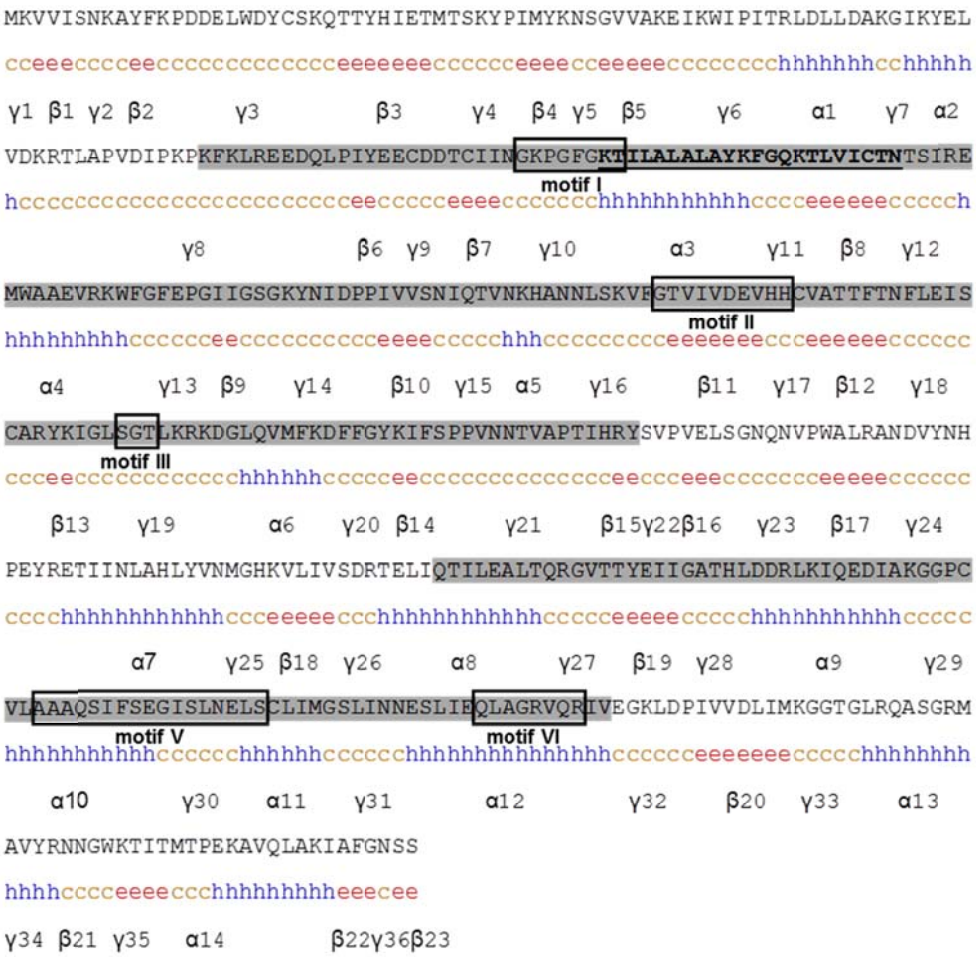


Figure 5.3 The predicted secondary structure of D10 analyzed by GOR IV and HTH programs. The secondary structure of each amino acid is represented by h (alpha helix), e (extended strand) and c (random coil) and the secondary structure elements are listed in order. Two putative helicase domains (SF1 and SF2 helicase ATP-binding type-1 domain as well as SF1 and SF2 helicase C-terminal domain) are shaded in grey. Five SF2 conserved helicase motifs are boxed and a helix-turn-helix DNA-binding motif is indicated in underlined bold script. Motifs I-III and motifs V-VI reside in putative SF1 and SF2 helicase ATP-binding type-1 domain as well as helicase C-terminal domain respectively.

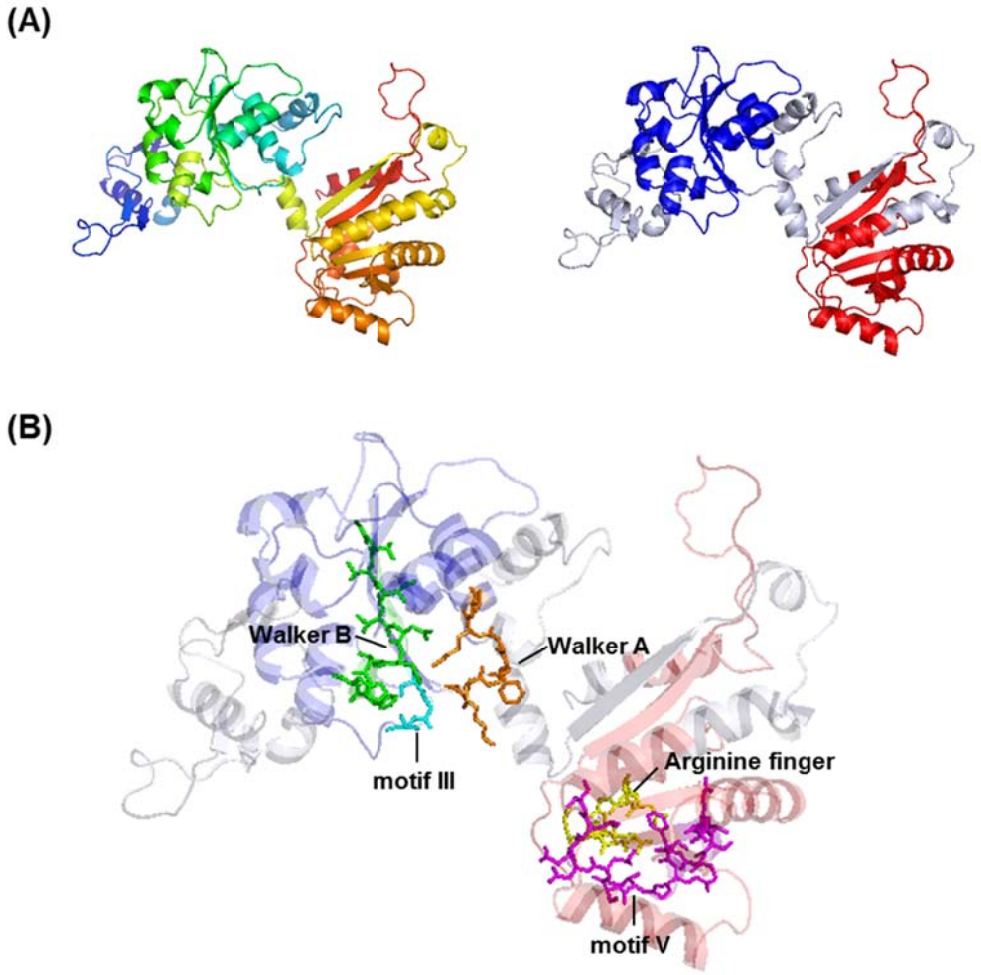


Figure 5.4 The modeled structure of D10 protein generated using PHYRE. (**A**) The right panel highlights the putative SF1 and SF2 helicase ATP-binding type-1 and helicase C-terminal domains in blue and red respectively. (**B**) Structural representation of five conserved motifs in D10 (Walker A motif, orange; Walker B motif, green; motif III, cyan; motif V, pink; Arginine finger, yellow). Walker A, Walker B and motif III reside in the SF1 and SF2 helicase ATP-binding type-1 domain. An arginine finger and motif V are located in the SF1 and SF2 helicase C-terminal domains.

5.1.2 Purification of T5 D10

The bioinformatic analysis described above revealed that the D10 protein has five conserved helicase motifs belonging to the helicase superfamily 2, including three universal helicase motifs, Walker A (motif I), Walker B (motif II) and an Arginine finger (motif VI), illustrated in Figure 5.5A. The expression constructs of D10 wild type (WT) and mutants (with one mutation in each of three conserved motifs I, II and VI) were generated and purification of WT and three mutant recombinant proteins was attempted. However, only D10 WT and R389N (motif VI mutant) proteins were successfully purified; the other mutant polypeptides were severely degraded or totally insoluble after removal of the GST tag. D10 proteins were purified using affinity chromatography (glutathione sepharose), anion-exchange chromatography and gel filtration chromatography. Figure 5.5B shows SDS-PAGE gels of wild-type D10 peak fractions from the final gel filtration step. The molecular weight of the purified protein, approximately 50 kDa, corresponds to that predicted for D10. Protein fractions contained a DNA-dependent ATP*ase* activity and the enzymatic activity was proportional to the D10 protein concentration of the fractions (Figures 5.5B, C and Figure 5.6A). The initial investigations showed that the DNA-dependent stimulation of D10 ATP*ase* activity was DNA-structure-dependent. ATP*ase* assays were performed in the presence of various linear DNA substrates, including ssDNA (T55), blunt-ended dsDNA (20 and 60 bp), and partially single- and double-stranded molecules (ss-dsDNA) with either a 5' or 3' ssDNA overhang (20 bp and T55 ssDNA). Interestingly, only substrates with a duplex DNA component stimulated D10 ATP*ase* activity, and the substrate with 20 bp of dsDNA and a T55 3' overhang (ds20-3'T55) was the most stimulatory (Figure 5.6B). Furthermore, ATP*ase* activity was almost completely abolished by the point mutation R389N in the conserved Arginine finger motif (Figure 5.6A), confirming that the DNA-dependent ATP*ase* activity is attributable to D10.

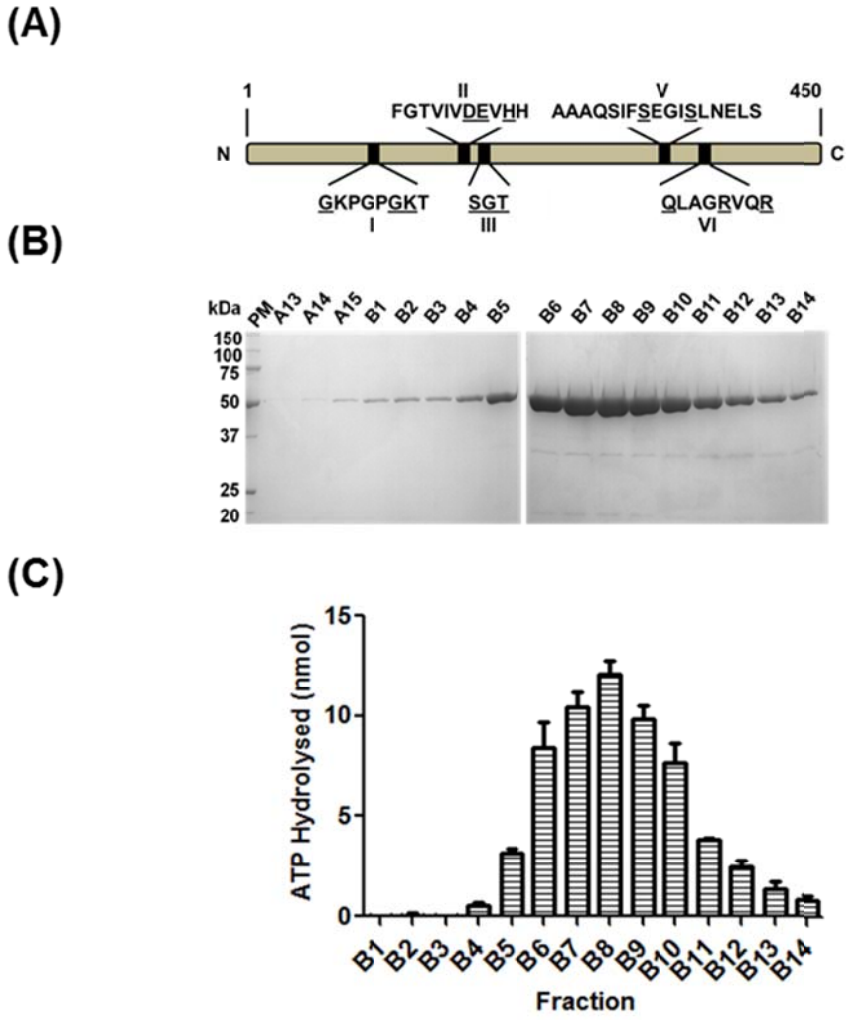


Figure 5.5 Purification and ATP*ase* activity of bacteriophage T5 D10. (**A**) Schematic diagram of putative helicase motifs in D10. Five conserved SF2 helicase motifs (I, II, III, V, VI) were identified in D10 and the conserved amino acid residues are underlined. (**B**) 12% SDS-PAGE gels showing D10 fractions from a Superdex 75 column. The molecular weight of D10 corresponded with ~50 kDa predicted. (**C**) ATP*ase* activity of D10 peak fractions was measured in the presence of ds20-3'T55, a partial single- and double-stranded DNA test substrate with 20 bp and a 3' 55-base oligo-dT overhang. The ATP*ase* activity was proportional to the D10 concentration of each fraction (peak, fraction B8). (n=3, mean and SEM)

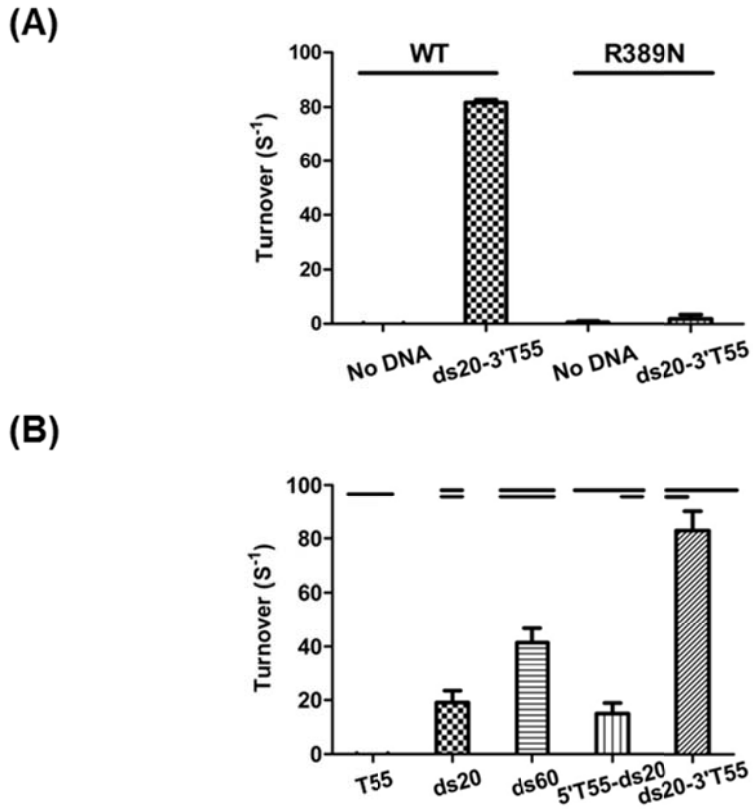


Figure 5.6 ATP*ase* activity of T5 D10. (**A**) ATP*ase* activities of D10 WT and R389N (5 nM) were determined at 22°C for 10 min in the absence or presence of ds20-3'T55 (5 nM), showing the activity was DNA-dependent and almost completely abolished by a mutation (R389N) in the Arginine finger. (**B**) At 100 mM NaCl, ATP*ase* activity of D10 (5 nM) was determined in the presence of various DNA substrates (5 nM) – ssDNA, T55; blunt-ended dsDNA, ds20 and ds60; partial duplexes with 20 bp and a 5' or 3' 55 oligo-dT overhang, 5'T55-ds20 and ds20-3'T55. (n=3)

5.1.3 D10 unwinds synthetic branched DNA substrates

Even though the ATP*ase* activity of D10 was stimulated by dsDNA and ss-dsDNA substrates, especially the one with a 3' T55 ssDNA overhang, D10 was not able to unwind linear ss-dsDNA test substrates (Figure 5.7). As indicated above, the UvsW helicase from bacteriophage T4 is the closest characterised homologue of D10 and D10 is 46 % similar in sequence to UvsW and shares similar conserved helicase motifs (Figure 5.2). Since D10 shares sequence similarity with UvsW, it is logical to suggest that D10 could unwind more

complex branched DNA substrates with ss- and dsDNA arms as having been shown for UvsW (Webb et al., 2007). D10 efficiently unwound a 20-bp duplex with two 20-base ssDNA arms (substrate Fork20, Figure 5.8A, lanes 1-5), but not a similar molecule with the duplex arm extended to 55 bp, even at the highest concentrations of protein tested (substrate Fork55, Figure 5.8A, lanes 6-10). Similarly, a dsDNA Y-junction with three 20-bp duplex arms was unwound efficiently, but not the substrate with one dsDNA arm extended to 55 bp (Figure 5.8A, substrate Y20, lanes 14-18; Y55, lanes 19-23). In addition, for Y20, D10 preferentially dissociated only one of the three strands (oligo A) from the substrate, as described in further detail below. The possibility that D10 has a strand annealing activity like that recently demonstrated for UvsW (Nelson and Benkovic, 2007) was also investigated. A strand annealing activity could confound the interpretation of the results if D10 behaved similarly to UvsW. However, D10 had no such activity as could be detected using the component oligonucleotides generated by heat-denaturartion of substrates Fork20 (Figure 5.9, lanes 6-10), Fork55 (lanes 16-20) and ds40 substrates (lanes 26-30), compared to the positive control of human PIF1 helicase (George et al., 2009). Thus, D10 can unwind fork and Y-junction DNA substrates, but dsDNA length appears to impose a restriction on the ability of D10 to unwind certain test substrates *in vitro*.

The ability of T5 D10 protein to unwind synthetic cruciform non-homologous Holliday junction substrates was also tested as these are unwound by UvsW (Figure 5.8B). The substrate with four 20 bp dsDNA arms, NHJ20, was unwound to generate two sets of products, forked DNA and ssDNA (lanes 11-15). However, there was only one kind of product (forked DNA with a 55 bp dsDNA arm, lanes 16-20) generated from the unwinding of the substrate with two 55 bp and two 20 bp dsDNA arms (NHJ55). Furthermore, the D10 mutant R389N without ATP*ase* activity failed to unwind NHJ20 as a test

substrate (Figure 5.10), confirming that the observed reaction products resulted from the enzymatic action of D10.

As mentioned above, one feature of these observations was that dsDNA length appears to impose a restriction on the ability of D10 to unwind certain test substrates (*i.e.* Fork20 compared to Fork55 and Y20 compared to Y55, Figure 5.8A). One possibility is that the enzyme has limited unwinding processivity and can only melt duplex DNA (≤ 20 bp) at, or close to, the substrate junction point. Alternatively, envisioning a branch migration activity, the non-homologous nature of the test substrate could result in re-annealing of long duplexes before they are completely dissociated. To investigate these possibilities, a close mimic of a natural Holliday junction substrate consisting of a pair of long (55 bp) homologous duplex arms and a pair of short (20 bp) non-homologous duplex arms (HJ55) was constructed. As shown in Figure 5.11, D10 unwound NHJ55, the Holliday junction substrate without any homologous regions, into forked DNA (lanes 1-5) as observed previously (Figure 5.8B). However, D10 converted HJ55 with two homologous 55-bp dsDNA arms into two different kinds of products, fork DNA and recombined linear duplex DNA (Figure 5.11, lanes 6-10) and the amounts of these two sets of products were almost equivalent (graph to the bottom). This infers that D10 has branch-migration activity and can unwind HJ55 in either of two possible directions.

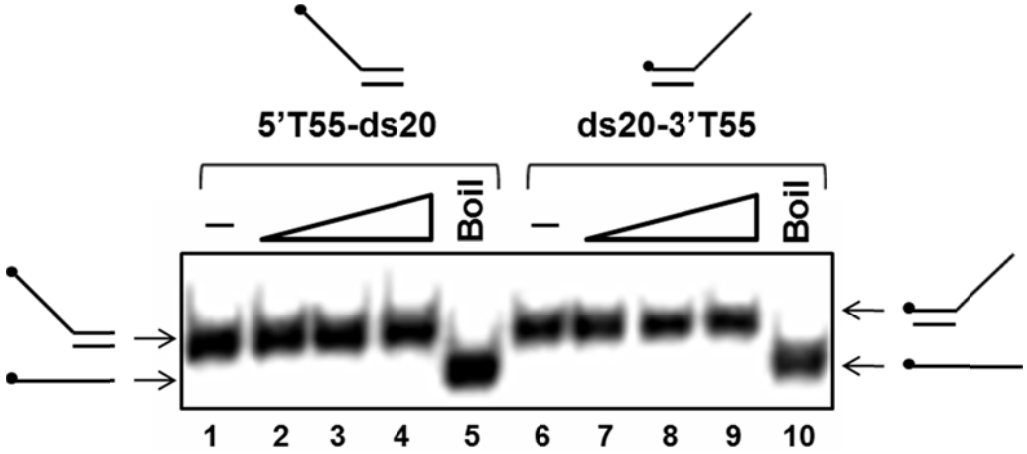
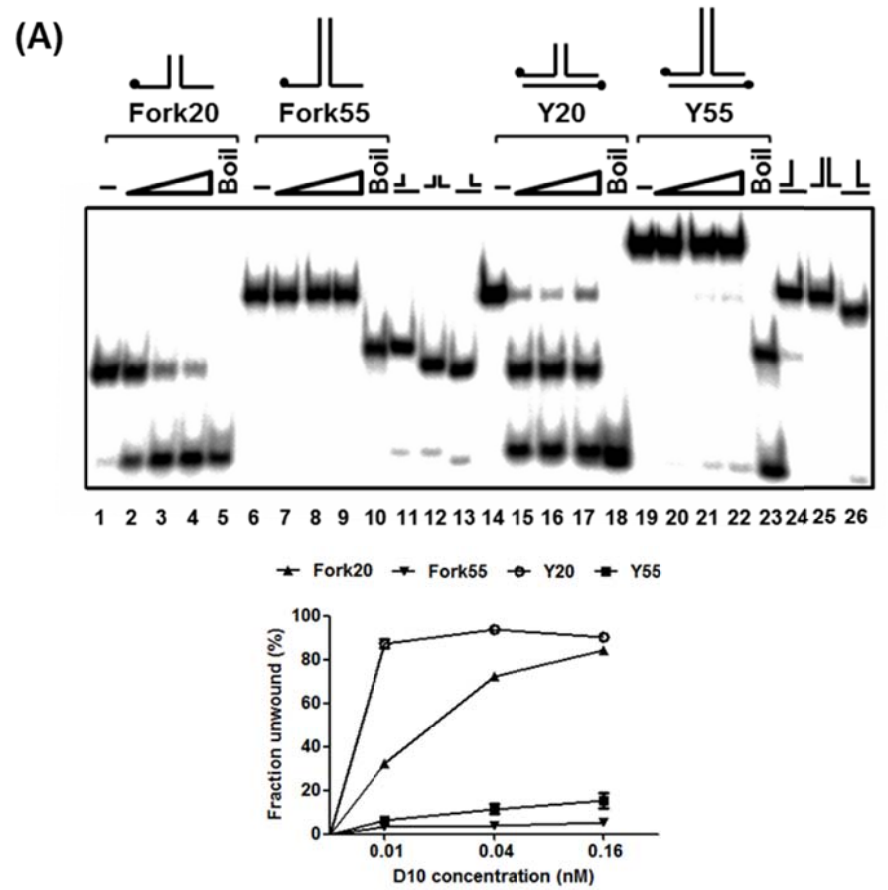


Figure 5.7 Unwinding activity of D10 on partially single- and double-stranded DNA substrates (5'T55-ds20 and ds20-3'T55). The two substrates (0.1 nM) were radiolabelled at their 5' ends and the position of the ^{32}P label in each substrate is indicated by a black circle. Reactions (0.01, 0.04, 0.16 nM D10) were incubated at 37°C for 20 min. The reaction products were separated on an 8% (19:1) polyacrylamide gel and visualized by phosphorimaging. Lanes 1 and 6, no protein control (-); Lanes 5 and 10, heat-denatured substrate control (Boil); Lanes 2-4 and Lanes 7-9, increasing D10 protein concentration. (three repeat experiments , n=3)


(A)
Fork20
Fork55
Y20
Y55
–
Boil
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
Fork20
Fork55
Y20
Y55
Fraction unwound (%)
100
80
60
40
20
0
0.01
0.04
0.16
D10 concentration (nM)

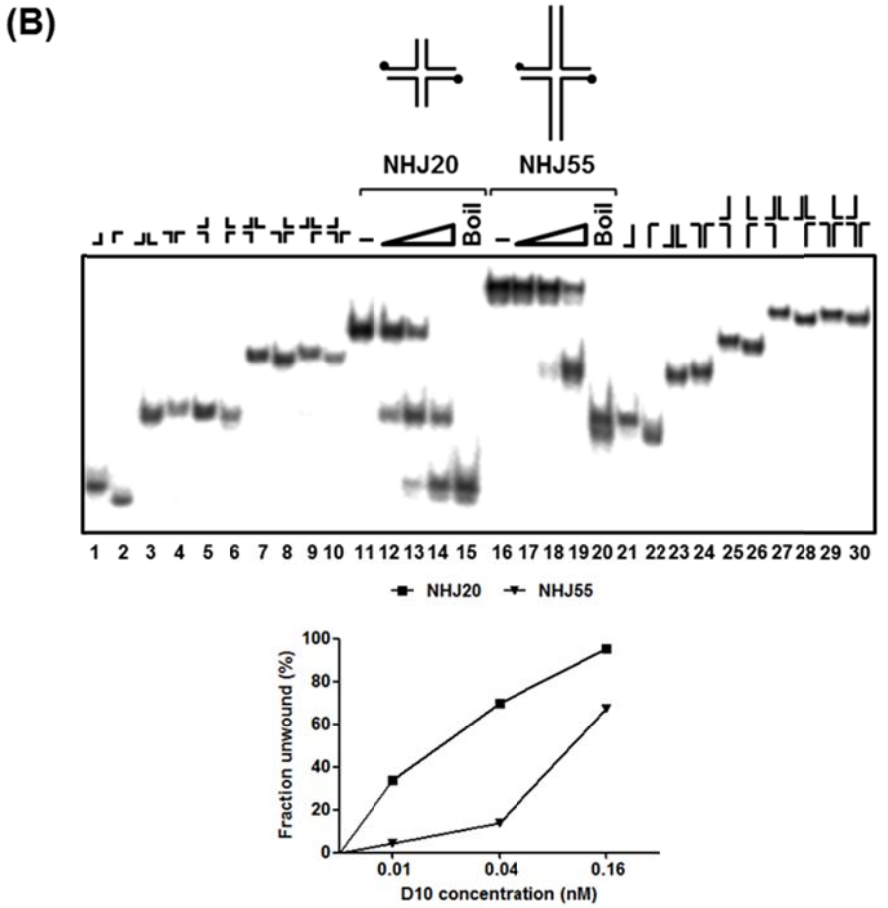

(B)
NHJ20
NHJ55
–
Boil
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
NHJ20
NHJ55
Fraction unwound (%)
100
80
60
40
20
0
0.01
0.04
0.16
D10 concentration (nM)

Figure 5.8 Bacteriophage T5 D10 unwinding of forked and non-homologous Holliday junction DNA. The constituent oligonucleotides of the six substrates are illustrated in Appendix A. All substrates (0.1 nM) were radiolabelled at their 5' ends as indicated by black circles. Reactions (0.1 nM ^{32}P-5'-end-labelled DNA substrate, 0.01, 0.04 and 0.16 nM D10) were performed at 37°C for 20 min and products were separated on polyacrylamide gels, visualized and quantified by phosphorimaging. **(A)** Unwinding of Y-shaped DNA molecules. Lanes 1, 6, 14 and 19 show native substrates/no protein (-) and lanes 5, 10, 18 and 23 are denatured (boiled) controls. Reactions for Fork20, lanes 2-4; Fork55, lanes 7-9; Y20, lanes 15-17 and Y55, lanes 20-22. Lanes 11-13 and 24-26 are markers for possible products. The bottom line graph shows quantitative data for D10 unwinding (n=3, mean and SEM). **(B)** Unwinding of cruciform DNA structures. Lanes 11-15 show unwinding of NHJ20 and controls, similarly, lanes 16-20 show unwinding of NHJ55. Lanes 1-10 and 21-30 are marker lanes for all possible products. Data are quantified in the graph at the bottom.

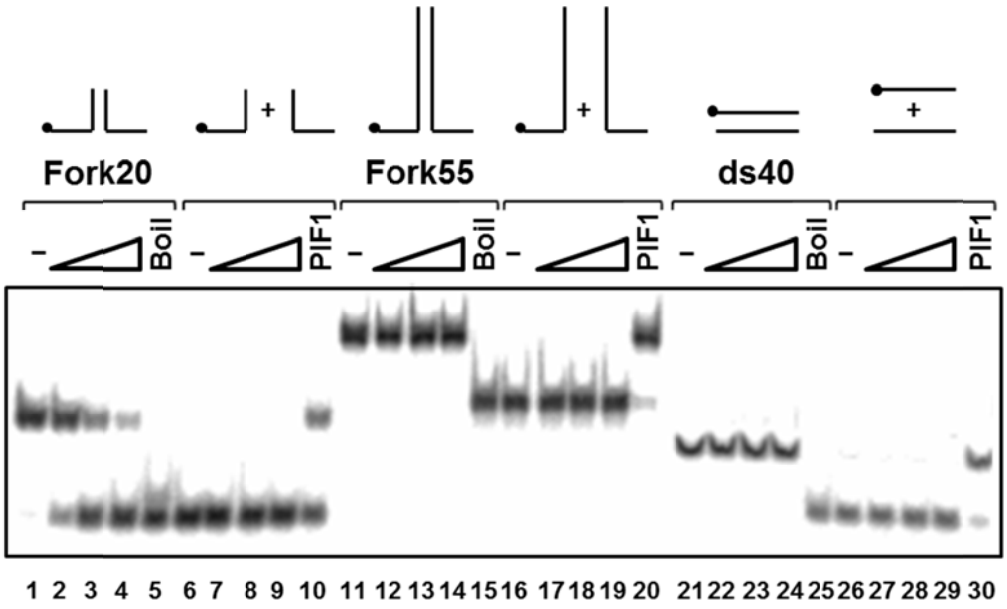


Figure 5.9 Assay for strand annealing activity of D10. Lanes 1-5 demonstrate the unwinding of substrate Fork20 (0.1 nM Fork20; 0.01, 0.04 and 0.16 nM D10; 37°C for 20 min). In lanes 6-10, the Fork20 substrate was heat denatured before addition to the reactions (0.1 nM). D10 did not induce strand annealing (lanes 7-9, 0.01, 0.04 and 0.16 nM), but strand annealing was catalysed by the human PIF1 helicase (lane 10). Lanes 11-20 and lanes 21-30 as 1-10 except using substrates Fork55 with the longer complementary arms and ds40 without unpaired tails, respectively.

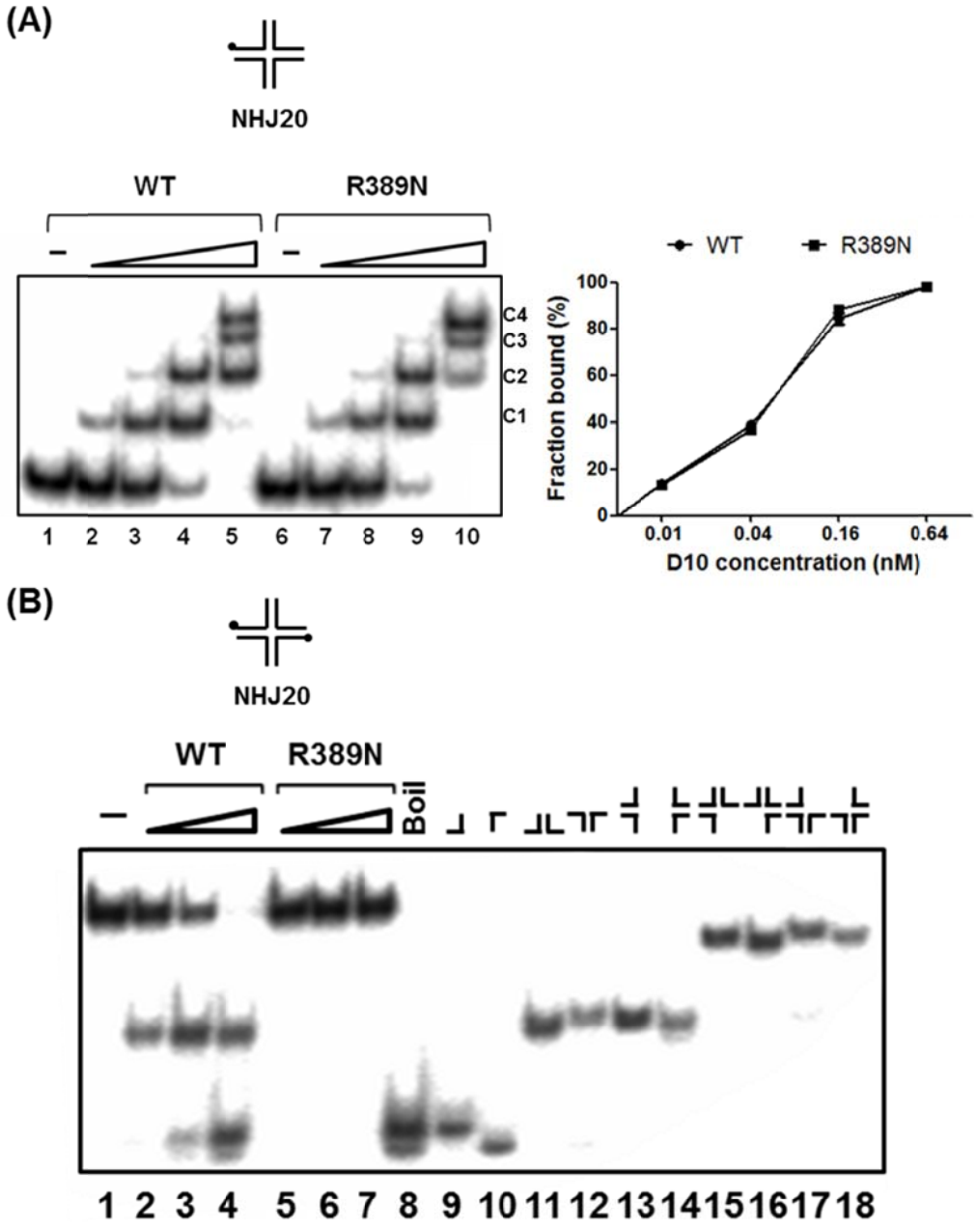

(A)
NHJ20
WT
R389N
C4
C3
C2
C1
1 2 3 4 5 6 7 8 9 10
WT
R389N
Fraction bound (%)
D10 concentration (nM)
(B)
NHJ20
WT
R389N
Boil
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
(C)
WT
R389N
Turnover (S-1)
No DNA
NHJ20
No DNA
NHJ20

Figure 5.10 Helicase, ATP*ase* and DNA-binding activities of D10 mutant (R389N) using the non-homologous Holliday junction substrate NHJ20, compared with D10 wild-type (WT). **(A)** D10 R389N had similar DNA binding pattern (4 predominant species, C1-C4) to that of D10 (WT), determined using 0.1 nM ^{32}P-5'-end-labelled NHJ20. Lanes 1 and 6, no protein control; lanes 2-5, D10 WT (0.01, 0.04, 0.16 and 0.64 nM); lanes 7-10, D10 R389N (0.01, 0.04, 0.16 and 0.64 nM). Helicase activity **(B)** and ATP*ase* activity **(C)** were completely abolished by the R389N point mutation in the Arginine finger motif. In panel **B**, the helicase reactions contained 0.1 nM substrate and the indicated proteins. Lanes 9-18 are labelled DNA markers for all possible products. Lane 1, no protein control; lane 8, heat-denatured substrate (Boil); lanes 2-4, 0.01, 0.04 and 0.16 nM D10 WT; lanes 5-7, with D10 R389N, 0.01, 0.04 and 0.16 nM. The ATP*ase* assays were performed as described in Materials and Methods, except the NaCl concentration in the reaction buffer was 20 mM NaCl and reactions were incubated at 37°C for 20 min. (n=3)

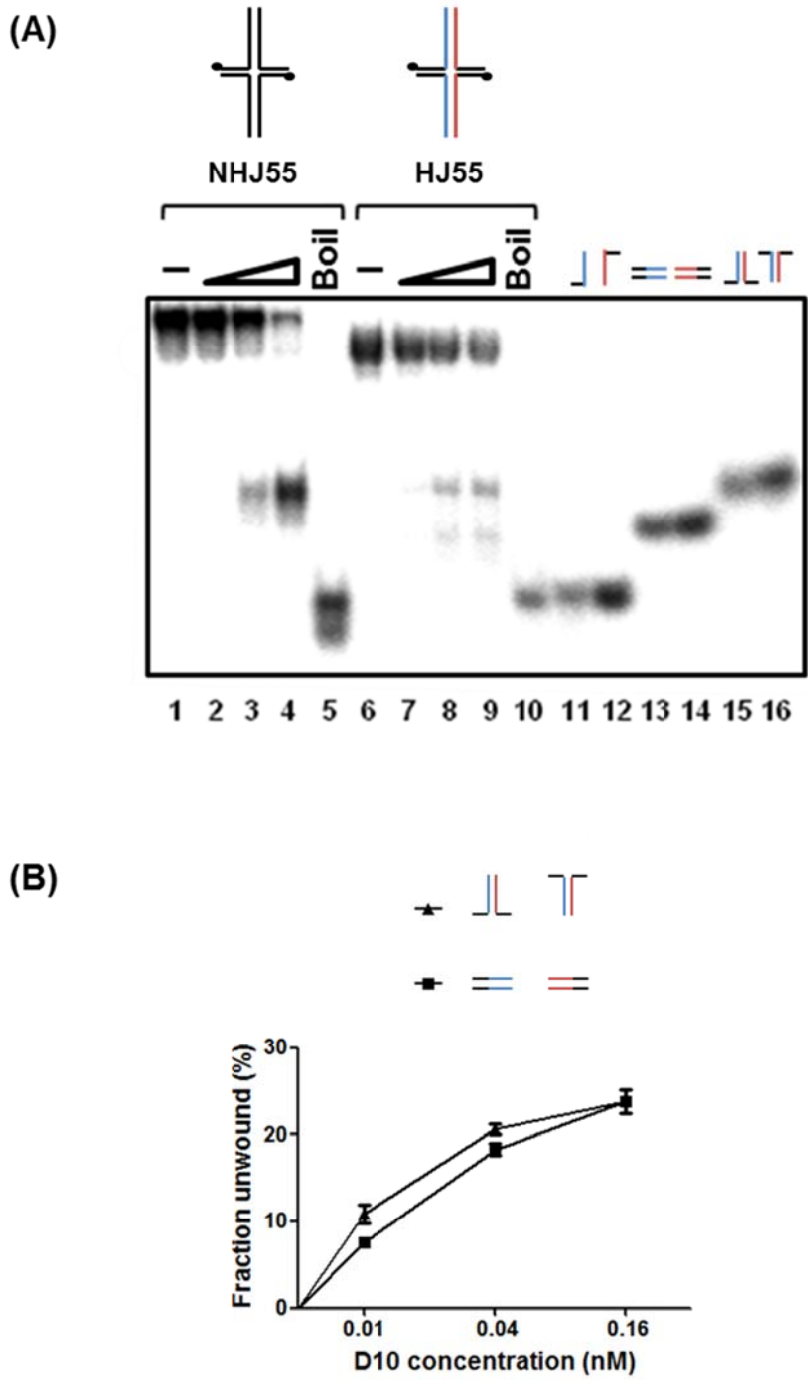


Figure 5.11 D10 unwinding of non-homologous and homologous Holliday junctions. (**A**) The component oligonucleotides of the non-homologous (NHJ55) and homologous (HJ55) Holliday junctions are shown in Appendix A. The homologous portions of HJ55 are coloured in red and blue. All the reactions were performed as above. In contrast to NHJ55 (lanes 1-5), D10 unwinding of HJ55 (lanes 6-10), resulted in two kinds of products, fork DNA and recombined duplex DNA. Lanes 1 and 6, no protein control (-); lanes 5 and 10, heat-denatured substrate control (Boil). Lanes 11-16 are markers for derivatives of substrate HJ55. The bottom graph (**B**) shows statistical data for unwinding of substrate HJ55, giving the percentage of each product class (duplex or fork) that D10 generated.

5.1.4 Binding of D10 to linear and branched DNA substrates

As described above, the ATP*ase* and helicase activities of D10 were tested using a variety of DNA substrates, including ssDNA, dsDNA, branched DNA substrates. The binding of D2 to these DNA substrates was examined using gel-shift assays with a series of DNA substrates (T20, 20-base dT ssDNA; ds20, 20-bp dsDNA; 5'T20-ds20 and ds20-3'T20, 20-bp dsDNA with either a 5' or 3' 20-base dT overhang; Fork20, a 20-bp duplex with two non-complementary 20-base ssDNA arms; Y20, a Y-junction with three 20-bp duplex arms; NHJ20, a Holliday junction with four 20-bp duplex arms) in the absence of ATP/Mg^2 (Figure 5.12A).

D10 was unable to bind T20 (lanes 1-5), ds20 (lanes 6-10) and 5'T20-ds20 (lanes 11-15) over a range of D10 concentrations (0.01-0.64 nM), correlating with its inability to unwind these substrates (Figure 5.12B, lanes 1-10). However, although D10 could not unwind 3'T20-ds20 (Figure 5.12B, lanes 6-15), this substrate was bound efficiently (Figure 5.12A, lanes 16-20), possibly reflecting the inherent translocation/unwinding polarity of D10. Compared to the linear DNA substrates, there is a significant increase in binding affinity for the more complex branched substrates such as, Fork20, Y20 and NHJ20 (Figure 5.12A, lanes 21-35 and graphed data). Also, at higher D10 protein concentrations, two, three and four protein-DNA complexes were observed with Fork20, Y20 and NHJ20 respectively. Although the composition of the unwinding complex of D10 with each of these three substrates remains unclear, the ability of D10 to form different numbers of complexes on these substrates may implicate the initiation of D10 helicase activity on these three substrates probably requires different protein stoichiometries.

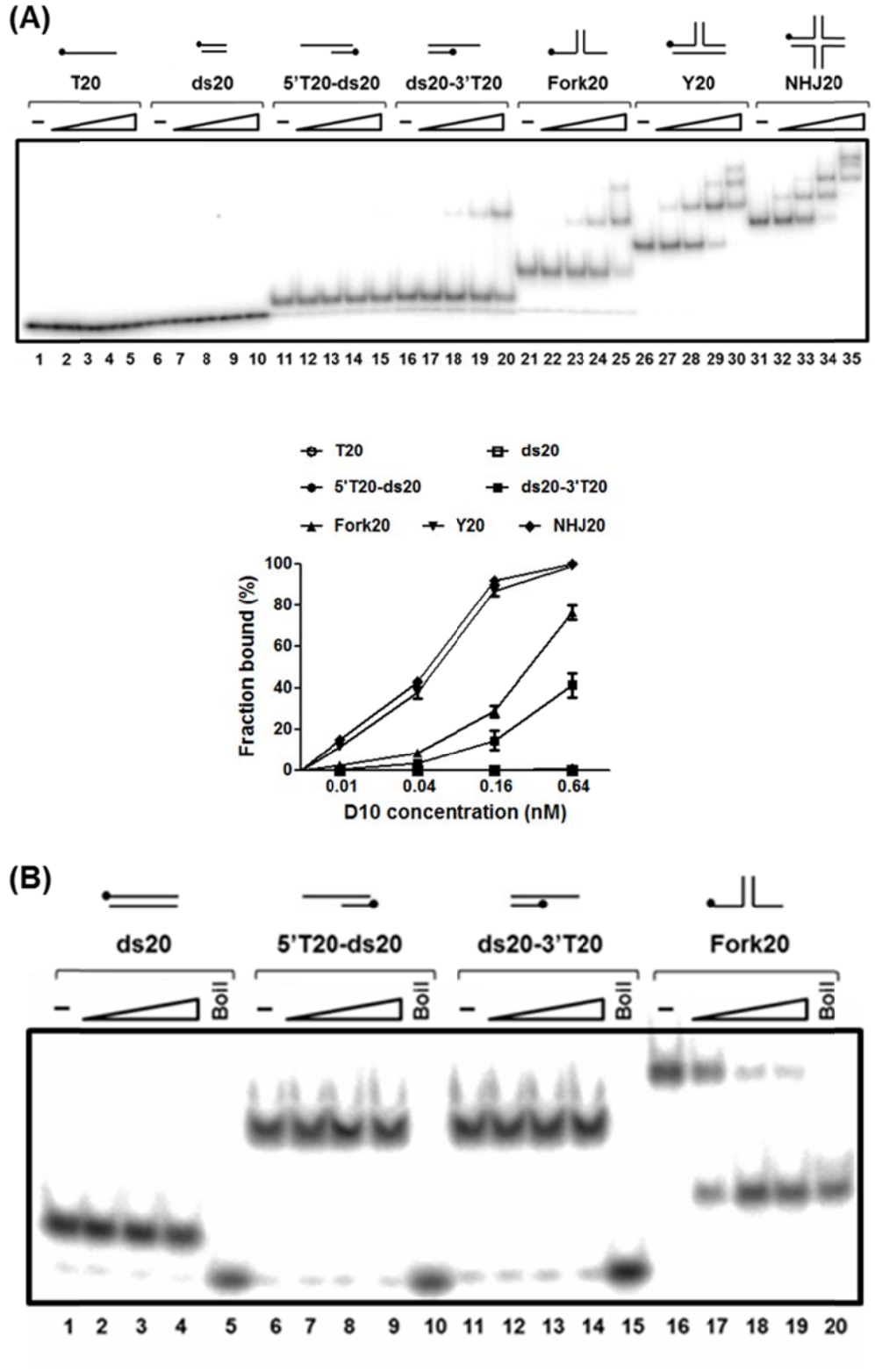


Figure 5.12 D10 unwinding and binding activities of various DNA substrates, including a 20-base dT ssDNA, T20; a 20-bp blunt-ended dsDNA, ds20; 20-bp duplexes with a 5' or 3' 20-base dT overhang, 5'T20-ds20 and ds20-3'T20; a 20-bp duplex with two non-complementary 20-base ssDNA arms, Fork20; a dsDNA Y-junction with three 20-bp duplex arms, Y20; a non-homologous Holliday junction with four 20-bp duplex arms, NHJ20.

Notes, the 20 base-pair duplex of ds20, 5'T20-ds20, ds20-3'T20 is identical and also a component of Y20 and NHJ20. (**A**) Gel-shift assay. Lanes 1, 6, 11, 16, 21, 26 and 31, no protein control (-); lanes 2-5, lanes 7-10, lanes 12-15, lanes 17-20, lanes 22-25, lanes 27-30 and lanes 32-35, with D10 protein titrations (0.01, 0.04, 0.16 and 0.64 nM). (**B**) Helicase assay. Lanes 1, 6, 11 and 16, no protein control (-); lanes 5, 10, 15 and 20, heat-denatured substrate control (Boil); lanes 2-4, Lanes 7-9, lanes 12-14 and lanes 17-19, D10 protein titrations (0.01, 0.04, 0.16 nM, respectively). (n=3)

5.1.5 DNA sequence context and unwinding activity of D10

The data described above identified three- and four-way branched structures (with ss- and dsDNA arms) as substrates for the DNA unwinding action of D10 and indicated that D10 has branch migration activity. However, these data also showed that D10 unwinding action is influenced by additional parameters other than DNA structure and length. For example, D10 could completely dissociate the substrate Fork20 (with a 20 bp duplex and two 20-base ssDNA arms, Figure 5.8A). It could also displace a 40 base oligonucleotide from substrate Y20 (with three 20 bp duplex DNA arms, Figure 5.8A), but the co-product that resembles Fork20 was only poorly unwound, even during an extended time course (Figure 5.14B). Furthermore, the relative mobility of the unwinding products indicated that there was a very strong bias as to which 40 mer oligonucleotide was displaced from the substrate (Figure 5.8A, lanes 11-13 compared to 15-17), which was unambiguously identified as shown in Figure 5.14A. Similar observations were made for substrate NHJ20, where there was a definite bias for unwinding the cruciform along one DNA axis (Figure 5.8B, lanes 11-15). A probable explanation for these data is that DNA sequence can influence D10 unwinding activity.

To address further the structure and sequence specificity of DNA unwinding by D10, the substrate Y20 was compared to dsDNA Y-junctions formed with the same sequence but with a nick at one of the three junction branch points, Y20-1 (oligo C nicked), Y20-2 (oligo B nicked) and Y20-3 (oligo A nicked, Figure

5.13A). Since Y20 has trilateral (three-fold rotational) symmetry, Y20-1, Y20-2 and Y20-3 should have the same structure, but differ only in the position of the nick relative to DNA sequence. Furthermore, these structures resemble stalled DNA replication forks, or intermediates that could arise during DNA recombination and repair.

Based on the ability of D10 to unwind Y20, one might expect all the nicked Y-junctions to be unwound to similar extents. Surprisingly, however, as shown in Figure 5.13B, compared to Y20 (lanes 1-5), only Y20-3 (oligo C nicked) was unwound efficiently (lanes 16-20). As summarised in the graph to the right, all of substrates were practically unwound at the lowest protein concentration tested, but the extents of unwinding of Y20-1 (lanes 6-10) and Y20-2 (lanes 11-15) were at least 20 fold lower than those of Y20 and Y20-3. The reaction products of Y20 and Y20-3 were also both consistent with displacement of the sequences corresponding to oligonucleotide A: two oligonucleotides (a1 and a2) in the case of substrate Y20-3. However, although poorly unwound, the reaction products of Y20-1 and Y20-2 were not consistent with preferential displacement of this sequence (Figure 5.14C).

To further explore the basis of this substrate-specific D10 unwinding, various Y-junctions were tested to see whether they differed in their binding affinity for D10 and ability to stimulate the DNA-dependent ATP*ase* activity. As shown in Figure 5.13C, there was little detectable difference in the binding affinities of D10 for substrates Y20 (lanes 1-5), Y20-1 (lanes 6-10), Y20-2 (lanes 11-15) and Y20-3 (lanes 16-20), as measured by gel-shift assay in the absence of ATP/Mg^{2+}. However, the binding pattern was characterised by the presence of three predominant species (C1, C2 and C3) and subtle differences were noted in the relative extents of each that formed with the different substrates. Furthermore, all Y-junctions with and without nicks stimulated the DNA-dependent ATP*ase* activity of D10 to similar extents (Figure 5.13D). Thus, as well as sequence, the

ability of D10 to initiate unwinding appears to be strongly influenced by other stereochemical features of the substrate, such as DNA strand discontinuity, relative to sequence.

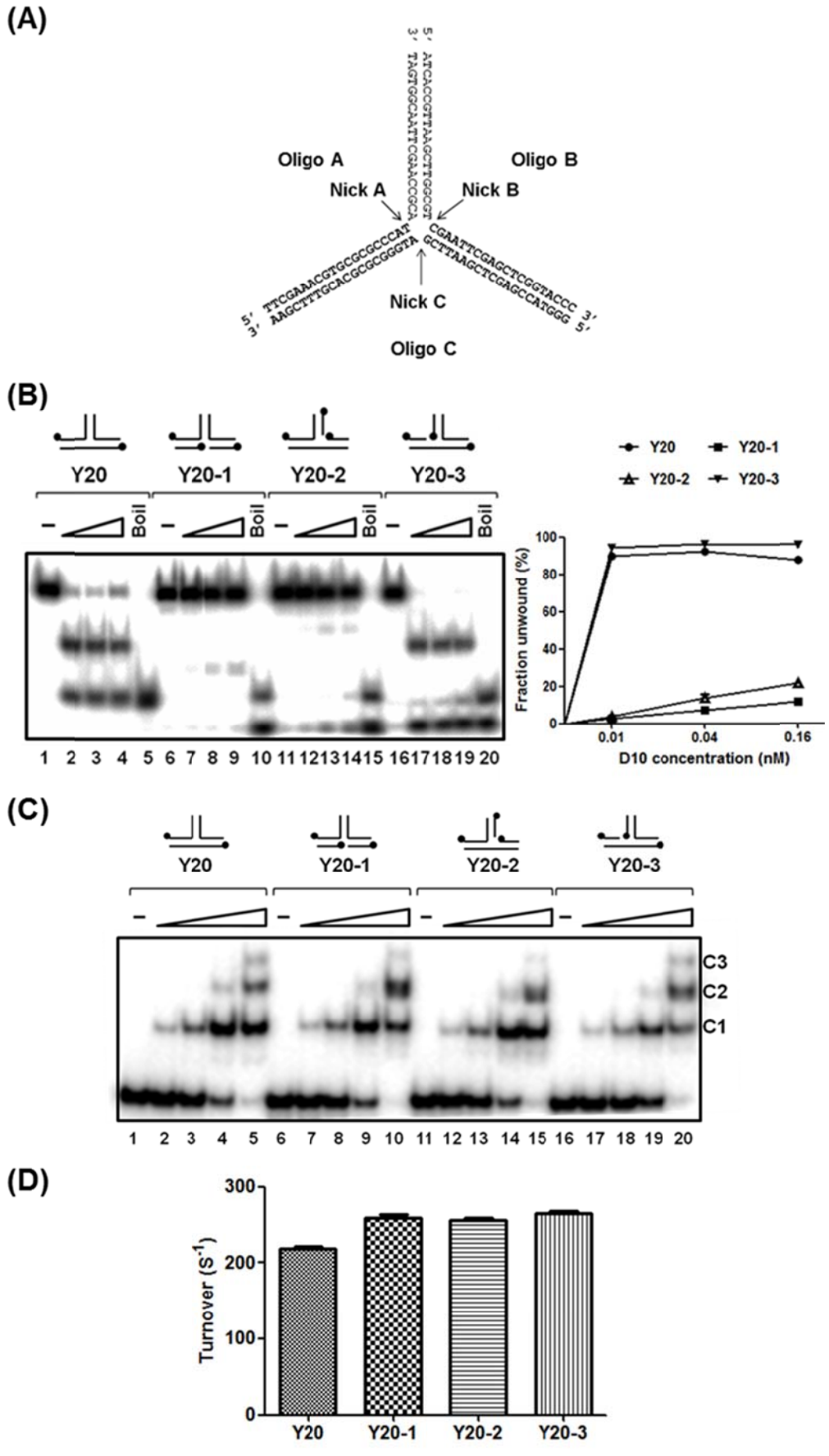

(A)
Oligo A
Oligo B
Nick A
Nick B
Nick C
Oligo C
(B)
Y20
Y20-1
Y20-2
Y20-3
Boil
Fraction unwound (%)
D10 concentration (nM)
(C)
C3
C2
C1
(D)
Turnover (S-1)

Figure 5.13 Helicase, DNA-binding and ATP*ase* activities of D10 on nicked DNA Y-junction substrates. **(A)** Sequence and nick position in the oligonucleotide substrates used. (**B**) Helicase reactions were performed as described above. Lanes 1, 6, 11 and 16, no protein control (-); lanes 5, 10, 15 and 20, boiled substrates (Boil); lanes 2-4, lanes 7-9, lanes 12-14 and lanes 17-19, substrates Y20, Y20-1, Y20-2 and Y20-3 with 0.01, 0.04, 0.16 nM D10, respectively. Statistical data (fraction unwound) are shown in the graph on the right. (**C**) Binding of D10 to Y-junction substrates. Binding assays (0.1 nM probe) were performed without ATP/Mg^{2+}, as described in Materials and Methods. D10 had almost identical binding affinity for all four substrates and three predominant DNA-protein complexes (C1, C2 and C3**)** formed. Lanes 1, 6, 11 and 16, no protein control (-), lanes 2-5 (Y20), lanes 7-10 (Y20-1), lanes 12-15 (Y20-2) and lanes 17-20 (Y20-3), with 0.01, 0.04, 0.16 and 0.64 nM D10. (**D**) D10 ATP*ase* activity stimulated by the four Y-junctions substrates. Reactions were performed as described in Materials and Methods, except the NaCl concentration in the reaction buffer was 20 mM NaCl and reactions were incubated at 37°C for 20 min. There was no significant difference in stimulation of ATP*ase* activity among the four substrates.

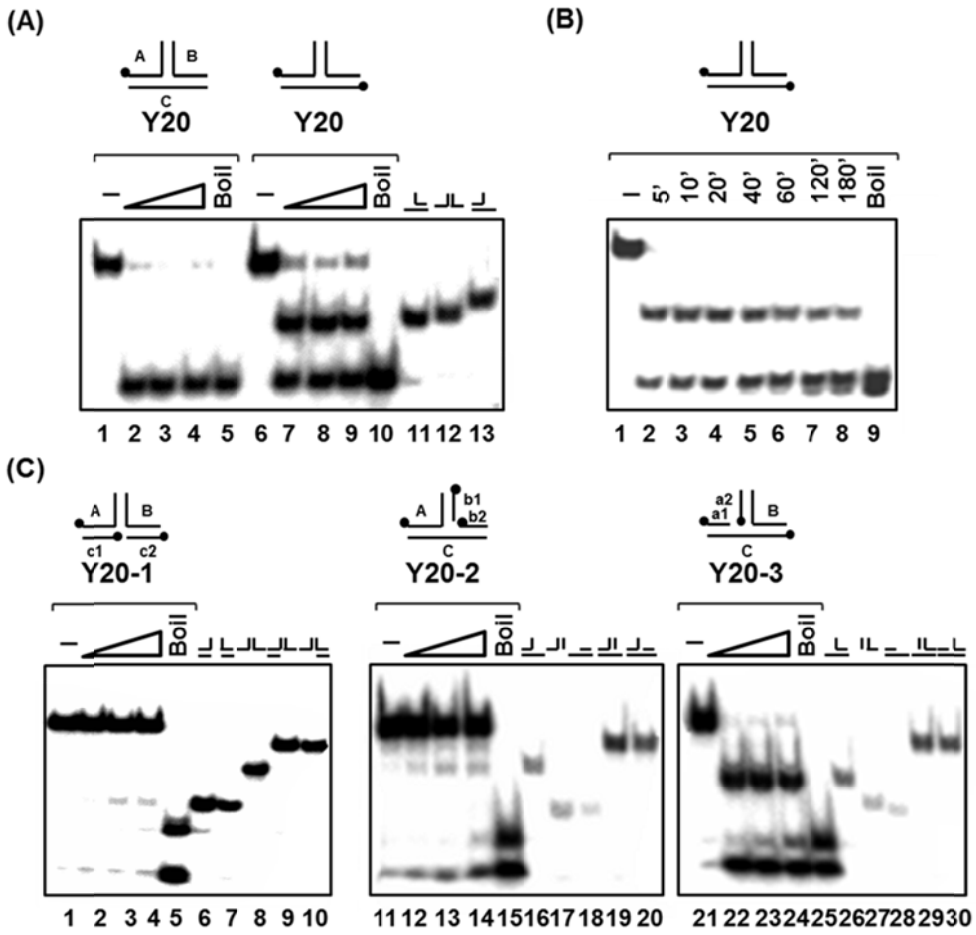


Figure 5.14 (A) Unwinding of Y20 by D10. The helicase reactions containing Y20 radiolabelled on strand A or two strands (A and C), were performed as in Figure 5.13 (20 minutes' incubation). Lanes 1 and 6, no protein control (-); lanes 5 and 10, heat-denatured

substrate control (Boil); lanes 2-4 and lanes 7-9, with D10 (0.01, 0.04 and 0.16 nM); lanes 11-13, markers for possible products. This unambiguously demonstrated that oligonucleotide A is preferentially displaced from substrate Y20. **(B)** Time-course experiment for D10 unwinding of Y20 with ^{32}P end-labelled oligonucleotides A and C. Lane 1, no protein control (-); lane 9, heat-denatured substrate control (Boil); lanes 2-8, time course of unwinding with 0.04 nM D10 (5, 10, 20, 40, 60, 120 and 180 min). (**C**) Unwinding of the nicked Y20 substrates Y20-1 (oligo C nicked), Y20-2 (oligo B nicked) and Y20-3 (oligo A nicked) by D10. Lanes 1, 11 and 21, no protein control (-); lanes 5, 15 and 25, heat-denatured substrate control (Boil); lanes 2-4, lanes 12-14 and lanes 22-24, with D10 (0.01, 0.04 and 0.16 nM); lanes 6-10, 16-20 and 26-30, markers for possible products. (n=3)

5.1.6 Structure and sequence substrate specificity of D10 unwinding

The above-mentioned work revealed the sequence and structure features of the substrate can affect D10 unwinding activity and the structure feature alone does not determine if a substrate is unwound by D10. To further understand whether the initiation of unwinding by D10 can be determined only by the sequence feature and strand discontinuity of the substrate. The nicked Y-junction substrates (Y20-1 and Y20-3) and corresponding linear duplex substrates with or without a nick (N20-1, ds40, formed with the sequence of oligonucleotide C (or c1 and c2) which D10 does not prefer to displace from Y20-1 or Y20; N20-3 and ds40', formed with the sequence of oligonucleotide A (or a1 and a2) which D10 prefers to displace from Y20-3 or Y20) was examined in detail (Figure 5.15).

Y20-1 was poorly unwound by D10 (lanes 5-9), whereas Y20-3 was efficiently unwound (lanes 24-28), as observed above. Moreover, D10 was unable to unwind the duplex linear DNA substrates with or without a nick or a specific sequence (lanes 10-19 and lanes 29-38). In this case, both of the sequence and structure of the substrate are required for determining the initiation of D10 unwinding, suggesting substrate specificity of D10 unwinding activity is both structure-dependent and sequence-dependent.

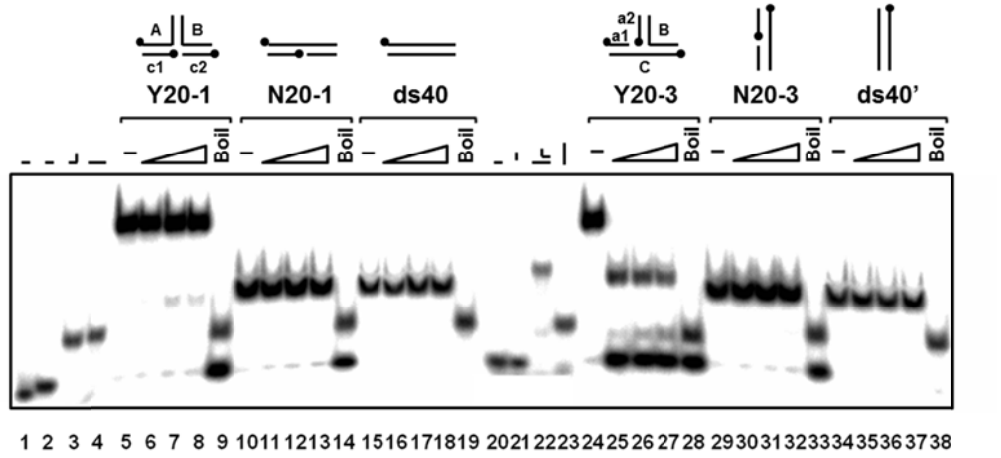


Figure 5.15 D10 unwinding of nicked duplex Y substrates (Y20-1 and Y20-3) and corresponding linear duplex substrates with or without a nick or a specific sequence (N20-1, ds40, N20-3 and ds40'). D10 was not able to unwind the linear duplex DNA substrates with or without a nick or a specific sequence, showing both of structure and sequence of the substrate contribute to determine the initiation of unwinding by D10. Lanes 5, 10, 15, 24, 29 and 34, no protein control (-); lanes 9, 14, 19, 28, 33 and 38, heat-denatured substrate control (Boil); lanes 6-8, lanes 11-13, lanes 16-18, lanes 25-27, lanes 30-32 and lanes 35-37, with D10 protein (0.01, 0.04 and 0.16 nM, respectively). Lanes 1-4 and lanes 20-23 are markers for possible products. (n=3)

5.1.7 High-resolution footprinting of D10 bound to a dsDNA Y-junction

Since the data described above indicated that sequence context can greatly influence unwinding of duplex Y-junctions by D10, the binding of D10 to such substrates was further investigated using a high-resolution hydroxyl radical (OH) footprinting assay. The OH radical is a small diffusible cleavage reagent that cuts with little inherent sequence bias. To determine the disposition of D10 bound to Y20, three substrates each with a different strand labelled were used in D10 binding reactions. The OH radical was generated by Fenton's reaction and after limited DNA cleavage, the DNA products were recovered and analyzed on a sequencing gel. As shown in Figure 5.16, D10 protected the 23-30 bases centred approximately around the junction point in each strand of Y20, as summarised in

Figure 5.16C. An analysis of the protection pattern with increasing protein concentration (Figure 5.16B) suggested that protein binding to the Y-junction was progressive, without demonstrating a significant bias for any strand or arm of the substrate. However, although the Y-junction substrate has trilateral symmetry, the protection pattern did not display such symmetry, with the protection of oligonucleotide C more extensive than that of A or B, as summarised in Figure 5.16C.

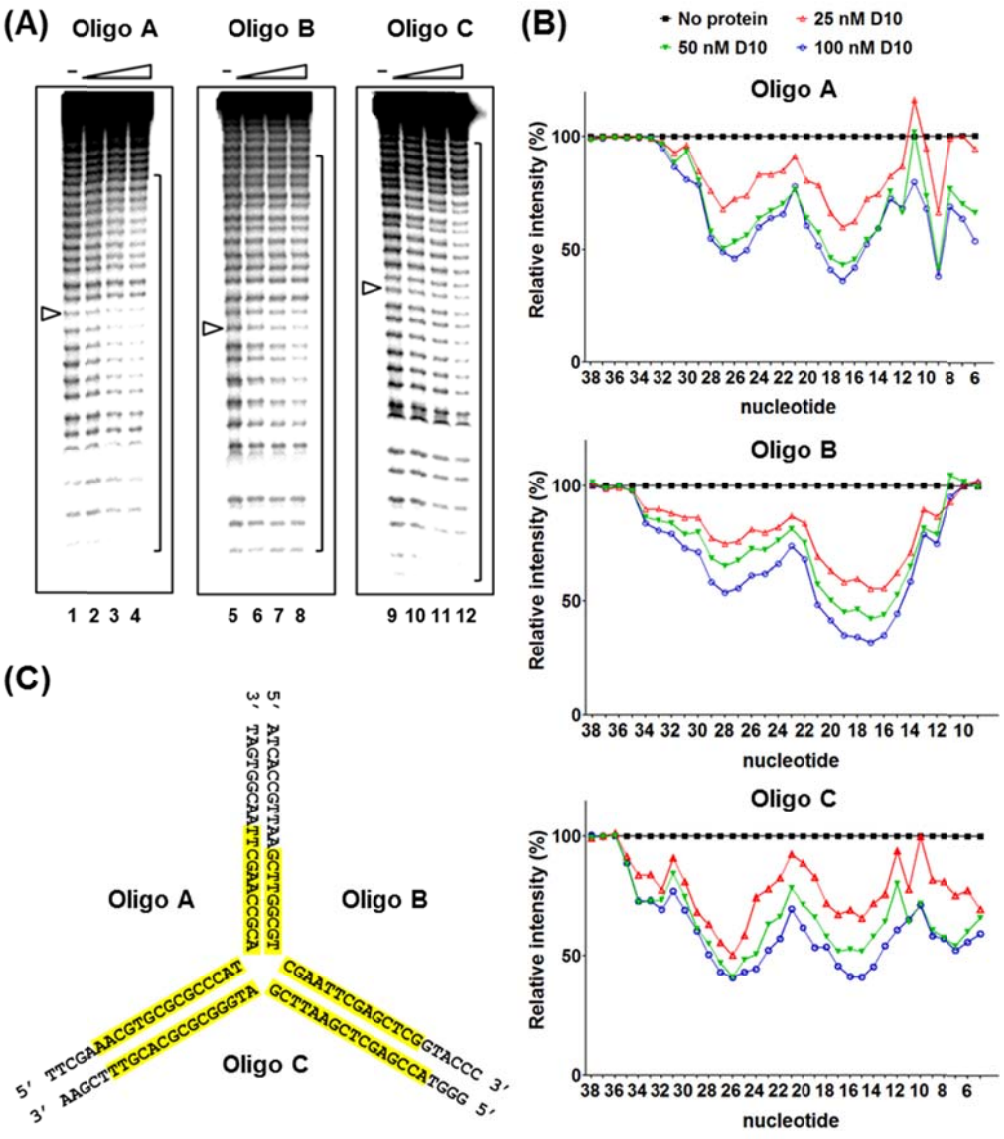


Figure 5.16 Hydroxyl radical footprinting of the D10-Y20 DNA complex in solution. **(A)** Three substrates were prepared by 5' ^{32}P end-labelling either oligonucleotide A, B or C of the Y20 substrate. Binding reactions (10 nM substrate; 0, 25, 50, 100 nM D10, lanes 1-4, 5-8 and 9-12 resepectively) were treated with the OH radicals and processed for analysis on a 15% sequencing gel. The 20th nucleotides at the branch point of each strand are indicated by arrows, and were determined from the mobility of the corresponding 20 mer oligonucleotide run on the same gel. **(B)** Regions of protection were determined by relative densitometric intensity of each band compared to the reaction without D10 as the datum (100%). The densitomertic tracings were generated from the data representing different D10 concentrations (no protein, black; 25 nM, red; 50 nM, green; 100 nM, blue); more than 5% change was scored as a nucleotide protected by D10. **(C)** The regions of protection (shaded in yellow) are schematically represented on the Y20 sequence. (n=2)

5.1.8 DEPC interference analysis of D10-Y20 DNA unwinding reactions

Diethyl pyrocarbonate (DEPC) carbethoxylates the N7 of nitrogen of purine bases (Appendix B) and the presence of this bulky adduct can interfere with any intimate protein-DNA base contacts required for substrate unwinding. Y20 substrates were generated with one strand end-labelled with γ-^{32}P and modified with DEPC such that there is very much fewer than one modification per molecule. To generate DEPC interference data for the unwinding reaction, helicase assays were set up so that only a small fraction of the substrate (~17%) was unwound. After electrophoresis, the unwound DNA (U) and the substrate that remained annealed in the helicase reaction (not unwound, NU) were both recovered, as was the total DNA from a control reaction that received no protein. After cleavage with piperidine, reaction products for each strand were analysed on a sequencing gel and quantified using a phosphorimager.

Several purine residues on each strand inhibited unwinding when carbethoxylated by DEPC and some residues promoted unwinding when modified (enhancement of activity). In Figure 5.17A, the positions and relative magnitude of interference are indicated with black circles. For the most part, these are contained within the regions that were protected in the hydroxyl radical footprinting assays (Figure 5.16C). Interestingly, the quantification of the data in

Figure 5.17B revealed that the number of purines and the magnitude of the resulting effects of DEPC modification are far greater for oligonucleotides B and C, compared to the displaced oligonucleotide A. Also, three purines at, or close to, the junction point of oligonucleotide A (indicated with arrows in figure 5.17C) did not alter the efficiency of the unwinding reaction when carbethoxylated, while the purines in similar positions in oligonucleotide B and C showed significant effects on unwinding when modified. These data imply that purine base contacts in oligonucleotide B and C, particularly those close to the junction point, influence the displacement of oligonucleotide A from the substrate in preference to oligonucleotides B and C. There is therefore a correlation with the data in Figure 5.13, where nicking of oligonucleotides B or C at the junction point inhibited the displacement of oligonucleotide A.

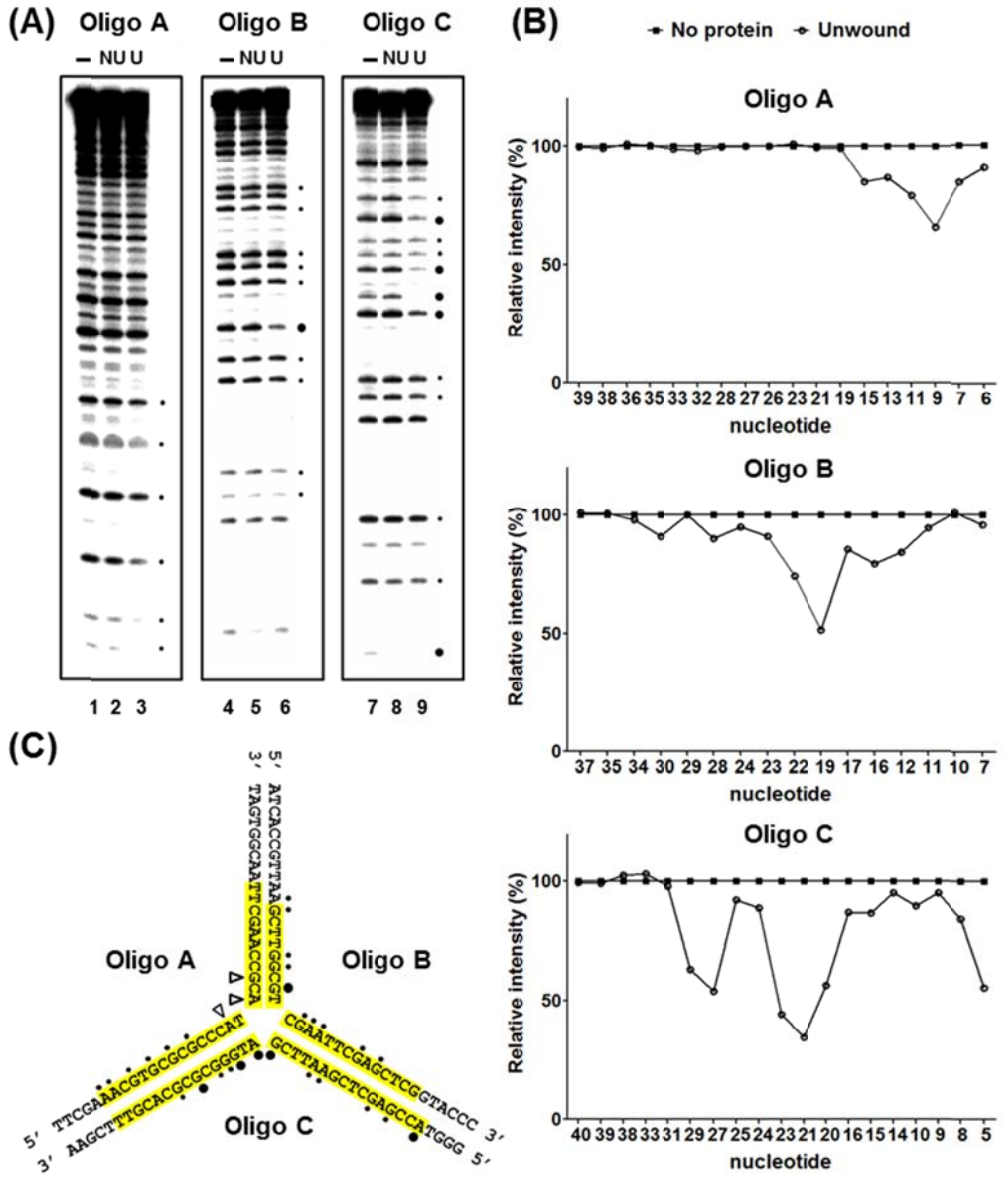


Figure 5.17 DEPC interference analysis of the unwinding of Y20 by D10. **(A)** The helicase reactions were performed with D10 and each of the three substrates ^{32}P end-labelled and DEPC-modified on either strand A, B or C. Reaction products were resolved on acrylamide gels and the desired DNA bands were recovered, cleaved with piperidine, and the products analysed on a 15% sequencing gel. The unwound DNA (U) and the substrate that remained annealed in the helicase reaction (not unwound, NU) were both recovered, as was the total DNA from a control reaction that received no protein (-). **(B)** The positions of modification-interference of unwinding were identified by measuring the changes in band relative intensity in densitometric tracings between the substrate and the unwound product. **(C)** Summary of the DEPC interference assay results. The OH protection (Figure 5.16) is shadowed in yellow and the positions of modification-interference within the Y20 substrate are

indicated (Black dots, interference; small symbol, less than 40% change; large symbol, 40-70% change). (n=2)

5.1.9 Nucleotides near the junction point and unwinding activity of D10

The above-mentioned DEPC and nicking data indicated that DNA sequences close to the junction points would be important for the unwinding activity of D10. To address if the nucleotides near the junction point affect DNA unwinding by D10, 7 base pairs around the junction point of Y20 were selected and divided into two sections of 3 bp (referred to Section 1, positions 1-3) and 4 bp (referred to Section 2, positions 4-7) for further investigation. Two additional Y-junction substrates (Y20' and Y20''), which originated from Y20 by rotating section 1 or section 2 120º clockwise (Figure 5.18A), were generated with either one strand radiolabelled and used to determine any changes of D10 unwinding activity.

The results show that D10 could displace oligonucleotide A', B' and C' from substrate Y20' (Figures 5.18B, lanes 1-15), but it could not unwind Y20'' (lanes 27-41). Furthermore, although D10 was able to displace all three component oligonucleotides from Y20', it appeared to favour the displacement of oligonucleotide B' (lanes 7-9) compared to oligonucleotides A' (lanes 2-4) and C' (lanes 12-14). Correspondingly, only one of the two possible Fork20-like co-products containing oligonucleotide B' (a 20 bp duplex with two 20-base ssDNA arms; the expected migration positions are indicated by lanes 17-18) was seen (lanes 2-4 and 7-9; lane 17, the expected migration position) and the amount of this co-product was much less than other Fork20-like co-products containing oligonucleotide A' and C' (lanes 2-4 and 12-14; lane 16, the expected migration position).

Furthermore, Y20, Y20' and Y20'' had similar binding affinity for D10, but different ability to stimulate D10 ATP*ase* activity (Figure 5.18C and D). As indicated in Figure 5.18C, there were only minimal differences in the D10

binding pattern (three predominant species, C1-C3) and affinity for substrates Y20 (lanes 1-5), Y20' (lanes 6-10) and Y20'' (lanes 11-15) in the absence of ATP/Mg^{2+}. The Y20 and Y20' stimulated D10 ATP*ase* activity to a similar extent, but the ATP*ase* activity in the presence of Y20'' that was not unwound by D10, was 25% lower compared with Y20 and Y20' (Figure 5.18D). All these observations confirm the existence of sequence substrate specificity of D10 for initiation of unwinding. Moreover, the recognition of nucleotide sequence in the specific region (here is 7 bp around the junction point) of the substrate by D10 is important for determining whether and how D10 unwinds the substrate.

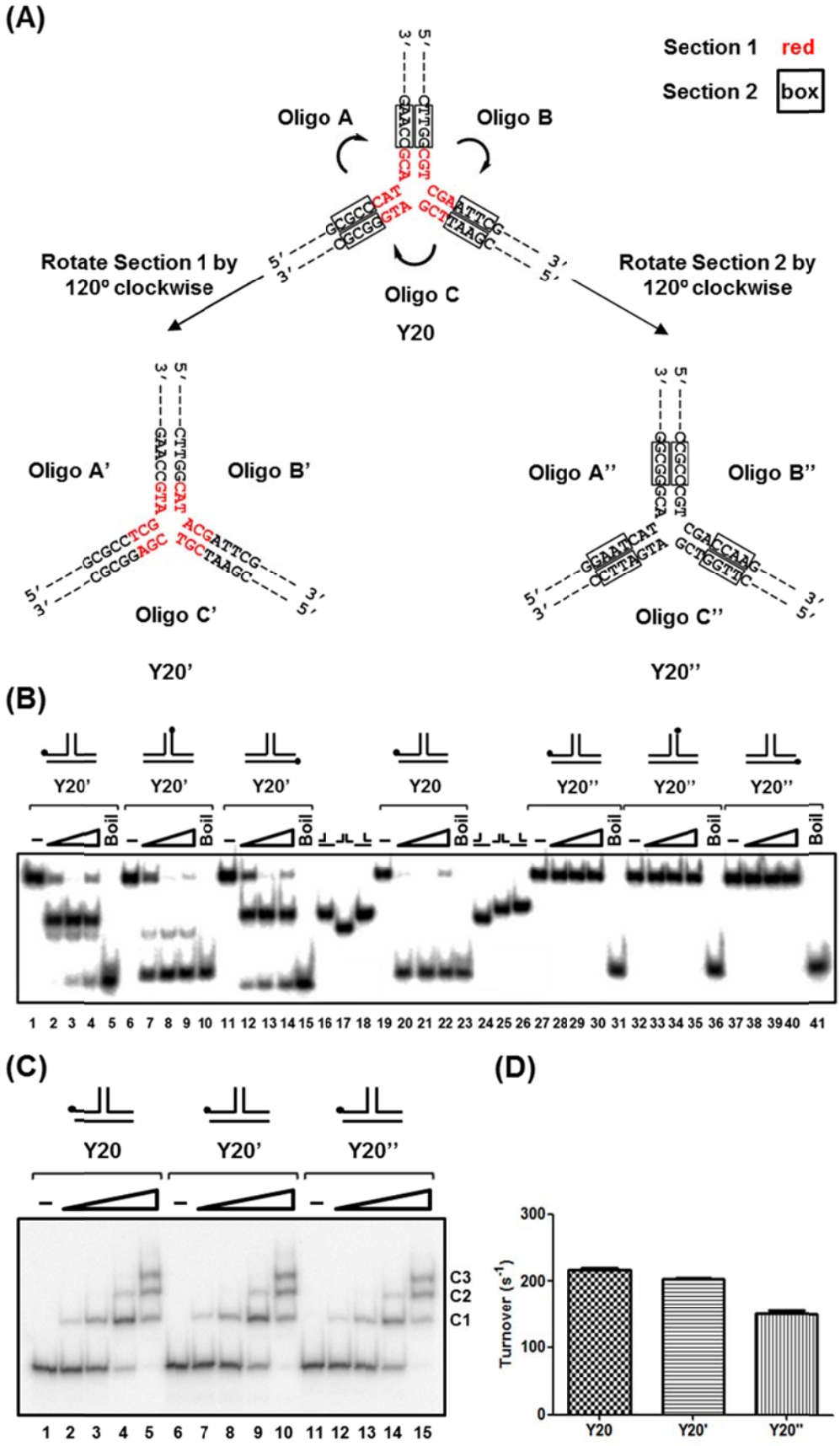

(A)
Section 1 red
Section 2 box
Oligo A
Oligo B
Oligo C
Y20
Rotate Section 1 by 120° clockwise
Rotate Section 2 by 120° clockwise
Oligo A'
Oligo B'
Oligo C'
Y20'
Oligo A''
Oligo B''
Oligo C''
Y20''
(B)
Y20'
Y20'
Y20'
Y20
Y20''
Y20''
Y20''
Boil
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41
(C)
Y20
Y20'
Y20''
C3
C2
C1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
(D)
Turnover (s-1)
300
200
100
0
Y20
Y20'
Y20''

Figure 5.18 Helicase, DNA-binding and ATP*ase* activities of D10 on Y-junction substrates with various sequences. **(A)** Y20' and Y20'' were constructed by rotating Section 1 (red; 3 bp; positions 1-3 from the junction point) and Section 2 (box; 4bp; positions 4-7) of Y20 120º clockwise. Some identical nucleotide sequences of these three substrates were omitted. **(B)** Unwinding of Y20' and Y20'' by D10. The helicase reactions containing 0.1 nM of Y20' or Y20'' radiolabelled on one of three strands were performed as described in Figure 5.13, using Y20 as a control. D10 was able to displace oligonucleotide A', B' and C' from Y20', but it showed a preference for displacing oligonucleotide B'. In stark contrast, D10 was not capable of unwinding Y20''. Lanes 1, 6, 11, 19, 27, 32 and 37, no protein control (-); lanes 5, 10, 15, 23, 31, 36 and 41, heat-denatured substrate control (Boil); lanes 2-4, lanes 7-9, lanes 12-14, lanes 20-22, lanes 28-30, lanes 33-35 and lanes 38-40, with D10 (0.01, 0.04 and 0.16 nM); lanes 16-18 and lanes 24-26, markers of possible products for substrates Y20' and Y20'' respectively. (**C**) Gel-shift assay showing D10 binding to Y20, Y20' and Y20''. The binding reactions (0.1 nM probe) were performed without ATP/Mg^{2+}, as described in Figure 5.13. D10 had similar binding affinity for all these three substrates and the binding patterns were all characterized by three predominant complexes (C1-C3). Lanes 1, 6 and 11, no protein control (-), lanes 2-5 (Y20), lanes 7-10 (Y20'), and lanes 12-15 (Y20''), with 0.01, 0.04, 0.16 and 0.64 nM D10. (**D**) ATP*ase* activity of D10 (5 nM) was determined in the presence of Y20, Y20' or Y20'' (5 nM). Reactions were performed at 37°C for 20 min as described in Figure 5.13. The Y20 and Y20' stimulated D10 ATPase activity to a similar extent, but the ATPase activity in the presence of Y20'' was 25% lower compared with Y20 and Y20'.

5.1.10 DNA sequence context and D10 ATP*ase* activity

As described in Section 5.1.2, it was revealed that D10 ATP*ase* activity is DNA-structure-dependent. However, the dependence of D10 ATP*ase* activity on the structure of the substrate cannot explain the observation that D10 ATP*ase* activity stimulated by Y20'' is lower than that stimulated by Y20 or Y20' (Figure 5.18C, Y20, Y20' and Y20'', with similar structure but different nucleotide sequence). This implicates D10 ATP*ase* activity is not only affected by the structure of the substrate, but also the sequence.

To further address the sequence effect of the substrate on D10 ATP*ase* activity, the structural complexity of the substrate was decreased. Blunt-ended dsDNA was selected as the test substrate since, relatively, it is the substrate with the simplest structure that is able to stimulate D10 ATP*ase* activity. The nucleotide sequences of these blunt-ended dsDNA molecules (20 bp and 40 bp)

derive from the duplex arms of the Y20 substrate (Figure 5.19A). Among all the 20-bp dsDNA substrates tested, there was no appreciable difference in the stimulation of D10 ATP*ase* activity by ds20 and ds20', but the ATP*ase* activity in the presence of ds20'' was significantly lower (30%) compared with ds20 and ds20' (Figure 5.19B). There may be a correlation between this data and the DEPC result of Y20 in Figure 5.17C, where D10 demonstrated more purine base contacts with the 20-bp duplex arms corresponding to ds20 and ds20' rather than that corresponding to ds20''. Furthermore, three 40-bp dsDNA substrates (ds40, ds40' and ds40'', each one is constituted by a combination of two 20-bp dsDNA substrates) stimulated D10 ATP*ase* activity to the similar extent (Figure 5.19B). This is possibly due to the fact that each 40-bp substrate has the sequence from either ds20 or ds20' that can efficiently stimulate D10 ATP*ase* activity. All these data suggested the sequence of the substrate could influence the ATP*ase* activity of D10.

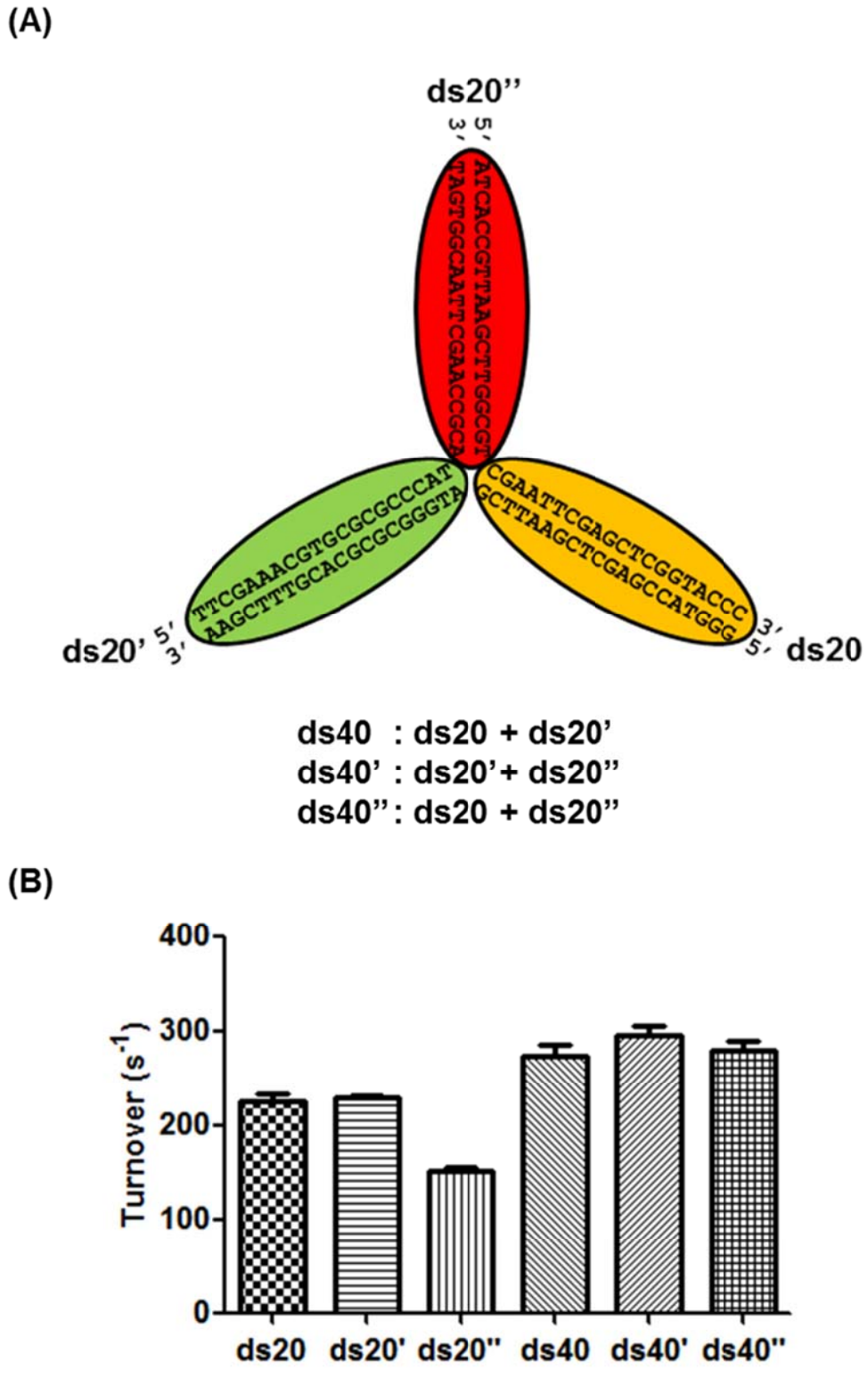


Figure 5.19 ATP*ase* activity of D10 (5 nM) was determined in the presence of various 20-mer and 40-mer blunt-ended dsDNA substrates (5 nM) – ds20, ds20', ds20'', ds40 (ds20+ds20'), ds40'(ds20'+ds20'') and ds40''(ds20+ds20''). All of them are originally from the Y20 substrate, indicated in panel **A**. The nucleotide sequences for ds20, ds20' and ds20'' are highlighted in orange, green and red ovals respectively. (n=3)

5.2 Discussion

This study shows that D10 is a DNA-dependent ATP*ase* (Figure 5.6A and Figure 5.10C) which is able to unwind branched DNA substrates *in vitro* (Figure 5.8), identifying the second helicase from bacteriophage T5. Under the conditions employed, D10 displayed no helicase activity on three linear DNA substrates, including blunt-ended dsDNA and dsDNA with either a 5' or 3' ssDNA overhang (Figure 5.7 and 5.12B). This suggests D10 is a branched-DNA specific helicase.

The bioinformatic analysis indicated that D10 shares high sequence similarity with its closely related homologue T4 UvsW (Figure 5.2). D10 showed unwinding activity on duplex Y-junction and Holliday junction DNA as UvsW (Webb et al., 2007) (Figure 5.8). However, the activities of D10 do not fully equate with those reported for UvsW. The results presented in this thesis revealed that D10 can unwind fork DNA substrates (a duplex with two non-complimentary ssDNA ends, Figure 5.8A) which UvsW was not able to unwind (Nelson and Benkovic, 2007; Webb et al., 2007). No strand annealing activity was detectable due to D10 protein using the substrates described (Figure 5.9). Notably, the helicase substrates between UvsW reports (Nelson and Benkovic, 2007; Webb et al., 2007) and this study are structurally similar, but not completely identical. For example, the free ssDNA ends of the fork substrates used in the UvsW reports contain some regions that are predicted to form stable secondary structures, which was not analysed and reported previously. Hence, further experiments are necessary to resolve the above-mentioned discrepancies between D10 and UvsW.

The ability of D10 to unwind Holliday junctions implies that D10 is responsible for branch migration of bacteriophage T5 DNA replication/repair intermediates since the Holliday junctions generally arise from recombination-dependent replication or recombinational repair. The extended dsDNA length of the Holliday junction substrate was shown to impose a restriction for D10 unwinding activity (Figure 5.8), which was also observed in a

previous study on UvsW functional homologue, RecG helicase that can catalyse branch migration of Holliday junctions in *E. coli* (Whitby and Lloyd, 1998). Because there are two possibilities to explain these data (limited unwinding processivity or non-homologous nature of the substrate), the unwinding activity of D10 was examined using a modified Holliday junction substrate, HJ55 containing long (55 bp) homologous duplex arms. It was demonstrated that D10 dissociated this substrate into two distinct products by branch migration of the junction point in either of two possible directions (Figure 5.11), as does the UvsW and RecG proteins (Webb et al., 2007; Whitby and Lloyd, 1998). Given that D10 has sequence specificity for Y-junction substrates (discussed in detail below), we cannot exclude the possibility that the sequences of the 55-bp arms could affect the unwinding activity of D10.

Furthermore, D10 helicase action on two sets of Y-junction substrates was also examined (Y junction with a nick, Figure 5.13; Y junction without a nick, Figure 5.18). The Y-junction substrates used in this study are similar to the replication fork intermediates. The helicase results with these substrates clearly indicated that not only DNA structure and length, but also sequence appears to affect the initiation of unwinding by D10. This is one interesting finding in this study and also a point of differentiation between D10 and other known helicases. The dependence of D10 unwinding activity on substrate sequence was initially observed in the unwinding of trilateral symmetrical Y20 substrate (a dsDNA Y-junction with three 20-bp duplex arms) by D10. D10 prefers to displace one component oligonucleotide from Y20 rather than the expectation of no preference to displace three oligonucleotides from Y20 (Figure 5.14). Moreover, the existence of sequence substrate specificity of D10 unwinding activity was further confirmed by the observation that unwinding activity of D10 on the two sets of Y-junction substrates with different sequence but similar structure was significantly different in terms of both patterns (oligonucleotides displaced) and extent (Y20-1, Y20-2 compared to Y20-3; each possesses a nick in one of the

three junction branch points of Y20, Figure 5.13; Y20' and Y20'' compared to Y20; Y20' and Y20'' are derived from Y20 by rotating section 1 or section 2 120º clockwise respectively, Figure 5.18). Now, it is quite clear that the nucleotide sequence and strand discontinuity of the substrate can affect the initiation of D10 helicase activity. However, they are not the only determinant, supported by the data in Figure 5.15 where D10 did not unwind the duplex substrate containing the sequence (nicked or not) of the oligonucleotide which D10 prefers to displace. All these data suggested that neither of the sequence and structure features of the substrate is dispensable to determine D10 unwinding activity, inferring substrate specificity of D10 unwinding activity is both structure-dependent and sequence-dependent.

Hydroxyl radical (OH) footprinting and DEPC interference analyses for D10 provide us with more details about D10 interaction with the Y20 substrate, showing the asymmetrical binding of D10 to Y20, where D10 showed more extensive contact with oligonucleotide C rather than oligonucleotide A or B (Figure 5.16) and more purine base contacts in oligonucleotide B and C rather than in oligonucleotide A (Figure 5.17). The biased purine base interactions (especially those close to the junction point) presumably correlate with the unwinding activity of D10. Subsequently, Y20 variants (Y20' and Y20'', Figure 5.18) were designed and generated to further investigate whether nucleotide sequences close to the junction point affect the unwinding activity of D10. Being different in that D10 showed a preference for displacing one oligonucleotide component from Y20, D10 could displace three oligonucleotides from Y20' and it could not unwind Y20'' (Figure 5.18). These data indicate that the recognition of nucleotides close to junction point (here is the region including 7 bp from the junction point) by D10 is essential for determining the initiation of unwinding by D10.

As described above, different D10 unwinding activity on the Y-junction substrates used (Y20-1 and Y20-2, compared with Y20-3; Y20' and Y20'', compared with Y20) was observed. Usually, the unwinding activity of many known helicases is correlated with their DNA binding activity (i.e. they display stronger helicase activity on the substrate which they have higher binding affinity for). It is therefore expected that D10 should have different affinity for each of the Y-junction substrates. Surprisingly, the binding patterns and extents of D10 binding to them are extensively similar (Figure 5.13C and Figure 5.18C). The similar binding to each set of Y-junction substrates with different sequence-related feature but similar structure (Y20-1, Y20-2 and Y20-3, with a nick at the junction point; Y20, Y20' and Y20'', without any nick) hints the binding activity of D10 on the substrate is principally structure-dependent and probably not affected by the sequence of the substrate.

Recently, the global structures of the duplex Y-junction DNA (with and without a nick at the junction point, 4SF and 3WJ) in solution were determined using a combination of high-resolution single-molecule fluorescence resonance energy transfer (SM-FRET) and molecular dynamics (MD) simulations. Both of the 4SF (with a nick) and 3WJ (without a nick) adopt Y-shaped structure with unpaired nucleotides (4SF, 1 bp in each arm; 3WJ, 2 bp in each arm) in the vicinity of the junction point (Figure 5.20), showing that the nick at the junction point of duplex Y junction would not hugely change the overall Y structure, but would alter the number of unpaired nucleotides near the junction point.

A high structural similarity between the duplex Y junction substrates (with and without a nick at the junction point) and no significant difference in the binding of D10 to both of two sets of Y-junction substrates (with and without a nick at the junction point, Figure 5.13C and Figure 5.18C) reinforce the proposal that D10 binding activity is structure-dependent. In general, the direct interactions with bases of DNA would be expected for a helicase that is sequence-specific.

Hence, the existence of unpaired nucleotides of duplex Y-junction DNA ignites an idea that D10 probably exerts its sequence specificity of unwinding activity by contact with the bases of these unpaired nucleotides. This idea may interpret the nicking and sequence-swapping data in Figure 5.13B and Figure 5.18B, where the alternations in the nucleotide sequence near the junction point lead to considerable changes in D10 unwinding activity. However, there is one unexplained exception that Y20'' could not be unwound by D10 since the modified nucleotides are outside the potential region of unpaired nucleotides (2 bp per arm), inferring that the nucleotides of the paired duplex arms may be transiently unpaired (spontaneously or induced by D10) or D10 may directly interact with the minor groove and/or major groove of duplex DNA to recognize the basepairs of the paired duplex arms.

All the above-mentioned data using Y-junction substrates indicated that substrate specificity of D10 unwinding activity is both structure-dependent and sequence-dependent, but the binding of D10 to the substrate is principally structure-dependent. Based on these findings, a model is proposed for the unusual substrate specificity of D10 helicase (Figure 5.21). Initially, D10 would select substrates for binding on the basis of its affinity for specific DNA structures (Figure 5.12). Upon binding, D10 would align its (unidentified) sequence recognition units with the relative region of the bound substrate. The recognition units would determine whether the substrate has the specific nucleotide sequence necessary to promote the initiation of unwinding. However, the results also showed that although a substrate could have both the specific structure and sequence, it would not be unwound by D10 if there were DNA strand discontinuity (a nick at the particular branch point) within the integrated sequence, suggesting sequence continuity could be required for sequence recognition by D10. This model fits all the data using Y-junction substrates presented here. However, further experiments are necessary for determining whether this model is also applicable to other branched-DNA substrates.

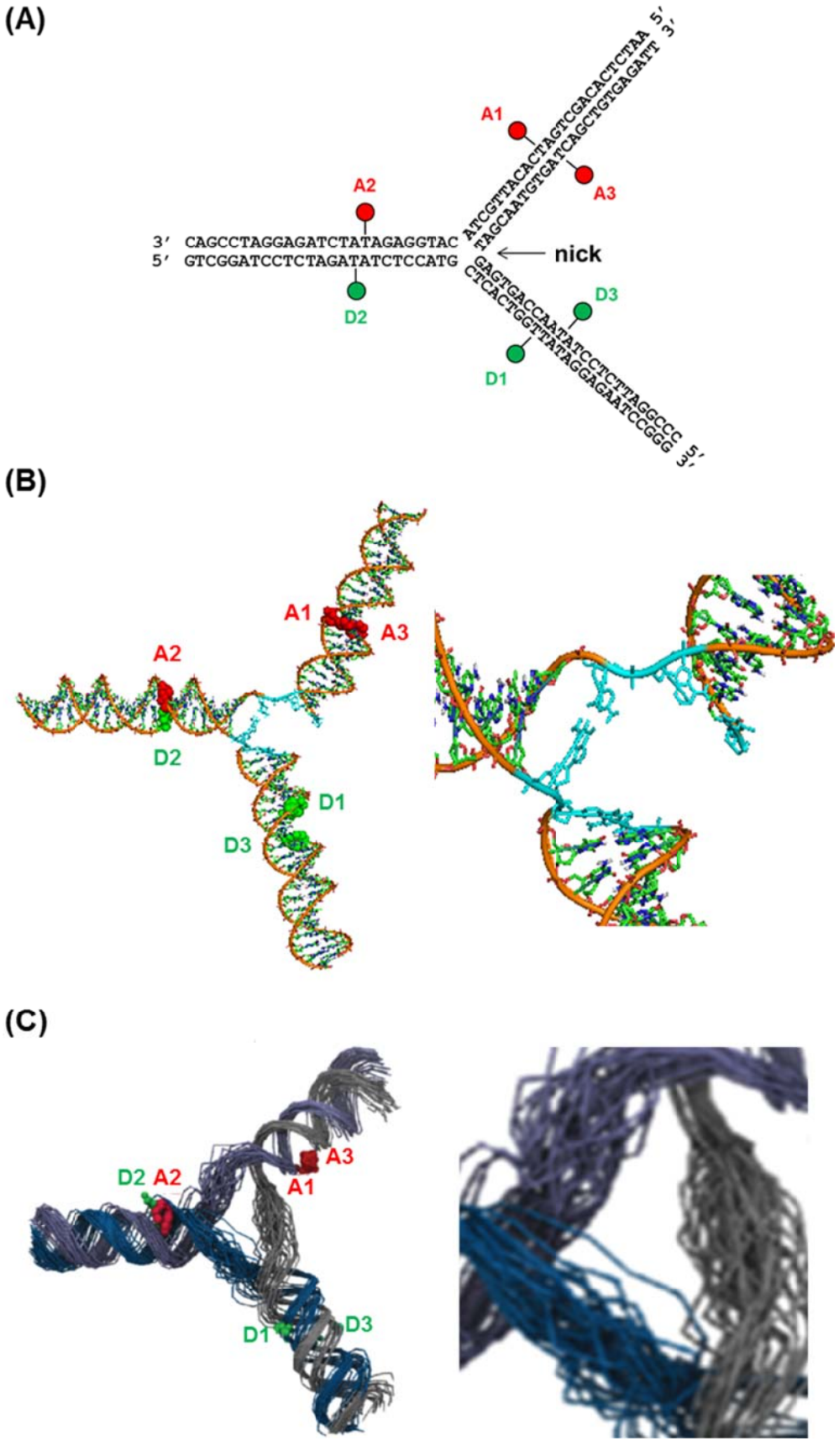

(A)
A1
A2
A3
D1
D2
D3
nick
3′ CAGCCTAGGAGATCTATAGAGGTAC
5′ GTCGGATCCTCTAGATATCTCCATG
ATCGTTACACTAGTCGACACTCTAA 5′
TAGCAATGTGATCAGCTGTGAGATT 3′
GAGTGACCAATATCCTCTTTAGGCCC 3′
CTCACTGGTTATAGGAGAAATCCGGG 5′
(B)
A1
A2
A3
D1
D2
D3
(C)
A1
A2
A3
D1
D2
D3

Figure 5.20 Single-molecule FRET for global structure determination of duplex Y-junction DNA without any nick (3WJ) and with a nick at the junction point (4SF). (**A**) Schematic representation for 3WJ and 4SF, showing the DNA sequence and the positions of fluorescent donor (D, green) and acceptor (A, red) dyes. (**B**) The solution structure of 3WJ in 1 mM $MgCl_2$ is displayed in cartoon and stick mode (PDB file is downloadable at http://pubs.acs.org). The close-up view of branchpoint region is on the right (1 bp was unpaired in each arm). (**C**) Superposition of 50 solution structures generated for 3WJ in 1 mM $MgCl_2$. The close-up view of branchpoint region is on the right (2 bp was unpaired in each arm). This figure is adapted with permissions from (Sabir et al., 2011; Sabir et al., 2012).

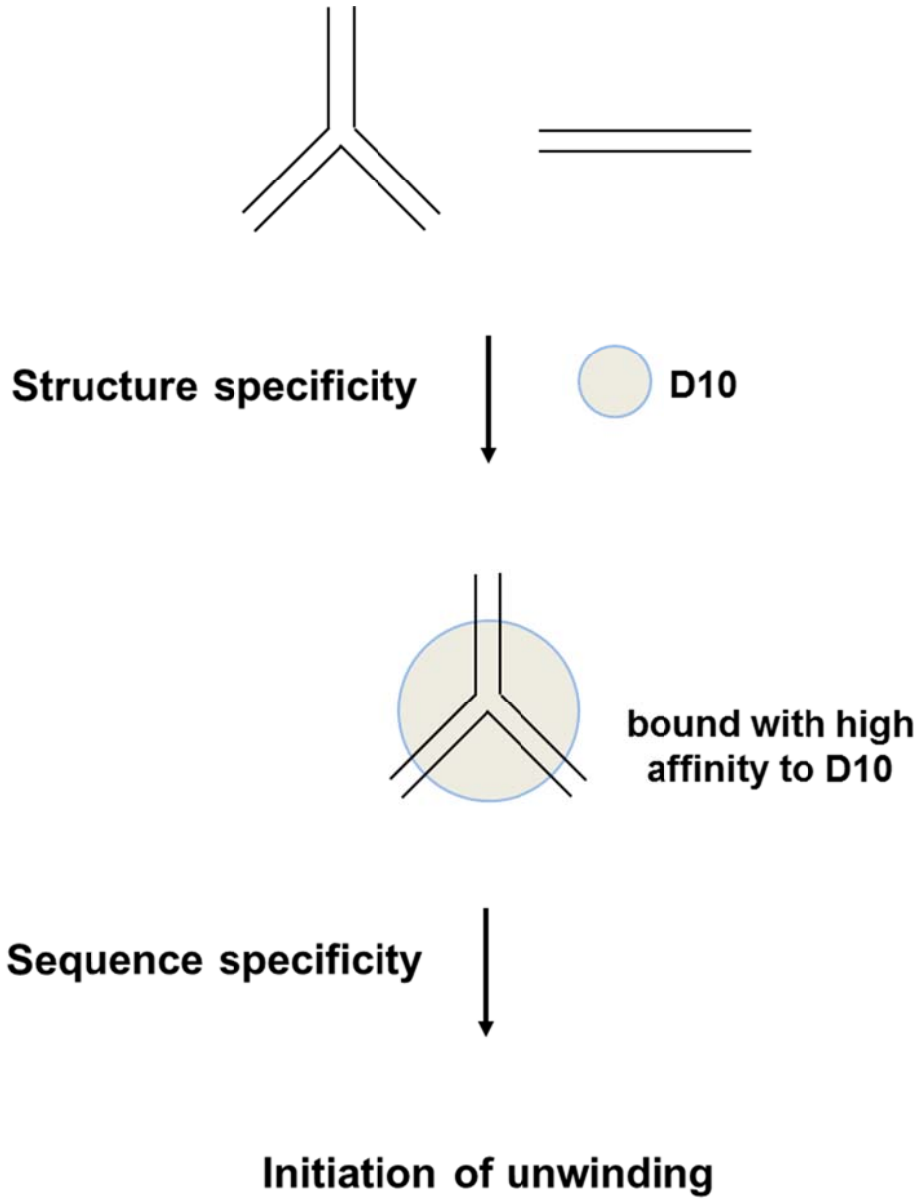


Figure 5.21 The proposed model for substrate specificity of D10 helicase. Assuming there are two kinds of substrates, duplex Y-junction and duplex blunt-ended DNA substrates. D10 would initially select substrates for binding on the basis of its affinity for specific DNA structures. Upon binding, D10 would align its (unidentified) recognition units with the relative region of the bound substrate and determine whether the substrate has the specific nucleotide sequence necessary for the initiation of unwinding. Also, the unwinding activity can be affected by the discontinuity of the integrated sequence (here is a nick at the particular branch point).

Due to lack of genetic data about D10, the precise function of D10 in the T5 phage life cycle is still unknown. However, based on the data presented here and some previous studies on the D10 homologues (UvsW and RecG) (Briggs et al., 2004; Long and Kreuzer, 2009; McGlynn and Lloyd, 2001; Nelson and Benkovic, 2007; Webb et al., 2007), some aspects about the potential roles of D10 in bacteriophage T5 may be envisioned. (a) The ability of D10 to unwind branched DNA suggested that D10 may be involved in some processes during DNA replication, recombination and repair, such as branch migration of Holliday junctions, fork regression and restart of stalled replication forks (Figure 5.22). (b) As mentioned in Section 1.1, the T5 genome has some mysterious nicks (five of them have consensus terminal sequences) in one strand of T5 dsDNA genome. The data in Figure 5.15, where D10 was unable to unwind the blunt-ended linear DNA substrate with a nick, assumed that D10 is not related to the unwinding of linear duplex region containing a nick of T5 genomic DNA. However, considering the sequence context of the substrate in detail, the terminal sequences near the nick of the test linear substrates (N20-1 or N20-3) were shown to be not completely identical to the above-mentioned consensus sequences, so additional experiments are required to confirm the above assumption. (c) D10 is the first helicase reported to have both structure and sequence substrate specificity, so there is no previous discussion about the functional role of this substrate specificity. T5 phage has the largest genome in all known T-odd viruses and keeps relatively rapid replication rate, so a lot of intermediates resembling the test substrates (forks, Y-junctions, Holliday junctions and nicked Y-junctions) could arise during replication. The intermediate with the same structure but different nucleotide sequences may have opposite (promoting or inhibitory) effects on T5 replication. Therefore, it is quite reasonable to propose that T5 has helicases which possess more than one selectivity criterion for their substrates, like D10, so that they could help T5 phage efficiently and precisely remove intermediates which hinder the replication.

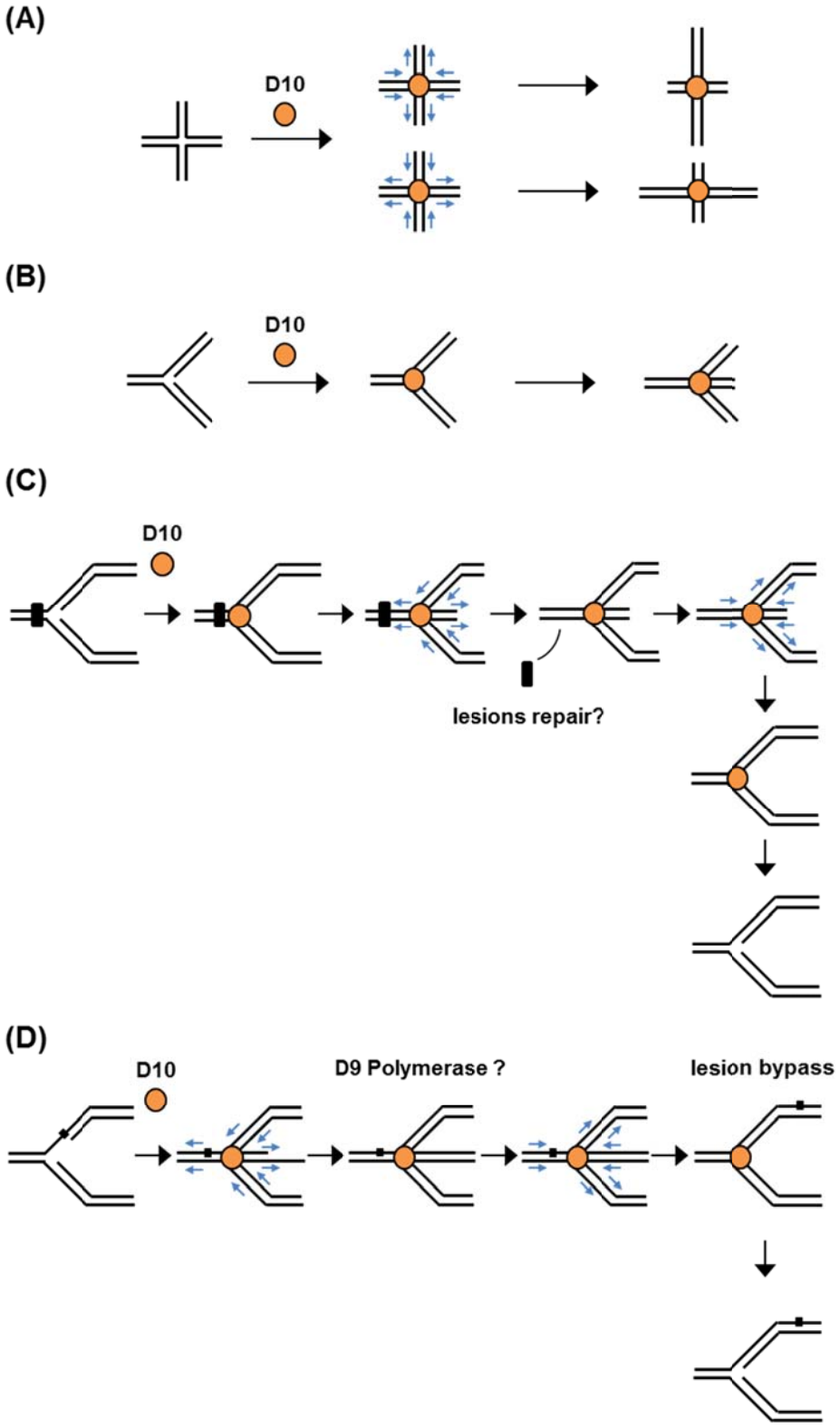


Figure 5.22 Schematic showing proposed functional roles of D10 in bacteriophage T5. (**A**) Branch migration of a Holliday junction in two directions. (**B**) Regression of a replication fork to form a Holliday junction. (**C**-**D**) Reactivation of stalled replication forks (**C**, lesions on both strands; **D**, a lesion on one strand) by helicase-mediated interconversion between replication fork and Holliday junction structures. The figure is adapted from (Briggs et al., 2004; Webb et al., 2007)

As indicated above, D10 possesses unusual substrate specificity of unwinding activity. Structural information could help to elucidate the underlying mechanism of this substrate specificity. The co-crystal structure of D10 with a DNA substrate (especially Y-junction substrate) would be the ideal structure to solve. A major practical barrier to this is the availability of large amounts of protein to facilitate extensive systematic crystallization trials. It was found to be difficult to obtain large amount of D10 wild-type protein (approx. 0.15 mg from 60 grams of cells). However, a possible solution emerged in that the work undertaken here has identified a highly expressed mutant (10 mg from 60 g cells) which has a mutation (R389N) in the arginine finger, almost eliminating both ATP*ase* and helicase activities, but retaining DNA-binding activity (Figure 5.6 and Figure 5.10). Whilst not ideal, structural studies on such a mutant could provide useful insight into the D10-DNA recognition mechanism.

Before this report comes out, the substrate specificity of all known helicases has been found to be either structure-dependent or sequence-dependent. The findings about the unusual substrate specificity of D10 in this thesis provide some new concepts for investigating helicase substrate specificity in the future. Hence, it is interesting to further study D10 to identify more unexplored details of this substrate specificity, probably related to the underlying mechanism of the helicase action.

Chaper 6. Conclusion and Future work

Bacteriophage T5 is a very efficient virus and it encodes many enzymes for its fast replication. Several enzymes involved in the replication have been identified, such as polymerase and flap endonuclease. However, one of the most important replication enzymes, a helicase, was not found in this organism. This is the first report on the identification and characterization of helicases from bacteriophage T5. In this study, a combined bioinformatic and biochemical approach was employed to identify two novel bacteriophage T5 helicases (D2 and D10) and characterize them.

6.1 Conclusion

The important findings and conclusions concerning D2 and D10 are as follows:

The D2 helicase

(i) The bioinformatic data have indicated that D2 could be an origin-binding protein involved in T5 origin-dependent replication since it shares some sequence similarity with many origin-binding proteins/helicases.

(ii) D2 displayed a rare bipolar helicase activity to unwind partial duplex DNA with either a 5' or 3' tail. The extent of 5'→3' or 3'→5' unwinding activity of D2 was found to be dependent on 5' or 3' tail length. One of the most interesting features of D2 is its biased polarity preference with its 3'→5' unwinding activity being greater than its 5'→3' unwinding activity when the 5' and 3' tail substrates have identical tail length. As discussed above, this feature could play a role in adjusting the speed of DNA replication.

(iii) Some structural features of D2 were also pointed out in this study. Except for a Walker A motif, no other conserved motifs of SF1-6 helicases were identified in the D2 protein sequence. This implicates D2 may possess some unconventional motifs or domains relevant to its helicase and ATP*ase* activities.

A K405E substitution mutation in the Walker A motif abolished D2 bipolar helicase activity, suggesting both of 5'→3' and 3'→5' motor units of D2 share the same catalytic motif (at least Walker A motif).

The D10 helicase

(i) D10 has been revealed to be a branched-DNA specific helicase which is able to unwind branched DNA substrates, including forks, Y-junctions and Holliday junctions, which resemble DNA replication, recombination and repair intermediates. This suggests D10 may play a role in processing DNA during replication, recombination and repair.

(ii) The extended dsDNA length of the Holliday junction substrate appeared to impose a restriction for D10 helicase activity. The observed restriction resulted from an artefact of using the non-homologous substrate whose long duplexes can be reannealed before they are completely dissociated. The ability of D10 to convert a close mimic of a natural Holliday junction substrate into two kinds of products infers D10 can branch-migrate and unwind Holliday junctions.

(iii) The substrate specificity of D10 unwinding has been shown to be structure-dependent and sequence-dependent using the Y-junction substrate, but the binding of D10 to the substrate is majorly structure-dependent. Also, the unwinding activity can be affected by the discontinuity of the integrated sequence (here is a nick at the particular branch point).

6.2 Future work

Based on the results presented in this thesis, more biochemical and structural biology work needs to be done for further characterization of D2 and D10 helicases. Since D2 probably has origin-binding activity, the identification of D2 binding site on T5 genomic DNA using McKay assays (McKay, 1981) and DNA footprinting assays could be used to determine the unknown sequence of the T5

replication origin. Furthermore, the same methods could also be employed to identify the actual biological substrate of D10 to facilitate understanding the function of D10 in T5 phage. In general, the activity of many helicases was found to be regulated by protein binding partners, so it would be a good idea to investigate the interactions of D2 and D10 with host cell or phage proteins using bacterial two-hybrid system (Euromedex, UK), pull-down assays and proteomic techniques.

Structural information of these two helicases is also important for elucidating the unusual characteristics of these two helicases. The co-crystal structure of helicase in complex with the DNA substrate would be the ideal structure to solve. Generally, at least 10 mg of the recombinant helicase need to be produced for the crystallization trials. The structural studies on these two helicases would provide more insights into the structural determinants for substrate specificity of D10 and the polarity preference of D2.

The challenges provided by this study involve a range of disciplines. Except the above-mentioned biochemical and structural biology work, it is also essential to achieve more genetic and biophysical information of D10 and D2 for confirming their precise functional roles and underlying mechanisms for the helicase actions.

Other than D2 and D10, the D6 protein was also thought to be a helicase. However, D6 protein did not show ATP*ase* activity probably due to the failure of protein purification or the reaction condition. The improvement for the protein purification strategy and the reaction condition is required for determining the function of D6.

Appendix A

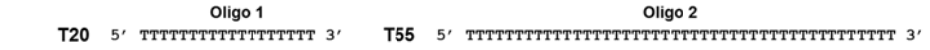
The sequences and structures of oligonucleotide substrates used in this study

ID	Sequence (5'-3')
Oligo 1	$(T)_{20}$
Oligo 2	$(T)_{55}$
Oligo 3 (Oligo c2)	GGGTACCGAGCTCGAATTCG
Oligo 4 (Oligo b2)	CGAATTCGAGCTCGGTACCC
Oligo 5	$(T)_n$CGAATTCGAGCTCGGTACCC
Oligo 6	GGGTACCGAGCTCGAATTCG$(T)_n$
Oligo 7	GGGTACCGAGCTCGAATTCGTTTTTTTTTTTTTTTTTTTATGGGCGCGCACGTTTCGAA
Oligo 8	TTCGAAACGTGCGCGCCCATAAAAAAAAAAAAAAAAAAAACGAATTCGAGCTCGGTACCC
Oligo 9 (Oligo A)	TTCGAAACGTGCGCGCCCATACGCCAAGCTTAACGGTGAT
Oligo 10 (Oligo B)	ATCACCGTTAAGCTTGGCGTCGAATTCGAGCTCGGTACCC
Oligo 11	GGGTACCGAGCTCGAATTCGTGGGCGGTGCCCAACGCATA
Oligo 12	TATGCGTTGGGCACCGCCCAATGGGCGCGCACGTTTCGAA
Oligo 13 (Oligo C)	GGGTACCGAGCTCGAATTCGATGGGCGCGCACGTTTCGAA
Oligo 14	TTCGAAACGTGCGCGCCCATACGCCAAGCTTAACGGTGATTTTGGTTTTCCCGCTTGAAAAACCGGCAACGGTGT
Oligo 15	ACACCGTTGCCGGTTTTTCAAGCGGGAAAACCAAAATCACCGTTAAGCTTGGCGTCGAATTCGAGCTCGGTACCC
Oligo 16	TGACGCGAAAACTGTTTGATACTCCGACAGTAATATATGCGTTGGGCACCGCCCAATGGGCGCGCACGTTTCGAA
Oligo 17	GGGTACCGAGCTCGAATTCGTGGGCGGTGCCCAACGCATATATTACTGTCGGAGTATCAAACAGTTTTCGCGTCA

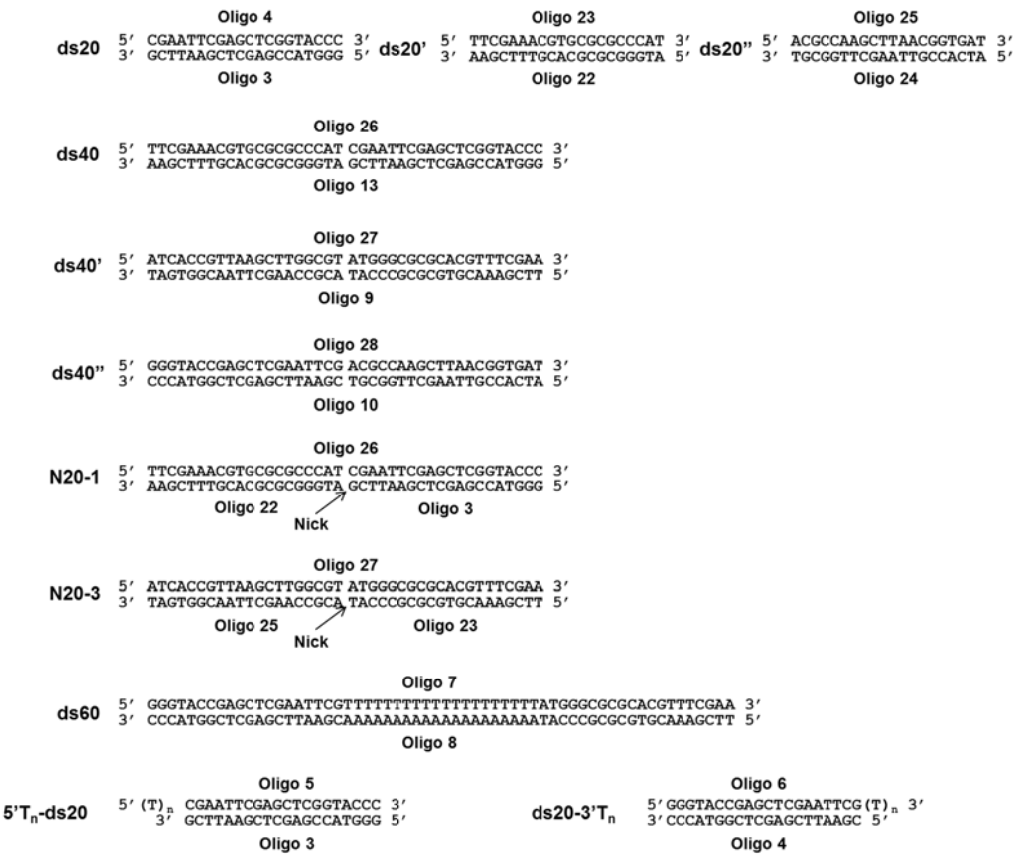
Oligo 18	TTCGAAACGTGCGCGCCCATATCGTTACATTAGAAGGATCCACTG GTTTCCCGCTTGAAAAACCGGCAACGGTGT
Oligo 19	ACACCGTTGCCGGTTTTTCAAGCGGGAAACCAGTGGATCCTTCTA ATGTAACGATCGAATTCGAGCTCGGTACCC
Oligo 20	GGGTACCGAGCTCGAATTCGATCGTTACATTAGAAGGATCCACTG GTTTCCCGCTTGAAAAACCGGCAACGGTGT
Oligo 21	ACACCGTTGCCGGTTTTTCAAGCGGGAAACCAGTGGATCCTTCTA ATGTAACGATATGGGCGCGCACGTTTCGAA
Oligo 22 (Oligo c1)	ATGGGCGCGCACGTTTCGAA
Oligo 23 (Oligo a1)	TTCGAAACGTGCGCGCCCAT
Oligo 24 (Oligo b1)	ATCACCGTTAAGCTTGGCGT
Oligo 25 (Oligo a2)	ACGCCAAGCTTAACGGTGAT
Oligo 26	TTCGAAACGTGCGCGCCCATCGAATTCGAGCTCGGTACCC
Oligo 27	ATCACCGTTAAGCTTGGCGTATGGGCGCGCACGTTTCGAA
Oligo 28	GGGTACCGAGCTCGAATTCGACGCCAAGCTTAACGGTGAT
Oligo 29 (Oligo A')	TTCGAAACGTGCGCGCCTCGATGCCAAGCTTAACGGTGAT
Oligo 30 (Oligo B')	ATCACCGTTAAGCTTGGCATACGATTCGAGCTCGGTACCC
Oligo 31 (Oligo C')	GGGTACCGAGCTCGAATCGTCGAGGCGCGCACGTTTCGAA
Oligo 32 (Oligo A'')	TTCGAAACGTGCGGAATCATACGGGCGGCTTAACGGTGAT
Oligo 33 (Oligo B'')	ATCACCGTTAAGCCGCCCGTCGACCAAGAGCTCGGTACCC
Oligo 34 (Oligo C'')	GGGTACCGAGCTCTTGGTCGATGATTCCGCACGTTTCGAA

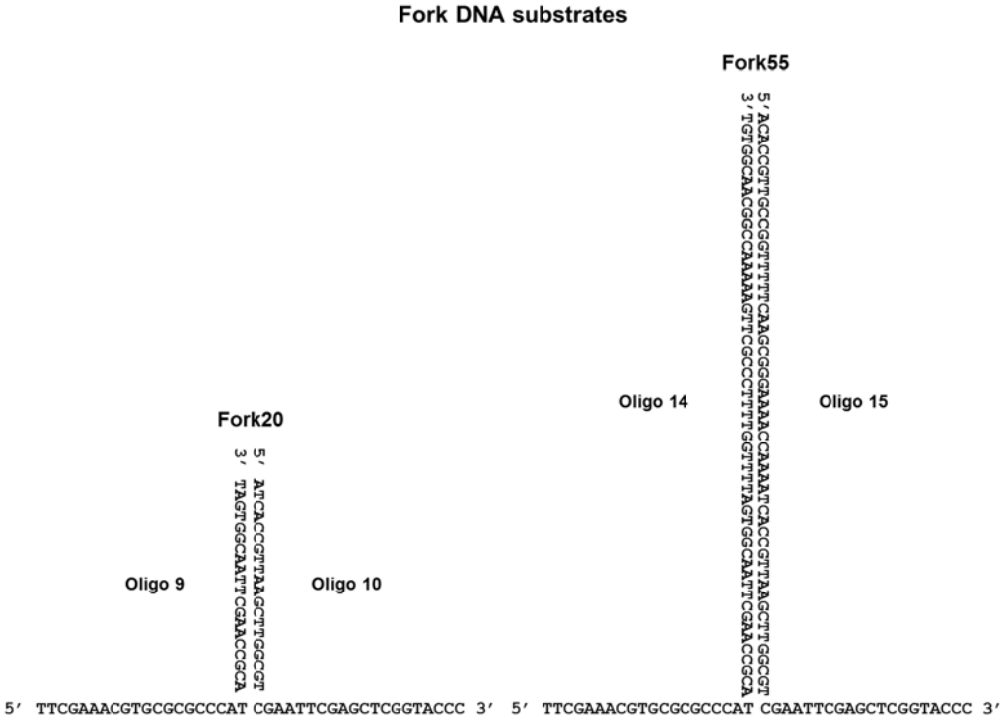
Linear DNA substrates

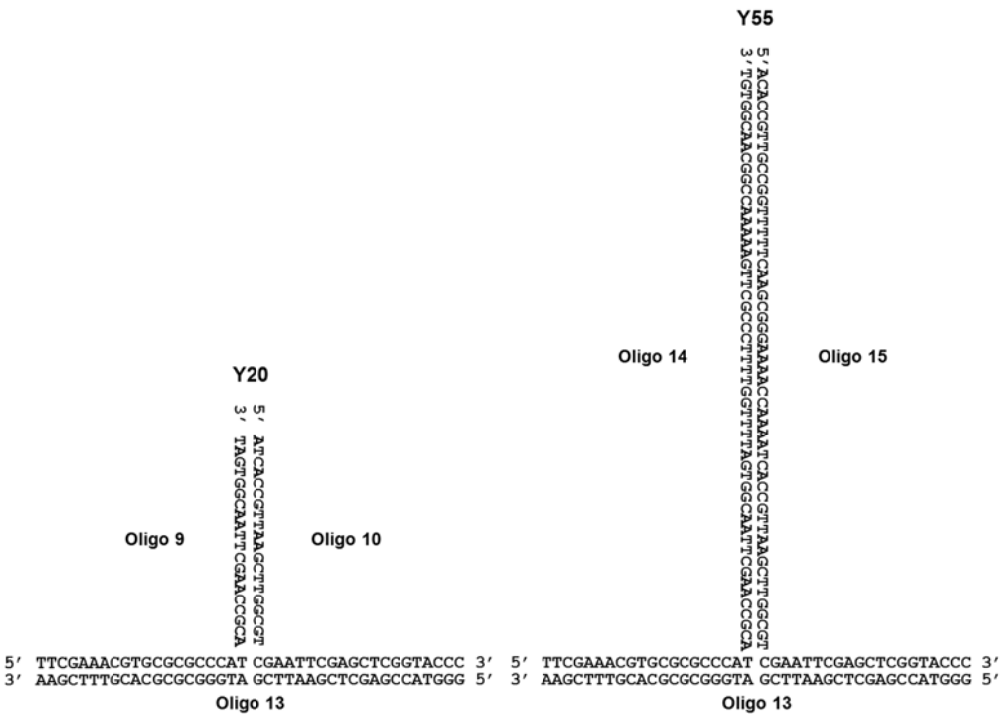
ssDNA



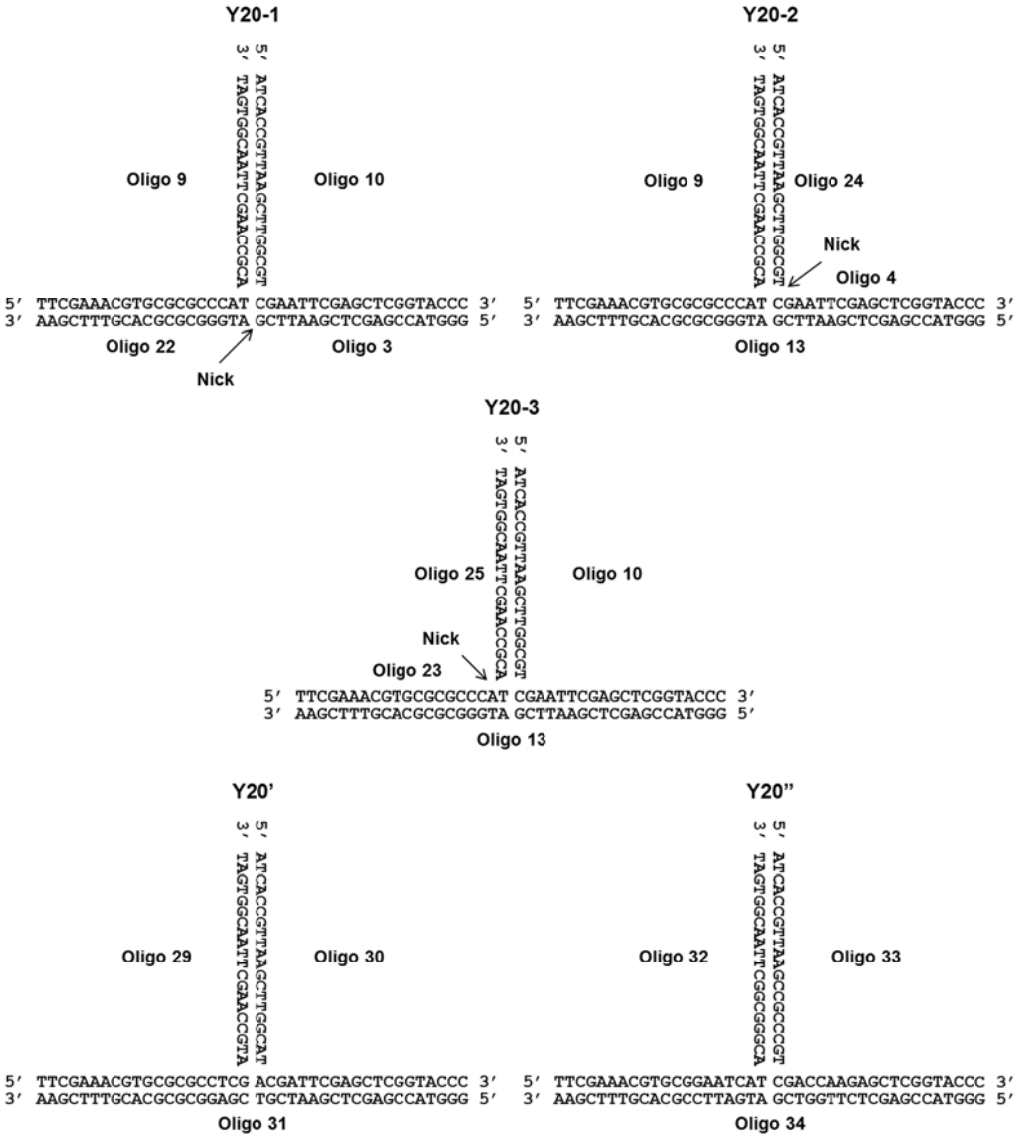
dsDNA



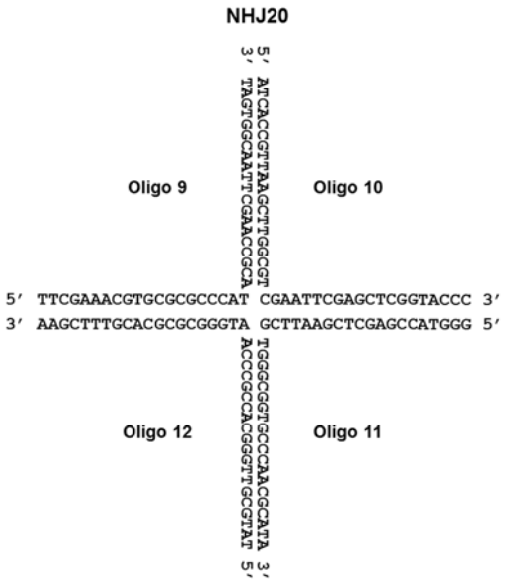

Fork DNA substrates
Fork55
Fork20
Oligo 9
Oligo 10
Oligo 14
Oligo 15
5′ TTCGAAACGTGCGCGCCCAT CGAATTCGAGCTCGGTACCC 3′
5′ TTCGAAACGTGCGCGCCCAT CGAATTCGAGCTCGGTACCC 3′

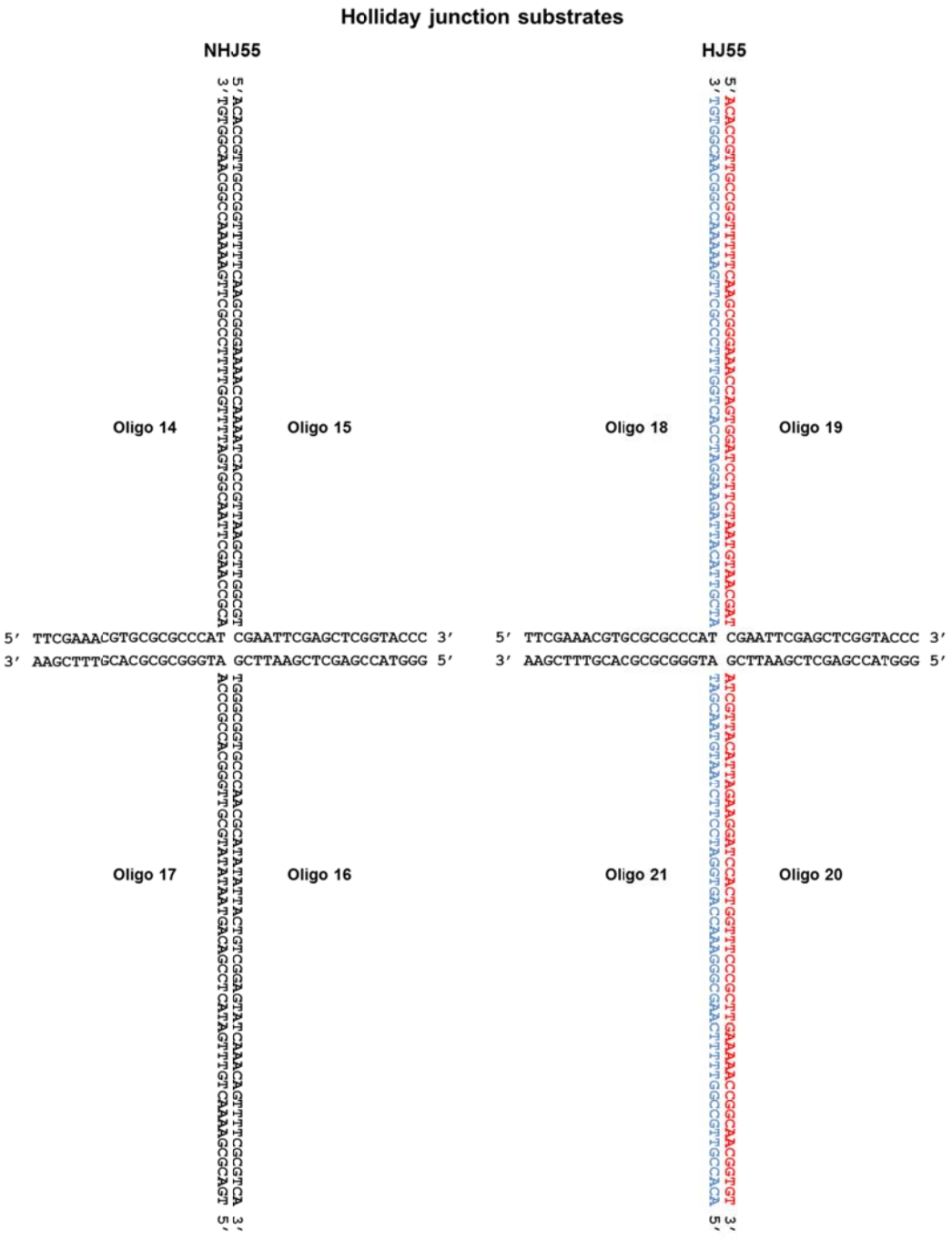

Y-junction substrates
Y55
Y20
Oligo 9
Oligo 10
Oligo 14
Oligo 15
5′ TTCGAAACGTGCGCGCCCAT CGAATTCGAGCTCGGTACCC 3′
3′ AAGCTTTGCACGCGCGGGTA GCTTAAGCTCGAGCCATGGG 5′
5′ TTCGAAACGTGCGCGCCCAT CGAATTCGAGCTCGGTACCC 3′
3′ AAGCTTTGCACGCGCGGGTA GCTTAAGCTCGAGCCATGGG 5′
Oligo 13
Oligo 13

Y-junction substrates

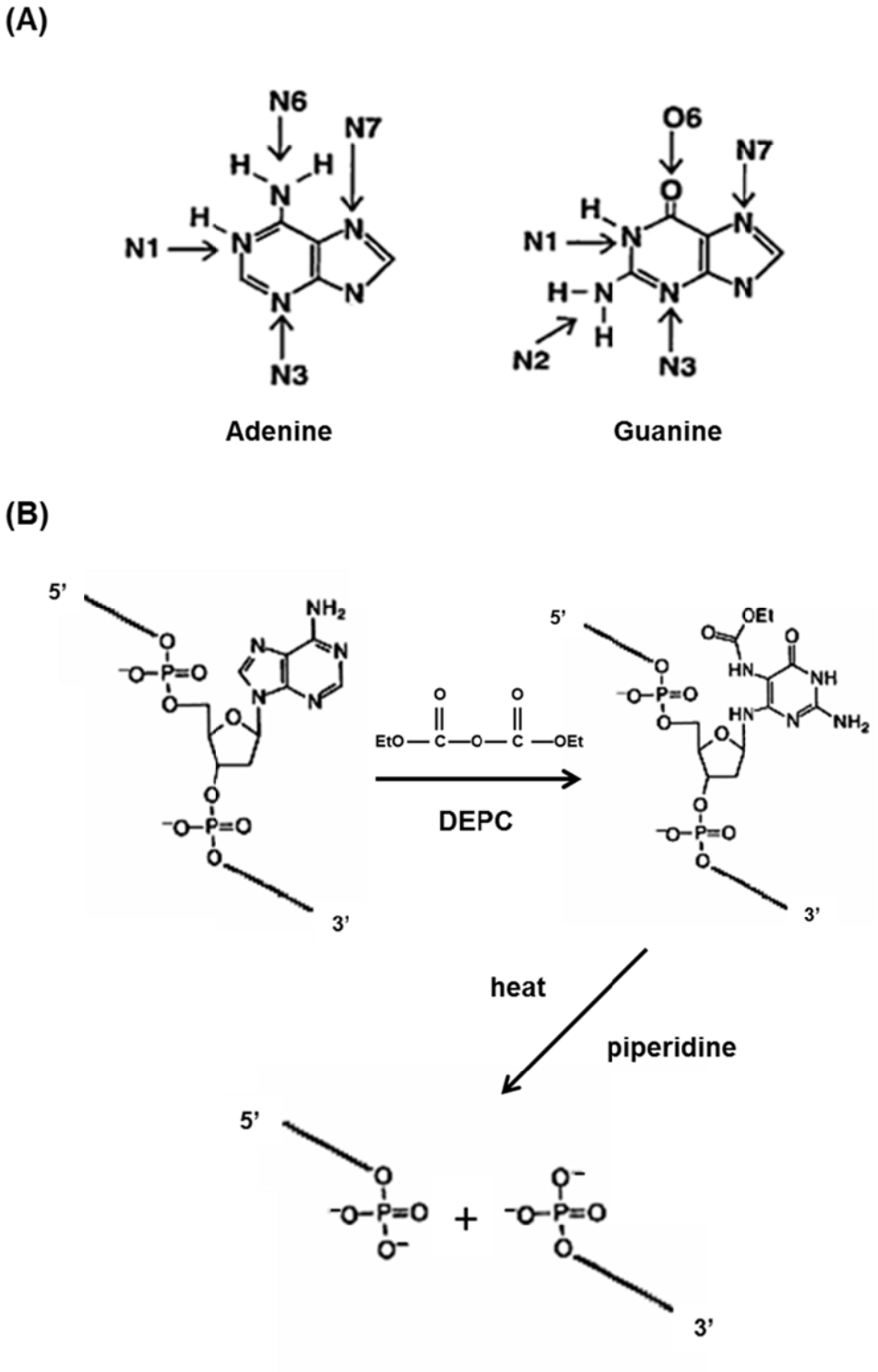


Holliday junction substrates




Holliday junction substrates
NHJ55
HJ55
Oligo 14
Oligo 15
Oligo 18
Oligo 19
5′ TTCGAAACGTGCGCGCCCAT CGAATTCGAGCTCGGTACCC 3′
3′ AAGCTTTGCACGCGCGGGTA GCTTAAGCTCGAGCCATGGG 5′
5′ TTCGAAACGTGCGCGCCCAT CGAATTCGAGCTCGGTACCC 3′
3′ AAGCTTTGCACGCGCGGGTA GCTTAAGCTCGAGCCATGGG 5′
Oligo 17
Oligo 16
Oligo 21
Oligo 20

Appendix B



(**A**) The chemical structures of purines (Adenine and Guanine). (**B**) The N7 position of purine (*i.e.* Adenine) is carbethoxylated and subsequent cleavage of the strand is promoted by heat and piperidine. This figure is adapted from (Rokita, 2001)

Bibliography

Aggarwal, M., and Brosh, R.M., Jr. (2009). Hitting the bull's eye: novel directed cancer therapy through helicase-targeted synthetic lethality. Journal of cellular biochemistry *106*, 758-763.

Aggarwal, M., Sommers, J.A., Shoemaker, R.H., and Brosh, R.M., Jr. (2011). Inhibition of helicase activity by a small molecule impairs Werner syndrome helicase (WRN) function in the cellular response to DNA damage or replication stress. Proc Natl Acad Sci U S A *108*, 1525-1530.

Ahnert, P., and Patel, S.S. (1997). Asymmetric interactions of hexameric bacteriophage T7 DNA helicase with the 5'- and 3'-tails of the forked DNA substrate. The Journal of biological chemistry *272*, 32267-32273.

Aiello, D., Barnes, M.H., Biswas, E.E., Biswas, S.B., Gu, S., Williams, J.D., Bowlin, T.L., and Moir, D.T. (2009). Discovery, characterization and comparison of inhibitors of Bacillus anthracis and Staphylococcus aureus replicative DNA helicases. Bioorganic & medicinal chemistry *17*, 4466-4476.

Alexandrov, A.I., Botchan, M.R., and Cozzarelli, N.R. (2002). Characterization of simian virus 40 T-antigen double hexamers bound to a replication fork. The active form of the helicase. The Journal of biological chemistry *277*, 44886-44897.

Amaratunga, M., and Lohman, T.M. (1993). Escherichia coli rep helicase unwinds DNA by an active mechanism. Biochemistry *32*, 6815-6820.

Amundsen, S.K., and Smith, G.R. (2007). Chi Hotspot Activity in Escherichia coli Without RecBCD Exonuclease Activity: Implications for the Mechanism of Recombination. Genetics *175*, 41-54.

Ariumi, Y., Kuroki, M., Abe, K., Dansako, H., Ikeda, M., Wakita, T., and Kato, N. (2007). DDX3 DEAD-box RNA helicase is required for hepatitis C virus RNA replication. J Virol *81*, 13922-13926.

Bailey, S., Eliason, W.K., and Steitz, T.A. (2007). Structure of hexameric DnaB helicase and its complex with a domain of DnaG primase. Science *318*, 459-463.

Beckman, L.D., Hoffman, M.S., and McCorquodale, D.J. (1971). Pre-early proteins of bacteriophage T5: structure and function. J Mol Biol *62*, 551-564.

Belon, C.A., and Frick, D.N. (2009). Helicase inhibitors as specifically targeted antiviral therapy for hepatitis C. Future virology *4*, 277-293.

Betterton, M.D., and Julicher, F. (2005). Opening of nucleic-acid double strands by helicases: active versus passive opening. Physical review. E, Statistical, nonlinear, and soft matter physics *71*, 011904.

Blinov, V.M., Koonin, E.V., Gorbalenya, A.E., Kaliman, A.V., and Kryukov, V.M. (1989). Two early genes of bacteriophage T5 encode proteins containing an

NTP-binding sequence motif and probably involved in DNA replication, recombination and repair. FEBS Lett *252*, 47-52.

Bochman, M.L., and Schwacha, A. (2009). The Mcm complex: unwinding the mechanism of a replicative helicase. Microbiology and molecular biology reviews : MMBR *73*, 652-683.

Boehmer, P.E., Craigie, M.C., Stow, N.D., and Lehman, I.R. (1994). Association of origin binding protein and single strand DNA-binding protein, ICP8, during herpes simplex virus type 1 DNA replication in vivo. The Journal of biological chemistry *269*, 29329-29334.

Borowski, P., Deinert, J., Schalinski, S., Bretner, M., Ginalski, K., Kulikowski, T., and Shugar, D. (2003). Halogenated benzimidazoles and benzotriazoles as inhibitors of the NTPase/helicase activities of hepatitis C and related viruses. European journal of biochemistry / FEBS *270*, 1645-1653.

Bowers, J.L., Randell, J.C., Chen, S., and Bell, S.P. (2004). ATP hydrolysis by ORC catalyzes reiterative Mcm2-7 assembly at a defined origin of replication. Molecular cell *16*, 967-978.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical biochemistry *72*, 248-254.

Briggs, G.S., Mahdi, A.A., Weller, G.R., Wen, Q., and Lloyd, R.G. (2004). Interplay between DNA replication, recombination and repair based on the structure of RecG helicase. Philosophical transactions of the Royal Society of London. Series B, Biological sciences *359*, 49-59.

Brosh, R.M., Jr., von Kobbe, C., Sommers, J.A., Karmakar, P., Opresko, P.L., Piotrowski, J., Dianova, I., Dianov, G.L., and Bohr, V.A. (2001). Werner syndrome protein interacts with human flap endonuclease 1 and stimulates its cleavage activity. The EMBO journal *20*, 5791-5801.

Bujalowski, W., Klonowska, M.M., and Jezewska, M.J. (1994). Oligomeric structure of Escherichia coli primary replicative helicase DnaB protein. The Journal of biological chemistry *269*, 31350-31358.

Byrd, A.K., and Raney, K.D. (2004). Protein displacement by an assembly of helicase molecules aligned along single-stranded DNA. Nat Struct Mol Biol *11*, 531-538.

Byrd, A.K., and Raney, K.D. (2005). Increasing the length of the single-stranded overhang enhances unwinding of duplex DNA by bacteriophage T4 Dda helicase. Biochemistry *44*, 12990-12997.

Caruthers, J.M., and McKay, D.B. (2002). Helicase structure and mechanism. Curr Opin Struct Biol *12*, 123-133.

Ceska, T.A., Sayers, J.R., Stier, G., and Suck, D. (1996). A helical arch allowing

single-stranded DNA to thread through T5 5'-exonuclease. Nature *382*, 90-93.

Chatterjee, D.K., Fujimura, R.K., Campbell, J.H., and Gerard, G.F. (1991). Cloning and overexpression of the gene encoding bacteriophage T5 DNA polymerase. Gene *97*, 13-19.

Chen, C.-S., Chiou, C.-T., Chen, G.S., Chen, S.-C., Hu, C.-Y., Chi, W.-K., Chu, Y.-D., Hwang, L.-H., Chen, P.-J., Chen, D.-S., *et al.* (2009). Structure-Based Discovery of Triphenylmethane Derivatives as Inhibitors of Hepatitis C Virus Helicase∞. Journal of Medicinal Chemistry *52*, 2716-2723.

Cheng, W., Brendza, K.M., Gauss, G.H., Korolev, S., Waksman, G., and Lohman, T.M. (2002). The 2B domain of the Escherichia coli Rep protein is not required for DNA helicase activity. Proc Natl Acad Sci U S A *99*, 16006-16011.

Chono, K., Katsumata, K., Kontani, T., Kobayashi, M., Sudo, K., Yokota, T., Konno, K., Shimizu, Y., and Suzuki, H. (2010). ASP2151, a novel helicase-primase inhibitor, possesses antiviral activity against varicella-zoster virus and herpes simplex virus types 1 and 2. The Journal of antimicrobial chemotherapy *65*, 1733-1741.

Crampton, D.J., Mukherjee, S., and Richardson, C.C. (2006). DNA-induced switch from independent to sequential dTTP hydrolysis in the bacteriophage T7 DNA helicase. Molecular cell *21*, 165-174.

Crute, J.J., Grygon, C.A., Hargrave, K.D., Simoneau, B., Faucher, A.M., Bolger, G., Kibler, P., Liuzzi, M., and Cordingley, M.G. (2002). Herpes simplex virus helicase-primase inhibitors are active in animal models of human disease. Nature medicine *8*, 386-391.

Davison, J., and Brunel, F. (1979). Restriction insensitivity in bacteriophage T5 I. Genetic characterization of mutants sensitive to EcoRI restriction. J Virol *29*, 11-16.

Decker, K., Krauel, V., Meesmann, A., and Heller, K.J. (1994). Lytic conversion of Escherichia coli by bacteriophage T5: blocking of the FhuA receptor protein by a lipoprotein expressed early during infection. Mol Microbiol *12*, 321-332.

Delagoutte, E., and von Hippel, P.H. (2001). Molecular mechanisms of the functional coupling of the helicase (gp41) and polymerase (gp43) of bacteriophage T4 within the DNA replication fork. Biochemistry *40*, 4459-4477.

Dillingham, M.S., and Kowalczykowski, S.C. (2008). RecBCD enzyme and the repair of double-stranded DNA breaks. Microbiology and molecular biology reviews : MMBR *72*, 642-671, Table of Contents.

Dillingham, M.S., Webb, M.R., and Kowalczykowski, S.C. (2005). Bipolar DNA translocation contributes to highly processive DNA unwinding by RecBCD enzyme. The Journal of biological chemistry *280*, 37069-37077.

Dillingham, M.S., Wigley, D.B., and Webb, M.R. (2000). Demonstration of

unidirectional single-stranded DNA translocation by PcrA helicase: measurement of step size and translocation speed. Biochemistry *39*, 205-212.

Dixon, W.J., Hayes, J.J., Levin, J.R., Weidner, M.F., Dombroski, B.A., and Tullius, T.D. (1991). Hydroxyl radical footprinting. Methods in enzymology *208*, 380-413.

Doherty, K.M., Sommers, J.A., Gray, M.D., Lee, J.W., von Kobbe, C., Thoma, N.H., Kureekattil, R.P., Kenny, M.K., and Brosh, R.M., Jr. (2005). Physical and functional mapping of the replication protein a interaction domain of the werner and bloom syndrome helicases. The Journal of biological chemistry *280*, 29494-29505.

Dong, F., Gogol, E.P., and von Hippel, P.H. (1995). The phage T4-coded DNA replication helicase (gp41) forms a hexamer upon activation by nucleoside triphosphate. The Journal of biological chemistry *270*, 7462-7473.

Dong, F., Weitzel, S.E., and von Hippel, P.H. (1996). A coupled complex of T4 DNA replication helicase (gp41) and polymerase (gp43) can perform rapid and processive DNA strand-displacement synthesis. Proc Natl Acad Sci U S A *93*, 14456-14461.

Donmez, I., and Patel, S.S. (2006). Mechanisms of a ring shaped helicase. Nucleic acids research *34*, 4216-4224.

Dumont, S., Cheng, W., Serebrov, V., Beran, R.K., Tinoco, I., Jr., Pyle, A.M., and Bustamante, C. (2006). RNA translocation and unwinding mechanism of HCV NS3 helicase and its coordination by ATP. Nature *439*, 105-108.

Edwards, M.C., Tutter, A.V., Cvetic, C., Gilbert, C.H., Prokhorova, T.A., and Walter, J.C. (2002). MCM2-7 complexes bind chromatin in a distributed pattern surrounding the origin recognition complex in Xenopus egg extracts. The Journal of biological chemistry *277*, 33049-33057.

Effantin, G., Boulanger, P., Neumann, E., Letellier, L., and Conway, J.F. (2006). Bacteriophage T5 structure reveals similarities with HK97 and T4 suggesting evolutionary relationships. J Mol Biol *361*, 993-1002.

Enemark, E.J., and Joshua-Tor, L. (2006). Mechanism of DNA translocation in a replicative hexameric helicase. Nature *442*, 270-275.

Feng, M., Patel, D., Dervan, J.J., Ceska, T., Suck, D., Haq, I., and Sayers, J.R. (2004). Roles of divalent metal ions in flap endonuclease-substrate interactions. Nat Struct Mol Biol *11*, 450-456.

Ficht, T.A., and Moyer, R.W. (1980). Isolation and characterization of a putative bacteriophage T5 transcription.replication enzyme complex from infected Escherichia coli. The Journal of biological chemistry *255*, 7040-7048.

Firman, K., and Szczelkun, M.D. (2000). Measuring motion on DNA by the type I restriction endonuclease EcoR124I using triplex displacement. The EMBO

journal *19*, 2094-2102.

Fischer, C.J., Maluf, N.K., and Lohman, T.M. (2004). Mechanism of ATP-dependent translocation of E.coli UvrD monomers along single-stranded DNA. J Mol Biol *344*, 1287-1309.

Gai, D., Zhao, R., Li, D., Finkielstein, C.V., and Chen, X.S. (2004). Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. Cell *119*, 47-60.

Geiselmann, J., Wang, Y., Seifried, S.E., and von Hippel, P.H. (1993). A physical model for the translocation and helicase activities of Escherichia coli transcription termination protein Rho. Proc Natl Acad Sci U S A *90*, 7754-7758.

Gentz, R., and Bujard, H. (1985). Promoters recognized by Escherichia coli RNA polymerase selected by function: highly efficient promoters from bacteriophage T5. J Bacteriol *164*, 70-77.

George, T., Wen, Q., Griffiths, R., Ganesh, A., Meuth, M., and Sanders, C.M. (2009). Human Pif1 helicase unwinds synthetic DNA structures resembling stalled DNA replication forks. Nucleic acids research *37*, 6491-6502.

Grandori, C., Robinson, K.L., Galloway, D.A., and Swisshelm, K. (2004). Functional link between Myc and the Werner gene in tumorigenesis. Cell Cycle *3*, 22-25.

Grandori, C., Wu, K.J., Fernandez, P., Ngouenet, C., Grim, J., Clurman, B.E., Moser, M.J., Oshima, J., Russell, D.W., Swisshelm, K.*, et al.* (2003). Werner syndrome protein limits MYC-induced cellular senescence. Genes & development *17*, 1569-1574.

Guan, K.L., and Dixon, J.E. (1991). Eukaryotic proteins expressed in Escherichia coli: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase. Analytical biochemistry *192*, 262-267.

Gueron, M., and Leroy, J.L. (1995). Studies of base pair kinetics by NMR measurement of proton exchange. Methods in enzymology *261*, 383-413.

Gupta, R., and Brosh, R.M., Jr. (2008). Helicases as prospective targets for anti-cancer therapy. Anti-cancer agents in medicinal chemistry *8*, 390-401.

Gyimesi, M., Sarlos, K., and Kovacs, M. (2010). Processive translocation mechanism of the human Bloom's syndrome helicase along single-stranded DNA. Nucleic acids research *38*, 4404-4414.

Ha, T., Rasnik, I., Cheng, W., Babcock, H.P., Gauss, G.H., Lohman, T.M., and Chu, S. (2002). Initiation and re-initiation of DNA unwinding by the Escherichia coli Rep helicase. Nature *419*, 638-641.

Hall, M.C., and Matson, S.W. (1999). Helicase motifs: the engine that powers DNA unwinding. Mol Microbiol *34*, 867-877.

Hanahan, D. (1983). Studies on transformation of Escherichia coli with plasmids. J Mol Biol *166*, 557-580.

Hanson, P.I., and Whiteheart, S.W. (2005). AAA+ proteins: have engine, will work. Nature reviews. Molecular cell biology *6*, 519-529.

Hausmann, R., and Gold, M. (1966). The enzymatic methylation of ribonucleic acid and deoxyribonucleic acid. IX. Deoxyribonucleic acid methylase in bacteriophage-infected Escherichia coli. The Journal of biological chemistry *241*, 1985-1994.

Heller, K.J., and Krauel, V. (1986). Cloning and expression of the ltf gene of bacteriophage T5. J Bacteriol *167*, 1071-1073.

Heusterspreute, M., Ha-Thi, V., Tournis-Gamble, S., and Davison, J. (1987). The first-step transfer-DNA injection-stop signal of bacteriophage T5. Gene *52*, 155-164.

Hickman, A.B., and Dyda, F. (2005). Binding and unwinding: SF3 viral helicases. Curr Opin Struct Biol *15*, 77-85.

Hickson, I.D. (2003). RecQ helicases: caretakers of the genome. Nat Rev Cancer *3*, 169-178.

Hingorani, M.M., Washington, M.T., Moore, K.C., and Patel, S.S. (1997). The dTTPase mechanism of T7 DNA helicase resembles the binding change mechanism of the F1-ATPase. Proc Natl Acad Sci U S A *94*, 5012-5017.

Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K., and Pease, L.R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene *77*, 51-59.

Hong, J., Kim, K.P., Heu, S., Lee, S.J., Adhya, S., and Ryu, S. (2008). Identification of host receptor and receptor-binding module of a newly sequenced T5-like phage EPS7. FEMS Microbiol Lett *289*, 202-209.

Hsieh, J., Moore, K.J., and Lohman, T.M. (1999). A two-site kinetic mechanism for ATP binding and hydrolysis by E. coli Rep helicase dimer bound to a single-stranded oligodeoxynucleotide. J Mol Biol *288*, 255-274.

Iggo, R.D., and Lane, D.P. (1989). Nuclear protein p68 is an RNA-dependent ATPase. The EMBO journal *8*, 1827-1831.

Ishaq, M., Hu, J., Wu, X., Fu, Q., Yang, Y., Liu, Q., and Guo, D. (2008). Knockdown of cellular RNA helicase DDX3 by short hairpin RNAs suppresses HIV-1 viral replication without inducing apoptosis. Molecular biotechnology *39*, 231-238.

Jezewska, M.J., Rajendran, S., and Bujalowski, W. (1998). Complex of Escherichia coli primary replicative helicase DnaB protein with a replication fork: recognition and structure. Biochemistry *37*, 3116-3136.

Johnson, D.S., Bai, L., Smith, B.Y., Patel, S.S., and Wang, M.D. (2007). Single-molecule studies reveal dynamics of DNA unwinding by the ring-shaped T7 helicase. Cell *129*, 1299-1309.

Kaliman, A.V., Kulshin, V.E., Shlyapnikov, M.G., Ksenzenko, V.N., and Kryukov, V.M. (1995). The nucleotide sequence of the bacteriophage T5 ltf gene. FEBS Lett *366*, 46-48.

Kaplan, D.L. (2000). The 3'-tail of a forked-duplex sterically determines whether one or two DNA strands pass through the central channel of a replication-fork helicase. J Mol Biol *301*, 285-299.

Kaplan, D.L., Davey, M.J., and O'Donnell, M. (2003). Mcm4,6,7 uses a "pump in ring" mechanism to unwind DNA by steric exclusion and actively translocate along a duplex. The Journal of biological chemistry *278*, 49171-49182.

Kawabe, T., Tsuyama, N., Kitao, S., Nishikawa, K., Shimamoto, A., Shiratori, M., Matsumoto, T., Anno, K., Sato, T., Mitsui, Y.*, et al.* (2000). Differential regulation of human RecQ family helicases in cell transformation and cell cycle. Oncogene *19*, 4764-4772.

Kim, D.E., Narayan, M., and Patel, S.S. (2002). T7 DNA helicase: a molecular motor that processively and unidirectionally translocates along single-stranded DNA. J Mol Biol *321*, 807-819.

Kim, S., Dallmann, H.G., McHenry, C.S., and Marians, K.J. (1996). Coupling of a replicative polymerase and helicase: a tau-DnaB interaction mediates rapid replication fork movement. Cell *84*, 643-650.

Kleymann, G., Fischer, R., Betz, U.A., Hendrix, M., Bender, W., Schneider, U., Handke, G., Eckenberg, P., Hewlett, G., Pevzner, V.*, et al.* (2002). New helicase-primase inhibitors as drug candidates for the treatment of herpes simplex disease. Nature medicine *8*, 392-398.

Korolev, S., Yao, N., Lohman, T.M., Weber, P.C., and Waksman, G. (1998). Comparisons between the structures of HCV and Rep helicases reveal structural similarities between SF1 and SF2 super-families of helicases. Protein science : a publication of the Protein Society *7*, 605-610.

Kwong, A.D., Rao, B.G., and Jeang, K.-T. (2005). Viral and cellular RNA helicases as antiviral targets. Nat Rev Drug Discov *4*, 845-853.

Lanni, Y. (1969). Functions of two genes in the first-step-transfer DNA of bacteriophage T5. J Mol Biol *44*, 173-183.

Laskey, R.A., and Madine, M.A. (2003). A rotary pumping model for helicase function of MCM proteins at a distance from replication forks. EMBO reports *4*, 26-30.

Lee, J.Y., and Yang, W. (2006). UvrD helicase unwinds DNA one base pair at a time by a two-part power stroke. Cell *127*, 1349-1360.

Lehoux, M., Fradet-Turcotte, A., Lussier-Price, M., Omichinski, J.G., and Archambault, J. (2012). Inhibition of human papillomavirus DNA replication by an E1-derived p80/UAF1-binding peptide. Journal of Virology.

Levin, M.K., Gurjar, M., and Patel, S.S. (2005). A Brownian motor mechanism of translocation and strand separation by hepatitis C virus helicase. Nat Struct Mol Biol *12*, 429-435.

Levin, M.K., Gurjar, M.M., and Patel, S.S. (2003). ATP binding modulates the nucleic acid affinity of hepatitis C virus helicase. The Journal of biological chemistry *278*, 23311-23316.

Levin, M.K., Wang, Y.H., and Patel, S.S. (2004). The functional interaction of the hepatitis C virus helicase molecules is responsible for unwinding processivity. The Journal of biological chemistry *279*, 26005-26012.

Li, D., Zhao, R., Lilyestrom, W., Gai, D., Zhang, R., DeCaprio, J.A., Fanning, E., Jochimiak, A., Szakonyi, G., and Chen, X.S. (2003). Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen. Nature *423*, 512-518.

Liao, J.C., Jeong, Y.J., Kim, D.E., Patel, S.S., and Oster, G. (2005). Mechanochemistry of t7 DNA helicase. J Mol Biol *350*, 452-475.

Lionnet, T., Spiering, M.M., Benkovic, S.J., Bensimon, D., and Croquette, V. (2007). Real-time observation of bacteriophage T4 gp41 helicase reveals an unwinding mechanism. Proc Natl Acad Sci U S A *104*, 19790-19795.

Lohman, T.M., and Bjornson, K.P. (1996). Mechanisms of helicase-catalyzed DNA unwinding. Annu Rev Biochem *65*, 169-214.

Long, D.T., and Kreuzer, K.N. (2009). Fork regression is an active helicase-driven pathway in bacteriophage T4. EMBO reports *10*, 394-399.

Maga, G., Gemma, S., Fattorusso, C., Locatelli, G.A., Butini, S., Persico, M., Kukreja, G., Romano, M.P., Chiasserini, L., Savini, L., *et al.* (2005). Specific targeting of hepatitis C virus NS3 RNA helicase. Discovery of the potent and selective competitive nucleotide-mimicking inhibitor QU663. Biochemistry *44*, 9637-9644.

Maluf, N.K., Fischer, C.J., and Lohman, T.M. (2003). A Dimer of Escherichia coli UvrD is the active form of the helicase in vitro. J Mol Biol *325*, 913-935.

Manosas, M., Xi, X.G., Bensimon, D., and Croquette, V. (2010). Active and passive mechanisms of helicases. Nucleic acids research *38*, 5518-5526.

Martin, A., Baker, T.A., and Sauer, R.T. (2005). Rebuilt AAA + motors reveal operating principles for ATP-fuelled machines. Nature *437*, 1115-1120.

McCorquodale, D.J. (1999). T5-LIKE PHAGES (SIPHOVIRIDAE). In Encyclopedia of Virology (Second Edition), G. Editors-in-Chief: Allan, and G.W. Robert, eds. (Oxford: Elsevier), pp. 1716-1722.

McCorquodale, D.J., Gossling, J., Benzinger, R., Chesney, R., Lawhorne, L., and Moyer, R.W. (1979). Gene D5 product of bacteriophage T5: DNA-binding protein affecting DNA replication and late gene expression. J Virol *29*, 322-327.

McCorquodale, D.J., Shaw, A.R., Shaw, P.K., and Chinnadurai, G. (1977). Pre-early polypeptides of bacteriophages T5 and BF23. J Virol *22*, 480-488.

McCorquodale, D.J., and Warner, H.R. (1988). Bacteriophage T5 and related phages. In The Bacteriophages, R. Calender, ed. (New York: Plenum Press), pp. 439-475.

McGeoch, A.T., Trakselis, M.A., Laskey, R.A., and Bell, S.D. (2005). Organization of the archaeal MCM complex on DNA and implications for the helicase mechanism. Nat Struct Mol Biol *12*, 756-762.

McGlynn, P., and Lloyd, R.G. (2001). Rescue of stalled replication forks by RecG: simultaneous translocation on the leading and lagging strand templates supports an active DNA unwinding model of fork reversal and Holliday junction formation. Proc Natl Acad Sci U S A *98*, 8227-8234.

McKay, R.D. (1981). Binding of a simian virus 40 T antigen-related protein to DNA. J Mol Biol *145*, 471-488.

Mikoulinskaia, G.V., Gubanov, S.I., Zimin, A.A., Kolesnikov, I.V., Feofanov, S.A., and Miroshnikov, A.I. (2003). Purification and characterization of the deoxynucleoside monophosphate kinase of bacteriophage T5. Protein Expr Purif *27*, 195-201.

Mondigler, M., Vogele, R.T., and Heller, K.J. (1995). Overproduced and purified receptor binding protein pb5 of bacteriophage T5 binds to the T5 receptor protein FhuA. FEMS Microbiol Lett *130*, 293-300.

Mozer, T.J., Thompson, R.B., Berget, S.M., and Warner, H.R. (1977). Isolation and characterization of a bacteriophage T5 mutant deficient in deoxynucleoside 5'-monophosphatase activity. J Virol *24*, 642-650.

Mozer, T.J., and Warner, H.R. (1977). Properties of deoxynucleoside 5'-monophosphatase induced by bacteriophage T5 after infection of Escherichia coli. J Virol *24*, 635-641.

Nanduri, B., Byrd, A.K., Eoff, R.L., Tackett, A.J., and Raney, K.D. (2002). Pre-steady-state DNA unwinding by bacteriophage T4 Dda helicase reveals a monomeric molecular motor. Proc Natl Acad Sci U S A *99*, 14722-14727.

Nelson, S.W., and Benkovic, S.J. (2007). The T4 phage UvsW protein contains both DNA unwinding and strand annealing activities. The Journal of biological chemistry *282*, 407-416.

Nelson, S.W., Perumal, S.K., and Benkovic, S.J. (2009). Processive and unidirectional translocation of monomeric UvsW helicase on single-stranded DNA. Biochemistry *48*, 1036-1046.

Nicholas, J. (1994). Nucleotide sequence analysis of a 21-kbp region of the genome of human herpesvirus-6 containing homologues of human cytomegalovirus major immediate-early and replication genes. Virology *204*, 738-750.

Nichols, B.P., and Donelson, J.E. (1977). Sequence analysis of the nicks and termini of bacteriophage T5 DNA. J Virol *22*, 520-526.

Nonin, S., Leroy, J.L., and Gueron, M. (1995). Terminal base pairs of oligodeoxynucleotides: imino proton exchange and fraying. Biochemistry *34*, 10652-10659.

Opresko, P.L., Cheng, W.H., and Bohr, V.A. (2004). Junction of RecQ helicase biochemistry and human disease. The Journal of biological chemistry *279*, 18099-18102.

Patel, S.S. (2009). Structural biology: Steps in the right direction. Nature *462*, 581-583.

Patel, S.S., and Donmez, I. (2006). Mechanisms of helicases. The Journal of biological chemistry *281*, 18265-18268.

Patel, S.S., and Hingorani, M.M. (1993). Oligomeric structure of bacteriophage T7 DNA primase/helicase proteins. The Journal of biological chemistry *268*, 10668-10675.

Pause, A., and Sonenberg, N. (1992). Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. The EMBO journal *11*, 2643-2654.

Perumal, S.K., Raney, K.D., and Benkovic, S.J. (2010). Analysis of the DNA translocation and unwinding activities of T4 phage helicases. Methods *51*, 277-288.

Pradhan, A., and Tuteja, R. (2006). Plasmodium falciparum DNA helicase 60. dsRNA- and antibody-mediated inhibition of malaria parasite growth and downregulation of its enzyme activities by DNA-interacting compounds. The FEBS journal *273*, 3545-3556.

Pyle, A.M. (2008). Translocation and unwinding mechanisms of RNA and DNA helicases. Annual review of biophysics *37*, 317-336.

Pyle, A.M. (2009). How to drive your helicase in a straight line. Cell *139*, 458-459.

Randell, J.C., Bowers, J.L., Rodriguez, H.K., and Bell, S.P. (2006). Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase. Molecular cell *21*, 29-39.

Rhoades, M. (1977). Localization of single-chain interruptions in bacteriophage T5 DNA. II. Electrophoretic studies. J Virol *23*, 737-750.

Rice, A.C., Ficht, T.A., Holladay, L.A., and Moyer, R.W. (1979). The purification and properties of a double-stranded DNA-binding protein encoded by the gene D5 of bacteriophage T5. The Journal of biological chemistry *254*, 8042-8051.

Rogers, S.G., Godwin, E.A., Shinosky, E.S., and Rhoades, M. (1979a). Interruption-deficient mutants of bacteriophage T5 I. Isolation and general properties. J Virol *29*, 716-725.

Rogers, S.G., Hamlett, N.V., and Rhoades, M. (1979b). Interruption-deficient mutants of bacteriophage T5. II. Properties of a mutant lacking a specific interruption. J Virol *29*, 726-734.

Rogers, S.G., and Rhoades, M. (1976). Bacteriophage T5-induced endonucleases that introduce site-specific single-chain interruptions in duplex DNA. Proc Natl Acad Sci U S A *73*, 1576-1580.

Rokita, S.E. (2001). Chemical reagents for investigating the major groove of DNA. Current protocols in nucleic acid chemistry / edited by Serge L. Beaucage ... [et al.] *Chapter 6*, Unit 6 6.

Sabir, T., Schroder, G.F., Toulmin, A., McGlynn, P., and Magennis, S.W. (2011). Global structure of forked DNA in solution revealed by high-resolution single-molecule FRET. Journal of the American Chemical Society *133*, 1188-1191.

Sabir, T., Toulmin, A., Ma, L., Jones, A.C., McGlynn, P., Schröder, G.F., and Magennis, S.W. (2012). Branchpoint expansion in a fully-complementary three-way DNA junction. Journal of the American Chemical Society.

Sakaki, Y. (1974). Inactivation of the ATP-dependent DNase of Escherichia coli after infection with double-stranded DNA phages. J Virol *14*, 1611-1612.

Sayers, J.R. (2005). Bacteriophage T5. In The Bacteriophages, R. Calender, ed. (New York: Oxford University Press), pp. 268-276.

Sayers, J.R., and Eckstein, F. (1990). Properties of overexpressed phage T5 D15 exonuclease. Similarities with Escherichia coli DNA polymerase I 5'-3' exonuclease. The Journal of biological chemistry *265*, 18311-18317.

Scheible, P.P., Rhoades, E.A., and Rhoades, M. (1977). Localization of single-chain interruptions in bacteriophage T5 DNA I. Electron microscopic studies. J Virol *23*, 725-736.

Schwartz, A., Rabhi, M., Jacquinot, F., Margeat, E., Rahmouni, A.R., and Boudvillain, M. (2009). A stepwise 2[prime]-hydroxyl activation mechanism for the bacterial transcription termination factor Rho helicase. Nat Struct Mol Biol *16*, 1309-1316.

Seidel, R., Bloom, J.G.P., Dekker, C., and Szczelkun, M.D. (2008). Motor step size and ATP coupling efficiency of the dsDNA translocase EcoR124I. The EMBO journal *27*, 1388-1398.

Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S., and Yokoyama, S. (2006). Structural Basis for RNA Unwinding by the DEAD-Box Protein Drosophila Vasa. Cell *125*, 287-300.

Sharma, S., Sommers, J.A., Gary, R.K., Friedrich-Heineken, E., Hubscher, U., and Brosh, R.M., Jr. (2005). The interaction site of Flap Endonuclease-1 with WRN helicase suggests a coordination of WRN and PCNA. Nucleic acids research *33*, 6769-6781.

Shlyapnikov, M.G., Ksenzenko, V.N., Kryukov, V.M., and Bayev, A.A. (1995). Specific properties of phage T5-encoded tRNAs. Molecular Biology *28*, 818-823.

Singleton, M.R., Dillingham, M.S., Gaudier, M., Kowalczykowski, S.C., and Wigley, D.B. (2004). Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. Nature *432*, 187-193.

Singleton, M.R., Dillingham, M.S., and Wigley, D.B. (2007). Structure and mechanism of helicases and nucleic acid translocases. Annu Rev Biochem *76*, 23-50.

Singleton, M.R., Sawaya, M.R., Ellenberger, T., and Wigley, D.B. (2000). Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. Cell *101*, 589-600.

Skordalakes, E., and Berger, J.M. (2003). Structure of the Rho transcription terminator: mechanism of mRNA recognition and helicase loading. Cell *114*, 135-146.

Snyder, C.E., Jr. (1984). Bacteriophage T5 gene A2 protein alters the outer membrane of Escherichia coli. J Bacteriol *160*, 1191-1195.

Snyder, C.E., Jr., and Benzinger, R.H. (1981). Second-step transfer of bacteriophage T5 DNA: purification and characterization of the T5 gene A2 protein. J Virol *40*, 248-257.

Soultanas, P., Dillingham, M.S., Wiley, P., Webb, M.R., and Wigley, D.B. (2000). Uncoupling DNA translocation and helicase activity in PcrA: direct evidence for an active mechanism. The EMBO journal *19*, 3799-3810.

Soultanas, P., and Wigley, D.B. (2000). DNA helicases: 'inching forward'. Curr Opin Struct Biol *10*, 124-128.

Soultanas, P., and Wigley, D.B. (2001). Unwinding the 'Gordian knot' of helicase action. Trends in biochemical sciences *26*, 47-54.

Spies, M., Amitani, I., Baskin, R.J., and Kowalczykowski, S.C. (2007). RecBCD Enzyme Switches Lead Motor Subunits in Response to χ Recognition. Cell *131*, 694-705.

Spurling, T.L., Eoff, R.L., and Raney, K.D. (2006). Dda helicase unwinds a DNA-PNA chimeric substrate: evidence for an inchworm mechanism. Bioorganic & medicinal chemistry letters *16*, 1816-1820.

Stankiewicz-Drogoń, A., Dörner, B., Erker, T., and Boguszewska-Chachulska, A.M. (2010). Synthesis of New Acridone Derivatives, Inhibitors of NS3 Helicase, Which Efficiently and Specifically Inhibit Subgenomic HCV Replication. Journal of Medicinal Chemistry *53*, 3117-3126.

Stano, N.M., Jeong, Y.J., Donmez, I., Tummalapalli, P., Levin, M.K., and Patel, S.S. (2005). DNA synthesis provides the driving force to accelerate DNA unwinding by a helicase. Nature *435*, 370-373.

Stitt, B.L., and Xu, Y. (1998). Sequential hydrolysis of ATP molecules bound in interacting catalytic sites of Escherichia coli transcription termination protein Rho. The Journal of biological chemistry *273*, 26477-26486.

Subramanya, H.S., Bird, L.E., Brannigan, J.A., and Wigley, D.B. (1996). Crystal structure of a DExx box DNA helicase. Nature *384*, 379-383.

Sun, B., Wei, K.J., Zhang, B., Zhang, X.H., Dou, S.X., Li, M., and Xi, X.G. (2008). Impediment of E. coli UvrD by DNA-destabilizing force reveals a strained-inchworm mechanism of DNA unwinding. The EMBO journal *27*, 3279-3287.

Szabo, C., Dharmgrongartama, B., and Moyer, R.W. (1975). The regulation of transcription in bacteriophage T5-infected Escherichia coli. Biochemistry *14*, 989-997.

Szabo, C., and Moyer, R.W. (1975). Purification and properties of a bacteriophage T5-modified form of Escherichia coli RNA polymerase. J Virol *15*, 1042-1046.

Tackett, A.J., Chen, Y., Cameron, C.E., and Raney, K.D. (2005). Multiple full-length NS3 molecules are required for optimal unwinding of oligonucleotide DNA in vitro. The Journal of biological chemistry *280*, 10797-10806.

Takahashi, T.S., Wigley, D.B., and Walter, J.C. (2005). Pumps, paradoxes and ploughshares: mechanism of the MCM2-7 DNA helicase. Trends in biochemical sciences *30*, 437-444.

Tanner, J.A., Zheng, B.J., Zhou, J., Watt, R.M., Jiang, J.Q., Wong, K.L., Lin, Y.P., Lu, L.Y., He, M.L., Kung, H.F.*, et al.* (2005). The adamantane-derived bananins are potent inhibitors of the helicase activities and replication of SARS coronavirus. Chemistry & biology *12*, 303-311.

Tanner, N.K., Cordin, O., Banroques, J., Doere, M., and Linder, P. (2003). The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis. Molecular cell *11*, 127-138.

Thoma, N.H., Czyzewski, B.K., Alexeev, A.A., Mazin, A.V., Kowalczykowski, S.C., and Pavletich, N.P. (2005). Structure of the SWI2/SNF2 chromatin-remodeling domain of eukaryotic Rad54. Nat Struct Mol Biol *12*, 350-356.

Thomsen, N.D., and Berger, J.M. (2009). Running in Reverse: The Structural Basis for Translocation Polarity in Hexameric Helicases. Cell *139*, 523-534.

Tomko, E.J., Fischer, C.J., Niedziela-Majka, A., and Lohman, T.M. (2007). A nonuniform stepping mechanism for E. coli UvrD monomer translocation along single-stranded DNA. Molecular cell *26*, 335-347.

Turley, H., Wu, L., Canamero, M., Gatter, K.C., and Hickson, I.D. (2001). The distribution and expression of the Bloom's syndrome gene product in normal and neoplastic human cells. British journal of cancer *85*, 261-265.

Tuteja, N., and Tuteja, R. (2004). Prokaryotic and eukaryotic DNA helicases. Essential molecular motor proteins for cellular machinery. European journal of biochemistry / FEBS *271*, 1835-1848.

Tuteja, R. (2007). Helicases - feasible antimalarial drug target for Plasmodium falciparum. The FEBS journal *274*, 4699-4704.

Valle, M., Chen, X.S., Donate, L.E., Fanning, E., and Carazo, J.M. (2006). Structural basis for the cooperative assembly of large T antigen on the origin of replication. J Mol Biol *357*, 1295-1305.

Velankar, S.S., Soultanas, P., Dillingham, M.S., Subramanya, H.S., and Wigley, D.B. (1999). Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. Cell *97*, 75-84.

von Hippel, P.H., and Delagoutte, E. (2001). A general model for nucleic acid helicases and their "coupling" within macromolecular machines. Cell *104*, 177-190.

Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. The EMBO journal *1*, 945-951.

Wang, H., Kim, S., and Ryu, W.S. (2009). DDX3 DEAD-Box RNA helicase inhibits hepatitis B virus reverse transcription by incorporation into nucleocapsids. J Virol *83*, 5815-5824.

Wang, J., Jiang, Y., Vincent, M., Sun, Y., Yu, H., Bao, Q., Kong, H., and Hu, S. (2005). Complete genome sequence of bacteriophage T5. Virology *332*, 45-65.

Warner, H.R., Drong, R.F., and Berget, S.M. (1975). Early events after infection of Escherichia coli by bacteriophage T5. Induction of a 5'-nucleotidase activity and excretion of free bases. J Virol *15*, 273-280.

Warner, H.R., Johnson, L.K., and Snustad, D.P. (1980). Early events after infection of Escherichia coli by bacteriophage T5. III. Inhibition of uracil-DNA glycosylase activity. J Virol *33*, 535-538.

Webb, M.R., Plank, J.L., Long, D.T., Hsieh, T.S., and Kreuzer, K.N. (2007). The phage T4 protein UvsW drives Holliday junction branch migration. The Journal

of biological chemistry *282*, 34401-34411.

Wessel, R., Schweizer, J., and Stahl, H. (1992). Simian virus 40 T-antigen DNA helicase is a hexamer which forms a binary complex during bidirectional unwinding from the viral origin of DNA replication. J Virol *66*, 804-815.

Whelan, F., Stead, J.A., Shkumatov, A.V., Svergun, D.I., Sanders, C.M., and Antson, A.A. (2011). A flexible brace maintains the assembly of a hexameric replicative helicase during DNA unwinding. Nucleic acids research.

Whitby, M.C., and Lloyd, R.G. (1998). Targeting Holliday junctions by the RecG branch migration protein of Escherichia coli. The Journal of biological chemistry *273*, 19729-19739.

Wong, I., and Lohman, T.M. (1992). Allosteric effects of nucleotide cofactors on Escherichia coli Rep helicase-DNA binding. Science *256*, 350-355.

Xu, H.Q., Deprez, E., Zhang, A.H., Tauc, P., Ladjimi, M.M., Brochon, J.C., Auclair, C., and Xi, X.G. (2003). The Escherichia coli RecQ helicase functions as a monomer. The Journal of biological chemistry *278*, 34925-34933.

Yao, N.Y., Georgescu, R.E., Finkelstein, J., and O'Donnell, M.E. (2009). Single-molecule analysis reveals that the lagging strand increases replisome processivity but slows replication fork progression. Proc Natl Acad Sci U S A *106*, 13236-13241.

Yedavalli, V.S.R.K., Neuveut, C., Chi, Y.-h., Kleiman, L., and Jeang, K.-T. (2004). Requirement of DDX3 DEAD Box RNA Helicase for HIV-1 Rev-RRE Export Function. Cell *119*, 381-392.

Yedavalli, V.S.R.K., Zhang, N., Cai, H., Zhang, P., Starost, M.F., Hosmane, R.S., and Jeang, K.-T. (2008). Ring Expanded Nucleoside Analogues Inhibit RNA Helicase and Intracellular Human Immunodeficiency Virus Type 1 Replication. Journal of Medicinal Chemistry *51*, 5043-5051.

Yu, J., Ha, T., and Schulten, K. (2006). Structure-based model of the stepping motor of PcrA helicase. Biophysical journal *91*, 2097-2114.

Zhou, Y., Duncan, T.M., and Cross, R.L. (1997). Subunit rotation in Escherichia coli FoF1-ATP synthase during oxidative phosphorylation. Proc Natl Acad Sci U S A *94*, 10583-10587.