1 Introduction

1.1 RepA DNA helicase

Helicases are ubiquitious motor proteins essential in key biological processes which require single-stranded DNA (ssDNA) such as DNA replication (Fig. 1.1), transcription, repair and recombination (Matson et al., 1990; Lohman et al., 1996). The basic reaction catalyzed by this family of enzymes is the unwinding of the duplex form of DNA or RNA, a process coupled to nucleoside triphosphate (NTP) hydrolysis. The unwinding of double-stranded DNA (dsDNA) by helicases is either in $5' \rightarrow 3'$ or in $3' \rightarrow 5'$ direction. Since the discovery of the first helicase from *Escherichia coli* (Abdel-Monem et al., 1976), a growing number of helicase proteins which are involved in many aspects of DNA metabolism in bacterial, viral and eukaryotic systems have now been characterized *in vitro*. In humans, malfunction of certain DNA helicases is associated with several severe diseases including Bloom´s syndrome, *Xeroderma Pigmentosum* and Werner´s syndrome (Friedberg, 1992; Hanawalt, 1994; Yu et al., 1996), with the development of cancer (Egelman, 1996), and with ageing (Bowles, 1998).

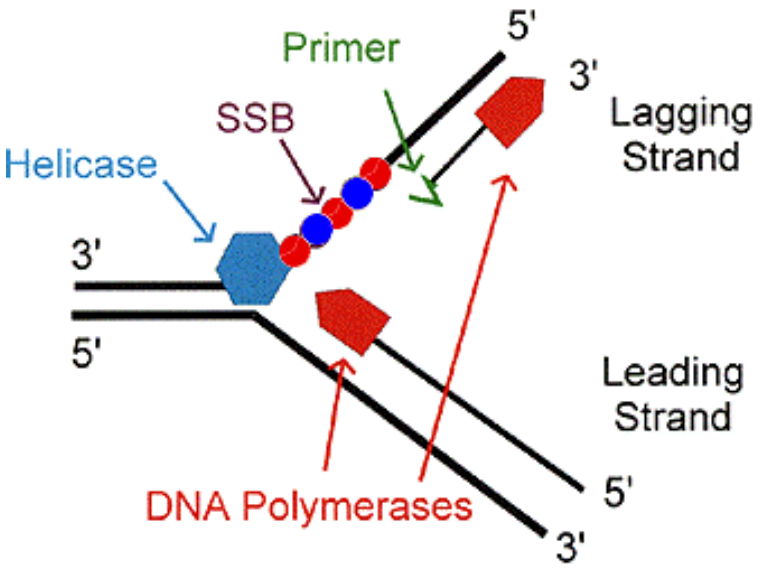

Figure 1.1: DNA replication fork. The two ssDNA stands unwound by DNA helicases serve as templates for the DNA polymerase to synthesize new, complementary DNA (leading and lagging strand).

The replicative hexameric helicase RepA is encoded by the broad host range plasmid RSF1010, an 8684-base-pair (bp) multicopy plasmid that can replicate in a wide variety of Gram-negative and also Gram-positive actinomyces (Scholz et al., 1989). Three plasmid-encoded proteins, RepA, RepB´, and RepC, exhibit RSF1010-specific helicase, primase and initiator protein activities, repectively, and are essential for the replication of this plasmid. *E. coli* DNA gyrase, SSB (single strand DNA binding protein) and the production of dnaZ (γ subunit of DNA polymerase III holoenzyme) are also required for replication of plasmid RSF1010, while the bacterial RNA polymerase and the DnaA, B, C, G, T proteins are not (Scherzinger et al., 1991). RepA is one of the smallest known homohexameric helicase enzymes with a total molecular mass of 180 kDa (Scherzinger et al., 1997). It unwinds dsDNA in 5'→ 3' polarity and prefers a tailed substrate with an unpaired 3'- tail mimicking a DNA replication fork. The RepA activity is fueled by ATP, dATP, GTP, and dGTP and less efficiently by CTP and dCTP while UTP and dTTP are poor effectors. Optimal unwinding activity was found at a narrow pH range around 5.5 (Scherzinger et al., 1997), similar as observed for yeast *Saccharomyces cerevisiae* RAD3 helicase (Sung et al., 1988). Below pH 5.6 and at low salts concentration, the RepA hexamers aggregate and form tubular structures (Röleke et al., 1997).

1.2 Structural features of DNA helicase

All DNA helicases for which the assembly state of the enzyme has been examined appear to function as oligomers, generally dimers or hexamers, thus providing multiple potential DNA binding sites, which are required for helicase function. The *E. coli* DnaB protein (San Martin et al., 1995; Yu et al., 1996), the bacteriophage T7 gp4 protein (Egelman et al., 1995), the *E.coli* branch migration RuvB protein (Stasiak et al., 1994), and the *E. coli* transcription termination protein Rho (Gogol et al., 1991) assemble as hexamers into a ring shape; by contrast, the *E. coli* proteins Rep (Wong et al., 1992) and UvrD (Runyon et al., 1993) (Helicase II) have been characterized as dimers. Detailed information is shown in Table1.

Table 1.1: Characteristics of DNA and RNA helicases.

Protein	Family group	Molecular weight	Direction of unwinding	Assembly state	Minimal requirement for oligomer formation
E.coli bacteriophage T7 gp4	DnaB-like	4A 62,655 4B 55,743	5' to 3'	Hexamer	dTTP, dTDP, dTMP-PCP, ATP, dATP
E. coli bacteriophage T4 gp41	DnaB-like	53,601	5' to 3'	Hexamer	ATP, GTP, ATPγS,GTPγS
RSF1010 RepA	DnaB-like	29,909	5' to 3'	Hexamer	None
E. coli DnaB	DnaB-like	52,390	5' to 3'	Hexamer	Mg^{2+}
E. coli RuvB	AAA^{+} family	37,174	5' to 3'	Hexamer	Mg^{2+}
E. coli rho	F1-ATPase	47,004	5' to 3'	Hexamer	RNA
B. subtilis phage SPP1 gene 40	DnaB-like	46,746	5' to 3'	Hexamer	ATP, Mg^{2+}
Simian virus large T antigen	SF III	81,907	3' to 5'	Hexamer	ATP, ADP, ATPγS
Bovine papillomavirus E1	papillomavirus family	68,246	3' to 5'	Hexamer	DNA
Human Bloom's syndrome helicase	SFII	159,000	3' to 5'	Hexamer	ATPγS, Mg^{2+}
Human MCM4	MCM family	96,606	3' to 5'	Hexamer	Not known
E. coli Rep	SF I	68,000	3' to 5'	Dimer	DNA
E. coli Helicase II (UvrD)	SF I	82,116	3' to 5'	Dimer	DNA

Concerning the primary amino acid sequences, all known helicases have been grouped into families and superfamilies (SF) (Gorbalenya et al., 1993; Ilyina et al., 1992). For example, *E. coli* Rep, PcrA, and *E. coli* UvrD belong to superfamily I (SF1); the BLM helicase belongs to the Rec-Q family in SF2; the viral large T antigens belong to the SF3 superfamily. The hexameric helicases do not fall within a single family, but are members of various families.

For example, the helicases RepA, DnaB, P1 Ban, SSP1 G40P, T7 gp4, and T4 gp41 belong to the DnaB-like family; and the rho protein belongs to the V- and F1-ATPase family. The smaller family of DnaB-like helicases shows five conserved motifs H1, H1A, H2, H3, H4 (Ilyina et al., 1992). The conserved H1 and H2 motifs contain the Walker A and B sequences (Walker et al., 1982). A highly conserved and essential lysine in H1 and aspartic acid in H2 have been identified from mutagenesis and structural studies. A conserved glutamic acid in H1a has been proposed to play a role in NTP hydrolysis (Sawaya et al.,199; Story et al., 1992; Abrahams et al., 1994). Little is known about the exact role of H3 and H4 may be involved in DNA binding (Washington et al., 1996). Additionally, there are several residues beyond H4 which show some sequence conservation, and they are involved in nucleotide binding and hydrolysis (Patel et al., 2000).

Several high-resolution structures have been reported for monomeric/dimeric helicases. A common helicase architecture has been revealed by crystal structures of two DNA helicases, B*acillus stearothermophilus* PcrA (Subramanya et al., 1996; Velankar et al., 1999) and E. *coli* Rep (Korolev et al., 1997), and an RNA helicase, the hepatitis C virus NS3 protein (Yao et al., 1997; Cho et al., 1998; Kim et al., 1997). The recently determined crystal structures of PcrA complexed with a DNA substrate have revealed details of the helicase mechanism (Soultanas et al., 2000). Of the intact, unmodified replicative hexameric helicases, information is only limited to low-resolution electron microscopic studies (Fig. 1.2). Crystal structure analyses of truncated domains of two hexameric helicases, DnaB (Fass et al., 1999) and T7 bacteriophage helicase-primase (Sawaya et al., 1999; Singleton et al., 2000) have been reported. In the structure of hexameric fragment of gene 4 helicase from bacteriophage T7, the deviation from expected six-fold symmetry of the hexamer indicates that the structure represents an intermediate on the catalytic pathway. The structural consequences of the asymmetry suggest a "binding change" mechanism to explain how cooperative binding and hydrolysis of nucleotides are coupled to conformational changes in the ring that most likely accompany duplex unwinding.

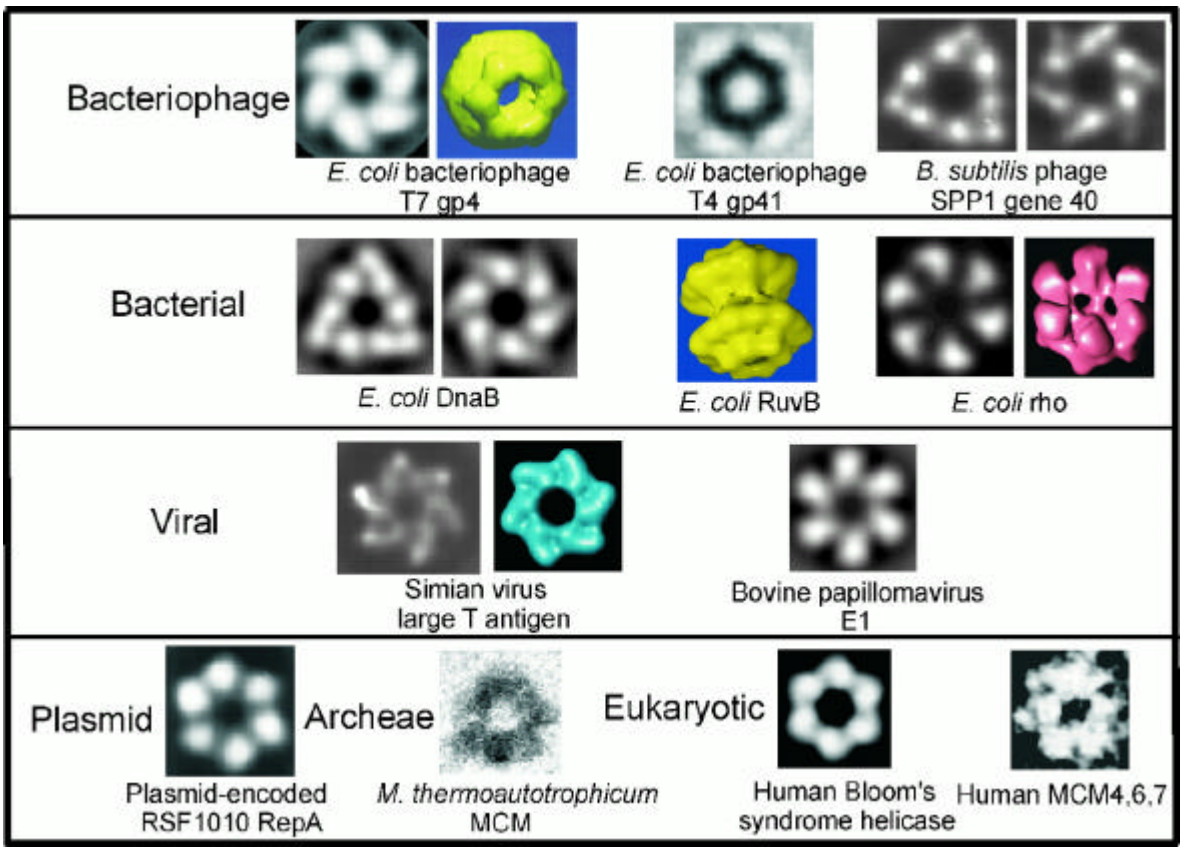

Figure 1.2: Electron microscopy images of hexameric helicases (Patel et al., 2000).

1.3 Nucleotide binding and NTP hydrolysis of helicases

All DNA helicases possess a consensus NTP binding site, as indicated by the presence of the conserved Walker A and B motif (I and II) (Gorbalenya et al., 1993). For helicase-catalyzed unwinding of dsDNA, NTP binding and hydrolysis are essential. The NTPs can function as switches that induce conformational changes necessary to promote DNA binding and release-steps required for translocation of helicase along the DNA double helix (Wong et al., 1992). Until now the molecular mechanism by which NTP binding and hydrolysis are coupled to DNA unwinding is poorly understood. Homo-oligomeric helicases feature at least one potential NTP binding site per subunit and nucleotides appear to bind to the oligomer with negative cooperativity or possibly two classes of “high” and “low” affinity sites (Patel et al., 2000). It is a common feature of the hexameric helicases that the six potential nucleotide binding sites are nonequivalent and can be clearly distinguished as three high affinity and three low affinity binding sites, as found for DnaB helicase (Bujalowski et al., 1993) and for transcription termination protein Rho (Geiselmann et al., 1992). In the hexameric

bacteriophage T7 DNA helicase, only three high affinity nucleotide binding sites per hexamer are observed, the low affinity binding sites being nearly undetectable (Hingorani et al., 1996).

The conversion of NTP to NDP and phosphate by helicase hydrolysis is necessary for helicase movement and DNA unwinding. Generally, Mg^{2+} is necessary for NTP hydrolysis. DNA or RNA greatly stimulates NTP hydrolysis from 10 to 100 fold. The K_m for NTP ranges from a few micromolar to a few millimolar, and the stimulated k_{cat} is determined to be as low as 0.6 s^{-1} and as high as 30 s^{-1} (Patel et al., 2000).

To understand how the NTPase reaction at the catalytic sites of helicases is coupled to movement along and unwinding of dsDNA, it is import to obtain a complete description of the NTPase pathway and it is necessary to identify the steps in the NTPase reaction that lead to nucleic acid duplex binding, release, movement, and unwinding (Patel et al., 2000). In hexamer helicases, determining the NTPase pathway is complicated because there are potentially six sites on the hexamer that can hydrolyse NTP. In T7 gp4, Rho, and F_1-ATPase, it was shown that three sites are noncatalytic and bind normally to NTP but impair hydrolysis. A model for NTP hydrolysis similar to that of F1 –ATPase was proposed for T7 gp4 and *E. coli* Rho in which three catalytic sites hydrolyse ATP in a sequential manner (Hingorani et al., 1997; Kim et al., 1999; Boyer, 1997). The important feature of this mechanism is that each active site sequentially switches between different conformations, and the changes at the three sites are all linked.

1.4 DNA binding to helicases

An understanding of helicase mechanisms requires studies of the interactions of helicase with its DNA substrate, especially the binding oligomeric states, binding stoichiometries, binding affinities and the influence of nucleotide cofactors which will switch the energetics of protein or protein-DNA assembly and thus the distribution of assembly states. Qualitative studies showed that helicases generally bind with higher affinity to ssDNA than dsDNA (Arai et al., 1981; Das et al., 1980).

Most known hexameric helicases bind ssDNA, except for Rho protein that binds ssRNA. Many hexameric helicases also bind dsDNA, but in most the affinity for dsDNA is much weaker relative to ssDNA (Liu et al., 1981; Hingorani et al., 1993); T7 gp4 binds dsDNA with an affinity about 50-fold lower than ssDNA (Hingorani et al., 1993). The exceptions are SV40 large T antigen, BPV E1 protein, and *E. coli* RuvB, which bind dsDNA with high affinity (Dean et al., 1987; Wilson et al., 1991; Muller et al., 1993). DNA modification studies with

SV40 large T antigen have shown that the interactions of the protein with DNA are mainly with the sugar-phosphate backbone. In SV40 large T antigen, modification of the phosphate backbone abolished DNA binding but modification of bases had only a small effect (SenGupta et al., 1992).

The interactions of ssDNA or RNA with *E. coli* DnaB, E. *Coli* rho protein, and T7 gp4 have been studied extensively using fluorimetric, nitrocellulose membrane binding, and ultrafiltration methods (Geiselmann et al., 1992; Hingorani et al., 1993; Bujalowski et al., 1995; Jezewska et al., 1996; Wang et al., 1993). These helicases bind one strand of DNA or RNA tightly with a K_d ranging from 60 nM to 200 nM. They are able to bind a second strand with a weaker K_d in the range of 4 - 40 µM. Most helicases show sequence-independent binding to ssDNA because they are involved in processes such as DNA replication and recombination that require movement on DNA in a sequence-independent manner. For the hexameric helicases, most studies are consistent with DNA binding in the central hole, as shown by EM, chemical cross-linking, fluorescence resonance energy transfer and crystal stucture studies. T7gp4, T4gp41, and *E.coli* DnaB are belived to encircle ssDNA (Yu et al., 1996; Egelman et al., 1995; Morris et al., 1995; Bujalowski et al., 1995), whereas SV40 large T antigen is thought to encircle either ssDNA or dsDNA, and *E.coli* RuvB protein has been proposed to encicle dsDNA (Egelman et al., 1995; Stasiak et al., 1994; Mastrangelo et al., 1989). *E.coli* Rho is an exception because it appears to bind RNA by wrapping it around the six N-terminal RNA binding domains (Richardson et al., 1996; Bogden et al., 1999).

1.5 Mechanism of DNA translocation and unwinding

The exact mechanism by which helicases accomplish unwinding of dsDNA is not yet clear: however, significant progress has been made and intense biochemical, genetic, and structural approaches are currently being pursued to gain insight into both the mechanism and role of helicases in various biological processes. The key to deciphering their action is to understand how NTP hydrolysis is coupled to movements along DNA and base-pair seperation.

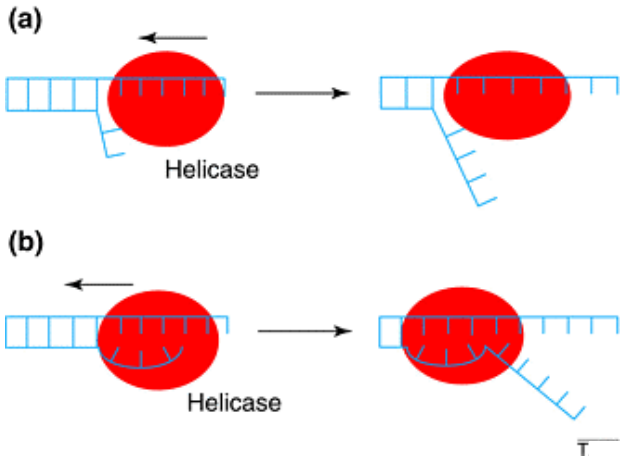

Figure 1.3: Passive and active mechanism for helicase unwinding. (a) In the passive mechanism, the helicase does not make direct contacts with the dsDNA. Instead, it operates by trapping ssDNA at a thermally fraying ss-dsDNA junction. (b) In the active mechanism the helicase interacts directly with the dsDNA and destabilizes the duplex, thus 'actively' unwinding the strands (Soultanas et al., 2001).

In general, mechanisms of helicase-catalyzed DNA unwinding can be classified as either "active" or "passive" (Fig. 1.3). In a passive mechanism, the helicase would faciliate unwinding indirectly by binding to ssDNA that becomes available through transient fraying of the duplex caused by thermal fluctuations at the ss/dsDNA junction. In an active mechanism, the helicase plays a direct role in destabilizing the duplex DNA. An active unwinding mechanism requires the functional helicase to feature at least two types of DNA binding. In one type, helicase would interact directly with the dsDNA at the junction fork and destabilize the base pairs actively through conformational changes of helicase triggered by NTP binding, hydrolysis, or product release. Sub-classes of this type are called "inch-worm" (Hill et al., 1981; Yarranton et al., 1979) and "rolling" models (Wong et al., 1992). A second type of this active mechanism is called "torsional" model in which the helicase does not interact with duplex DNA, but binds simultaneously to both of the single strands at the ss/dsDNA junction and unwinds dsDNA by distorting the adjacent duplex region through an NTP-induced conformational change (Lohman, 1992). Crystal structures of PcrA helicase complexed with a DNA substrate give a hint of the general model linking ATP binding and hydrolysis to DNA translocation for many helicases (Soultanas et al., 2000). In this model, the translocation process can be a "Mexican wave" in a soccer ground with spectators (DNA bases) jumping up and down as the enzyme creeps along the linear ssDNA axis (Fig. 1.4).

In general, the unwinding of dsDNA by the catalytic action of helicases involves unidirectional translocation and base-pair separation processes. The translocation along DNA involves a series of cyclical binding and release events, which implies that the affinity of

helicase for the DNA transiently changes during translocation. The helicase may cycle between DNA binding states of tight, weak, and no affinity using specific steps in the ATP hydrolysis cycle as "switches". However, too little is known about the base-pair separation process to propose models for NTPase-coupled DNA unwinding. Two general models are proposed based on possible interactions with the duplex DNA for the hexameric helicases (Patel et al., 2000).

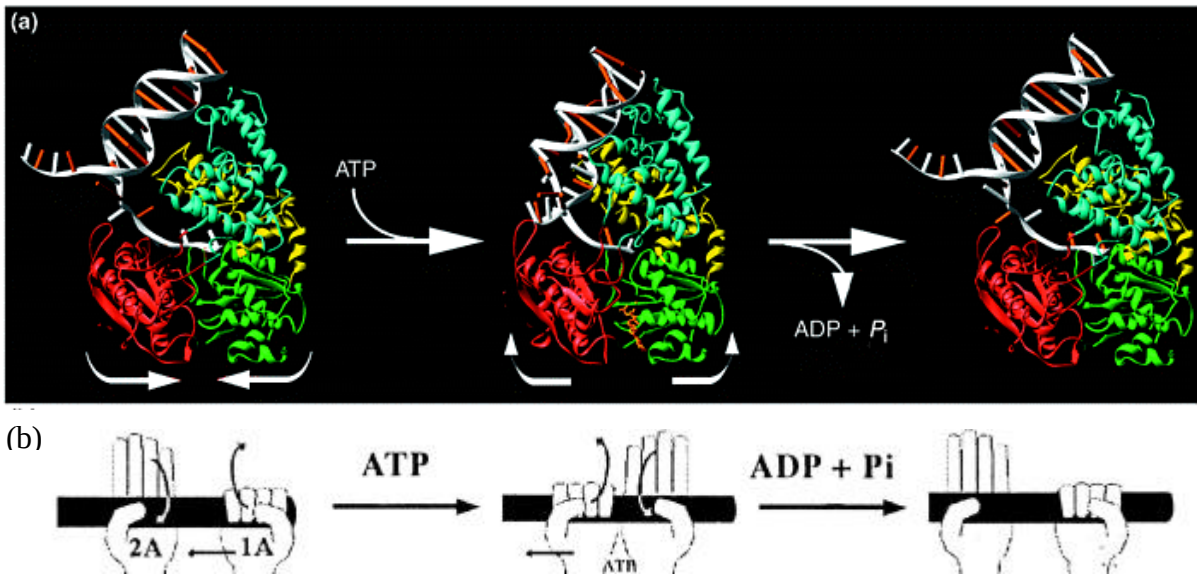

Figure 1.4: A model showing the conformational states of PcrA helicase during translocation giving rise to 'active inchworm' helicase activity. (a) Cartoons demonstrating the alteration in affinity for ssDNA of domains 1A (green) and 2A (red). (b) A cartoon representation of the above mechanism (Soultanas et al., 2000).

1.6 RepA – the first known intact hexameric helicase in 3D structure

RepA is the first intact hexameric replicative DNA helicase whose three-dimensional structure is known in detail (at 2.4Å resolution) (Niedenzu et al., 2001). The RepA hexamer is pot shaped and shows an annular structure with 6-fold rotational symmetry and a 17Å wide central hole, suggesting that only ssDNA can pass through during unwinding reaction (Fig. 1.5). Homologs of all five conserved sequence motifs of the DnaB-like helicase family are all found in RepA (Fig. 1.6). All the residues essential for triphosphate binding and hydrolysis in RepA are conserved in helicases and related enzymes utilizing NTP hydrolysis (Walker et al., 1982; Schulz, 1992). In a modeled ATP-RepA complex, the six ATP binding sites of RepA

are each located at the interfaces between two adjacent monomers and belong to the conserved Walker A and B motif. They are defined by the consensus sequence for the P loop (Saraste et al., 1990) ^{40}GAGKS^{44} and residues Asp140, Glu77, His179 (belonging to the same monomer) and Arg207 (from the adjacent monomer). The adenine base is found sandwiched between Arg86 of the same (in cis) and Tyr243 of the adjacent (in trans) monomer (Fig. 1.5c). Like the helicase domain of the bacteriophage T7 gp4 protein, the "arginine finger" Arg207 contributes to the active site in the adjacent monomer and suggests cooperativevity between monomers in ATP hydrolysis and helicase activity of RepA.

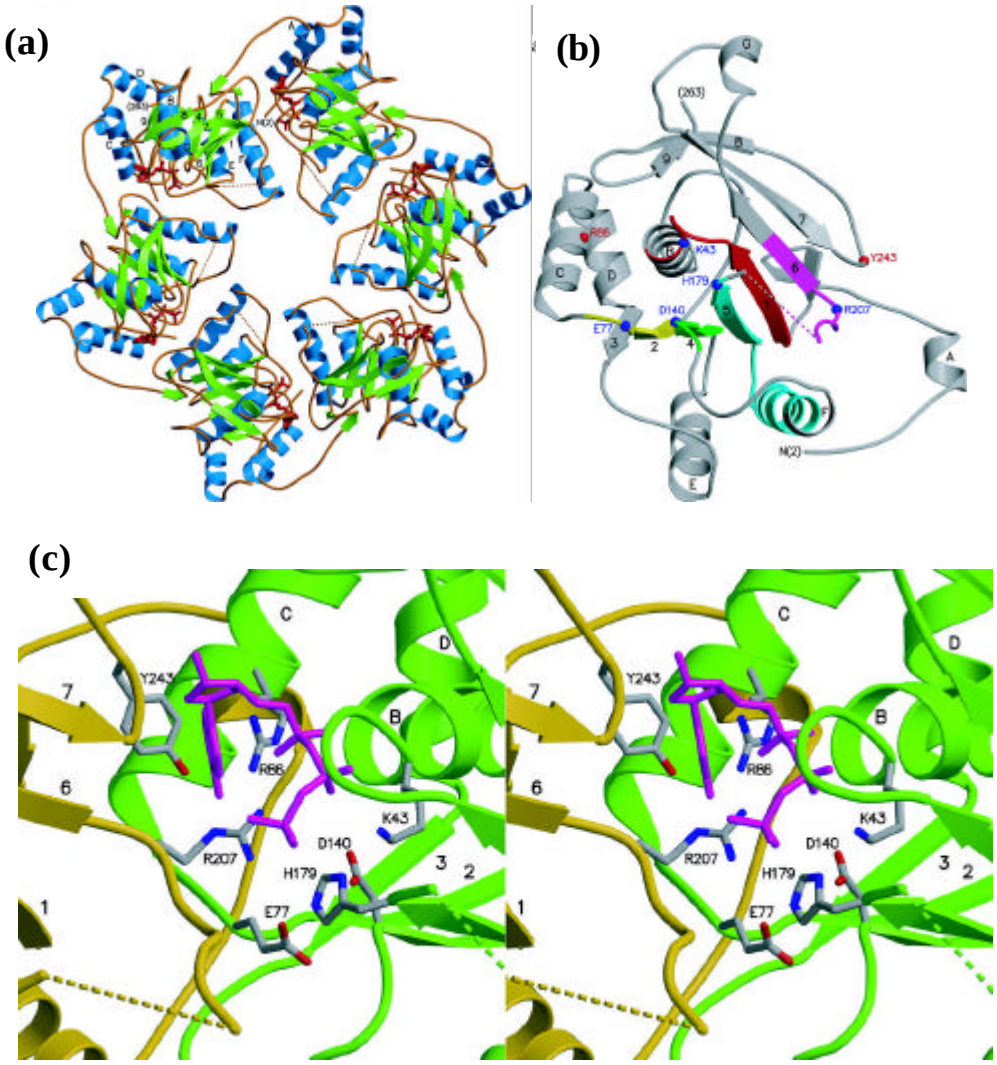

Figure 1.5: Crystal structure of the RepA homohexamer. (a) A view of the RepA hexameric ring. (b) A view of monomer structure of RepA. (c) Stereoview of ATP binding site of RepA located between two monomers drawn in yellow and green, respectively. see details in Niedenzu et al., 2001.

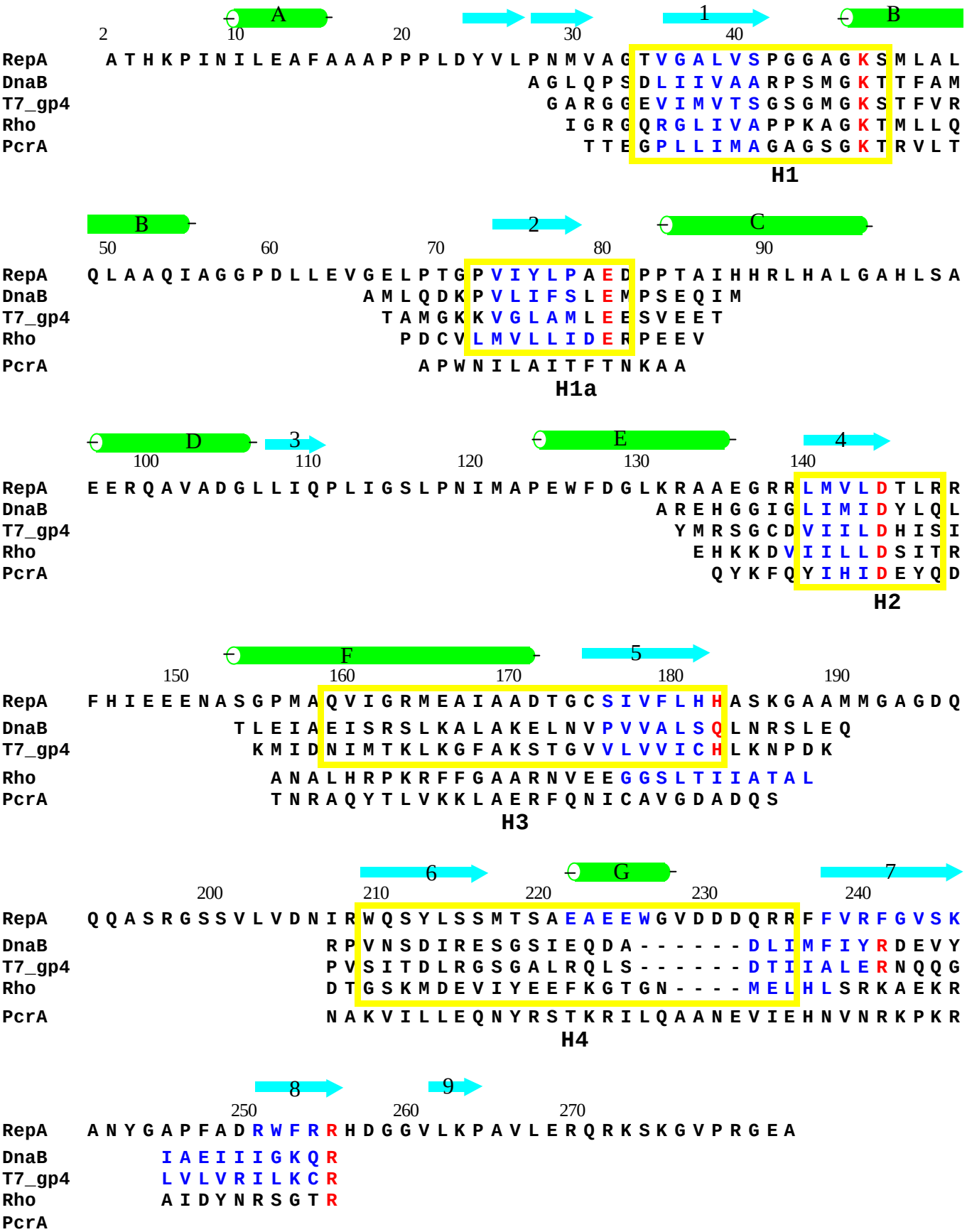

Figure 1.6: Primary sequence alignment of RepA and several other hexameric helicases (Niedenzu et al., 2001).

1.7 The aim of this work

Since DNA helicases are essential enzymes in all aspects of DNA metabolism, a detailed understanding of the mechanism(s) by which helicases function at the molecular level is important. Furthermore, the rapid rise in the number of human-pathogenic bacteria, which have acquired resistance to almost all commonly used drugs, requires action to develop novel drugs. Since helicases play essential roles in the metabolism of DNA and RNA and the replicative helicases of bacteria and eukaryotes differ substantially, helicase inhibitors may offer a feasible route towards this goal. Therefore, replicative prokaryotic DNA helicases constitute an attractive and yet unexplored target for development of new antibiotics as conditionally lethal DnaB mutants lead to stop of DNA synthesis and therefore unviable cells (Carl, 1970; Wechsler et al., 1971). Since the three-dimensional structure of the hexameric DNA helicase RepA of the broad host plasmid RSF1010 is the only one that has been determined from full-length subunits, it is an ideal tool to study inhibition of helicases by natural compounds and their synthetically derived analogues with the aids of structural information.

Although biochemical work has been done previously to characterize this enzyme (Scherzinger et al., 1997), the molecular details of its function and mechanism are only in the early stages. To understand helicase-catalyzed DNA unwinding at the molecular level requires information on the coupling of NTP binding and hydrolysis to DNA unwinding as well as the identification of the intermediate helicase-DNA states that occur during unwinding. Such understanding requires quantitative studies of the enengetics (thermodynamics) and kinetics of helicase binding to nucleotide cofactors (NTP, NDP, Pi) and DNA, as well as structural information (Lohman et al., 1996). Furthermore, all of these interactions (protein assembly, DNA binding, NTP binding and hydrolysis) are coupled (one process influences the other), making an examination of their linkage essential.

The aim of this work is to investigate how this molecular motor RepA carries out this important process, to obtain insight into the mechanism of inhibition and to help clarifying in general how DNA is unwound by helicases.