Stony Brook University

OFFICIAL COPY

The official electronic file of this thesis or dissertation is maintained by the University Libraries on behalf of The Graduate School at Stony Brook University.

© All Rights Reserved by Author.

DNA melting, unwinding and E1 complex assembly during BPV1 replication initiation

A Dissertation Presented

By

Xiaofei Liu

To

The Graduate School

In Partial Fulfillment of the

Requirements

For the Degree of

Doctor of Philosophy

In

Molecular and Cellular Biology

Stony Brook University

August 2009

Stony Brook University

The Graduate School

Xiaofei Liu

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend the acceptance of this dissertation.

Dr. Arne Stenlund, Thesis Advisor

Associate Professor, Cold Spring Harbor Laboratory

Dr. Leemor Joshua-Tor, Chair of Defense

Dean, Watson School of Biological Sciences

Professor, Cold Spring Harbor Laboratory

Dr. Adrian R. Krainer, Committee Member

Professor, Cold Spring Harbor Laboratory

Dr. Patrick Hearing, Committee Member

Professor, Department of Molecular Genetics and Microbiology

Stony Brook University

Dr. Hiro Furukawa, Outside Member

Assistant Professor, Cold Spring Harbor Laboratory

This dissertation is accepted by the Graduate School

Lawrence Martin

Dean of the Graduate School

Abstract of the Dissertation

DNA melting, unwinding and E1 complex assembly during BPV1 replication initiation

By

Xiaofei Liu

Doctor of Philosophy

In

Molecular and Cellular Biology

Stony Brook University

2009

The initial separation of two strands, DNA melting, is one of the essential steps during replication initiation. DNA melting provides ssDNA template for the helicase to assemble and to initiate dsDNA unwinding. Papillomavirus replication initiator protein E1 carries out both dsDNA melting and unwinding activities during viral replication initiation. One specific complex of E1, E1 double trimer (DT) has been implicated to be responsible for the melting of the ori template DNA. I have analyzed two residues in the helicase domain of the E1. These residues are part of a highly conserved structural motif, the β-hairpin, which is present in the helicase domain of all papovavirus initiator proteins. We demonstrate that the β-hairpin has two functions. First, it is the tool used by the E1 double trimer (DT) to pry open and melt double-stranded DNA. Second, it is required for the unwinding activity of the hexameric E1 helicase. The fact that the same structural element, but not the same residues, contacts both dsDNA in the DT for melting and ssDNA in the double hexamer (DH) for helicase activity provides a link between local

origin melting and DNA helicase activity and suggests how the transition between these two states comes about. E1 DT assembly requires two DNA binding activities of E1. One is the specific binding of E1 DBD to E1 binding site (BS) and another one is the nonspecific binding of E1 oligomerization and helicase domain to the DNA sequences flanking the E1 BS. I identified a large number of residues involved in the nonspecific DNA binding activity of E1 and demonstrated that this nonspecific DNA binding activity is required for both DT assembly and viral DNA replication. Finally, I studied the ATP binding and hydrolysis pocket of E1 and found out the residues that are required for nucleotide binding and DT assembly. Together, these data made substantial contributions towards the understanding of the mechanism of dsDNA melting during replication initiation.

Table of Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Acknowledgements

First, I want to thank my advisor Dr. Arne Stenlund. During the past four years, I received unreserved help and patient guidance from him, which are absolutely invaluable to me. More importantly, he taught me how to think and work independently, which is critical for being a scientist.

I would also want to thank my committee members: Dr. Leemor Joshua-Tor, Dr. Patrick Hearing, Dr. Adrian Krainer and Dr. Hiro Furukawa. I very much appreciate all the time, help, and advice they have given me. Thanks especially to Pat for driving to CSHL to attend my committee meetings. Lots of my work is based the beautiful X-ray crystal structures solved by Leemor's lab.

There are also many other faculties who have helped along the way. I am thankful for the encouragement and advice of MCB program director Dr. Rolf Sternglanz and Dr. Robert Haltiwanger.

I must also express my appreciation to our lab member Stephen Schuck for his support. I also want to thank James building manager Martha Daddario for her generous help during my stay in James building.

Special thanks to my friends Zhan Yao, Wen Xue, Xingyue He for their help and suggestions.

Finally, I would like to thank my parents. It is hard for them to have their only kid to be as distant as half the earth, but they understand my pursuit and always encourage me. I could never have gone through these years without their love.

Chapter 1 Introduction

1.1 DNA replication

DNA replication is one of the most important biological processes as it is the basis for the maintenance of genetic material and biological inheritance. The genetic material of most organisms is double strand DNA (dsDNA), which consists of two complementary DNA strands. The complementary nature of DNA provides the basis for semiconservative replication, where each of the two DNA strands serves as the template for synthesis of a new DNA strand. The dsDNA is melted and unwound, followed by the assembly of replication fork, which contains the replication factors required for DNA synthesis. The synthesis of DNA is catalyzed by DNA polymerase. DNA polymerase adds new nucleotides to the 3'-OH of the preexisting DNA or RNA. One strand, termed the leading strand, is synthesized continuously. The other strand, the lagging strand, is synthesized in short segments, known as Okazaki fragments. Other replication factors in the replication fork carry out specific functions as well. The primase makes the short RNA primers for the DNA polymerase which requires a primer with a free 3'-OH to initiate the DNA synthesis. RNase H removes the RNA primer after DNA synthesis starts. DNA polymerase fills in the gap left by RNA primer removal and the nick between fragments is repaired by DNA ligase. DNA helicase is dedicated to dsDNA unwinding, while the single strand DNA (ssDNA) generated from this unwinding is stabilized by ssDNA binding proteins (SSB, RPA). Topoisomerase is required to remove the

supercoils produced by dsDNA unwinding. A sliding clamp is also required to increase the processivity of the DNA polymerase. In the termination state of circular dsDNA replication, topoisomerase II is involved in separating the two daughter DNA molecules linked to each other topologically.

1.2 Initiation of DNA replication

Because the initiation process provides the primary control of DNA replication, it has been extensively studied in both prokaryotic and eukaryotic systems such as *E.coli*, double-stranded DNA viruses and eukaryotic cells [1-4]. Studies show that four key steps are required to happen in an orderly manner in these systems. The initiation of DNA replication begins with the recognition of the replicator, the origin of replication (ori), by the initiator protein which binds specifically to the ori according to the replicon model [5]. The DNA duplex in the ori is then destabilized and melted to generate the ssDNA. This melting activity is required because a single strand DNA template is needed for the DNA helicase to assemble and to initiate dsDNA unwinding. After unwinding of the two strands of the DNA duplex by the helicase, other replication factors are recruited to the ori either by protein-protein interaction or protein-DNA interaction to form the replication fork (Figure 1-1).

1.3 Comparison of the replication initiation process in different organisms

Table 1-1. Comparison of the replication initiation processes in different systems. The DNA replication initiation process is compared within representative prokaryotic and eukaryotic model systems.

		Prokaryotic Systems		DNA Viruses		Eukaryotic systems	
		E.coli	Phage λ	Simian Virus 40 (SV40)	Papillomavirus (PV)	S.cerevisiae	Homo Sapien
Replication origin		*oriC*	*oriλ*	SV40 ori	PV ori	ARS1	Unknown
Replication initiation steps	**Step1** Ori recognition	DnaA recognizes *oriC*	λ O recognizes *oriλ*	T-ag recognizes SV40 ori	E1 recognizes PV ori with the help of E2	ORC binds to ARS1 with modest specificity	ORC binds to dsDNA with low specificity
	Step 2 Ori melting	DnaA forms a large helical nucleoprotein complex to melt *oriC*	λ O melts *oriλ* (complex unknown)	T-ag melts ori, (complex not identified)	E1double trimer melts ori	Unknown	Unknown
	Step 3 dsDNA unwinding	DnaB double hexamer unwinds dsDNA	DnaB double hexamer unwinds dsDNA	T-ag double hexamer unwinds dsDNA	E1 double hexamer unwinds dsDNA	MCM in complex with CDC45 and GINS unwinds dsDNA	MCM in complex with CDC45 and GINS unwinds dsDNA
	Step 4 Replication factors recruitment and replication fork assembly <u>Replication fork components</u>	DnaB double hexamer, Primase DnaG, SSB, Gyrase, Topo I, DNA pol III holoenzyme, DNA pol I, RNase H, DNA ligase	DnaB double hexamer, Primase DnaG, SSB, Gyrase, Topo I, DNA pol III holoenzyme, DNA pol I, RNase H, DNA ligase	T-ag double hexamer. RPA, PCNA, DNA pol α-primase complex, RFC, DNA pol δ, Topo I & II, RNase I, DNA ligase I, 3'-5' exonuclease	E1 double hexamer, RPA, PCNA, DNA pol α-primase complex, RFC, DNA pol δ, Topo I & II, RNase I, DNA ligase I, 3'-5' exonuclease	ORC, Cdt1, CDC6, MCM, CDC45, CDK, DDK, GINS, DNA pol α/ε/δ, DNA primase, RPA, RFC, CDK inhibitors, etc.	ORC, Cdt1, CDC6, MCM, CDC45, CDK, DDK, GINS, DNA pol α/ε/δ, DNA primase, RPA, RFC, CDK inhibitors, etc.
Replication initiation regulation		Cell cycle independent	Cell cycle independent	Unknown	Unknown	Cell cycle dependent	Cell cycle dependent

Prokaryotic and eukaryotic replication systems use analogous, rather than homologous, mechanisms to initiate replication. The factors involved in prokaryotic and eukaryotic replication initiation processes share low sequence homology but carry out similar functions such as ori identification, local template melting and dsDNA unwinding (Table 1-1). The similarities and differences among systems are compared below.

1.3.1 Representative model systems

Many systems have been well established for the study of DNA replication initiation.

1. Prokaryotic systems

   1.1 *E. coli*. *E.coli* is the best studied prokaryotic system and provides the most detailed knowledge of replication initiation. The minimum factors required for *E.coli* replication initiation consist of the bacterial initiator DnaA, DNA helicase DnaB and the helicase loader DnaC [2].

   1.2 Bacteriophage λ, another representative prokaryotic replication system, uses a similar mechanism as *E. coli* for replication initiation. The viral initiator O and helicase loader P are encoded by the viral genome while the helicase for DNA unwinding is DnaB [6, 7].

2. Double-stranded DNA viruses.

   Eukaryotic DNA viruses have long been used as simplified model systems for eukaryotic DNA replication because viruses depend largely on the host cellular replication machinery for DNA synthesis. Simian virus 40(SV40), a member of the polyomavirus family, and papillomavirus (PV) are among the most well

studied viruses for replication. The replication initiators SV40 large tumor antigen (T-ag) and papillomavirus E1 are encoded by their respective viral genomes. T-ag and E1 carry out most of functions required for viral replication initiation such as ori identification, local template melting, unwinding and replication factor recruitment [1].

3. Eukaryotic cells

   Initiation of DNA replication in eukaryotic cells is less well characterized than prokaryotic systems and eukaryotic DNA viruses. Most of the study of eukaryotic replication has been carried out in budding yeast *S.cerevisiae* and fission yeast *S.pombe*. Other systems such as Xenopus, Drosophila and human cell lines have also been used to study the eukaryotic replication initiation [3, 4]. The initiation of eukaryotic replication is much more complex than the prokaryotic and viral systems in that a large number of proteins and protein complexes are involved and the process is subject to cell cycle regulation.

1.3.2 Origin recognition

Replication origin identification is carried out by different proteins or protein complexes in different species.

The initiation of chromosomal replication in *E.coli* begins with the binding of bacterial initiator DnaA to the replication origin. The single replication origin *oriC* is

about 250 bp long and contains multiple 9-bp repeat elements called DnaA boxes that constitute the specific DnaA binding sites.

Similar to *E. coli*, the DNA viruses use a single initiator protein for ori recognition. The SV 40 T-ag and papillomavirus E1 are both capable of ori identification. The DNA binding domains (DBD) of T-ag and E1 bind to their unique binding sites. The SV40 replication origin contains a palindromic array of four GAGGC pentanucleotides, which serve as T-ag binding sites. The T-ag binding sites are flanked by an easily denatured imperfect palindrome termed early palindrome (EP) on one side and an A-T rich sequence on the other side [8-10]. The papillomavirus replication origin is similar in arrangement to SV40. Different from T-ag, the loading of E1 to the papillomavirus origin requires the help of an auxiliary factor E2 [1].

In eukaryotic cells, a six subunit complex called the origin recognition complex (ORC) is involved in ori recognition. The complexity of eukaryotic ori recognition is significantly increased compared to *E.coli* and DNA viruses because multiple subunits are involved and regulated in eukaryotic ori binding. ORC was first identified in the budding yeast *S.cerevisiae* [11]. This complex is well conserved in other eukaryotic cells. The yeast ORC binds the replication origin with modest specificity while the replication origins in higher eukaryotic cell are less well defined [3, 4].

1.3.3 Ori melting

The mechanism of ori melting in *E. coli* is not clear. Biochemical and structural data suggested that DnaA forms a large helical nucleoprotein complex in the ori and destabilizes the local ori DNA [12-14]. Viral initiators T-ag and E1 are capable of ori melting. The binding of T-ag and E1 to the ori causes the distortion and untwisting of the ori [15-17]. DNA melting activity has not been characterized in any of the eukaryotic replication factors.

1.3.4 Helicase loading and DNA unwinding

In *E.coli,* the loading of the helicase DnaB to melted ssDNA bubble requires the help of a helicase loader, DnaC. Single-stranded DNA and DnaA interact with the DnaB-DnaC complex and recruit DnaB [18, 19].

Different from *E.coli*, the unwinding of dsDNA in DNA viruses is carried out by the initiator. The viral helicase, which is a double hexamer complex for both T-ag and E1, assembles directly on the melted ssDNA [17, 20-22].

The helicase loading process of eukaryotic cells in analogous to that of *E. coli*. ORC recruits the MCM2-7 complex, which functions as the eukaryotic helicase, with the help of Cdt1 and CDC6. The unwinding activity of the MCM complex requires help from

additional factors and complexes. Recently, a CDC45/MCM2-7/GINS complex has been shown to have *in vitro* helicase activity [23].

1.3.5 Cell cycle regulation

As opposed to prokaryotic systems, cell cycle regulation is required for eukaryotic DNA replication to initiate the genome duplication once and only once during the S-phase of each cell cycle. Cell cycle regulation is achieved through two protein kinases, CDK (cyclin-dependent kinase) and DDK (Dbf4-dependent kinase). CDK and DDK activate the pre-RC for DNA replication in S-phase and inhibit new pre-RC assembly. The pre-RC can only form during G1-phase when the level of CDK and DDK is low [3, 4].

1.4 Melting is an important but undefined step in the replication initiation

In most replication systems, the initiator proteins have been characterized and the loading process can be recapitulated by binding of the initiator to the replication origin. The replication fork components have been dissected by protein purification and their biochemical activities have largely been defined by *in vitro* assays [4, 11, 18, 24-26]. The mechanism of dsDNA unwinding by DNA helicases has also been extensively studied by both biochemical and structural approaches [27-30]. Compared to these fairly well understood initiation steps, much less is known about the mechanism of dsDNA melting.

One limitation in studying melting is that only a few proteins are known to contain dsDNA melting activity. DnaA from *E.coli* has been shown to be able to melt the ori [12, 13, 31]. However, only very recently has a mechanism of melting been proposed based on the X-ray crystal structure of the DnaA monomer [14]. The viral initiators Simian virus 40 (SV40) large tumor antigen (T-ag) and papillomavirus E1 are the only other proteins that have been shown to be capable of dsDNA melting.

1.5 Advantages of using DNA viruses to explore the mechanism of dsDNA melting

Besides the fact that only a few elements (initiator, replication origin) are involved in viral replication initiation, another advantage of SV40 and papillomaviruses as model systems for replication initiation is that the two viral initiator proteins can be overexpressed and purified from baculovirus (T-ag) or *E.coli* (E1). This facilitates the establishment of *in vitro* assays for testing the biochemical activities of the two initiators, such as melting and unwinding. Initiator complex assembly on the ori can also be simulated by *in vitro* systems containing initiator and ori fragments [17, 21, 32]. This provides the possibility of identifying the specific complex form that is responsible for dsDNA melting.

Although it is well established that T-ag has ori distortion and melting activity, the active complex has not been identified. In contrast, in bovine papillomavirus type 1

(BPV1), one specific complex, E1 DT, has been identified as responsible for melting. However, it is not known how E1 DT assembles, and the molecular details of the mechanism of the melting are not clear [17].

1.6 Papillomavirus as model system to study the replication initiation

Papillomaviridae are a group of small dsDNA tumor viruses with genomes of 7000-8000bp. The viral genome encodes at least eight early proteins with the name (E1-E8) that are responsible for viral gene expression, viral DNA maintenance and host immune response regulation, together with two late (capsid) proteins that are expressed in more superficial, differentiated cells. Papillomaviruses infect cutaneous or mucosal epithelial cells of the host. Most of the infections are clinically unapparent or benign lesions such as common warts, anogenital warts, and laryngeal papillomatosis. However, some of infections have been shown to be associated with malignant carcinomas. In humans, infections with some high-risk papillomavirus strains (HPV16, HPV18, HPV31 and HPV45) have been demonstrated to be responsible for cervical cancers. Papillomaviruses have also been implicated in a subset of head and neck cancers [33, 34]. The replication of papillomavirus has been explored extensively because of the potential clinical applications. Of all the papillomaviruses, bovine papillomavirus type 1(BPV1) and human papillomavirus type 11(HPV11) are most frequently used as DNA replication models.

Upon entry into the host cell, the viral genome is transiently amplified to approximately 50 to 200 copies. It then establishes a latent infection, existing as an extrachromosomal plasmid [35, 36]. In the latent state, the viral genome replicates on the average once per cell cycle and maintains a low copy number that is estimated to be less than 20 copies per cell [37, 38]. These characteristics make it a good model to study plasmid replication. Both *in vivo* and *in vitro* replication assays of BPV1 have been established. Three elements from the viral genome have been demonstrated to be required for initiation of BPV1 replication: two viral peptides encoded by the viral genome, E1 and E2, and the replication origin [39-41]. This requirement applies to other papillomaviruses as well.

1.6.1 The replication origin of papillomavirus

The replication origin of papillomavirus is located in a region that is upstream of the coding region called Long Control Region (LCR) or Upstream Regulatory Region (URR). It contains the binding sites of E1 and E2 together with an A/T rich region [40-43].

In BPV1, the sequence that E1 binds to is an 18bp imperfect palindrome, 5'-ATTGTTGTTAACAATAAT-3', which consists of four overlapping hexanucleotide E1 binding sites (E1 BS). E1 BS is recognized by E1 DNA binding domain (E1 DBD). Two of the E1 BS have a consensus sequence of 5'-ATTGTT-3'while the other two have the sequence 5'-ATTATT-3' and 5'GTTGTT-3', respectively[44]. The four sites are

separated by three nucleotides from the adjacent half of the palindrome [17, 44] (Figure 1-2).

The binding site for E2 is also an imperfect inverted repeat sequence as 5'ACCN$_6$GGT-3'. E2 has the highest affinity to sequence ACCGN$_4$CGGT while it also binds to AACN$_6$GGT, ACCN$_6$GTT and ACAN$_5$CGGT with lower affinity [45-48]. The E2 binding site in the BPV1 ori is 5'ACCATCACCGTT3'. E2 has a low affinity to this binding site, which facilitates the E2 dissociation after ori identification. There are 17 E2 binding sites in the BPV1 genome, but other papillomaviruses may contain as few as four consensus E2 binding sites [47]. For most papillomaviruses, at least one E2 binding site is required for replication. In some cases, two E2 binding sites are able to initiate the replication without the presence of an E1 BS [49-53].

The A-T rich region is a common feature of both prokaryotic and eukaryotic replication origins. The A-T rich region is distorted and melted readily during the initiation process to generate a single-stranded template for the helicase. In BPV1, the A-T rich region contains six A-T base pairs. It is the only element in the DNA sequences flanking E1 BS required for replication[40]. The A-T rich region is melted by E1 DT during replication initiation [17, 54].

1.6.2 Papillomavirus replication initiator protein E1

1.6.2.1 Structure and function of E1

The viral initiator protein E1 plays a key role in the initiation process of replication. It is encoded by the E1 open reading frame (ORF) of the viral genome and has a molecular weight of about 70KD[55]. Papillomavirus E1 is distantly related to SV40 T-ag. Structural data suggests the DNA binding and helicase domains of these two proteins are quite similar to each other [28, 56-61]. These two initiators are AAA+ (ATPases associated with various cellular activities) proteins. E1 provides similar functions as T-ag during replication initiation: ori recognition, local template melting, helicase activity on dsDNA and recruitment of other replication factors to form the replication fork. Both of E1 and T-ag belong to the super family 3 (SF3) helicase family and they unwind the DNA with a 3' to 5' polarity using the energy provided by ATP hydrolysis.

1.6.2.2 Functional domains of E1

The full-length E1 protein consists of four domains: the N-terminal domain, the DNA binding domain (DBD), the oligomerization domain and the helicase domain.

BPV1 E1 DBD is located between aa 150-300. It has a high specificity for E1 BS in the ori. The crystal structure of E1 DBD shows that the E1 DBD is very similar to the SV40 T-ag DBD structurally, though they share a low sequence homology (about 6% identity) [57, 61, 62]. The essential residues for DNA binding are located on an extended loop and an α helix [57, 62]. E1 DBD binds to ori as a dimer by the interaction of two α3 helices (Figure 1-3) [62]. Disruption of this dimerization by mutation of one residue

A206 involved in the interaction caused defects in both DNA binding *in vitro* and DNA replication *in vivo*[63].

The E1 helicase domain contains an AAA+ module for ATP binding and hydrolysis. It carries out both dsDNA melting and unwinding during replication initiation using the energy provided by ATP hydrolysis. In the hexamer structure of E1 oligomerization and helicase domains in complex with ssDNA, oligomerization domains contact each other to form a rigid collar to hold six subunits together while six helicase domains form a channel for the ssDNA to go through. Each helicase domain projects a specific motif called the β-hairpin into the channel in a right-handed staircase pattern. Each β-hairpin contacts adjacent nucleotides in the ssDNA and escorts ssDNA through the channel by ATP hydrolysis and subsequent ADP release (Figure 1-4) [28].

The N-terminal domain of E1 is not well conserved among papillomaviruses and is only required for replication *in vivo* [64, 65]. This region contains both nuclear import and export signals and has been shown to be regulated by cell-cycle related kinases [66-70].

1.6.2.3 Modification and regulation of E1

E1 has been shown to be posttranslationally modified by phosphorylation, unbiquitination and sumoylation in vivo [71-75]. Working separately or in combination, those modifications are thought to regulate E1's activity, stability or location [66, 72, 74, 76]. Numerous phosphorylation sites in E1 have been identified and some indications

exist that these sites may be important for the DNA replication [74]. E1 interacts with the CyclinE-CDK2 complex, and Cylin E-CDK2 can stimulate the replication of BPV plasmids in *Xenopus* egg extracts[71]. Binding of the Cyclin E-CDK2 complex to BPV1 E1 also prevents E1 from being ubiquitinated and degraded [72]. HPV11 E1 cellular localization is regulated by phosphorylation [66, 70]. CDK phosphorylation can inactivate a nuclear export sequence (NES) of HPV11 E1[66]. However, not all papillomaviruses E1 contain such NES. Sumoylation is also required for E1 intranuclear accumulation[76]. Molecular chaperones have been suggested to be capable of stimulating formation of large oligomeric E1 complexes on ori [77].

1.6.3 The auxiliary factor E2

Full length E2 is a polypeptide of 350-500 amino acids encoded by the E2 ORF. E2 is a transcription factor and plays a vital role in expression of viral genes. It has also been identified as a vital factor for viral DNA replication [39, 41]. E2 is absolutely required for *in vivo* viral DNA replication[39]. DNA replication *in vitro* is dependent on E2 only when the E1 concentration is very low or when non-specific competitor DNA is introduced to challenge the DNA binding specificity of E1 [41, 78-80].

Full length E2 consists of two conserved domains, the N terminal activation domain and the C terminal DNA binding domain (DBD). Those two domains are linked by a poorly conserved proline rich hinge region. In the crystal structure of BPV1 E2 DBD bound to its DNA binding site, the E2 DBD forms a dimeric β-barrel structure that positions two α recognition helices in contact with the DNA. This contact causes the

bending of the DNA to approximately 40°[81]. The structure of the E2 activation domain has also been solved by X-ray crystallography [82, 83]. The E2 activation domain consists of a curved anti-parallel β-sheet domain and a helical domain containing three anti-parallel helices.

The E2 ORF also encodes two truncated forms of E2, E2C and E8/E2, either by transcription from an internal promoter and/or alternative RNA splicing [84-88]. Both contain the E2 DBD but lack the activation domain. E2C and E8/E2 serve as repressors of replication and transcription in contrast to full length E2 which is an activator in both processes [88-92]. However, heterodimers formed between the full-length E2 and truncated E2 proteins are activators of E2-dependent transcription and papillomavirus DNA replication [93, 94].

1.6.4 Identification of the replication origin by the E1-E2 complex

The E1 DBD has a high specificity for E1 BS in the ori. This high specificity is masked by the nonspecific DNA binding activity of its helicase domain. E1 can initiate DNA replication *in vitro* from non-specific DNA sequence in the absence of E2 [78, 95, 96]. *In vivo* DNA replication assay has shown that E1 needs the auxiliary factor E2 to identify the ori [39]. By interacting with each other, E1 and E2 bind cooperatively to the ori in the form of an $E1_2E2_2$ complex that has high specificity for ori (Figure 1-5) [78, 97, 98]. In BPV1, interactions between the DBDs of E1 and E2 bound to their proximal sites produce a significant bend in DNA which further facilitates the interaction between the activation domain of E2 and the helicase domain of E1 [99-101]. The latter interaction

prevents nonspecific DNA binding activity of E1 helicase domain [1, 102, 103]. In other papillomaviruses that contain an ori with distal E1 and E2 BS, the interactions between DBDs of E1 and E2 are dispensable for the cooperative binding [101].

1.6.5 Replication origin melting, unwinding and further initiation complex assembly

After the $E1_2E2_2$ complex binds to ori, E2 proteins are displaced by the second dimer of E1 by ATP hydrolysis[104]. In the next step, more E1 molecules are recruited to ori to form a head-to-head E1 double trimer (DT). DT formation needs the cooperation of specific DNA binding activity of E1 DBD to E1 BS with the nonspecific DNA binding activity of E1 helicase domain to the flanking sequence of E1 BS. In DT, the E1 molecules are positioned in a helical arrangement at dsDNA with an offset of 3bp relative to each other. This arrangement is determined by the positions of E1 BS in the ori. The dsDNA in the ori is distorted and melted by E1 DT to generate ssDNA. E1 DT also serves as a precursor for the E1 double hexamer (DH) assembly, which unwinds dsDNA (Figure 1-5). ATP hydrolysis is required for DH assembly [17]. The E1 DH recruits the host cell replication machinery by direct interactions. The same set of cellular factors that support SV40 replication are also able to support BPV1 replication (RPA, PCNA, DNA polymerase α-primase complex, RFC, DNA polymerase δ, Topoisomerase I & II, RNase I, DNA ligase I, 3'-5' exonuclease) [26].

1.7 Overview of thesis

Double strand DNA melting is an essential step in the initiation of replication. The melting is generated by two different activities intrinsic to the E1 protein and carried out by the E1 DT, which untwists the DNA, and the E1 DH, which has DNA helicase activity. In order to study the melting activity of E1, the melting activity of E1 needs to be distinguished from its helicase activity. In chapter 2 a mutagenesis approach is used to identify residues in the β-hairpin that are specifically required to act as part of the either untwisting or helicase activity.

The assembly of the E1 DT requires the cooperation of the specific DNA binding activity of E1 DBD to the E1 BS with the nonspecific DNA binding activity of the E1 oligomerization and helicase domain to the DNA sequences flanking the E1 BS[17]. E1 DBD and its binding to E1 BS have been studied extensively by biochemical, genetic and structural biology approaches and in general is well understood [44, 57, 62, 105, 106]. However, much less is known about the nonspecific DNA binding activity of the E1 [17]. Chapter 3 studies the residues of E1 oligomerization and the helicase domain involved in nonspecific DNA binding activity of E1 and its role in DT assembly and replication initiation.

Nucleotide binding and hydrolysis by E1 is central to all the functions of E1, including DT and DH assembly as well as melting and DNA helicase activities [17, 55]. Chapter 4 provides a systematic study of the residues in the ATP binding pocket of E1 and tries to define their roles in ATP binding and hydrolysis.

Taken together, this thesis addresses dsDNA melting, unwinding and E1 DT assembly during the BPV1 replication initiation process.

Figure 1-1. **Four key steps in the initiation process of DNA replication.** The initiation process of can generally be considered to involve four steps: recognition of replicator, the origin of replication, by the initiator protein; local template DNA melting to prepare the ssDNA; DNA unwinding by helicase; recruitment of replication factors, such as DNA polymerase and ssDNA binding proteins, to initiate DNA synthesis.

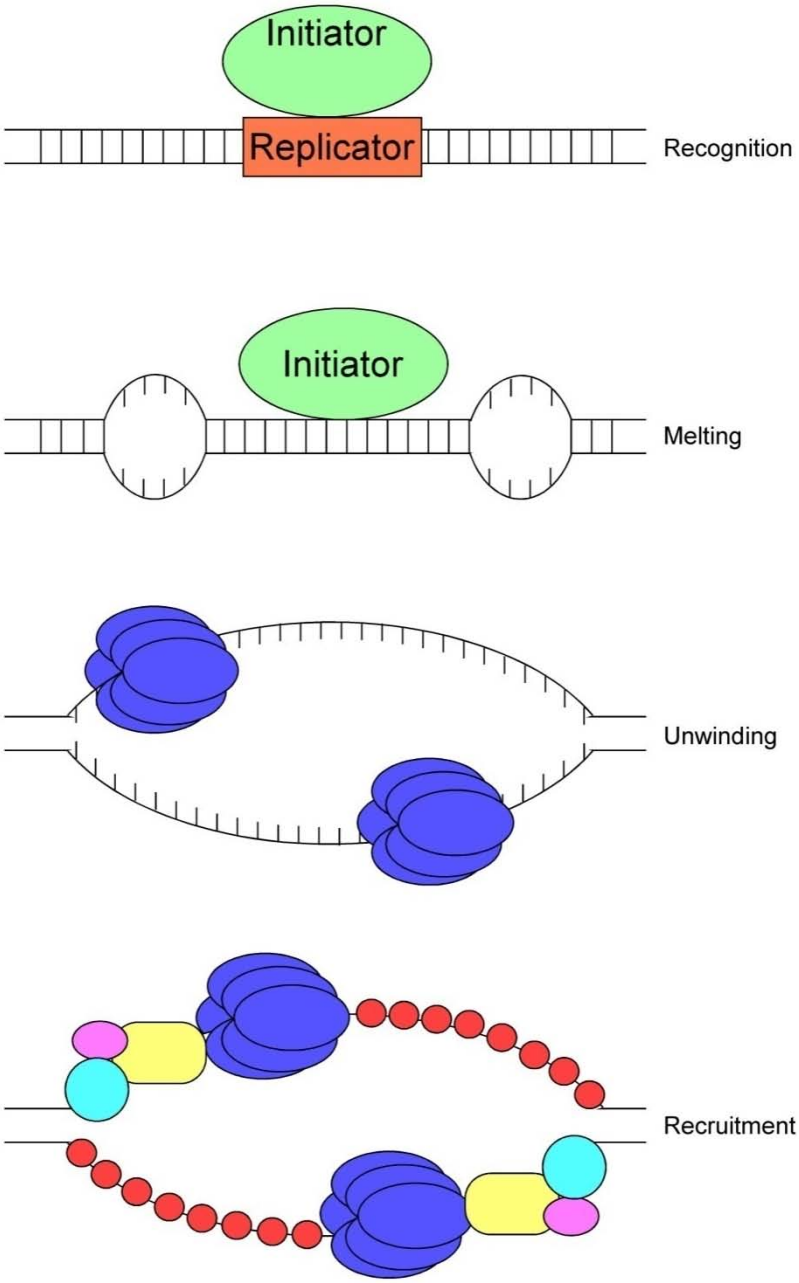

Initiator
Replicator
Recognition
Initiator
Melting
Unwinding
Recruitment

Figure 1-2. **BPV1 replication origin.** The BPV1 minimal origin is contained within nucleotides 7914-27 in the BPV genome. It consists of four overlapping E1 binding sites, an E2 binding site and an A-T rich region in the DNA sequences flanking E1 binding sites.

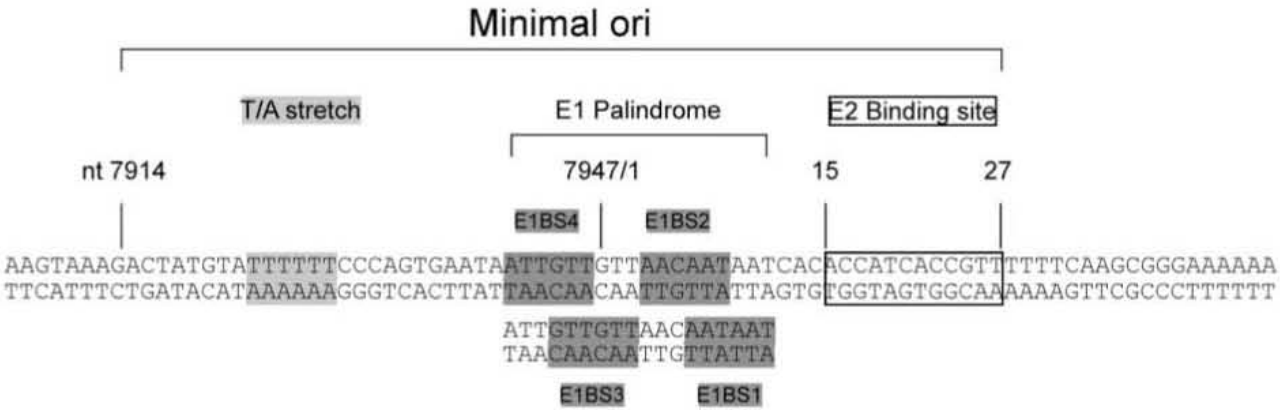
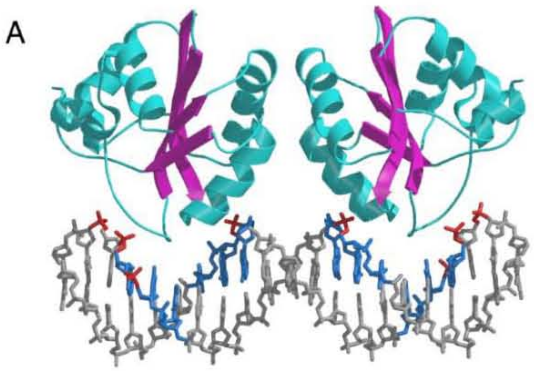

Minimal ori
T/A stretch
E1 Palindrome
E2 Binding site
nt 7914
7947/1
15
27
E1BS4
E1BS2
AAGTAAAGACTATGTATTTTTTCCCAGTGAATAATTGTTGTTAACAATAATCACACCATCACCGTTTTTTCAAGCGGGAAAAAA
TTCATTTCTGATACATAAAAAAGGGTCACTTATTAACAACAATTGTTATTAGTGTGGTAGTGGCAAAAAAGTTCGCCCTTTTTT
ATTGTTGTTAACAATAAT
TAACAACAATTGTTATTA
E1BS3
E1BS1

Figure 1-3. Structures of two complexes of E1 DBD with ori DNA containing four E1 binding sites (BS). A. Structure of the dimer of E1 DBD bound to E1 BS 2 and BS4 on the ori DNA. **B.** Structure of the tetramer of E1 DBD bound to E1 BS 1-4 on the ori DNA viewed perpendicular and parallel to the DNA helical axis. Each DBD is individually colored. E1 BS2 and BS4 are colored blue while E1 BS 1 and BS3 are colored yellowed. Adapted from Enemark, E.J., A. Stenlund, and L. Joshua-Tor, Crystal structures of two intermediates in the assembly of the papillomavirus replication initiation complex. Embo J, 2002. 21(6): p. 1487-96 [62].


A

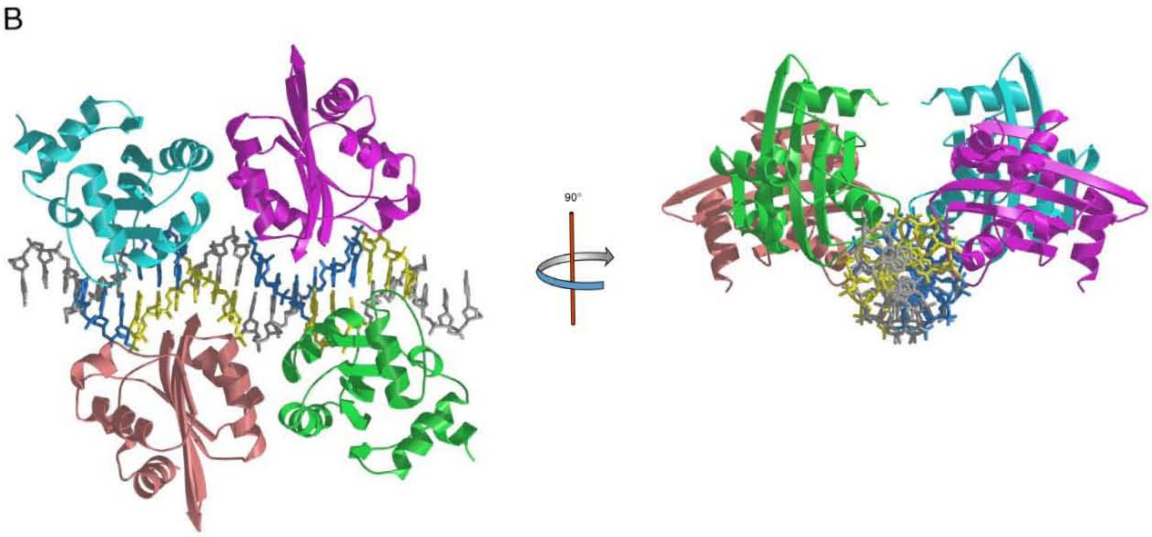

B
90°

Figure 1-4. Structure of the hexamer of E1 oligomerization and helicase domain in complex with ssDNA and ADP. A. Ribbon representations of the two hexamers of the E1oligomerization and helicase domain viewed perpendicular to the channels. **B.** Ribbon representations of the two hexamers of the E1 oligomerization and helicase domains viewed parallel to the channels. Single-stranded DNA goes through the channel formed by six subunits that are color-coded individually. The β-hairpin on the E1 helicase domain projects into the channel and contacts ssDNA. Adapted from Enemark, E.J. and L. Joshua-Tor, Mechanism of DNA translocation in a replicative hexameric helicase. Nature, 2006. 442(7100): p. 270-5 [28].

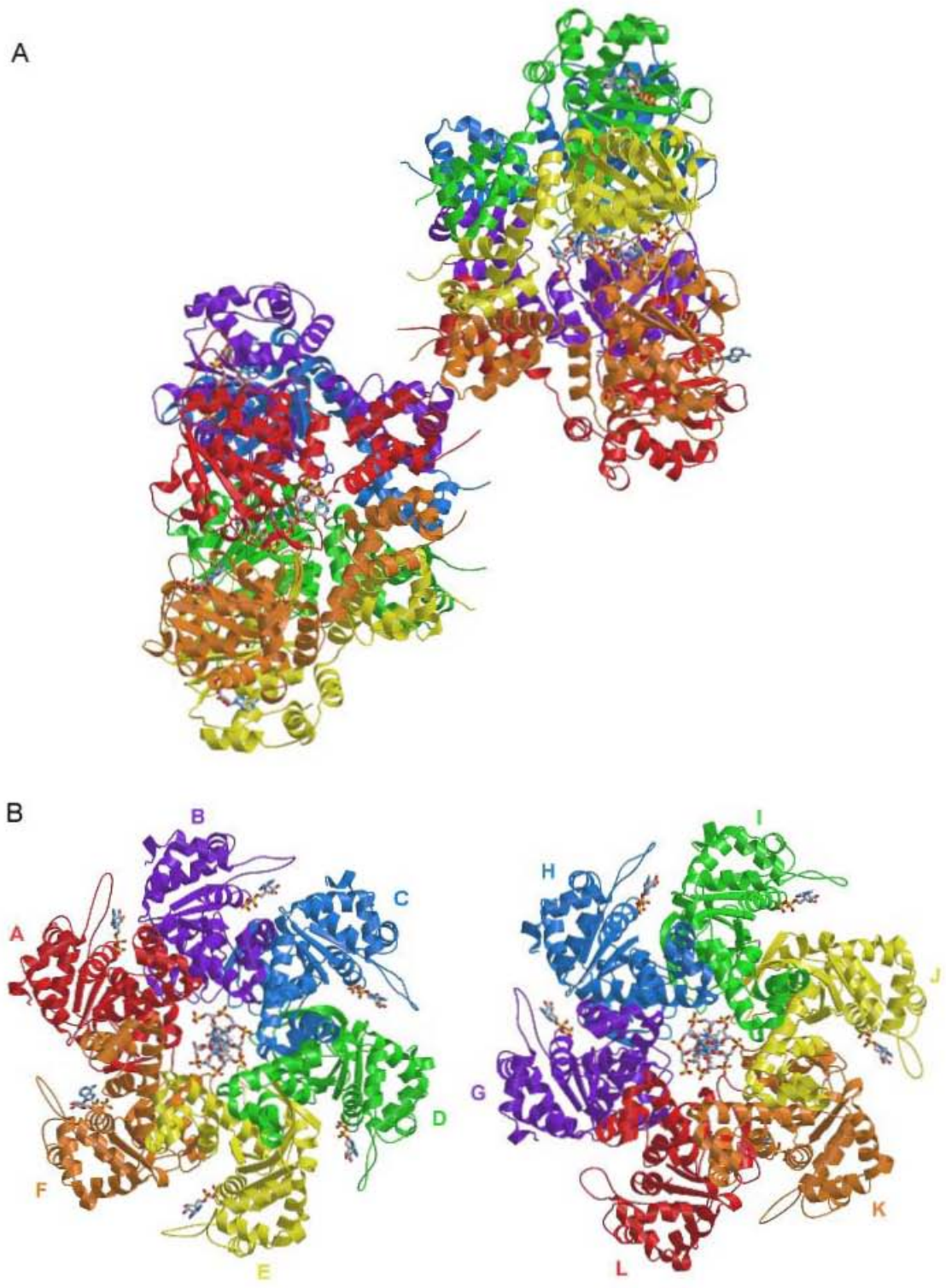
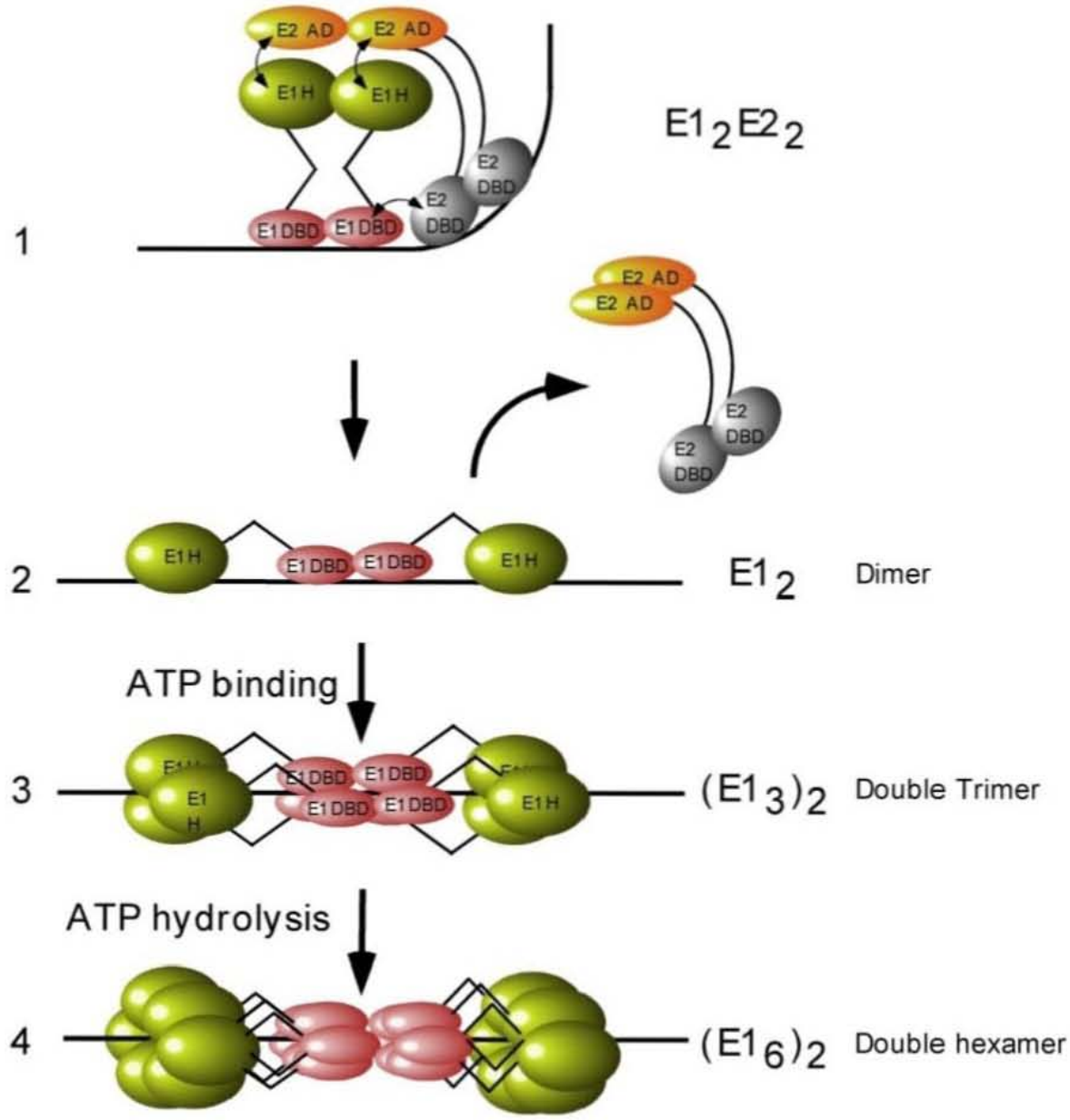

A
B
B
C
A
F
D
E
I
H
J
G
K
L

Figure 1-5. BPV1 DNA replication initiation. E1 and E2 bind cooperatively to the ori via the $E1_2E2_2$-ori complex, which has a high specificity to the ori. Then E2 is displaced and additional E1 molecules are recruited to form the E1 double trimer complex, which is responsible for the local template melting. Ori melting provides the single strand DNA template for the E1 double hexamer, the functional helicase, assembly. Adapted from Stenlund, A., E1 initiator DNA binding specificity is unmasked by selective inhibition of non-specific DNA binding. Embo J, 2003. 22(4): p. 954-63 [102].


E2 AD
E2 AD
E1 H
E1 H
E2 DBD
E2 DBD
E1 DBD
E1 DBD
$E1_2E2_2$
1
E2 AD
E2 AD
E2 DBD
E2 DBD
E1 H
E1 DBD
E1 DBD
E1 H
2
$E1_2$
Dimer
ATP binding
E1 DBD
E1 DBD
E1 DBD
E1 DBD
E1 H
3
$(E1_3)_2$
Double Trimer
ATP hydrolysis
4
$(E1_6)_2$
Double hexamer

Chapter 2 Adjacent residues in the E1 initiator β-hairpin define different roles of the β-hairpin in ori melting, helicase loading, and helicase activity

2.1 Introduction

The preparation of template DNA for replication involves several basic steps that are obligatory in all organisms. One step is the initial separation of the two DNA strands at the origin of DNA replication (ori), local melting, to generate a single stranded region of DNA. This step is essential because most known DNA helicases, which are required for unwinding of the template, require a single stranded region to initiate unwinding (For a review see[30]). In spite of the obvious importance of local melting for the replication process, only a few activities are known that can perform this task. In eukaryotes, no such activity has been identified. In bacteria, DnaA is responsible for local ori melting, but the mechanism involved is unknown except that the process involves a large complex of DnaA and is dependent on bound ATP [12, 13]. Viral initiator proteins such as T-ag and E1 from the papovavirus family (papillomaviruses, polyomavirus and SV40) are the only other proteins known to melt double-stranded ori DNA, although the mechanism employed by these proteins is unknown [15, 104, 107, 108].

From recent studies, there is evidence that the local melting of the ori is initiated by a specific form of the E1 protein, a double trimer (DT), which forms on the E1 binding sites present in the ori. Formation of the DT is dependent on ATP binding, but not hydrolysis. The DT is a required precursor for the formation of the active DH helicase and correct melting appears to be a prerequisite for the formation of a DH that can unwind the ori [17].

Recent crystal structure determination of fragments of both SV40 T-ag and papillomavirus E1 has identified a highly conserved structural element, a β-hairpin, in the helicase domain of these proteins [28, 56, 58, 103]. In both the E1 and T-ag structures, this β-hairpin is oriented towards the central channel of the hexamer (Figure 2-1A and 2-1B) and in the E1 structure, which also includes ssDNA, the hairpin is making contacts with the ssDNA. Mutational analysis of T-ag has suggested that this β-hairpin is required for the DNA helicase activity[109]. Biochemical and mutational analysis of the corresponding β-hairpin in the bovine papillomavirus E1 protein has indicated that the β-hairpin instead may be required for local ori melting activity. Mutation of the highly conserved H507 at the tip of the β-hairpin results in a protein that fails to melt ori DNA in the presence of ADP and fails to form a DT, the precursor form of E1 required for the assembly of the DNA helicase [17].

Here we demonstrate that the E1 β-hairpin is required for both local ori melting and DNA helicase activity, but that these activities are associated with different residues

in the β-hairpin (Figure 2-1B). K506 is required for both local ori melting and DNA helicase activities, but H507 is only required for local ori melting. We also show that the β-hairpin likely interacts with DNA directly and that, although either of the aromatic side chains H, Y, or F at position 507 is functional for DT formation, melting of the ori requires either H or F at this position in the β-hairpin. Together, these results identify the β-hairpin as the instrument that E1 uses for both the local melting of dsDNA and for DNA helicase activity, although these activities reside in different E1 complexes.

2.2 Results

2.2.1 DNA replication activity of substitution mutants at H507.

To define the function of the tip of the β-hairpin, we substituted H507 with a range of amino acids with different properties, including A, V, L, R, N, M, F and Y. We generated these substitutions in the *E. coli* expression vector pET E1 and expressed and purified full-length E1 with the respective substitutions. We also generated the same substitutions in the context of the mammalian expression vector pCGE1. We first tested these substitutions in transient DNA replication assays *in vivo* to determine whether they affected DNA replication.

We transfected an expression vector encoding the viral E2 protein, which is required for DNA replication *in vivo,* together with an ori plasmid and expression vectors

for either WT E1 or the individual 507 substitutions into CHO cells using electroporation. Two, three and four days after transfection, low molecular weight DNA was harvested and analyzed by southern blotting after digestion with DpnI, which digests unreplicated (methylated) DNA and HindIII, which linearizes the plasmid (Figure 2-1C). We observed robust replication, detectable as a prominent 2.9 kb band, in the presence of the WT E1 and H507F expression vectors (lanes 1-3 and 22-24, respectively) as well as a faint trace of replication (>10-fold reduced) with the H507Y vector (lanes 25-27). The remaining H507 substitutions did not support detectable DNA replication.

We next tested the substitutions for their ability to support DNA replication *in vitro* (Figure 2-1D). In such an *in vitro* DNA replication assay, a plasmid containing the ori is incubated in the presence of E1 in a H293 cell extract in the presence of radiolabeled nucleotide. The only substitutions that had detectable activity were H507Y and H507F (Figure 2-1D, compare lanes 1, 8 and 9). The other substitutions generated no detectable replication product (lanes 2-7). These results demonstrate that H507 is important for DNA replication and that F or Y, which have side chains similar to that of H, are also active for replication.

2.2.2 Complex formation by substitution mutants at H507.

Preparation of a template for initiation of DNA replication can be divided into several steps. The first step, sequence-specific binding of the initiator to the ori, is followed by local melting of the ori, which can be detected by treatment with

permanganate, which reacts with unbasepaired thymine. After melting, the DNA helicase unwinds the template. A defect in any of these steps would result in a defect for DNA replication. To determine whether the H507 substitutions could generate the appropriate E1 DNA complexes required for these different activities, we first tested them for their ability to form a DH on a short (32bp) probe (Figure 2-2A). This complex, which is non-functional and does not unwind the template, provides a control for the ability of E1 to oligomerize into a DH and to bind and hydrolyze ATP[17]. All mutants with the exception of H507V (lanes 7-9) were capable of DH formation on the 32 bp probe, albeit at different levels. This results demonstrates that the β-hairpin is not directly involved in E1 oligomerization or ATP binding and hydrolysis. The H507V protein (lanes 7-9) was prone to aggregation, which is the likely cause for its failure to function in EMSA.

We have previously demonstrated that E1 has the intrinsic propensity to form E1 BS-independent trimers on dsDNA in the presence of ADP and that formation of the trimer relies on the E1 β-hairpin[17]. We next tested the mutant proteins for their ability to form this trimer. Using a short probe where the E1 BS had been mutated, the WT E1 (Figure 2-2B, lanes 1-3) and two of the mutants, H507F (lanes 10-12) and H507Y (lanes 19-21) formed the trimer while the other substitutions failed to do so. Clearly, the failure to form the trimer did not result from a general DNA binding defect, since all of the substitutions formed other complexes such as dimers on this probe. H507V (lanes 25-27) was the only mutant with a general DNA binding defect, likely due to the aggregation problem mentioned above.

We next determined whether the H507 substitutions could form the functional DT that melts the ori and the functional DH that unwinds the ori. Formation of these complexes requires an 84bp ori probe with four E1 BS (Figure 2-2C). In the presence of ADP, the progression from DT to DH is arrested at the DT stage because DH formation requires ATP hydrolysis. We can therefore analyze DT formation in the presence of ADP, and analyze DH formation in the presence of ATP for each mutant protein. Our expectation was that the same mutants that were unable to form the trimer would also be unable to form the DT. Furthermore, since the DT is a precursor for the DH, the mutants defective for trimer and DT formation are also expected to be defective for DH formation on the 84 bp probe. Only the two mutants H507Y (lanes 8-10) and H507F (lanes 14-16), which could form the trimer, were also capable of forming the DT. H507F also formed a DH (lanes 17-19), while H507Y failed to do so (lanes 11-13). The rest of H507 substitutions all failed to form both DT and DH (Lanes 20-25 and lanes 27-50). We can therefore distinguish three types of mutants. The majority of the substitutions are defective for DNA replication as well as for trimer, DT and DH formation. The two substitutions H507F and H507Y are exceptions in that they have replication activity and also are able to form a trimer and DT, however only H507F forms the DH on the ori probe.

2.2.3 Template melting by H507 substitutions.

The results presented above demonstrated that the majority of the H507 substitutions (H507A, V, L, R, N, and M) are defective for DT formation, which may account for their replication defect. Since the DT melts the ori, we would expect that the mutants defective for DT formation would also have melting defects. We therefore performed permanganate reactivity assays, which detect melted DNA (Figure 2-3). We incubated an ori probe labeled on the top strand with the WT or mutant E1 proteins in the presence of ADP and treated with $KMnO_4$ for 2 min at room temperature. WT E1 generates a characteristic melting pattern consisting of melting at three positions, one thymine to the left of the E1 BS (L), one thymine to the right of the E1 BS (R) and the T6-stretch to the left of the E1 BS (Figure 2-3, lane 2). The right hand reactivity was observed with all the mutants with the exception of H507V. The left reactivity and the reactivity in the T6 stretch were only observed with the mutant H507F, albeit at lower levels than for WT E1 (lane 9). These results demonstrate that all substitutions at H507 with the exception of H507F show melting defects and that although H507Y can form a DT, this DT does not melt the DNA correctly. This result was surprising in light of the activity of H507Y for DNA replication *in vitro* (Figure 2-1D). We have provided an explanation for the activity of H507Y in *in vitro* replication later in the paper.

2.2.4 H507 is not required for DNA helicase activity.

An ori-fragment unwinding assay is a comprehensive assay that measures the ability of E1 to bind DNA correctly, to locally melt the ori, and to provide DNA helicase activity and, most importantly, to couple these activities to each other. In such an assay,

E1 is incubated with an ori fragment in the presence of *E. coli* SSB and generation of ssDNA can be detected by the appearance of ssDNA as an ssDNA/SSB complex by EMSA (Figure 2-4A). The H507F substitution, which had near-WT activity for DNA replication *in vivo*, showed significant ori fragment unwinding activity, indicating that it can bind, melt and unwind the ori fragment nearly as well as the WT protein (compare lanes 3 and 4 to lanes 17 and 18), consistent with the *in vivo* and *in vitro* replication results. In contrast, H507Y (lanes 19 and 20) showed only trace amounts of activity in this assay (~10% of WT), similar to H507L and H507N (lanes 9-10 and 13-14), while the rest of the substitutions lacked detectable activity. These results are completely consistent with the melting results above and indicate that the lack of melting observed for all mutants except H507F in the permanganate reactivity assays also results in a defect for ori fragment unwinding.

To ascertain that defects in helicase activity were not the cause of the replication defect, we measured DNA helicase activity of the H507 substitutions using several different helicase assays, one of which is shown in Figures 2-4B and 2-4C, where we used a time-resolved fluorescence helicase assay. In this assay, E1 is incubated with a fluorescently labeled oligonucleotide substrate in which the fluorescence is quenched by the presence of the complementary strand. As the substrate is unwound, the quencher is removed, resulting in a dramatic increase in fluorescence. Thus, the fluorescence and hence the unwinding in a single sample can be monitored in real time. We chose a low ratio of E1 to substrate (8:1), approximately one E1 hexamer per helicase substrate, to reveal differences in all steps of the unwinding process including the binding of E1 to the

substrate and formation of the hexameric helicase on ssDNA[110]. We then measured unwinding every two min for 32 min (Figure 2-4B). Interestingly, the five substitutions (A, M, F, Y and R) that showed the lowest levels of unwinding compared to WT E1 had only a 2- fold reduction in the level of unwound substrate after 30 min. H507N and H507L showed levels of unwinding similar to that of WT E1 after 30 min. We also in a separate experiment tested WT E1, H507N and H507 in triplicate and obtained virtually identical results (Figure 2-4C). Clearly, these modest helicase defects are not responsible for the replication defects of the H507 substitutions. H507F, which has among the lowest activities in the helicase assay, is the only substitution that has close to WT activity for DNA replication *in vivo* and *in vitro*. Similarly, H507N, which shows no defect for helicase activity, has no activity for DNA replication. A position in which such a wide range of substitutions only has modest effects on the helicase activity is clearly not a critical residue for the helicase activity of the E1 protein.

An interesting aspect of these helicase assays is the slow rate of unwinding that is observed in all cases. This effect is observed in all of our helicase assays and is not related to the fluorescent substrate (data not shown). Since the substrate is very short, it is unlikely that this reflects the actual unwinding step. Instead, this slow step likely reflects the formation of the hexameric helicase on the substrate. It is interesting that some of the mutants, e.g., H507R, H507A, and H507M, plateau significantly earlier (7–10 min) than the rest of the mutants (~15 min), possibly reflecting a more rapid formation of the active helicase on the substrate. This may indicate that binding to these artificial helicase substrates is affected by the mutations at H507. Nevertheless, these results demonstrate

that a defect for helicase activity is not likely to account for the DNA replication defect of substitutions at H507.

2.2.5 Another residue in the β-hairpin, K506, is required for DNA helicase activity.

These results demonstrate that the function of the conserved histidine at the tip of the β-hairpin is to melt dsDNA and that clearly this histidine is not required for helicase activity. We have previously generated a mutation in the residue adjacent to H507 (K506A)[17]. This mutant has similar defects as H507A and fails to form a trimer and a DT (Figure 2-5A, compare lanes 2-4 and 5-7) and also fails to melt the template (data not shown), although this mutant can form the nonfunctional DH on a short probe as well as WT E1 (Figure 2-5A, compare lanes 9-11 and 12-14) and therefore is capable of oligomerization and ATP hydrolysis. However, in contrast to the substitutions at position H507, which maintained significant helicase activity, K506A has no detectable helicase activity compared to WT E1 (Figure 2-5 B, compare lanes 1-5 and 6-10). Thus, template melting is affected by substitution of either of K506 or H507, but the helicase activity is significantly affected only by the K506A substitution. This result demonstrates that the β-hairpin is involved in at least two different biochemical activities, template melting and helicase activity and that these two functions can be distinguished by substitutions at these two adjacent residues.

2.2.6 H507Y shows altered specificity for initiation of DNA replication *in vitro*

The lack of melting and unwinding activity of H507Y was curious, given the high levels of *in vitro* DNA replication for this mutant (Figure 2-1D). To understand this discrepancy, we performed a form-U assay, which detects unwinding as a fast-migrating unwound form of an input plasmid [111] (Figure 2-6A). The substrate for this assay is generated by incubating an ori plasmid with topoisomerase to generate a relaxed form of the plasmid (compare lanes 2 and 3). During unwinding process, positive supercoils are introduced in front of the unwinding fork while the negative supercoils are left behind. The eukaryotic topoisomerase I in the unwinding system relax the positive supercoils to release the tension for the helicase to move forward. The complete unwinding of a circular dsDNA plasmid will end up with two interlinked circular ssDNA molecules with negative supercoils (form-U). This form-U plasmid migrates at a similar position as supercoiled form I plasmid because of the negative supercoils [111]. In the absence of added E1, incubation of this substrate in the presence of SSB, ATP and topoisomerase I results in loss of some topoisomers due to further relaxation (compare lanes 3 and 4). In the presence of E1, SSB, ATP and topoisomerase I, form-U plasmid is generated (lane 5). In this assay, we observed significant unwinding with H507F (lane 12) and importantly now also with H507Y (lane 13) which had ~2-fold reduced activity compared to WT E1, consistent with the *in vitro* DNA replication result. The remaining mutants showed unwinding levels 3- to 10-fold lower than those of WT E1 (lanes 6-11).

These data demonstrate that H507Y is unable to unwind the ori fragment but is active for DNA synthesis *in vitro* and for unwinding in the form-U assay. A distinction between these assays is that the fragment unwinding assay uses a small (~100 bp) ori

fragment, while both the *in vitro* DNA replication assay and the form-U assay use the whole ori plasmid. It is unlikely that topological differences are responsible for the differing results since the substrate for the form-U assay is relaxed with topoisomerase I and therefore is topologically similar to a linear DNA fragment. An interesting possibility is that H507Y can use other sequences present in the plasmid backbone for unwinding in the form-U assay and for replication *in vitro*.

To determine whether the H507Y mutant showed an "altered specificity" phenotype we compared WT E1, H507F and H507Y for the ability to initiate replication on plasmids with and without an ori (Figure 2-6B). Significant levels of replication can be observed on plasmids lacking an ori due to the low ori specificity of the E1 protein in the absence of E2 [41, 78]. We compared the ability of WT E1, H507F and H507Y to replicate pUC 19 and the ori plasmid, which contains a 60 bp ori fragments inserted into the polylinker of pUC 19. The replicated DNA was digested with Pvu II and Dde I and analyzed by PAGE. WT E1 showed ~3-fold reduced levels of replication with the pUC 19 template compared to the ori template, as expected (Figure 2-6B, compare lanes 1 and 4). H507F showed ~ 2-fold reduction with the pUC 19 template compare to the ori template (lanes 2 and 5). Strikingly, H507Y showed identical levels of replication in reactions containing pUC 19 or the ori plasmid (compare lanes 3 and 6), demonstrating that H507Y has no preference for the ori sequence and uses other sequences in the plasmid backbone for replication. This result is consistent with both the lack of activity in the fragment unwinding experiment (Figure 2-4A) and the substantial activity in the form-U assay (Figure 2-6A).

Importantly, although this replication result would be consistent with a reduction in ori specificity of H507Y, the failure to unwind the ori in the fragment unwinding assay (Figure 2-4A) demonstrates that the effect is more likely to result from an altered specificity. An altered specificity of H507Y also provides an explanation for the *in vivo* DNA replication results. Initiation of DNA replication *in vivo* requires cooperative binding of E1 and E2, which relies on binding sites for E2[1]. Since E2 binding sites are present only at the ori, H507Y cannot initiate at other sites in the plasmid and is confined to the suboptimal ori sequence for initiation, resulting in a severely reduced level of DNA replication *in vivo* compared to WT E1 and H507F (Figure 2-1C).

To determine whether a particular sequence in the plasmid backbone was required for initiation of DNA replication by H507Y, we performed *in vitro* replication assays in which we terminated the replication reaction early (at 8 min) to trap intermediates. After termination of the replication reactions, we digested the DNA with the restriction enzymes Pvu II and Dde I and analyzed the samples by PAGE (Figure 2-6C). Incorporation into each DNA fragment was quantitated and total incorporation was divided by fragment size to yield the relative incorporation/bp, which was plotted as a function of the position in the ori plasmid (Figure 2-6D). WT E1 and the H507F mutant gave rise to identical patterns, with ~5-fold enrichment at the origin of DNA replication compared to sequences distal to the ori. In contrast, the level of replication at the ori was reduced significantly for H507Y, consistent with an altered sequence specificity of

H507Y. The ability of H507Y to utilize a different target sequence is equivalent to restoration of the interaction by a compensatory mutation in the target. Such data is generally considered to be good evidence for a direct interaction. We have not been able to identify the sequence(s) that is used instead of ori, most likely because multiple such sequences are present.

2.3 Discussion

2.3.1 The β-hairpin is required for both melting and DNA helicase activity, but the two activities can be separated by mutation of conserved residues

We have analyzed two highly conserved residues in the E1 initiator protein. These residues, K506 and H507, are part of a highly conserved structural motif, the β-hairpin, which is present in the helicase domain of all papovavirus initiator proteins. These viral initiator proteins have been considered unique among helicases in that they are capable of initiating unwinding from completely double stranded DNA [111]. This ability is not due to the properties of the helicase, but results from the ability to generate an activity distinct from the helicase, the DT, which melts dsDNA prior to formation of the helicase[17]. Significantly, none of the substitutions at H507 had a severe defect for DNA helicase activity. This result rules out a role for the conserved histidine in DNA helicase activity. This particular residue therefore appears to be dedicated to the formation of the DT and melting of the template. The adjacent residue, K506, is required for both melting and helicase activity since K506A lacks helicase activity entirely, in addition to showing the

same defects as H507A for DT formation and melting. This demonstrates that the β-hairpin plays different roles for melting in the DT and for helicase activity in the DH. The effects of the substitutions at H507 on helicase activity are modest. Therefore it is clear that this particular residue does not play an important role in helicase activity of the E1 protein, consistent with the near WT activity of for example H507F.

2.3.2 H507 is required for local ori melting, but not for DNA helicase activity of E1

The majority (six out of eight) of the substitutions at H507 are severely defective for all the activities that we can measure, with the exception of DNA helicase activity. These defects include trimer and DT formation, melting, unwinding and DNA replication *in vivo* and *in vitro*. None of the substitutions appear to have general defects such as defects for DNA binding, oligomerization or ATP binding/hydrolysis since they all can form the non-productive DH on the short (32 bp) probe. The underlying cause for the defect for unwinding and DNA replication for these substitutions appears to be the failure to form the correct protein-DNA complex, i.e. the DT. This in turn is likely caused by the failure of the substituted side chains to interact correctly with DNA, as will be discussed below. The excellent correlation between the H507 β-hairpin substitutions that support DNA replication and also support trimer and DT formation provides good evidence that the recently characterized DT is an essential functional complex that melts the ori DNA, as has been proposed [17]. The fact that H507 and K506 play different roles in the helicase activity of E1 is interesting in light of the cocrystal of a hexamer of the E1 oligomerization and helicase domains bound to ssDNA[28]. In this structure, K506

interacts through its side chain with ssDNA, providing an explanation for the importance of this residue for the helicase activity of E1. H507 in this structure interacts both through the main chain and through its side chain with the ssDNA. The side-chain interaction is clearly of limited importance for the helicase activity of E1 based on the mutational results presented here.

2.3.3 The conservation of the β-hairpin

The E1 proteins belong to the SF3 family of helicases, which also includes the initiator proteins from polyoma viruses (T-ag) and parvoviruses (Rep). Interestingly, when these three groups of helicases are aligned (Figure 2-7), the lysine at the position corresponding to K506 in BPV1 E1 is completely conserved between these three virus groups, consistent with the importance of this residue for the helicase activity of the E1 protein. In contrast, although the papilloma and polyomaviruses have a virtually completely conserved His at position 507(all of the 214 T-ag sequences have a His at this position, and out of 168 E1 proteins only 13 [12 tyrosine, 1 phenylalanine] do not have the conserved histine), the parvovirus group contains a variety of residues (C, N, G, H, and M) at this particular position. Since it is well established that the parvovirus Rep proteins are functional helicases, this clearly demonstrates that a histidine at this particular position is not an essential feature of SF3 helicases. We believe that the reason that the histidine is not conserved in the parvoviruses is that the parvoviruses have ssDNA genomes and therefore do not require an activity that locally melts the ori.

2.3.4 H507Y shows altered specificity

H507Y is capable of forming the trimer and DT, and is fully functional for *in vitro* DNA replication. However, it fails to melt and unwind the ori fragment (Figure 2-3 and Figure 2-4). This indicates that, while the Tyr side chain interacts with DNA equally well as the histidine or phenylalanine to form the DT, there is a structural aspect of tyrosine side chain that is incompatible with local ori melting and for DH formation. H507Y apparently is capable of using other DNA sequences for unwinding in the form-U assay and in the *in vitro* DNA replication assays, indicating that the consequence of the H507Y substitution is a change in the sequence that can be melted.

The altered specificity of H507Y indicates that the histidine side chain may contact DNA directly. F and Y both have similar aromatic six-membered ring structures, and the flat ring structure is shared with the histidine (an aromatic five-membered ring). A possible mode of interaction with DNA based on these side-chains is intercalation of the planar rings into DNA, as has been observed for binding of TBP, in which a phenylalanine in each stirrup intercalates into the minor groove [112, 113]. This intercalation results in major distortion of the DNA, including unstacking of neighboring basepairs.

The likely target for the β-hairpin is the stretch of 6 T-A bp, which is the only sequence flanking the E1 BS that is essential for DNA replication. These are the same 6 thymines that are melted by the E1 DT (Figure 2-3). Mutants affecting this T-A stretch

have a melting defect, but such templates can still support DT formation[54]. The dependence on the T-A stretch for melting but not for DT formation, mirrors the distinction that we observe between H507F and H507Y. While the phenylalanine and tyrosine substitutions both can generate the DT on the ori fragment, the tyrosine substitution cannot melt this template. An interesting possibility, therefore, is that the histidine and phenylalanine at the tip of the β-hairpin both are capable of engaging the 6 T-A bp flanking the E1 BS, while the extra hydroxyl of tyrosine, although it allows DT formation, precludes the β-hairpin from interacting with DNA in a manner that is required for melting.

2.4 Material and Methods

E1 Protein: Expression and Purification

WT E1 and E1 mutants were expressed in E. coli as N-terminal GST fusions, purified by affinity chromatography, cleaved and isolated by ion exchange chromatography as described [97]. E1 purified in this manner is monomeric, as determined by glycerol gradient sedimentation and gel filtration [110]. The concentrations of the WT and mutant E1 proteins were determined with the fluorescent dye SYPRO Red using 532nm as the excitation wave-length in a FujiImager FLA 5000.

Plasmid constructs

The template used for *in vivo* and *in vitro* DNA replication assays was a 60 bp minimal ori fragment (7914-27) cloned between the XbaI and HindIII sites in pUC 19 [40]. The template for generation of ori probes is a 110-base pair BPV-1 sequence (nucleotides 7894-57, centered on the E1-binding site) cloned between the XbaI and HindIII sites in pUC19 generating the plasmid 11/12/X [114].

***In vivo* DNA Replication Assay**

In vivo DNA replication assays were performed as described [39, 40]. Briefly, CHO cells were transfected by electroporation with an expression vector for the E2 protein (pCGE2), an ori plasmid (7914–7927), and an expression vector for either WT E1 (pCGE1) or the respective H507 substitutions. Two, three, and four days after transfection, low molecular weight DNA was prepared using alkaline lysis, digested with DpnI and HindIII, and analyzed by Southern blotting.

***In vitro* DNA Replication**

In vitro DNA replication assays were performed essentially as described [41, 78]. *In vitro* replication was performed in 25 µ1 reaction mixtures containing the following: 40 mM HEPES-KOH (pH7.5), 8 mM $MgCl_2$, 0.5 mM DTT, 3 mM ATP, 0.2 mM each of GTP, UTP and CTP; 0.1 mM each of dATP, dGTP and dTTP, 10 µM of [α-^{32}P]dCTP (2 µCi;

3000 Ci/mmol), 40 mM creatine phosphate, 400 ng creatine kinase, 10 µl S100 extract and 0.5 µl high salt nuclear extract from H293 cells. The concentration of template in the *in vitro* reactions was 2ng/µ1. Reactions were incubated for 60 min at 37°C unless stated otherwise. The reactions were stopped by addition of SDS to 1% and EDTA to 10 mM, treated with proteinase K followed by phenol/chloroform extraction and precipitation with ethanol and ammonium acetate. The products were analyzed by electrophoresis on 1% agarose gels in TAE buffer.

For the *in vitro* replication assays shown in Figure 2-6, a slightly different protocol was used. To alleviate the long lag preceding initiation that is observed in these assays, the replication mix was preincubated for 20 min at 32°C in the presence of ATP but in the absence of rNTPs and dNTPs. Upon addition of rNTPs and dNTPs, initiation of DNA replication is instantaneous.

DNA helicase Assays

Oligonucleotide displacement assays were performed using two methods. For the time course experiments in Figure 2-4, a helicase kit from Perkin Elmer (Trupoint) was used. Briefly, an europium-labeled 44-mer annealed to a 26-mer containing a fluorescence quencher was incubated with E1 in a buffer containing 50 mM Tris-HCl (pH 7.9), 5 mM $MgCl_2$, 2 mM DTT, 1 mM ATP, and 0.2 mg of BSA/ml and incubated at 37C. Time-resolved fluorescence was measured every 2 min using 1420 Victor software in a PerkinElmer fluorometer. The substrate concentration was 4 nM and the capture strand

concentration was 15 nM and the reactions were carried out in a volume of 50 µl. WT E1 was used at two concentrations, 30nM and 60 nM (2 × E1), while the H507 substitutions were tested at 30 nM. At 30nM of WT E1, maximally 50% of the substrate is unwound.

Oligonucleotide displacement assays in Figure 2-5 were performed essentially as described [79, 110]. A 50-mer oligonucleotide with partial complementarity to M13mp18 was synthesized, generating a substrate with a 28 nucleotide long double-stranded region and a 22 nucleotide long single-stranded 3′ tail. E1 was incubated with substrate in a buffer containing 50 mM Tris-HCl (pH 7.9), 3 mM $MgCl_2$, 2 mM DTT, 1 mM ATP, and 0.2 mg of BSA/ml at 37°C for 15 min. After incubation, SDS was added to 0.1%, and the samples were analyzed by PAGE.

EMSA

Four percent acrylamide gels (39:1 acrylamide:bis) containing 0.5 × TBE, lacking EDTA, were used for all EMSA experiments. E1 was added to the probe (~2 fmol) in 10 µl binding buffer, BB (20 mM HEPES [pH 7.5], 100 mM NaCl, 0.7 mg/ml BSA, 0.1% NP40, 5% glycerol, 5 mM DTT, 5 mM $MgCl_2$, and 2 mM ATP or ADP). After incubation at room temperature for 1 hr, the samples were loaded and run for 2 hr at 9V/cm. The ability to generate discrete complexes, especially the DT and DH, was critically dependent on high-purity acrylamide, freshly prepared APS solution, overnight polymerization of the gels, and precise prerunning time (9V/cm for 4 hr).

Combined EMSA and unwinding assays

Unwinding assays were performed by incubating 2 fmol of probe with E1 at 32°C for 30 min under EMSA conditions (see above), but in the presence of 10 mg/ml *E. coli* SSB. Before loading samples on the EMSA gel, the concentration of NaCl was increased to 500 mM to disrupt E1 DNA complexes. The ssDNA was detected as an SSB/ssDNA complex.

Form-U assays

Relaxed DNA substrate was prepared by incubation of plasmid DNA with human topoisomerase I (Sigma) in 10 mM Tris HCl (pH 7.5), 100 mM KCl, 1 mM PMSF, and 1mM 2-mercaptoethanol. After incubation overnight at 37°C, the relaxed products were deproteinized, ethanol precipitated, and resuspended in 10 mM Tris HCl (pH 7.5) containing 1 mM EDTA. The conditions for unwinding were modified from those described previously[111]. Reaction mixtures (30µ1) containing 20 mM HEPES (pH 7.5), 5% glycerol, 40 mM creatine phosphate, 7 mM $MgCl_2$, 1 mM dithiothreitol, 4 mM ATP, 33 µg of creatine kinase/ml, 125 ng of relaxed ori plasmid DNA, 2 units of human topoisomerase I, 450 ng of *E. coli* SSB, and 400 ng of E1 were incubated for 3 hr at 37°C, and the reaction was terminated by the addition of 15 mM EDTA, 0.3% SDS, 1 µg of tRNA, 0.3 µg of proteinase K and further incubated for 15 min at 37°C. Reactions were extracted with phenol/chloroform, and ethanol precipitated. The DNA was dissolved in 10 mM TrisHCl (pH 7.5), 1 mM EDTA, and electrophoresed in 1% agarose gels at 9V

per cm. Gels were stained with ethidium bromide, photographed and quantitated under UV irradiation.

Permanganate Reactivity Assays

Permanganate reactivity assays were performed as described [110]. Briefly, binding reactions (20 mM HEPES [pH 7.9], 100 mM NaCl, 0.1% NP40, 5% glycerol, 5 mM DTT, 5 mM $MgCl_2$, and 5 mM ATP) containing ~10 fmol of end-labeled probe were assembled and incubated with E1 at room temperature. After 30 min., KMnO4 was added to a final concentration of 6 mM and reactions was incubated for a further 2 min. Modification was terminated by adding β-mercaptoethanol to 80 mM, SDS to 0.3% and EDTA to 10 mM. Reactions were then digested with proteinase K (20 µg/ml) for 60 min at 37°C and the DNA was recovered by phenol–chloroform extraction and ethanol precipitation. Cleavage at modified bases was achieved with piperidine (30 min at 90°C).

Accession codes

Protein Data Bank: E1 oligomerization and heliase domain hexamer structure coordinates can be accessed with PDB code 2GXA[28].

Figure 2-1. DNA replication by substitution mutants in the β-hairpin

A. Image representing the structure of a hexamer of the BPV E1 helicase domain highlighting the positions of two residues, K506 and H507, which form the tip of the β-hairpin and generate an inner constriction in the hexameric ring. Adapted from Enemark, E.J. and L. Joshua-Tor, Mechanism of DNA translocation in a replicative hexameric helicase. Nature, 2006. 442(7100): p. 270-5 [28].

B. Image of the β-hairpin structure in BPV E1 showing the highly conserved residues K506 and H507.

C. *In vivo* DNA replication. The eight substitutions at residue H507 were introduced into the mammalian E1 expression vector pCGE1 by site-directed mutagenesis and tested for activity in transient DNA replication assays. Ori plasmid (100ng), 0.5 µg E2 expression vector (pCGE2) and 2 µg expression vectors for the WT or mutant E1 proteins were co-transfected into CHO cells using electroporation, and low molecular weight DNA was harvested 48, and 72 and 96 hours after transfection. The DNA was digested with HindIII, which linearizes the plasmid and DpnI, which digests unreplicated, methylated DNA and analyzed by Southern blotting.

D. Cell-free DNA replication. The ability of WT E1 and the eight H507 substitutions to support DNA replication in a cell-free replication system was compared. Each substitution mutant protein (6 pmol), purified from *E.coli*, was incubated with 50 ng of ori plasmid in replication mix (see Material and Methods) in the presence of radiolabeled dCTP. After 60 min at 37°C, the replication products were isolated and analyzed by agarose gel electrophoresis and quantitated.

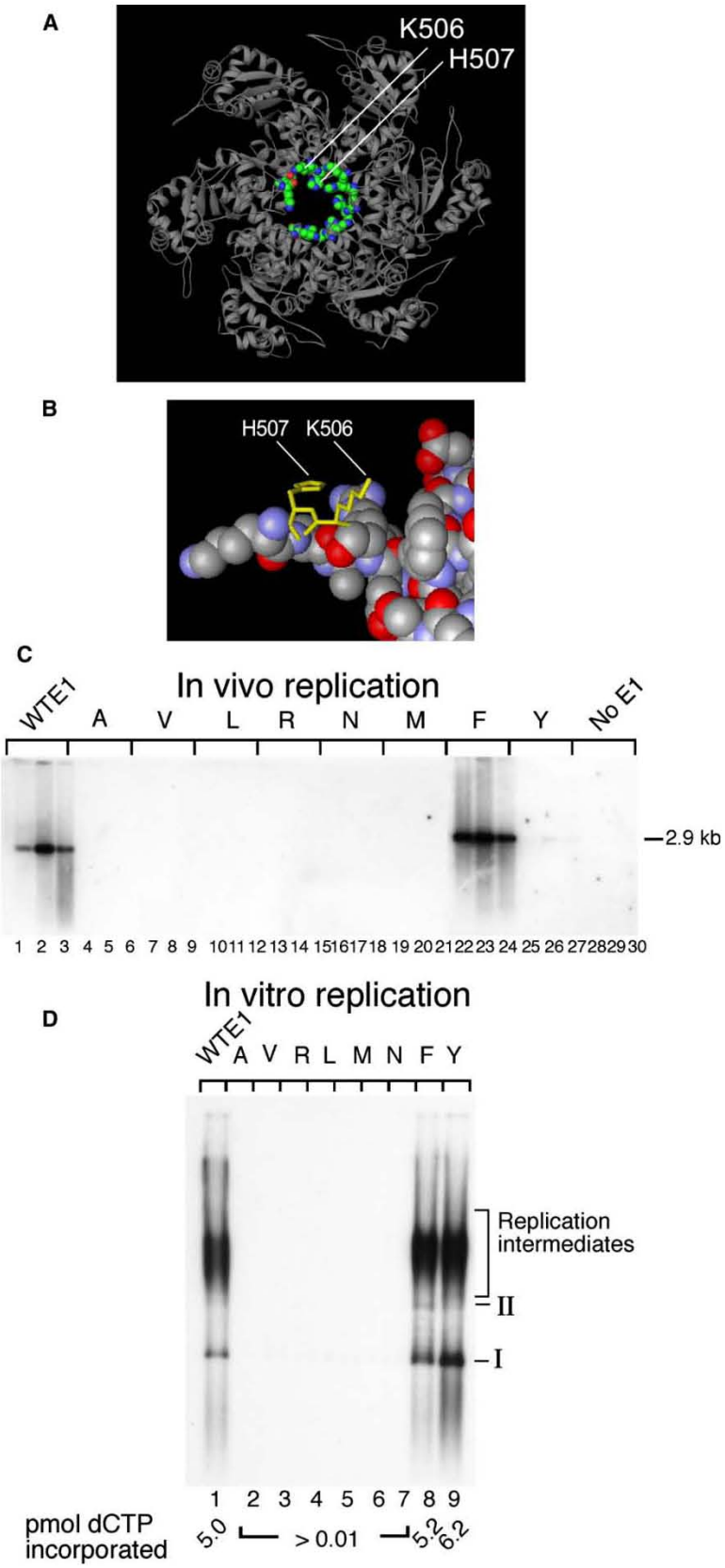

A
K506
H507
B
H507 K506
C
In vivo replication
WTE1
A V L R N M F Y
No E1
2.9 kb
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
D
In vitro replication
WTE1
A V R L M N F Y
Replication intermediates
II
I
1 2 3 4 5 6 7 8 9
pmol dCTP incorporated
5.0
> 0.01
5.2
6.2

Figure 2-2. Complex formation by substitution mutants at H507

A. Formation of the nonproductive DH by H507 substitutions. EMSA was performed using the 32 bp ori probe shown below the panel. Three quantities of the WT and mutant E1 (60, 120 and 240 fmol) were used in the presence of ATP as indicated in the figure. Lane 28 contained probe alone**.**

B. Trimer formation by H507 substitutions. EMSA was performed using a 39 bp probe with a mutated E1 BS as shown below. Three quantities (15, 30 and 60 fmol) of the WT E1 (lanes 1-3) and the respective mutant proteins were used in the presence of ADP. Lane 28 contained probe alone.

C. DT and DH formation by H507 substitutions. EMSA was performed using the 84 bp ori probe shown below. Three quantities (30, 60 and 120 fmol) of the WT E1 and the respective mutant proteins were used in the presence of ADP or ATP as indicated in the figure. Lane 1 and lane 26 contained probe alone.

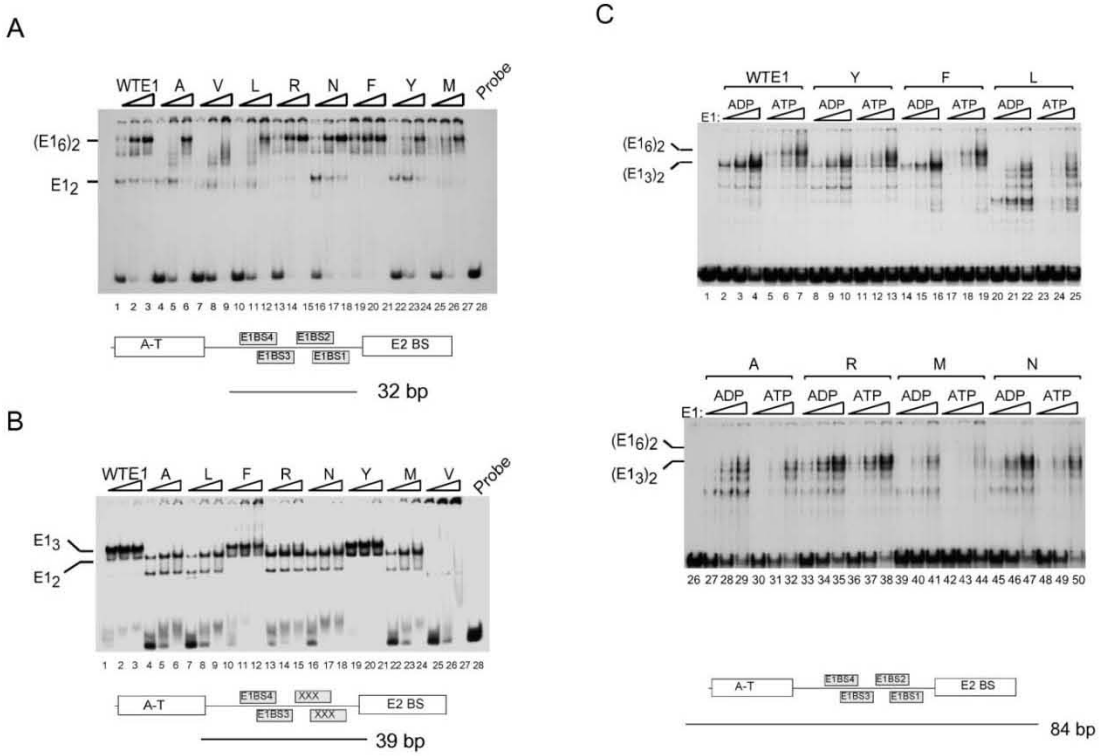

A
WTE1 A V L R N F Y M Probe
(E1$_6$)$_2$
E1$_2$
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
A-T
E1BS4
E1BS2
E1BS3
E1BS1
E2 BS
32 bp
B
WTE1 A L F R N Y M V Probe
E1$_3$
E1$_2$
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
A-T
E1BS4
XXX
E1BS3
XXX
E2 BS
39 bp
C
WTE1 Y F L
E1: ADP ATP ADP ATP ADP ATP ADP ATP
(E1$_6$)$_2$
(E1$_3$)$_2$
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
A R M N
E1: ADP ATP ADP ATP ADP ATP ADP ATP
(E1$_6$)$_2$
(E1$_3$)$_2$
26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
A-T
E1BS4
E1BS2
E1BS3
E1BS1
E2 BS
84 bp

Figure 2-3. Permanganate reactivity of substitutions at residue 507

Permanganate reactivity assays were performed by incubating the ori probe with 1 pmol of WT E1 (lane 2) or of the respective H507 substitutions (lanes 3-10) in the presence of ADP. Following treatment with $KMnO_4$, modified DNA was cleaved with piperidine and analyzed by denaturing PAGE. Lane 1 shows the permanganate reactivity of the probe in the absence of E1. Positions of prominent permanganate reactivity (R, L, and T6) are indicated. Below is a schematic for the positions of permanganate reactivity for WT E1 relative to the ori sequence.

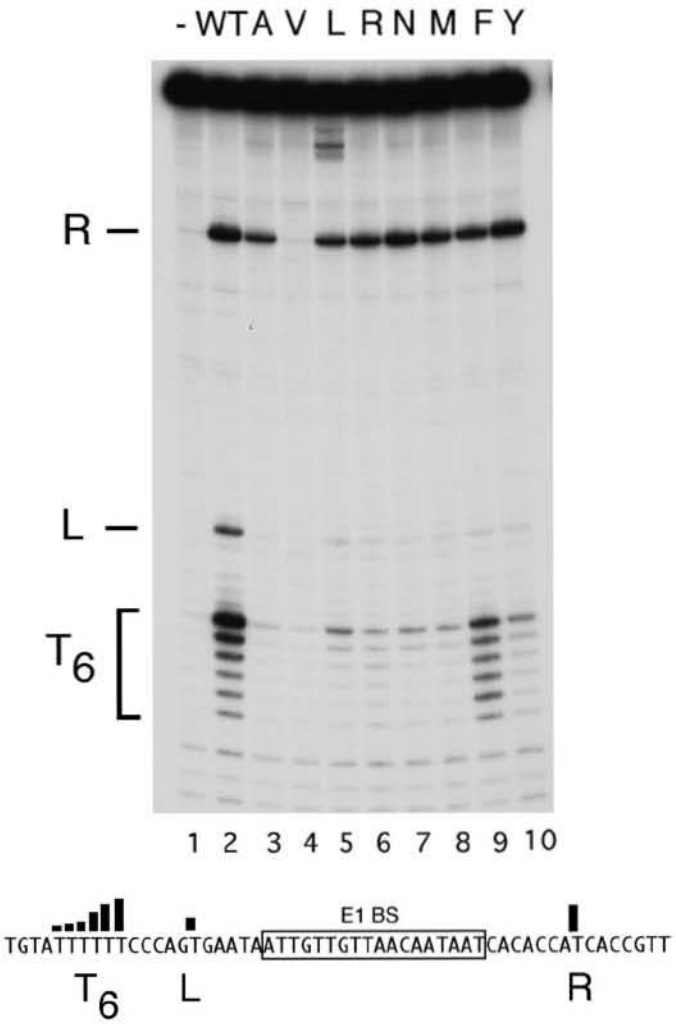

- WT A V L R N M F Y
R
L
T6
1 2 3 4 5 6 7 8 9 10
E1 BS
TGTATTTTTTCCCAGTGAATAATTGTTGTTAACAATAATCACACCATCACCGTT
T6
L
R

Figure 2-4. Ori Fragment Unwinding and Helicase Activity of Substitution Mutants at H507

A. Ori fragment unwinding by substitution mutants at H507. Ori fragment unwinding assays were performed by incubating an ori probe with 80 or 160 fmol of WT E1 (lanes 3 and 4), or of the respective H507 substitutions (lanes 5-20) in the presence of *E. coli* SSB followed by analysis by PAGE. Lane 1 contained probe alone; in lane 2, SSB was added to denatured probe, providing a marker for the ssDNA-SSB complex.

B. DNA helicase activity of substitution mutants at H507. The substitutions at H507 were tested for DNA helicase activity using a time-resolved fluorescence based oligonucleotide displacement assay. WT E1 (1.5 pmol) or the respective E1 mutants were incubated with 200 fmol of fluorescent substrate at 37°C, and fluorescence was measured every 2 min for 32 min. In one sample (2 × WT E1) 3 pmol of WT E1 was used.

C. WT E1, H507F, and H507N were tested for DNA helicase activity in triplicate using the same procedure as in (B). The error bars represent the standard deviation for each point.

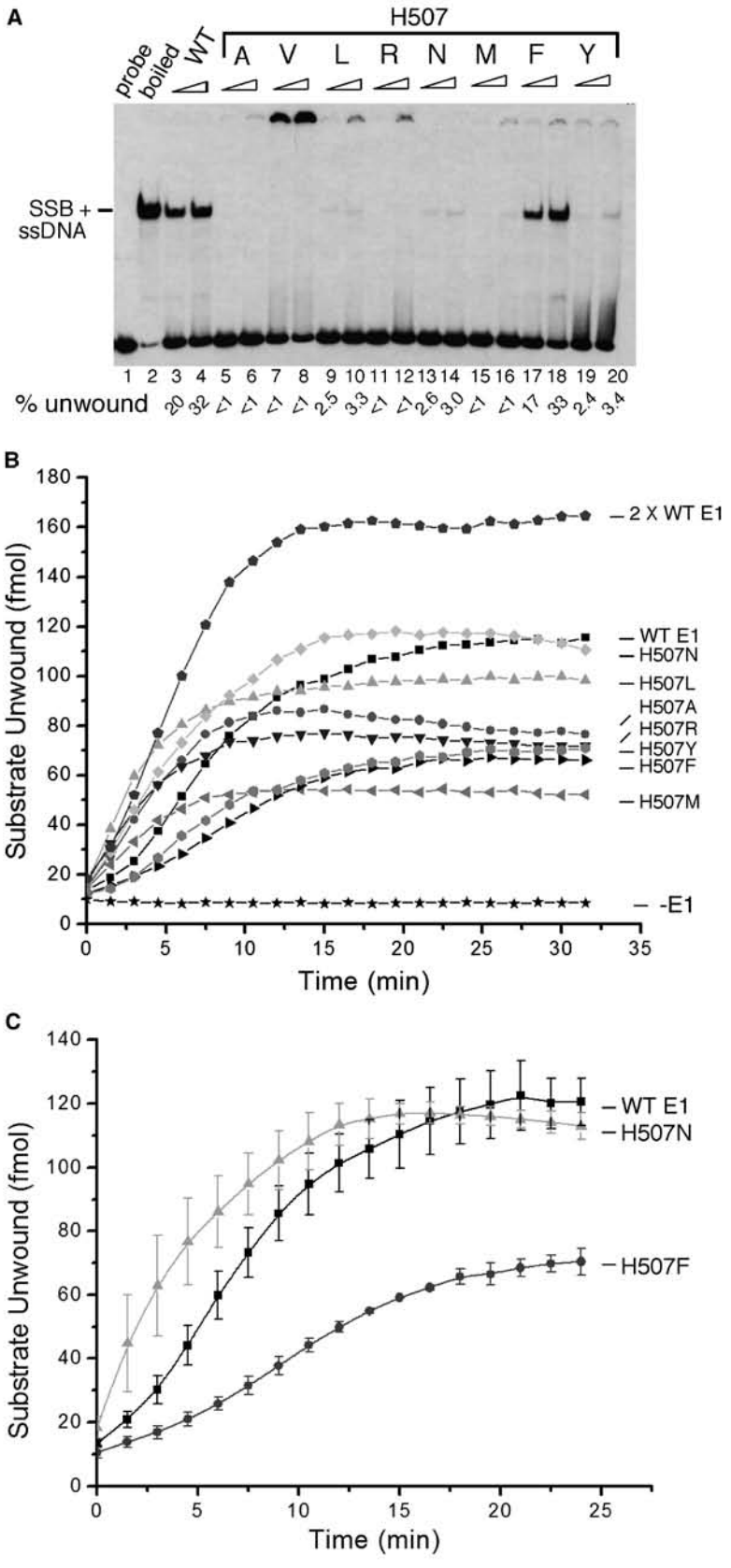

A
H507
probe
boiled
WT
A V L R N M F Y
SSB +
ssDNA
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
% unwound 20 32 <1 <1 <1 <1 2.5 3.3 <1 <1 2.6 3.0 <1 <1 17 33 2.4 3.4
B
Substrate Unwound (fmol)
Time (min)
2 X WT E1
WT E1
H507N
H507L
H507A
H507R
H507Y
H507F
H507M
-E1
C
Substrate Unwound (fmol)
Time (min)
WT E1
H507N
H507F

Figure 2-5. K506A is defective for double trimer formation and for DNA helicase activity

A. The β-hairpin substitution mutant K506A was tested by EMSA for the ability to form the DT (left panel) and the DH (right panel) in parallel with WT E1. In the left panel three quantities (30, 60, and 120 fmol) of WT E1 (lanes 1-3) and K506A (lanes 4-6) were used in the presence of ADP. In the right panel, 60, 120 and 240 fmol of WT E1 (lanes 1-3) or K506A (lanes 4-6) were used with the short 32 bp ori probe shown below.

B. K506A was tested for DNA helicase activity in parallel with WT E1. Five quantities (15, 30, 60, 120, 240 fmol) of WT E1 (lanes 1-5) or K506A (lanes 6-10) were incubated with the helicase substrate and analyzed by PAGE. In lane 11, no E1 was added, and lane 12 shows a boiled sample.

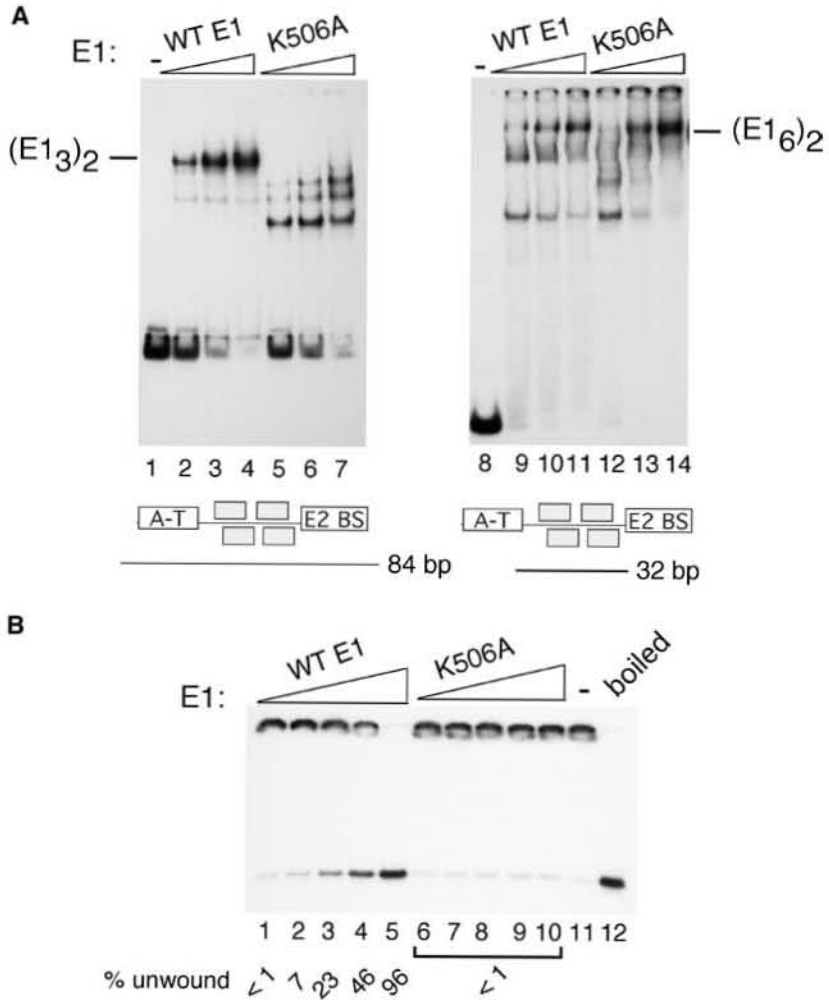

A
E1:
-
WT E1
K506A
(E1$_3$)$_2$
(E1$_6$)$_2$
1 2 3 4 5 6 7
8 9 10 11 12 13 14
A-T
E2 BS
84 bp
32 bp
B
E1:
WT E1
K506A
- boiled
1 2 3 4 5 6 7 8 9 10 11 12
% unwound
<1 7 23 46 96 <1

Figure 2-6. Specificity of initiation of DNA replication *in vitro*

A. Plasmid unwinding by substitution mutants at H507. WT E1 and H507 substitutions were tested for the ability to unwind an ori plasmid using a form-U assay. The ori plasmid DNA (lane 2) was relaxed by incubation with topoisomerase I (lane 3) and then incubated in the absence of E1 (lane 4) or in the presence of 6 pmol of WT E1 (lane 5) or of the indicated H507 substitutions, respectively (lanes 6-13). The samples were analyzed by agarose gel electrophoresis and stained with ethidium bromide, and the level of form-U DNA was quantitated. Lane 1 contains a DNA ladder.

B. *In vitro* DNA replication assays were performed using either pUC 19 (lanes 1–3) or an ori plasmid containing a 60 bp ori fragment cloned into the polylinker of pUC 19 (lanes 4–6). The plasmids were incubated in the presence of 6 pmol of WT E1, (lanes 1 and 3), H507F (lanes 2 and 4), or H507Y (lanes 3 and 6). The replication products for each template were digested with PvuII and DdeI and analyzed by PAGE. The level of incorporation of 32P dCTP was determined for each sample and compared.

C. *In vitro* DNA replication assays were performed using the ori+ template and either WT E1, H507F, or H507Y (lanes 1–3), respectively, and terminated after 8 min. As a control, WT E1 was used to replicate the ori− plasmid pUC 19 under the same conditions (lane 4). The replicated material was digested with PvuII and DdeI and analyzed by PAGE.

D. Each band in (C) was quantitated, and the incorporation was divided by fragment size to derive the relative labeling/nucleotide. The relative labeling efficiency for each restriction fragment from the plasmid is plotted in the graph.

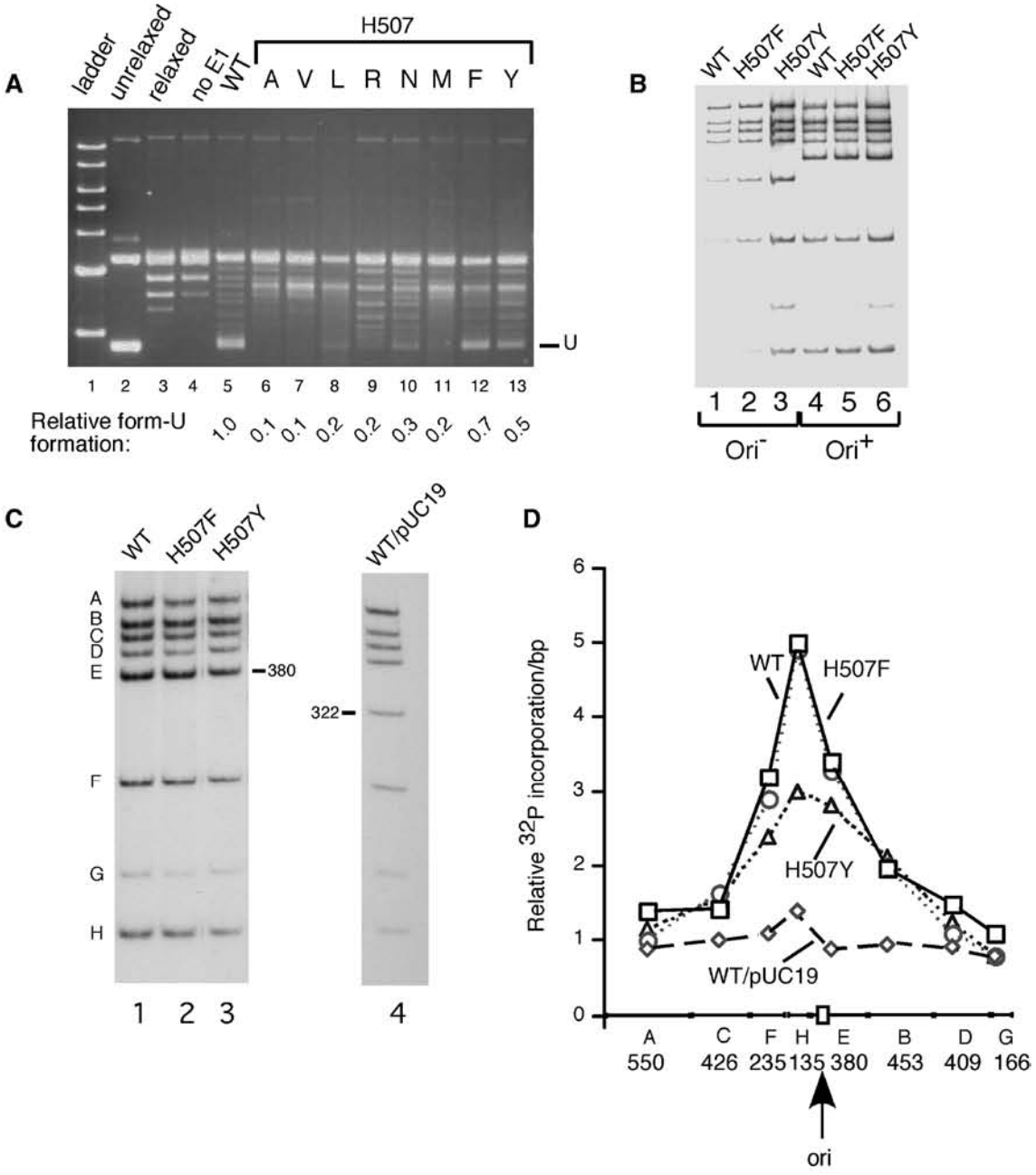
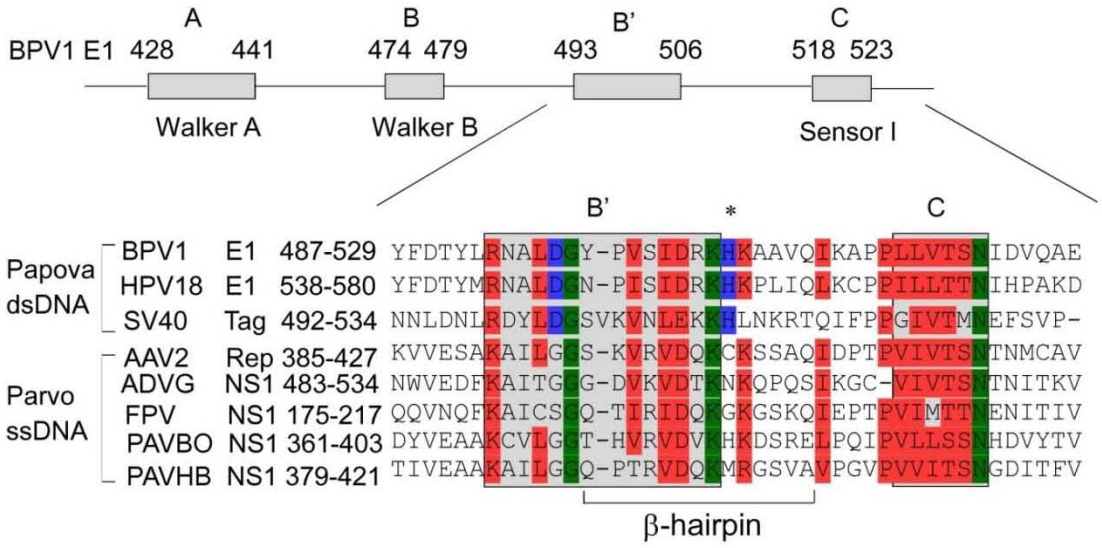

A
ladder
unrelaxed
relaxed
no E1
WT
H507
A V L R N M F Y
U
1 2 3 4 5 6 7 8 9 10 11 12 13
Relative form-U formation:
1.0 0.1 0.1 0.2 0.2 0.3 0.2 0.7 0.5
B
WT H507F H507Y WT H507F H507Y
1 2 3 4 5 6
Ori⁻ Ori⁺
C
WT H507F H507Y WT/pUC19
A
B
C
D
E
F
G
H
380
322
1 2 3 4
D
Relative 32P incorporation/bp
WT
H507F
H507Y
WT/pUC19
A C F H E B D G
550 426 235 135 380 453 409 166
ori

Figure 2-7. Sequence alignment of the B and C motifs of representative members of the SF3 helicase family

Shown above the alignment is the relative location of SF3 signature motifs in BPV E1. The number above each motif corresponds to the starting and end residue number of each motif in BPV E1. The residue at the tip of the β-hairpin (H507) is marked with an asterisk. Green, invariant residues in papova and parvoviruses. Red, conserved residues in papovaviruses and parvoviruses. Blue, highly conserved residues in papova viruses only. Included in the comparison are the following: BPV1 E1, bovine papillomavirus type 1 E1 protein; HPV18 E1, human papillomavirus type 18 E1 protein; SV40 TAg, simian virus 40 large T antigen; AAV2 Rep, adeno-associated virus 2 Rep 40 protein; ADVG NS1, Aleutian mink disease parvovirus (strain G) NS 1 protein; FPV NS1, feline panleukopenia virus NS1 protein; PAVBO NS1, bovine parvovirus NS 1 protein; and PAHVB NS1, human parvovirus B19 NS1 protein.


A
B
B'
C
BPV1 E1 428 441 474 479 493 506 518 523
Walker A
Walker B
Sensor I
B'
*
C
Papova dsDNA
BPV1 E1 487-529 YFDTYLRNALDGY-PVSIDRKHKAAVQIKAPPLLVTSNIDVQAE
HPV18 E1 538-580 YFDTYMRNALDGN-PISIDRKHKPLIQLKCPPILLTTNIHPAKD
SV40 Tag 492-534 NNLDNLRDYLDGSVKVNLEKKHLNKRTQIFPPGIVTMNEFSVP-
Parvo ssDNA
AAV2 Rep 385-427 KVVESAKAILGGS-KVRVDQKCKSSAQIDPTPVIVTSNTNMCAV
ADVG NS1 483-534 NWVEDFKAITGGG-DVKVDTKNKQPQSIKGC-VIVTSNTNITKV
FPV NS1 175-217 QQVNQFKAICSGQ-TIRIDQKGKGSKQIEPTPVIMTTNENITIV
PAVBO NS1 361-403 DYVEAAKCVLGGT-HVRVDVKHKDSRELPQIPVLLSSNHDVYTV
PAVHB NS1 379-421 TIVEAAKAILGGQ-PTRVDQKMRGSVAVPGVPVVITSNGDITFV
β-hairpin

2.5 Author Contributions

Chapter 2 is adapted from: Mol Cell, 2007. **25**(6): p. 825-37. Xiaofei Liu, Stephen Schuck and Arne Stenlund, Adjacent residues in the E1 initiator β-hairpin define different roles of the β-hairpin in Ori melting, helicase loading, and helicase activity[115]. Author contributions: X. L. is responsible for the bulk of the experiments; S.S. for some of the EMSA and *in vitro* replication assays (Figure 2-2C, 2-5A, 2-6 BCD); A.S. contributed some of the *in vitro* replication assays (Figure 2-6 BCD) together with the overall design and interpretation of the studies.

Chapter 3 Nonspecific DNA binding activity of E1 for E1 double trimer assembly

3.1 Introduction

The first step in initiation of DNA replication is the binding of an initiator protein to the replication origin (ori)[5]. The papillomavirus (PV) initiator E1 is an AAA+ (ATPases associated with various cellular activities) protein and belongs to SF3 helicase family. E1 provides at least four functions required for replication initiation, specific and nonspecific DNA DNA binding activities, local DNA melting of ori, and unwinding activity carried out by the E1 DNA helicase in front of the replication fork.

It is well established that E1 is capable of both sequence specific and sequence nonspecific DNA binding. The E1 DNA binding domain (DBD) has a high specificity for E1 binding sites (E1 BS) in the ori. However, this specific DNA binding activity is masked by a nonspecific DNA binding activity present in the E1 helicase domain. The combination of these two DNA binding activities results in a modest specificity of E1 for binding to the ori, since the non-specific DNA binding activity is dominant.

Loading of E1 to the ori requires a specific factor E2, which is a transcription factor encoded by the viral genome. E1 and E2 bind cooperatively to the ori in the form of an $E1_2E2_2$ ori complex that has high specificity for ori[78, 97, 98]. In BPV1, interactions between DBDs of E1 and E2 bound to their proximal sites produces a significant bend in DNA which further facilitates the interaction between activation domain of E2 and helicase domain of E1[99-101]. This latter interaction prevents nonspecific DNA binding by the E1 helicase domain [1, 102, 103]. After the $E1_2E2_2$ complex binds to ori, the E2 dimer is displaced and further E1 molecules are added to the complex, eventually resulting in the formation of a complex containing six E1 molecules bound to the ori as a head-to-head E1 double trimer (DT) [104]. In the DT, the E1 molecules are positioned in a helical arrangement encircling the dsDNA with each E1 molecule offset by 3bp relative to each other. This arrangement is determined by the arrangement of E1 BS in the ori. The E1 DT serves as a required precursor for the functional helicase, which is an E1 double hexamer (DH)[17].

The formation of the DT requires the cooperation of the sequence specific DNA binding activity present in the E1 DBD and the nonspecific DNA binding activity present in the E1 helicase domain [17]. E1 DBD and its binding to E1 BS have been studied extensively by biochemical, genetic and structural biology approaches and in general is well understood [44, 57, 62, 105]. Much less is known about the non-specific DNA binding activity and how the E1 helicase domain contacts dsDNA in a sequence non-specific manner.

The structural information that exists is that of a fragment of E1 containing the oligomerization and helicase domains of E1 as a hexamer assembled around a ssDNA, and this structural information therefore does not provide direct information about the interaction between the E1 helicase domain and dsDNA. Oligomerization domains contact each other to form a rigid collar that holds the six subunits together. The six helicase domains form a channel enclosing ssDNA. Each helicase domain projects a specific motif the β-hairpin into the channel in a right-handed staircase pattern[28]. Two conserved residues K506 and H507 on the tip of the β-hairpin have been implicated in the binding of E1 helicase domain to E1 BS flanking sequence[17]. Later studies suggested H507 intercalates into the DNA to pry open the dsDNA while K506 is required both for DT formation and helicase activity[115], but it is not known if the β-hairpin is the only element involved in the dsDNA binding of the helicase domain. Melting result suggests multiple sites are melted in the E1 BS flanking sequence during initiation[54], so it is possible that other residues besides β-hairpin are also involved in the nonspecific binding. On the other hand, two specific lysine residues K356 and K359 on the oligomerization domain point in the same direction as β-hairpin in the hexamer structure. It is possible that these two residues also interact with dsDNA in the DT. To answer these questions, we performed a mutagenesis screen of the surface residues of the E1 helicase domain together with some possible residues in the oligomerization domain for the trimer and DT formation. We found two patches of residues, one in helicase domain and one in oligomerization domain, are involved in the nonspecific DNA binding activity of the E1 helicase domain for DT assembly.

We have previously demonstrated two adjacent residues in the β-hairpin plays two different roles in initiation of DNA replication. K506 is required for both template melting and for DNA halicase activity, while H507 is only required for template melting and does not take part in the helicase activity[115]. Here we are investigating other residues such as the highly conserved F464, which is positioned adjacent to the β-hairpin. What we found is that this residue, which is absolutely invariant in all E1 proteins, is essential both for non-specific DNA binding activity and DT formation, and for DNA helicase activity of the E1 protein. These results demonstrate that the reason that F464 is completely conserved is that this residue is essential for DNA helicase activity.

3.2 Results

3.2.1 Mutagenesis of inner surface residues in the E1 hexamer

It is now well established that a particular structure, the β-hairpin, in the E1 helicase domain is required for formation of the E1 double trimer on the ori, most likely because the β-hairpin is directly involved in non-sequence specific binding to the dsDNA flanking the E1 binding sites in the BPV origin of DNA replication [17, 115]. Structural data demonstrates that the β-hairpin lines the inner surface of the channel formed by six subunits of the E1 helicase and oligomerization domains. We reasoned that if residues other than the β-hairpin are involved in binding to dsDNA, they would likely line up with

the residues in the hairpin in the hexamer structure. Based on this prediction, we selected a set of residues on the inner surface of the E1 helicase domain as candidates for residues that might interact with dsDNA. These residues (S456, N459, K461, S462, F464, T490, and Y491) were mutated to alanine. We also expanded the set of mutations in the β-hairpin by generating D504A, R505A, and K508A, which together with the previously characterized mutants K506A and H507A, accounts for the majority of the residues in the β-hairpin. In addition, we generated a few mutations in the E1 oligomerization domain (N352A, S353A, K356A and K359A) based on the X-ray crystal structure of the hexamer of E1 oligomerization and helicase domain[28]. We changed all these residues to alanine by site-directed mutagenesis in full length E1, and expressed and purified the mutant proteins from *E. coli*. Because the alanine substitution at K508 could not be expressed, a leucine substitution was generated instead (K508L).

3.2.2 Screening the inner surface substitutions for E1 trimer formation

DNA binding by the full length E1 protein requires cooperation of two different DNA binding activities, one located in the E1 DBD and another present in the E1 helicase domain and depending on the β-hairpin. A minimal complex that reflects the requirement for both of these activities is an E1 trimer, which can form on short DNA fragments in the presence of ADP, without apparent sequence dependence and which can be detected by EMSA. We used this assays to screen the mutants in the helicase and oligomerization domain.

As we have previously described, two substitutions at the tip of the β-hairpin, K506A and H507A, are defective for trimer formation in this assay. Here we found that also R505A in the β-hairpin was defective for trimer formation, while K508L could still form the trimer. Two of the mutants in the oligomerization domain, K356A and K359A, showed partial defects for trimer formation and two of the mutants in the helicase domain, K461A and F464A, were completely defective for trimer formation. Because F464 is completely conserved and K461 and R505 are not (Figure 3-7B), we chose F464 for further mutational analysis. Because this residue, based on the hexamer structure clearly is required for the helicase activity of the E1 hexamer we decided to perform a more extensive mutagenesis at this position to try to distinguish between the role of this residue in helicase activity and in template melting. We therefore generated a set of additional substitutions at this position (F464L, M, T, N, K, Y and H).

3.2.3 Two patches of residues on the E1 oligomerization and helicase domain, respectively, are required for the DT assembly

A defect for trimer formation could be due to a number of defects in the protein, including defects in nucleotide binding since nucleotide binding is required for trimer formation[17]. We therefore decided to test the trimer defective mutants for DT formation, since in these assays intermediate phenotypes can be observed. We used the WT 84bp ori probe and performed the EMSA in the presence of ADP. We first compared WT E1 and the mutants K356A, K359A, K461A and R505. We also included K506A and H507A, which we have previously demonstrated are defective for DT formation

(Figure 3-1A). All the mutants, with the exception of K359A and K461A (Figure 3-1A, lanes 8-10 and 38-40, respectively) showed severe defects for DT formation, forming ladders and failing to form the DT band. The fact that all of these mutants were capable of DNA binding demonstrated that they are capable of nucleotide and DNA binding, indicating that the defect is not caused by a gross overall structural defect of these proteins. K359A (Figure 3-1A, lanes 8-10) and K461A (Figure 3-1A, lanes 11-13) showed less of a defect for DT formation, since the mutants, in addition to the ladder, could from the DT complex.

In panel B, we tested the F464 substitutions in the same assay. Here, all the substitutions with the exception of F464Y (Figure 3-1B, lanes 22-24) showed severe defects for DT formation, although all the mutants were capable of DNA binding and only fail to form the correct DT complex.

3.2.4 The residues on E1 oligomerization and helicase domain required for DT assembly are involved in the nonspecific DNA binding activity

As discussed above, E1 utilizes two different DNA binding activities to form a functional DT complex. The specific DNA binding activity of the E1 DBD is used to bind to the E1 BS and the nonspecific DNA binding activity which resides in the E1 helicase and oligomerization domain, which binds to dsDNA nonspecifically and binds to the DNA sequences flanking the E1BS. We have previously demonstrated that in DNA binding assays where sequence specific DNA binding is challenged by the use of non-

specific competitor DNA, the nonspecific DNA binding activity is dominant, and therefore that full length E1, which contains both DNA binding activities, is very sensitive to non-specific competitor DNA, while the DBD is completely insensitive to competitor DNA due to its high degree of sequence specificity.

A prediction, based on these observations is that mutations that affect the non-specific DNA binding activity of E1 would be less sensitive to nonspecific competitor DNA than the WT E1 protein. This would provide us with an assay to measure the level of non-specific DNA binding activity in context of the full length E1 with an intact DBD, a task that would be difficult using any other method. Now because the mutants that we wished to test are defective for trimer and DT formation, we took advantage of the ability of these mutants to form E1 dimers on short ori probes containing E1 BS. To perform such assays, the radiolabeled probe was mixed with nonspecific competitor DNA (pUC19) at three different ratios. WT E1 or E1 mutants were then added to the probe mixes, incubated and then loaded onto EMSA gels. As observed in panel A, WT E1 in the absence of competitor DNA forms a robust dimer band with a trace of $E1_3$ (lane 1). In the presence of competitor DNA, lanes 2-4, the level of binding is greatly reduced, reflecting the sensitivity to the competitor DNA. The mutants showed a different pattern. The mutant that was least sensitive to competitor DNA K506A, (Figure 3-2A, lanes 21-24), showed a less that 2-fold reduction in complex formation even at the highest level of competitor DNA indicating that K506 contributes the most to nonspecific DNA binding. This level of competition was similar to that observed with the E1DBD (Figure 3-2B, lanes 13-16). The remaining mutants, K356A, K359A, K461A, H507A, F464A and

R505A showed reductions in binding 2-3-fold at the highest level of competitor DNA. Complex formation for WT E1 was reduced ~20-fold at the same concentration of non-specific competitor DNA.

These results demonstrate that mutation of any of the residues that we have identified as showing defects for DT formation renders E1 a more sequence specific DNA binder. We interpret these results to mean that all of these residues contribute to the non-specific DNA binding activity of E1.

3.2.5 Multiple residues on E1 oligomerization and helicase domain required for DT assembly are involved in the binding of the E1 helicase domain to the DNA sequences flanking E1 BS

These results indicate that, based on the indirect assay described above, multiple residues on the inner surface of the E1 helicase domain play a role in non-sequence specific DNA binding. We have previously demonstrated that we can distinguish between the DNA binding by the E1 DBD and the E1 helicase domain, using a DNAse footprinting assay since the E1 DBD which gives rise to the sequence specific binding binds to the E1 DBD and the helicase domain which gives rise to the non-sequence specific DNA binding binds to the sequences flanking the E1 BS. Consequently, E1 mutants such as H507A and K506A fail to protect the sequences flanking the E1 BS. To test if the other residues are also involved in binding to the flanking sequences we tested

all the alanine substitutions at K356, K359, K461, F464, R505, K506 and H507 in the DNase I footprinting assay (Figure 3-3).

Comparison of the footprints generated by the E1 DBD (Figure 3-3, lanes 1-2) and full length E1 (Figure 3-3, lanes 3-4) illustrate the difference between protections of the E1 BS and the additional protection of the sequences flanking the E1 BS. The E1 BS protection, indicated by a box with diagonal lines extends ~ 20 nucleotides. The extended footprint, indicated by a grey box upstream E1 BS and a black box downstream E1 BS, respectively, is clearly detectable on both sides of the E1 BS protection. As described previously, the two β-hairpin substitutions K506A and H507A(Figure 3-3, lanes 15-18), both give rise to robust protection over the E1 BS, but very weak or non-existent protection over the flanking sequences (compare lanes 15-18 to lane 19). The β-hairpin substitution R505A (lanes 13-14), the oligomerization domain mutant K356A (lanes 5-6), and the helicase domain mutant K461A (lanes 9-10) all behaved similarly providing robust protection of the E1 BS and weak protection of the flanking sequences. The two substitutions K359A and F464A behaved differently. K359A (lanes 7-8) gave rise to a protection similar to that of WT E1, with protection of the sequences on both sides of the E1 BS. F464A (lanes 11-12) generated flanking protection on the upstream side of the E1 BS, but not on the downstream side of the E1 BS. These results demonstrate that all the substitutions, with the exception of K359A, show defects for protection of the sequences flanking the E1BS. K359A is the substitution that showed the least defect for trimer and DT formation (see Figure 3-1). The reason for the asymmetric protection observed with

F464A is unclear, but may indicate that the sequences on the up-stream side of the E1BS provide better binding sites for the E1 helicase domain.

3.2.6 Sequences flanking E1 BS are required for DT assembly

A peculiar feature of E1 DT formation on the ori is that formation of this complex is strongly dependent on the absolute length of the probe. The DT depends completely on the presence of four E1 BS in the ori. However, in addition to the four E1 BS, which occupy 18 bp, ~33 bp of flanking sequences on either side of the E1 BS is required for DT formation. These flanking sequences are bound by the E1 helicase domain but without apparent sequence-specificity. We have shown that deletion of 17 bp from either the left or the right end of the probe resulted in complexes containing 5 E1 molecules instead of 6 as in the DT. This result was interpreted to mean that E1 in the DT is bound to the DNA in a helical arrangement and that deletion of sequences from either end therefore results in the loss of one monomer of E1. To further this observation we generated a series of probes with different size deletions from the left hand (upstream) end (Figure 3-4B). We tested complex formation of WT E1 on truncated probes in the presence of ADP (Figure 3-4A). On the WT ori E1 formed a discrete complex containing six E1 molecules corresponding to a DT. Deletion of 7 bp had only a slight effect on complex formation observed as an increase in the $E1_5$ band relative to the $E1_6$ complex (Figure 3-4A, lanes 16-18). Deletion of 10 bp however, (Figure 3-4A, lanes 13-15) showed a clear effect. With this probe, the $E1_5$ complex now becomes predominant over the $E1_6$ complex. Further deletion results in the appearance of smaller complexes such

that deletion of 22 bp results in a predominant complex containing four E1 molecules (Figure 3-4, lanes 1-3). These results suggested that DT assembly requires the E1 BS flanking sequences in a length dependent manner and that interestingly, removal of the first 13 bp result in formation of an $E1_5$ complex. Removal of another 9 bp results in an $E1_4$ complex. These results are difficult to explain by any other mode of DNA binding than a helical arrangement of the E1 molecules around the ds DNA.

3.2.7 Nonspecific binding activity of E1 is required for the viral DNA replication

To determine whether the nonspecific DNA binding activity of E1 is required for viral DNA replication, substitutions of K356 (lane 12), K359 (lane 13), K461 (lane 14), F464 (lane 2-9 and 15), R505 (lane 16), K506 (lane 17), H507 (lane 18) were tested in the *in vitro* DNA replication assay (Figure 3-5). In this assay, a plasmid containing the ori is incubated in the presence of E1 in a H293 cell extract in the presence of radiolabeled nucleotide. WT E1 (lane 1 and 11) concentration was adjusted to be optimum for the replication while the substitutions were used as the same quantity as WT E1. The level of DNA synthesis was quantitated after separation of the replication products by agarose gel electrophoresis. K359A supported *in vitro* replication at a similar level to WT E1 (compare lane 11 and 13) while K356A (lane 12) had half activity of WT E1. A trace of activity of K461A (lane 14) could be detected, but no activity was detected for the rest of substitutions.

3.2.8 F464 is also required for helicase activity of E1

As mentioned earlier we were interested in determining whether the highly conserved residue F464, which is important for DT formation is conserved because of its role in DT formation and therefore in template melting, or because of its role in the helicase activity of E1. As we demonstrated in Figure 3-1, F464Y had modest activity in the DT formation assay, while the rest of the substitutions at this position were defective for DT formation. In the cocrystal structure of a hexamer of the BPV1 E1 oligomerization and helicase domains bound to ssDNA, F464 and K506 on the β-hairpin contacts the ssDNA with their side chains [28] and we had shown that K506 is required for the helicase activity of E1 [115]. To try to distinguish between a critical role of F464 in DT formation and DNA helicase activity, we tested the F464 substitutions for DNA helicase activity (Figure 3-6). In this assay, E1 is incubated with a fluorescently labeled oligonucleotide substrate in which the fluorescence is quenched by the presence of the complementary strand. As the substrate is unwound, the quencher is removed, resulting in a dramatic increase in fluorescence. Thus, the fluorescence and hence the unwinding in a single sample can be monitored in real time. We measured unwinding every 1.5 min for 27 min (Figure 3-6). Interestingly, all of the substitutions at F464 had severe defects for DNA helicase activity. These results indicate that the absolute conservation of F464 among all papovaviruses, is likely due to the requirement of this particular residue for DNA helicase activity. This conclusion is supported by the fact that this residue is also conserved in adeno-associated viruses, which are unlikely to require template melting, and consequently DT formation, since these viruses have a single-stranded genome (Figure 3-7B). In a set of similar helicase assays, we found K356A, K359A, K461A and

R505A were not defective for helicase activities (data not shown), which indicate these residues are not required for the helicase activity of E1.

3.3 Discussion

3.3.1 Nonspecific DNA binding activity of BPV1 E1 is contributed by multiple motifs from helicase domain and oligomerization domain, respectively

Our screening identified multiple residues required for nonspecific dsDNA binding activity of E1 oligomerization and helicase domain. These residues can be divided into two patches, one on E1 oligomerization domain and one on E1 helicase domain (Figure 3-7A).

There are five residues on the BPV1 E1 helicase domain involved in nonspecific dsDNA binding and DT assembly: R505, K506 and H507 on the β-hairpin, F464 on a hydrophobic loop that contains F464 and K461 on a charged loop. Alanine substitution of any of these five residues will cause a severe defect of E1 in supporting viral replication. K356 and K359 are on the E1 oligomerization domain. Alanine substitution of K359 causes little defect in E1 in supporting viral replication, which is consistent with the DT assembly and E1 BS flanking sequence protection result. K356A fails to form DT but it still has half activity of WT E1 in supporting viral replication (Figure 3-5). We will provide an explanation in the following discussion.

Of all the residues that are involved in the nonspecific dsDNA binding of BPV1 E1, F464 and K506 are invariant while K356, K359and H507 are highly conserved in all PV initiator E1 protein (Figure 3-7B and data not shown). This suggested all PV E1 proteins bind to dsDNA nonspecifically using a similar mechanism as BPV1 E1. PV E1 contains a variety of residues in the positions of K461 (G, A, S, C, R, H, N) and R505 (A, V, I, C, M). It is not known whether other motifs are required for the nonspecific DNA binding activity of the rest E1 initiators.

Most residues involved in the nonspecific dsDNA binding activity of E1 are positively charged (K356, K359, K461, R505 and K506). These residues are mostly likely to bind to backbone phosphates through electrostatic interactions based on previous studies of nonspecific dsDNA binding [116]. We have previously proposed that H507 on the tip of the β-hairpin intercalates into the dsDNA to pry open the DNA. F464 shares a flat ring side chain structure with H507. So F464 is likely to use the same mechanism to interact with dsDNA. This interaction mode of H507 and F464 also provides the indication of why K356A substitution on E1 oligomerization domain and the alanine subsitutions on E1 helicase domain (K461, F464, R505, K506 and H507) have different activity in supporting viral replication. All these alkaline substitutions fail to form DT. However, K356A still support viral replication in about half activity of WT E1 and the rest alanine substitutions had severe defect in supporting viral replication (Figure 3-5). One possibility is that the residues on E1 helicase domain destabilize the dsDNA upon

binding while the binding of K356 just provides an architectural role for DT assembly. Although K356A fails to form DT, it still can form a series of complexes up to $E1_5$ (Figure 3-1). $E1_5$ consists of four E1 subunits that bind to E1 BS and another E1 subunit that can bind to either half of the ori, most likely to be recruited through protein-protein interaction between E1 subunits. The intact trimer half of this $E1_5$ may still melt the dsDNA by the residues on the helicase domain and initiates DNA replication. This interprets the lowered activity of K356A in supporting viral replication both *in vivo* [117] and *in vitro* (Figure 3-5). For the alanine substitutions of K461, F464, R505, K506 and H507 on E1 helicase domain, although they still can bind to ori through E1 DBD and protein-protein interaction, the complexes formed by these substitutions cannot generate the conformation change in dsDNA required for DNA melting, which attribute to their failure in replication. The different effect of alanine substitutions of K356 and the residues on E1 helicase domain (K461, F464, R505, K506 and H507) in replication represents a qualitative difference of their roles on dsDNA binding. Noteworthy, some of these residues are involved in multiple activities of E1 in the replication initiation process such as K506, which is required for both DNA melting and helicase activity. Mutations of such residues may cause defects in various activities of E1 that attribute to the DNA replication failure.

Our result is consistent with a recent report that K356 of BPV1 E1 was involved in dsDNA binding and melting[117]. Furthermore, our result suggested that K356 is involved in the DT assembly. DT is the intermediate complex that melts dsDNA during

replication initiation [17, 115]. This interprets why K356 is required for the melting activity.

It is not clear how the residues those are involved the nonspecific binding activity of E1 are placed the dsDNA flanking E1 BS. In principle, these residues can bind to any dsDNA sequence. However, these nonspecific DNA bindings are restricted by the distances of these residues to E1 DBD which binds to E1 BS in the ori specifically. In BPV1, a T-A stretch with six T-A base pairs is the only element in the E1 BS flanking sequence required for replication[40]. This T-A stretch is melted after DT assembles and this melting is important for helicase assembly [17, 54, 115]. This T-A stretch is dispensable for DT assembly because it can be converted to other sequences and DT still can form on the changed ori probe[54]. These facts together with the result of the complexes formation of WT E1 on truncated ori probes suggested that the nonspecific binding of these residues to the DNA sequences flanking E1 BS is determined by their distances to E1 DBD. In light of the role of adjacent F464 and β-hairpin in DNA melting, the T-A stretch is likely to be contacted and melted by these two motifs.

3.3.2 Nonspecific DNA binding activity of E1 is required for viral replication initiation

E1 initiator protein contains both specific DNA binding activity in its DNA binding domain and nonspecific DNA binding activity in its oligomerization and helicase domain. Specific DNA binding activity of E1 is required for the loading of initiator to the

ori. However, the specific DNA binding of E1 DBD to E1 BS in the ori is compromised by the nonspecific binding activity of E1 oligomerization and helicase domain. The nonspecific DNA binding activity of E1 needs to be masked by the formation of an $E1_2E2_2$ complex for the E1 initiator loading [102]. Our results showed that alanine substitutions of the residues involved in the nonspecific dsDNA binding activity of E1 disrupted the assembly of DT. DT has been shown to be an important replication intermediate and is responsible for melting the dsDNA to initiate DNA replication [17, 115]. Most substitutions of the residues involved in the nonspecific dsDNA binding activity of E1 (with the exception of K359A) had severe defects in supporting viral replication compared to WT E1. These results suggested the nonspecific dsDNA binding activity of E1 is crucial for replication initiation. This also explains why PV developed such a very complicated mechanism for viral replication initiator loading.

3.3.3 Invariant F464 residue is required for both nonspecific dsDNA binding for DT assembly and helicase activity of E1

Our data suggested that F464 in E1 is required for its helicase activity (Figure 3-6). This piece of data is consistent with the observation that in the cocrystal structure of a hexamer of the BPV1 E1 oligomerization and helicase domains bound to ssDNA, the aromatic side chain of F464 forms van der Waals interaction with the sugar moiety of the ssDNA [28]. PV E1 proteins belong to SF3 family helicases, which large T antigen of polyomaviruses and Rep proteins of parvoviruses also belongs to. Both papillomaviruses and polyomaviruses have dsDNA genomes while parvorviruses have ssDNA genomes.

The alignment of SF3 helicases shows that the phenylalanine residue at the position corresponding to F464 in BPV1 E1 is invariant in SF3 helicases (Figure 3-7B). Base on our helicase assay result, we concluded that this residue is absolutely conserved in SF3 helicases because of its important role in the helicase activity.

Our data also suggested that F464 plays another important role in the nonspecific dsDNA binding of E1 for DT assembly. All F464 substitutions of E1 with the exception of F464Y failed to form the DT, an important replication intermediate complex that is responsible for the local template melting. F464Y mainly forms an $E1_5$ complex although it still can form a trace of an $E1_6$ complex that corresponds to DT (Figure 3-1). Base on our previous discussion, a proper binding of F464 is like to be required to destabilize the dsDNA for the DNA melting. Because of the subtle difference of phenylalanine and a tyrosine side chain, which has an extra hydroxyl group compare to phenylalanine, the DT formed by F464Y may not be able the melt the DNA properly. This is analogous to the difference effects of phenylalanine and tyrosine substitutions of H507 in the DT assembly and DNA melting that we observed in our previous study of the β-hairpin [115]. The role of F464 in DNA melting explains the discrepancy of the result of the replication assay (Figure 3-5) and of helicase activity assay (Figure 3-6). In the *in vitro* DNA replication assay, all the F464 substitutions failed to support DNA replication completely (Figure 3-5), while in the helicase assay, some of F464 substitution mutants still have residue helicase activities such as F464H and F464M (Figure 3-6).

A similar role of F464 in dsDNA melting and helicase activity has been reported but the conclusion was based on an incomplete set of substitutions of F464 that were completely inactive in helicase assay[118]. Our data showed that some F464 substitutions are not completely defective for helicase activities such as F464H and F464M (Figure 3-6). Another study showed that F459 in SV40 T-ag, which corresponds to F464 of BPV1 E1, is involved in helicase activity and dsDNA binding [109]. However, our data suggested that F464 substitutions are still capable of ori binding through the E1 DBD. In the T-ag study, a truncated T-ag form was used while we used full length BPV1 E1 in our study. This difference of the length of the peptides might attribute to the discrepant results in two studies.

The dual role that F464 plays in DNA melting and helicase activity is consistent with our previous suggested helicase assembly model[115]. In this model, E1 double trimer melts the local template dsDNA to generate an ssDNA template for more E1 to be recruited to form a functional helicase. In this model, the β-hairpin and F464 remain in contact with the DNA throughout melting and the transition to the helicase. The fact that both F464 and K506 are involved in nonspecific dsDNA binding and helicase activity while K356, K359, K461, R505 and H507 are only required for nonspecific dsDNA binding and melting suggested melting of the dsDNA and dsDNA unwinding by helicase are two different but highly related processes.

3.4 Materials and Methods

E1 Protein: Expression and Purification

WT E1 and E1 mutants were expressed in E. coli as N-terminal GST fusions, purified by affinity chromatography, cleaved and isolated by ion exchange chromatography as described [97]. E1 purified in this manner is monomeric, as determined by glycerol gradient sedimentation and gel filtration [110]. The concentrations of the WT and mutant E1 proteins were determined with the fluorescent dye SYPRO Red using 532nm as the excitation wave-length in a FujiImager FLA 5000.

Plasmid constructs

The template used for *in vitro* DNA replication assays was a 60 bp minimal ori fragment (7914-27) cloned between the XbaI and HindIII sites in pUC 19 [40]. The template for generation of ori probes is a 110-base pair BPV-1 sequence (nucleotides 7894-57, centered on the E1-binding site) cloned between the XbaI and HindIII sites in pUC19 generating the plasmid 11/12/X [114].

Probes

All probes were generated by PCR using primers end labeled with [γ-^{32}P] ATP and T4 polynucleotide kinase. Probes were purified by PAGE, eluted by diffusion, and precipitated.

EMSA

Four percent acrylamide gels (39:1 acrylamide:bis) containing 0.5 × TBE, lacking EDTA, were used for all EMSA experiments. E1 was added to the probe (~2 fmol) in 10 μl binding buffer, BB (20 mM HEPES [pH 7.5], 100 mM NaCl, 0.7 mg/ml BSA, 0.1% NP40, 5% glycerol, 5 mM DTT, 5 mM $MgCl_2$, and 2 mM ATP or ADP). After incubation at room temperature for 1 hr, the samples were loaded and run for 2 hr at 9V/cm. The ability to generate discrete complexes, especially the DT and DH, was critically dependent on high-purity acrylamide, freshly prepared APS solution, overnight polymerization of the gels, and precise prerunning time (9V/cm for 4 hr).

DNase footprints

DNase footprints were carried out as described previously [97]. Proteins and probe (15,000 c.p.m.) were incubated at room temperature in binding buffer [20mM potassium phosphate pH 7.4, 0.1M NaCl, 1 mM dithiothreitol (DTT), 1 mM EDTA, 0.7 mg/ml bovine serum albumin (BSA), 0.1% NP-40 and 5% glycerol, 2 ng dI-dC] in a final volume of 10 μ1. After 20 min, 10 μ1 of a solution containing 5 mM $CaCl_2$ and 10 mM $MgCl_2$ was added together with 1 μ1 of DNase I. After 60 s, cleavage was terminated by the addition of 130 ml of STOP solution (0.2 M NaCl, 10 mM EDTA, 1% SDS) followed by phenol/chloroform extraction. The aqueous phase was precipitated by the addition of 350 μ1 of 0.45 M ammonium acetate in ethanol.

Competitor Assay

Competitor assays were performed by incubating 2 fmol of probe with E1 under EMSA conditions. The 10 μ1 reaction system contained 20 mM HEPES [pH 7.5], 100 mM NaCl, 0.7 mg/ml BSA, 0.1% NP40, 5% glycerol, 5 mM DTT, 5 mM $MgCl_2$, and 2 mM ADP. PUC 18 plasmids were added to the system as nonspecific DNA competitor. Then the reactions were incubated at RT for 15 min. After that, the samples were loaded and run for 2 hr at 9V/cm.

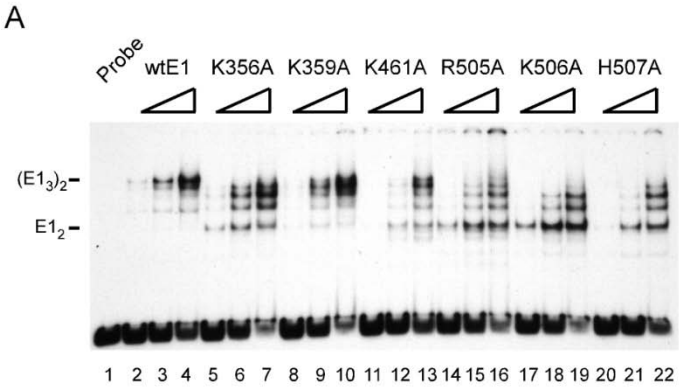
***In vitro* DNA Replication Assay**

In vitro DNA replication assays were performed essentially as described [41, 78]. *In vitro* replication was performed in 25 μ1 reaction mixtures containing the following: 40 mM HEPES-KOH (pH7.5), 8 mM $MgCl_2$, 0.5 mM DTT, 3 mM ATP, 0.2 mM each of GTP, UTP and CTP; 0.1 mM each of dATP, dGTP and dTTP, 10 μM of [α-^{32}P]dCTP (2 μCi; 3000 Ci/mmol), 40 mM creatine phosphate, 400 ng creatine kinase, 10 μl S100 extract and 0.5 μl high salt nuclear extract from H293 cells. The concentration of template in the *in vitro* reactions was 2ng/μ1. Reactions were incubated for 60 min at 37°C unless stated otherwise. The reactions were stopped by addition of SDS to 1% and EDTA to 10 mM, treated with proteinase K followed by phenol/chloroform extraction and precipitation with ethanol and ammonium acetate. The products were analyzed by electrophoresis on 1% agarose gels in TAE buffer.

Accession codes

Protein Data Bank: E1 oligomerization and heliase domain hexamer structure coordinates can be accessed with PDB code 2GXA[28].

Figure 3-1. Double trimer formation of WT E1 and E1 point mutants. WT E1 and the E1 point mutations were tested by EMSA for DT formation using the 84 bp ori probe. Three protein concentrations (3, 6, 12ng) for each protein (labeled at the top of the gels) were used in the presence of ADP. In panel A, WT E1, K356A, K359A, K461A, R505A, K506A, and H507A were tested. In panel B, WT E1, F464A, F464L, F464M, F464T, F464N, F464K, F464Y and F464H were tested.


A
Probe
wtE1
K356A
K359A
K461A
R505A
K506A
H507A
(E1$_3$)$_2$
E1$_2$
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22

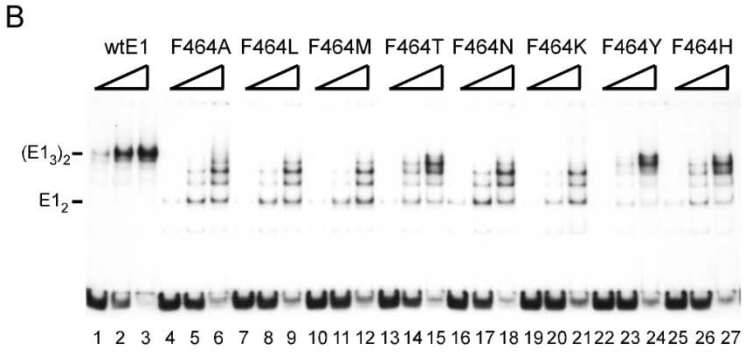

B
wtE1
F464A
F464L
F464M
F464T
F464N
F464K
F464Y
F464H
(E1$_3$)$_2$
E1$_2$
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

Figure 3-2. Residues on E1 oligomerization and helicase domain required for DT assembly are required for nonspecific DNA binding activity. EMSA was performed using a 39 bp ori probe where E1 binds as a dimer. In each set of four, EMSA was performed in the absence or in the presence of 4, 8, or 16 ng of pUC 18 as nonspecific dsDNA competitor. In panel (**A**) WT E1 (lanes 1-4), K356A (lanes 5-8), K359A (lanes 9-12), K461A (lanes 13-16), R505A (lanes 17-20), K506A (lanes 21-24) were tested. Lane 25 contained probe alone. In panel (**B**) F464A (lanes 5-8), H507A (lanes 9-12) and E1 DBD (lanes 13-16) were tested together with WT E1 (lanes 1-4).

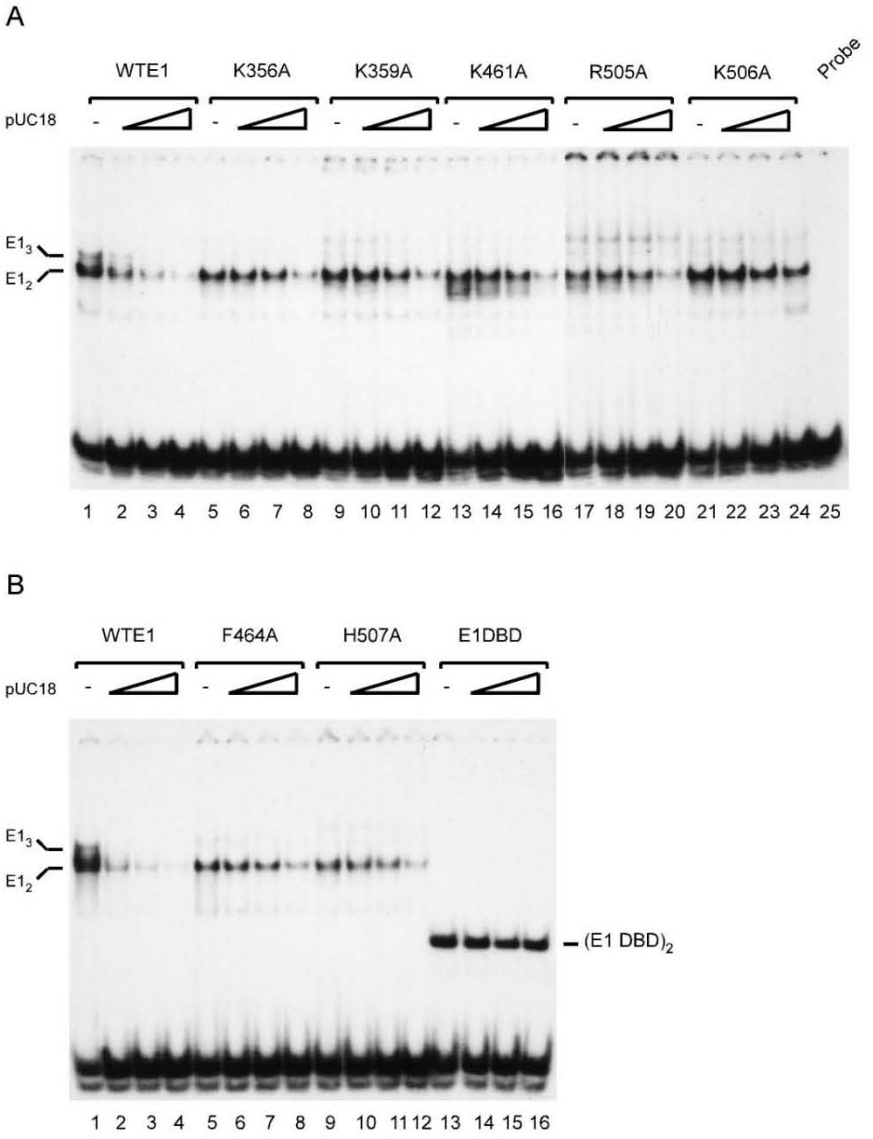

A
WTE1
K356A
K359A
K461A
R505A
K506A
Probe
pUC18
E1$_3$
E1$_2$
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
B
WTE1
F464A
H507A
E1DBD
pUC18
E1$_3$
E1$_2$
(E1 DBD)$_2$
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Figure 3-3. E1 mutants with defects for non-specific DNA fail to protect sequences flanking the E1 BS. DNase I footprints were performed on the top strand of an ori fragment comparing the protection observed with E1 DBD, WT E1, and E1 point mutants. E1 DBD (120, and 240 fmol) (lanes 1-2) and 140 and 280 fmol of WT E1 (lanes 3-4), K356A (lanes 5-6), K359A (lanes 7-8), K461A (lanes 9-10), F464A (lanes 11-12), R505A (lanes 13-14) K506A (lanes 15-16), and H507A (lanes 17-18) were used, respectively. Lanes 19 contained probe alone.

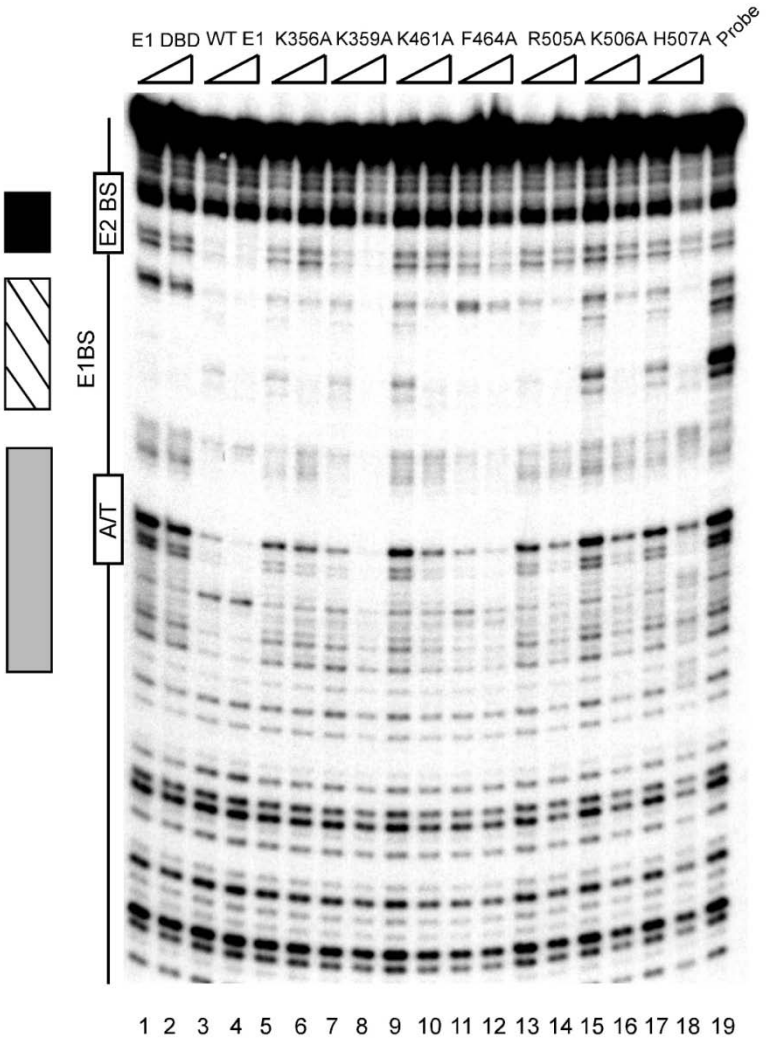

E1 DBD
WT E1
K356A
K359A
K461A
F464A
R505A
K506A
H507A
Probe
E2 BS
E1BS
A/T
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

Figure 3-4. DNA sequences flanking the E1 BS are required for DT assembly. EMSA was performed using WT E1 and the 84 bp ori probe, and progressive deletions in this probe, as indicated in panel B. **A.** Three quantities of E1 (3, 6, 12ng) were used for EMSA in the presence of ADP utilizing the 7 probes indicated in panel B. The mobility of the different E1 complexes are indicated. **B.** Schematic description of the 84bp ori probe, and the 6 progressive deletions used in panel A. **C.** Cartoon depicting the predicted arrangement of the E1 molecules in the E16 (E1DT), $E1_5$ and $E1_4$ complexes.

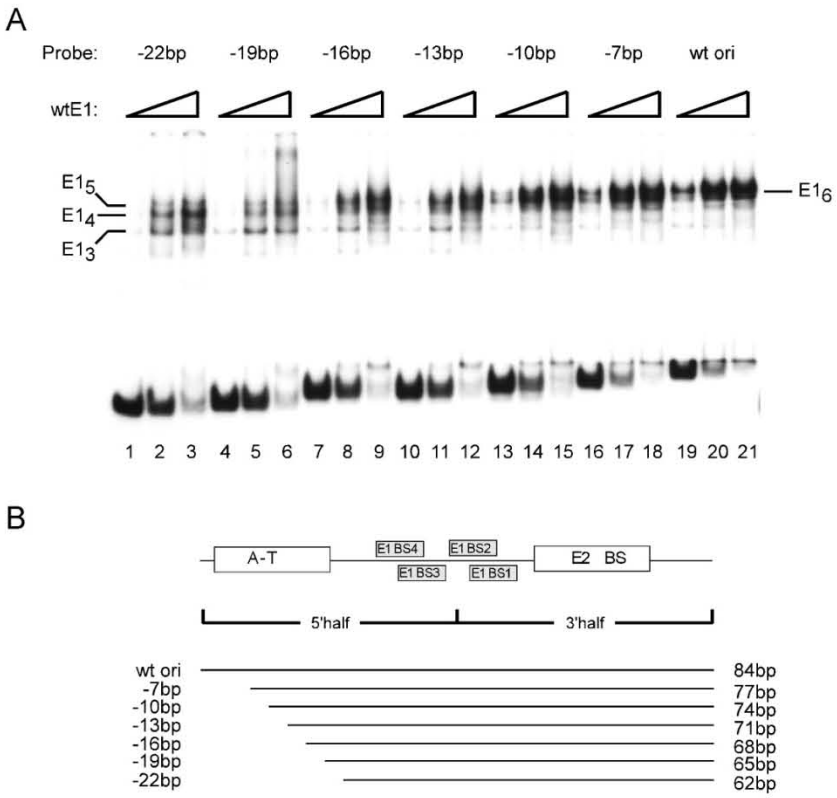

A
Probe:
-22bp
-19bp
-16bp
-13bp
-10bp
-7bp
wt ori
wtE1:
E1 5
E1 4
E1 3
E1 6
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
B
A-T
E1 BS4
E1 BS2
E1 BS3
E1 BS1
E2 BS
5'half
3'half
wt ori 84bp
-7bp 77bp
-10bp 74bp
-13bp 71bp
-16bp 68bp
-19bp 65bp
-22bp 62bp
C
E1 6
E1 5
E1 4
1 2 3 4 5 6

Figure 3-5. Mutants defective for non-specific DNA binding are defective for *in vitro* DNA replication. The ability of WT E1 and E1 point mutations to support DNA replication in a cell-free replication system were compared. 400 ng of each E1 protein was used in an *in vitro* DNA replication assay in the presence of radiolabeled dCTP and the products were analyzed by agarose gel electrophoresis. The WT E1 and the mutant proteins were used as indicated above the lanes. The lanes 10 and 19 contained no E1. I and II refer to form I and II of the replicated ori plasmid, respectively.

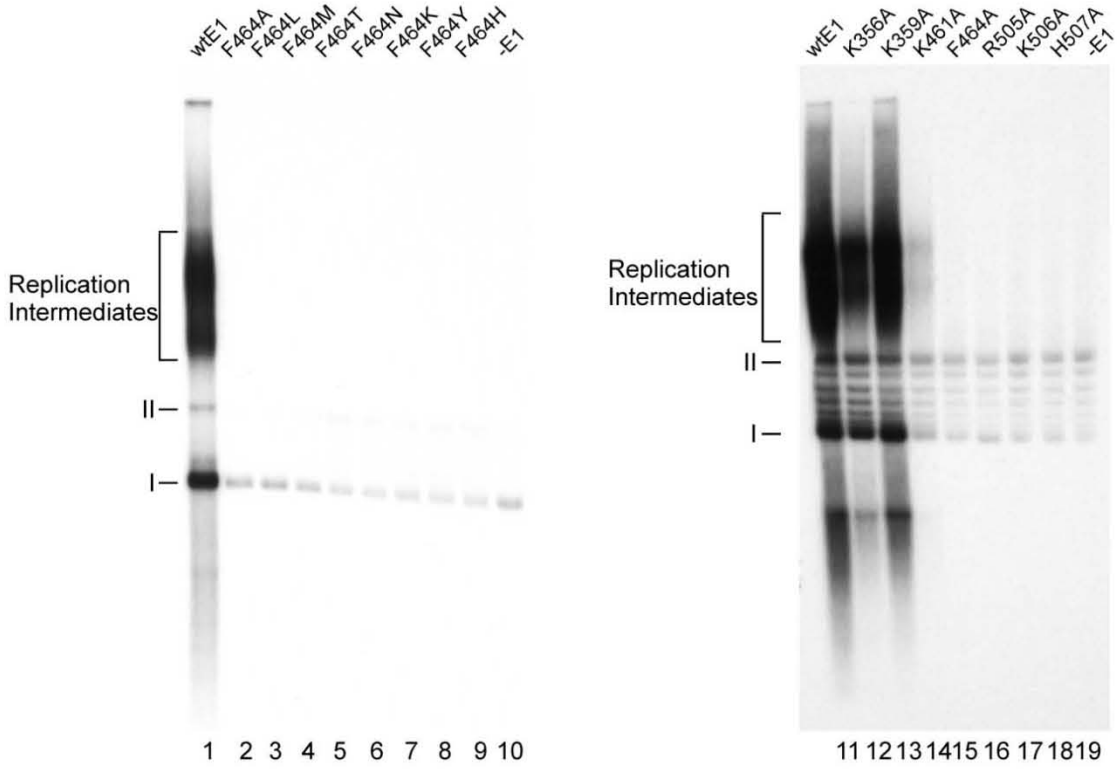

wtE1
F464A
F464L
F464M
F464T
F464N
F464K
F464Y
F464H
-E1
Replication Intermediates
II
I
1 2 3 4 5 6 7 8 9 10
wtE1
K356A
K359A
K461A
F464A
R505A
K506A
H507A
-E1
Replication Intermediates
II
I
11 12 13 14 15 16 17 18 19

Figure 3-6. Helicase activity of substitution mutations at F464. The substitutions at F464 were tested for DNA helicase activity using a time-resolved fluorescence based oligonucleotide displacement assay. WT E1 (1.5 pmol) or the respective E1 mutants were incubated with 200 fmol of fluorescent substrate at 37 °C, and fluorescence was measured every 2 min for 26 min.

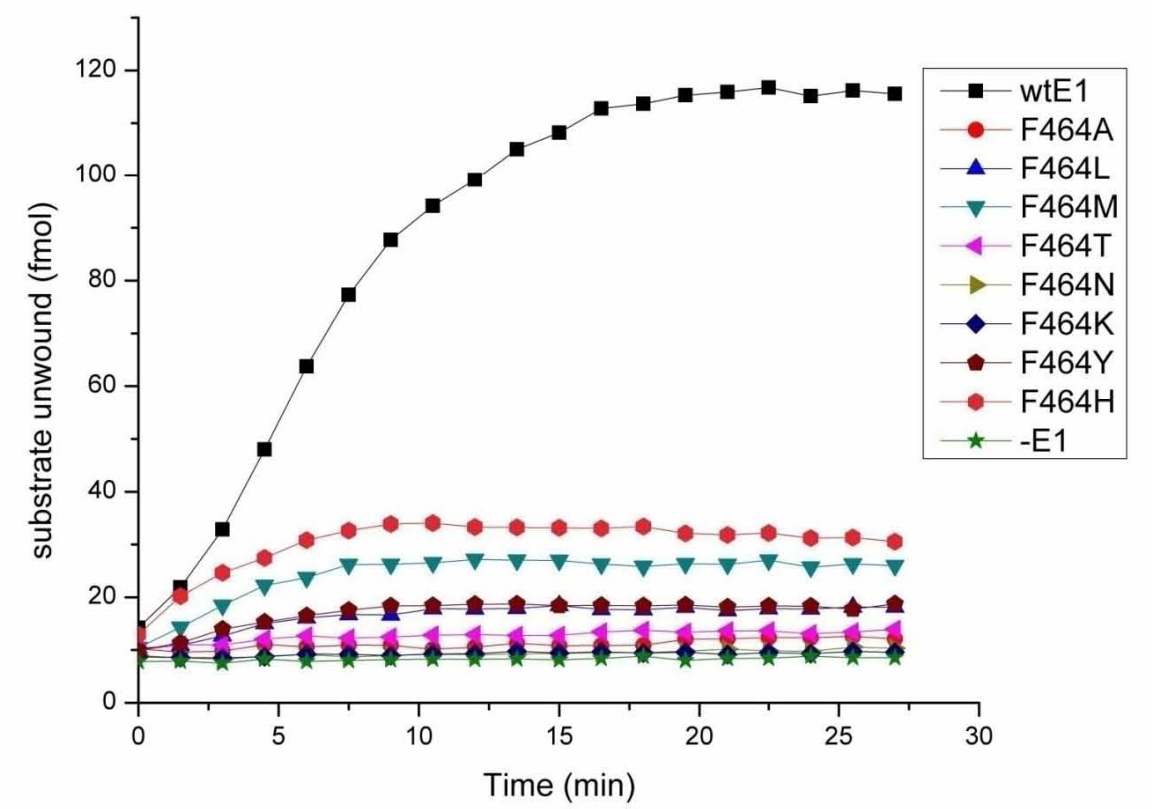
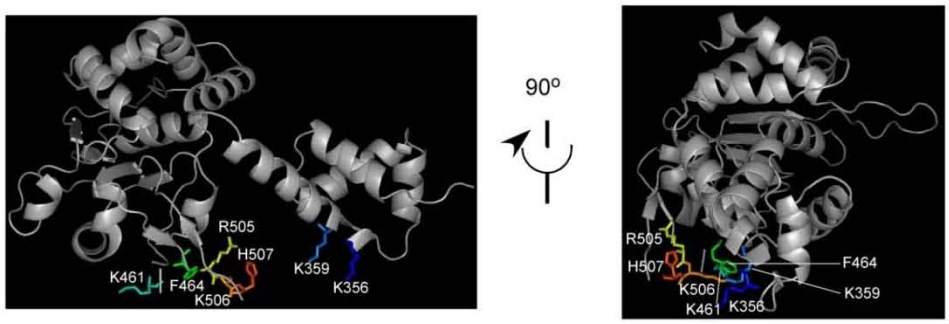

substrate unwound (fmol)
Time (min)
0
20
40
60
80
100
120
0
5
10
15
20
25
30
wtE1
F464A
F464L
F464M
F464T
F464N
F464K
F464Y
F464H
-E1

Figure 3-7. Residues of E1 oligomerization and helicase domain involved in nonspecific DNA binding activity of E1.

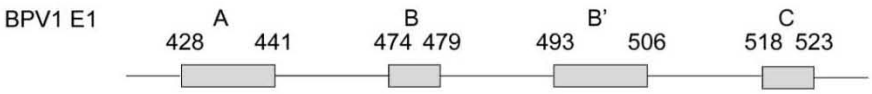
A. Images representing the structure of a monomer of BPV1 E1 oligomerization and helicase domain highlighting the position of seven residues, K356, K359 on E1 oligomerization domain and K461, F464, R505, K506, H507 on E1 helicase domain, that are involved in the nonspecific dsDNA binding activity of E1 protein. The monomer on the right is the same monomer on the left after a clockwise rotation around z-axis by 90 degrees. Adapted from Enemark, E.J. and L. Joshua-Tor, Mechanism of DNA translocation in a replicative hexameric helicase. Nature, 2006. 442(7100): p. 270-5 [28].

B. Sequence alignment of the B and B' motifs of representative members of the SF3 helicase family. Shown above the alignment is the relative location of SF3 signature motifs in BPV E1. The number above each motif corresponds to the starting and end residue number of each motif in BPV E1. The positions corresponding to K461, F464, R505, K506 and H507 in BPV1 E1 are marked with asterisks. Dark grey, invariant residues in SF3 helicases; light grey, highly conserved residues in SF3 helicases. Includes in the comparision are the following: BPV1 E1, bovine papillomavirus type 1 E1 protein; HPV4, 5, 9, 16, 17, 18, 27, 29, 38, 40, 41 E1, human papillomavirus type 4, 5, 9, 16, 17, 18, 27, 29, 38, 40, 41 E1 protein; SV40 T-ag, simian virus 40 large T antigen; AAV2 Rep, adeno-associated virus 2 Rep 40 protein.

A



B



461 464 505/506/507

```
                          *  *                                                  ***
BPV1_E1 460-509   HKSHFWLASLADTRAALVDDATHACWRY----------FDTYLRNALDG-YPVSIDRKHKA
HPV4_E1 455-504   RSSHFWLQPLMDCKVGFMDDATYVCWTY----------IDQNLRNALDG-NPMCIDAKHRA
HPV5_E1 461-510   SKSQFWLQPLSECKIALLDDVTDPCWIY----------MDTYLRNGLDG-HYVSLDCKYRA
HPV9_E1 460-509   CKSTFWLQPIADTKLALIDDVTHVCWEY----------IDQYLRNGLDG-NYVCLDMKHRA
HPV16_E1 504-553  SKSHFWLQPLADAKIGMLDDATVPCWNY----------IDDNLRNALDG-NLVSMDVKHRP
HPV17_E1 464-513  CRSNFWLQPLADTKLALIDDVTFVCWDY----------IDQYLRNGLDG-NVVCLDLKHRA
HPV18_E1 511-560  STSHFWLEPLTDTKVAMLDDATTTCWTY----------FDTYMRNALDG-NPISIDRKHKP
HPV27_E1 498-547  SSSHFWLQPLSDSKIGLLDDATPQCWSY----------IDTYLRNLLDG-NPVSIDRKHKT
HPV28_E1 517-566  SNSHFWLQPLADAKIGLLDDATSQCWCY----------IDTYLRNALDG-NQVCIDRKHRA
HPV29_E1 515-564  SHSHFWLQPLSEAKMGLLDDATSQCWSY----------VDTYLRNALDG-NVMCIDRKHRS
HPV38_E1 459-508  SKSQFWLQPLADGKIGLLDDATDVCWQY----------IDSFLRNGLDG-NLVSLDIKHKA
HPV40_E1 504-553  SCSHFWLQSLADAKVAMLDDVTAACWGY----------MDTHMRNLLDG-NPTSIDRKHKP
HPV41_E1 470-519  NGSHFWLSPLRGARCCLIDDATLTFWRY----------ADQNMRALLDG-YEISIDAKHRN
SV40_Tag 455-515  DRLNFELGVAIDQFLVVFEDVKGTGGESRDLPSGQGINNLDN-LRDYLDGSVKVNLEKKHLN
AAV2_Rep 360-408  TNENFPFNDCVDKMVIWWEEGKMTAK----------VVESAKAILGG-SKVRVDQKCKS
                                  B                                  B'
```

Chapter 4 Role of ATP binding pocket of E1 in DT assembly

4.1 Introduction

Viral initiator proteins from DNA viruses belong to the superfamily 3 (SF3) helicases[119, 120]. Well-studied members of this group include the T-antigens from the polyomaviruses, the E1 proteins from the papillomaviruses, and the Rep proteins from the adeno-associated viruses. These proteins are multifunctional proteins that utilize ATP binding and hydrolysis by their AAA+ modules for multiple functions in the replication of their viral genomes and are therefore of particular interest for understanding how AAA+ proteins carry out multiple ATP driven functions (For reviews, see [121, 122]. For example, the E1 protein, which plays an essential role in papillomavirus DNA replication, has multiple functions that are affected by binding or hydrolysis of ATP[55, 104, 110, 123-125]. E1 is a DNA binding protein, which binds specifically to E1 binding sites (E1 BS) in the origin of DNA replication [17, 43, 44, 126, 127]. DNA binding activity requires nucleotide binding by E1[17]. In the presence of ATP or ADP, E1 can form a specific double trimer (DT) complex on the ori, and through ATP hydrolysis, this complex can melt the ori DNA[17, 54, 115]. In a process that requires ATP hydrolysis, the DT is then converted into a double hexamer (DH), which has ATP dependent DNA helicase activity and is the replicative DNA helicase [17, 55]. Consequently, the E1 AAA+ module is utilized for ATP binding and hydrolysis in a number of different E1

complexes with different functions. An interesting question is how the same motif for ATP binding and hydrolysis is used in these different complexes to achieve their differing functions.

Structural studies of representatives from all three groups, E1 proteins, T-antigens, and Rep proteins, have provided important information about how ATP is bound and hydrolyzed by these proteins and the structural consequences that result [56, 58, 103, 122, 128, 129]. For example, in the recent crystal structure of a hexamer of the E1 oligomerization and helicase domains formed on ssDNA, an ATP binding pocket is formed by ~11 residues from two adjacent monomers of the E1 helicase domain (Figure 4-1A) [28]. The structural data available for SV40 T-ag demonstrate that all of these positions, with two exceptions, are conserved (Figure 4-1B) [56]. Because most of the residues thought to be involved in ATP binding and hydrolysis in these AAA+ proteins are highly conserved, the specific function of the individual residues have been predicted for these proteins[121].

To gain a more precise understanding of the role of these particular residues in ATP binding and hydrolysis, and because a systematic analysis of such residues has not been performed for SF3 type helicases, we performed a mutational analysis of these 11 residues. Based on the behavior of mutants in these residues, the residues can be classified into three groups. Most of the mutations (7) result in a protein that fails to bind nucleotide and consequently also fail to hydrolyze nucleotide. Three mutants can still

bind nucleotide but fail to hydrolyze ATP. Surprisingly, two of these mutants mimic the ATP bound state and can bind DNA in the absence of nucleotide.

4.2 Results and Discussion

Based on the structure of the hexamer of the E1 oligomerization and helicase domains (Figure 4-1A), we mutated all the residues predicted to contact ATP in a BPV1 E1 hexamer [28]. These substitutions were generated in the *E. coli* expression vector pETGST-E1 by site-directed mutagenesis. Ten of the 11 residues (K425, K439, S440, D478, D479, D497, Y499, N523, Y534 and R538) were changed into alanine. Because the alanine substitution at R493 was not expressed this residue was mutated into Leu, Met and Glu. After expression and purification of the mutant proteins they were first tested for the ability to bind to the origin of DNA replication together with the BPV E2 protein. The E1 and E2 proteins bind cooperatively to adjacent sites in the origin of DNA replication [78, 126, 130, 131]. The resulting complex, $E1_2E2_2$-ori, does not rely on nucleotide binding or hydrolysis for formation and therefore serves as a convenient control for the structural integrity of the E1 mutants and for the intrinsic ability of the mutant proteins to bind to the E1 BS in the ori.

The alanine substitutions were tested for $E1_2E2_2$ complex formation to ascertain that they had no structural defects due to the mutations (Figure 4-1C). As is well

established, E2 alone binds to the E2 binding site present in the ori and forms an E2 dimer complex (Figure 4-1C, lane 23). In the presence of WT E1 a larger complex $E1_2E2_2$ forms through cooperative binding of E1 and E2 to adjacent E1 and E2 binding sites (Figure 4-1B, lanes 1-2). All of the mutants, with one exception, could form this complex as well as WT E1 (Figure 4-1C, lanes 3-20). The exception, R538A, showed a slight defect in formation of the $E1_2E2_2$ complex (Figure 4-1C, lanes 21-22). R493E, M, and L, also formed WT levels of $E1_2E2_2$-ori complex formation (data not shown).

We next tested the mutants in an ATPase assay (Figure 4-1D). In this assay, ATP with a radioactively labeled γ-phosphate was incubated with either the WT or the mutant E1 proteins. Hydrolysis of the ATP would result in the appearance of radioactively labeled free phosphate, which can be separated from radioactively labeled ATP by thin layer chromatography (Figure 4-1D). The 10 alanine substitutions were, with one exception devoid of detectable ATPase activity. The exception (Y534, lane 10) had ~ 25% of the ATPase activity of the WT E1 (compare lanes 1 and 13). The residues in the position of Y534 are not well conserved while most of the rest residues are highly conserved (except Y499) in SF3 helicases (Figure 4-1B). This result demonstrated that all of the 10 residues play a role in ATP hydrolysis. As expected, the mutants R493E, M, and L was also devoid of ATPase activity (Figure 4-1D, lanes 15-17).

Measurement of nucleotide binding for proteins such as E1 are complicated by the fact that ATP binds between two subunits, and the level of ATP binding would therefore

depend greatly on whether the protein is monomeric or oligomeric. We have demonstrated that E1 in the absence of DNA is monomeric and methods such as filter-binding assays to measure ATP binding are therefore not practical [110]. Instead we used an indirect method to measure nucleotide binding. In the absence of nucleotide E1 binds to DNA weakly or not at all depending on the conditions[17]. The basis for this dependence on nucleotide binding for DNA binding is unknown, but is likely the result of conformational changes in E1 that exposes the E1 DNA binding domain.

To determine whether the 10 alanine substitutions were also defective for nucleotide binding, we utilized EMSA. E1 can, in the presence of ADP or ATP form a trimer complex on any short DNA sequence[17]. We incubated the WT and mutant E1 proteins with a 39bp ori probe that lacks E1 BS. In the absence of nucleotide, no complex is observed (Figure 4-2A, lanes 1-3), while in the presence of ADP (lanes 4-6) the trimer band is observed. Two of the mutants, D479A and Y534A similar to WT E1, bound as a trimer in the presence of ADP (Figure 4-2A, lanes 10-12 and 16-18, respectively), but not in its absence (Figure 4-2A, lanes 7-9 and 13-15, respectively). Two other mutants, D478A and N523A showed a different behavior. These two mutants showed significant trimer formation both in the absence (Figure 4-2B, lanes 7-9 and 13-15, respectively), and presence of ADP (Figure 4-2B, lanes 10-12 and 16-18, respectively). The remaining 6 alanine substitutions, K425A, K439A, S440A, D497A, Y499A, and R538A, failed to form the trimer complex (Figure 4-2C) as did the three R493 substitutions (Figure 4-2D, lanes 2-12).

These results placed the mutants in three different categories. The largest group, seven mutants, failed to form a trimer in the absence and presence of ADP indicating that these mutants are defective for nucleotide binding. The second group with the two mutants D479A and Y534A behaved as WT E1, i.e. these mutants formed the trimer but only in the presence of ADP. Since D479A was defective for ATP hydrolysis, this demonstrates that this mutant can bind but not hydrolyze ATP. Y534A can also clearly bind ATP, but is only slightly defective for ATP hydrolysis (Figure 4-1B). The third group, consisting of the mutants D478A and N523A, are the most interesting mutants, since they behaved as if they had nucleotide bound even in the absence of nucleotide. One of these (N523A) was clearly able to bind nucleotide, since DNA binding was stimulated by the addition of ADP.

We focused on the four mutants that were capable of DNA binding in the E1 trimer assay above and tested these mutants for the ability to form the functional double trimer (DT) and double hexamer (DH)[17] (Figure 4-3). As expected, WT E1 failed to form a complex in the absence of nucleotide (Figure 4-3A, lanes 1-2) formed a DT ($E1_6$) in the presence of ADP (lanes 3-4) and a DH ($E1_{12}$) in the presence of ATP (lanes 5-6). The mutant D478A (Figure 4-3A, lanes 7-12) formed ladders in the absence of nucleotide (lanes 7-8). In the presence of ADP and ATP a DT was formed (Figure 4-3A, lanes 9-10 and 11-12, respectively). N523A (Figure 4-3B, lanes 10-18) behaved similarly. These results clearly indicate that D478A and N523A are incapable of hydrolyzing ATP,

consistent with the ATPase assays, but because an effect of ADP addition is observed these proteins can both bind nucleotides. However, since DNA binding is observed even in the absence of nucleotide these mutations clearly have an effect on the DNA binding properties of the protein in the absence of nucleotide, mimicking some aspects of nucleotide binding, as observed above.

The mutants D479A (Figure 4-3A, lanes 13-14) and Y534A (Figure 4-3A, lanes 19-20) behaved as predicted from the trimer formation and failed to form DT without nucleotide. D479A also formed the DT very weakly in the presence of ADP (Figure 4-3A, lanes 15-18) and instead formed a larger complex of unknown composition. This indicates that this mutant may have a defect in ADP binding, however the DT was formed more efficiently in the presence of ATP (lanes 17-18). Y534A (Figure 4-3A, lanes 21-24), however, formed DT in the presence of both ADP and ATP, showing no indication that the residual ATPase activity of this mutant allows the formation of the DH.

The DNA binding properties of the two mutants D478A and N523A clearly indicate that they are defective for ATP hydrolysis since the DH does not form in the presence of ATP. This result is consistent with the ATPase assays, which also demonstrated a lack of ATPase activity for these mutants (Figure 4-1C). Furthermore, these two mutants have substantial DNA binding activity in the absence of nucleotide, indicating that the mutations may mimic the nucleotide bound state. Interestingly

however, both of these mutants are also affected by addition of nucleotide, indicating that they can still bind nucleotide.

The behavior of D478A and N523A indicated that these particular mutations mimic a conformational change that normally is induced by nucleotide binding, giving rise to DNA binding activity in the absence of nucleotide. Another possibility is that these particular mutant proteins already have nucleotide stably bound that survives the purification procedure. To distinguish between these two possibilities, we performed the experiment shown in Figure 4-4. We first incubated the E1 protein under EMSA conditions with 2mM ADP at room temperature for 10 min. The sample was then diluted 20-fold to provide a final ADP concentration of 0.1 mM. Probe was then added to the reaction mixture and the sample was divided in two. To one half no ADP was added (lane 2), to the other half 2mM ADP was added (lane 3). WT E1 failed to form trimer without nucleotide (lane 1) and also failed to form trimer in the sample diluted without nucleotide (lane 2) but formed a robust complex in the presence of added ADP (lane 3). This result demonstrates that 0.1mM ADP is not sufficient for trimer formation. It also demonstrates that WT E1 does not bind ADP sufficiently well to allow dilution and therefore that the half-life of nucleotide bound to WT E1 is very short since the duration of the dilution is less than one minute.

We performed the same experiment with N523A and D478A (Figure 4-4, lanes 4-9). These mutants, which bind DNA without nucleotide, showed no difference in trimer

formation without ADP and with the diluted ADP (Figure 4-4, lanes 4-5 and 7-8, respectively), demonstrating either that they like WT E1 bind nucleotide with a very short half-life or that they cannot bind nucleotide. Since we have already shown that both of these proteins can bind nucleotide in Figure 4-3, this result demonstrates that the half-life for bound nucleotide is very short. That N523A is capable of binding ADP is also clearly demonstrated by the incubation with 2mM ADP, which resulted in a substantial increase in binding (Figure 4-4, lane 6). The D478A mutant behaved similarly except that addition of ADP did not stimulate trimer formation (Figure 4-4, lanes 7-9) consistent with the results in Figure 4-2.

These results demonstrate that N523A clearly can bind ADP, but that the resulting E1-ADP complex has a very short half-life, since after dilution, which takes less than 1 min. the effect of the ADP is completely lost (compare lanes 4 and 5 in Figure 4-4). Since the E1 purification procedure does not include ADP and takes about 48 hours, it is very unlikely that ADP could still be bound to E1 after this procedure. This result indicates that the most likely possibility is that these mutations induce conformational changes, similar to those induced by nucleotide binding.

Our expectation was that the majority of the residues that we mutated in E1 would affect both ATP binding and hydrolysis. This was clearly the case, seven of the residues (K425A, K439A, S440A, D497A, Y499A, R538A and R493 M, L, E) when mutated, resulted in a protein that was defective for both ATP binding and hydrolysis. These

include the expected Walker A mutations (K439A and S440A), the Sensor 2 mutation (K425A) and the arginine finger (R538A), plus an additional three residues (R493M/L/E, D497A and Y499A). The only surprises in this group were the Sensor 2 and the arginine finger mutations, which generally affect ATP hydrolysis but not ATP binding.

We also hoped that some mutations would affect ATP hydrolysis only, since these mutant proteins would be useful reagents. Three of the mutants, D478A, D479A and N523A are completely defective for ATP hydrolysis (Figure 4-1B), but still are affected by the addition of ATP indicating that they can bind nucleotide (Figure 4-3A and B). Interestingly, these mutants represent the Walker B motif (D478, D479) and the Sensor 1 (N523), three residues that are in close proximity to each other. Two of these mutants (D478A and N523A) also showed a completely unexpected phenotype since these proteins, even in the absence of nucleotide, are capable of binding DNA. A particularly striking result is the trimer formation in Figure 4-2B which clearly demonstrates that D478A and N523A are fully competent for trimer formation in the absence of nucleotide. It is also interesting that the phenotype that we observe in Figure 3, in contrast to the results in Figure 4-2, only partially mimics nucleotide binding. Here, although D478A and N523A clearly show that they can bind DNA in the absence of nucleotide, a ladder is formed instead of the DT, and DT formation requires addition of ADP. This result indicates that the nucleotide binding has at least two functions in these assays, one is to activate DNA binding in a generic way and the other is specifically to allow or to stimulate DT formation and that these two functions can be separated. Only the first of these functions is mimicked by the D478A and N523A mutations.

Currently, we cannot propose a structural model for how the D478A and N523A mutations result in a conformation that mimics the ATP bound state for either of these mutants. It is well established that the conserved residues in Walker B (D478 and D479) are involved in hydrolysis of ATP and consistent with such a function neither of the alanine substitutions can hydrolyze ATP although both can bind nucleotide. The Sensor 1 residue, N523, is generally involved in contacting Walker B and the γ-phosphate of ATP. The normal function of these residues in AAA+ proteins does not explain the ability of these mutations to mimic the nucleotide bound state. Furthermore, neither of these residues is in a radically different conformation in the structures generated in the absence or presence of nucleotide [28]. It is possible that this is due to fact that the structural information is derived exclusively from hexamer structures and that these two residues would be in a different conformation in a monomer, for example. Resolution of these questions will require further analysis into the nature of activation of DNA binding that nucleotide binding obviously can confer upon the E1 protein.

4.3 Materials and Methods.

Recombinant proteins

Expression and purification of E1 and E2 proteins were carried out as described[97].

ATPase assays

ATPase assays were performed in a 20µl reaction containing 30mM HEPES, pH 7.5, 30mM NaCl, 1 mM DTT, 7 mM $MgCl_2$, 100 µg/ml bovine serum albumin, 100 µM ATP and 40,000 cpm [γ-^{32}P]ATP and E1. Reactions were incubated for 1 hour at room temp and stopped by the addition of EDTA to a final concentration of 10 mM. 2 µl of the reactions was spotted onto a polyethyleneimine-cellulose plate and the plate was then developed in 1M formic acid and 0.5 M $LiCl_2$ for 40 minutes. After drying, the plate was exposed to a Fuji imaging plate and the level of free phosphate was determined by scanning the plate using a Fuji BAS Imager.

EMSA

Four percent acrylamide gels (39:1 acrylamide:bis) containing 0.5 × TBE, lacking EDTA, were used for all EMSA experiments. E1 was added to the probe (~2 fmol) in 10 µl binding buffer, BB (20 mM HEPES [pH 7.5], 100 mM NaCl, 0.7 mg/ml BSA, 0.1% NP40, 5% glycerol, 5 mM DTT, 5 mM $MgCl_2$, and 2 mM ATP or ADP). After incubation at room temperature for 1 hr, the samples were loaded and run for 2 hr at 9V/cm. The ability to generate discrete complexes, especially the DT and DH, was critically dependent on high-purity acrylamide, freshly prepared APS solution, overnight polymerization of the gels, and precise prerunning time (9V/cm for 4 hr).

Accession codes

Protein Data Bank: E1 oligomerization and heliase domain hexamer structure coordinates can be accessed with PDB code 2GXA[28].

Figure 4-1. Residues of E1 involved in ATP binding and hydrolysis

A. **Residues in E1 involved in nucleotide binding and hydrolysis**. A schematic image of the interface between two E1 monomers that constitute the ATP binding pocket of BPV E1 with the residues that are predicted to be involved in ATP binding and hydrolysis. The Walker A, Walker B and Sensor 1 and Sensor 2 motifs, and the Arginine finger are indicated. Adapted from Enemark, E.J. and L. Joshua-Tor, Mechanism of DNA translocation in a replicative hexameric helicase. Nature, 2006. 442(7100): p. 270-5 [28].

B. **Sequence alignment of A, B, B' and C motifs of representative members of the SF3 helicase family.** Shown above the alignment is the relative location of SF3 signature motifs in BPV E1. The number above each motif corresponds to the starting and end residue number of each motif in BPV E1. The residues invovled in ATP binding and hydrolysis in BPV1 E1 are marked with asterisks. Dark grey, invariant residues in papovaviruses and parvoviruses. Light grey, conserved residues in papovaviruses and parvoviruses. Included in the comparison are the following: BPV1 E1, bovine papillomavirus type 1 E1 protein; HPV11 E1, human papillomavirus type 11 E1 protein; SV40 TAg, simian virus 40 T antigen; AAV2 Rep, adeno-associated virus 2 Rep 40 protein.

C. **Formation of the $E1_2E2_2$-ori complex.** EMSA was performed using an 84 bp ori probe. Two quantities (1.5 and 3 ng) of WT E1 and of each E1 substitution, as indicated at the top of the gels, were used in the presence of 0.1ng of full-length

E2. In lane 23, E2 alone was added. The mobility of the $E2_2$ and $E1_2E2_2$ complexes are indicated.

D. ATPase activity of E1 substitution mutations of residues involved in ATP binding and hydrolysis. Eighty ng of WT E1 or of each respective E1 substitution mutant as indicated were tested for ATPase activity using ^{32}P labeled γ-ATP. After the reaction the free phosphate was separated from ATP by thin layer chromatography and quantitated using a FujiImager. Lane 12 and lane 14 contained [γ-^{32}P] ATP only.

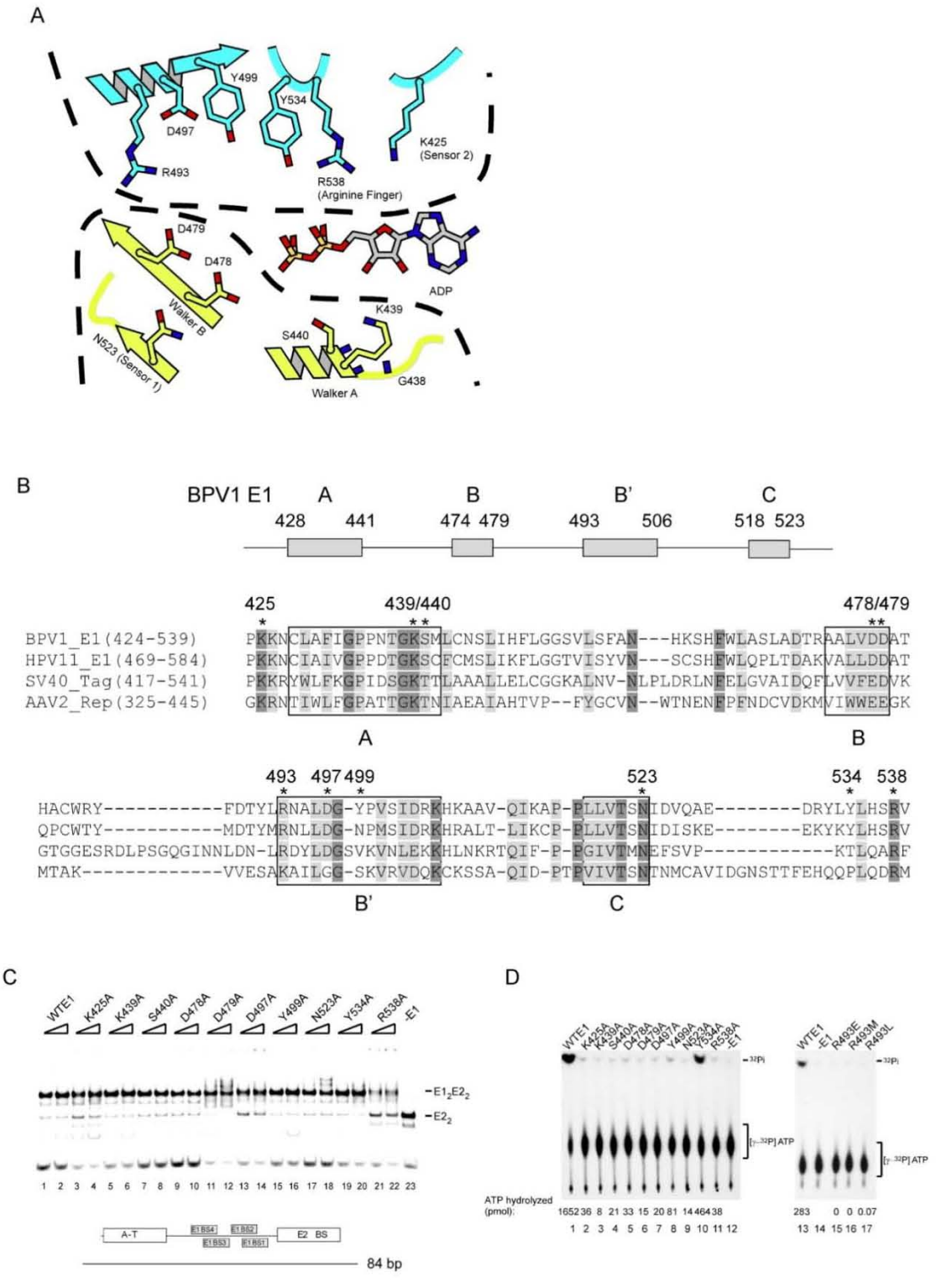

A
Y499
Y534
D497
K425
(Sensor 2)
R493
R538
(Arginine Finger)
D479
D478
ADP
K439
Walker B
S440
N523 (Sensor 1)
G438
Walker A
B
BPV1 E1
A
B
B'
C
428
441
474 479
493
506
518 523
425
439/440
478/479
BPV1_E1(424-539)
HPV11_E1(469-584)
SV40_Tag(417-541)
AAV2_Rep(325-445)
493 497 499
523
534 538
C
WTE1
K425A
K439A
S440A
D478A
D479A
D497A
Y499A
N523A
Y534A
R538A
-E1
-E1E2 2
-E2 2
A-T
E2 BS
84 bp
D
ATP hydrolyzed
(pmol):
R493E
R493M
R493L

Figure 4-2. Trimer formation of WT E1 and E1 substitution mutants

WT E1 and the E1 point mutants were tested for complex formation by EMSA on a 39 bp ori probe on which the WT E1 forms a trimer. Three quantities (6, 12 and 24 ng) of WT E1 and the respective E1 mutants were used in the absence or presence of 2 mM ADP as indicated at the top of the panels. In panel **A**, D479A, and Y534A were tested, in panel **B**, D478A and N523A were tested. In panel **C**, K425A, K439A, S440A, D497A, Y499A and R538A were tested in the presence of ADP, and in panel **D**, R493L, R493M and R493E were tested in the presence of ADP.

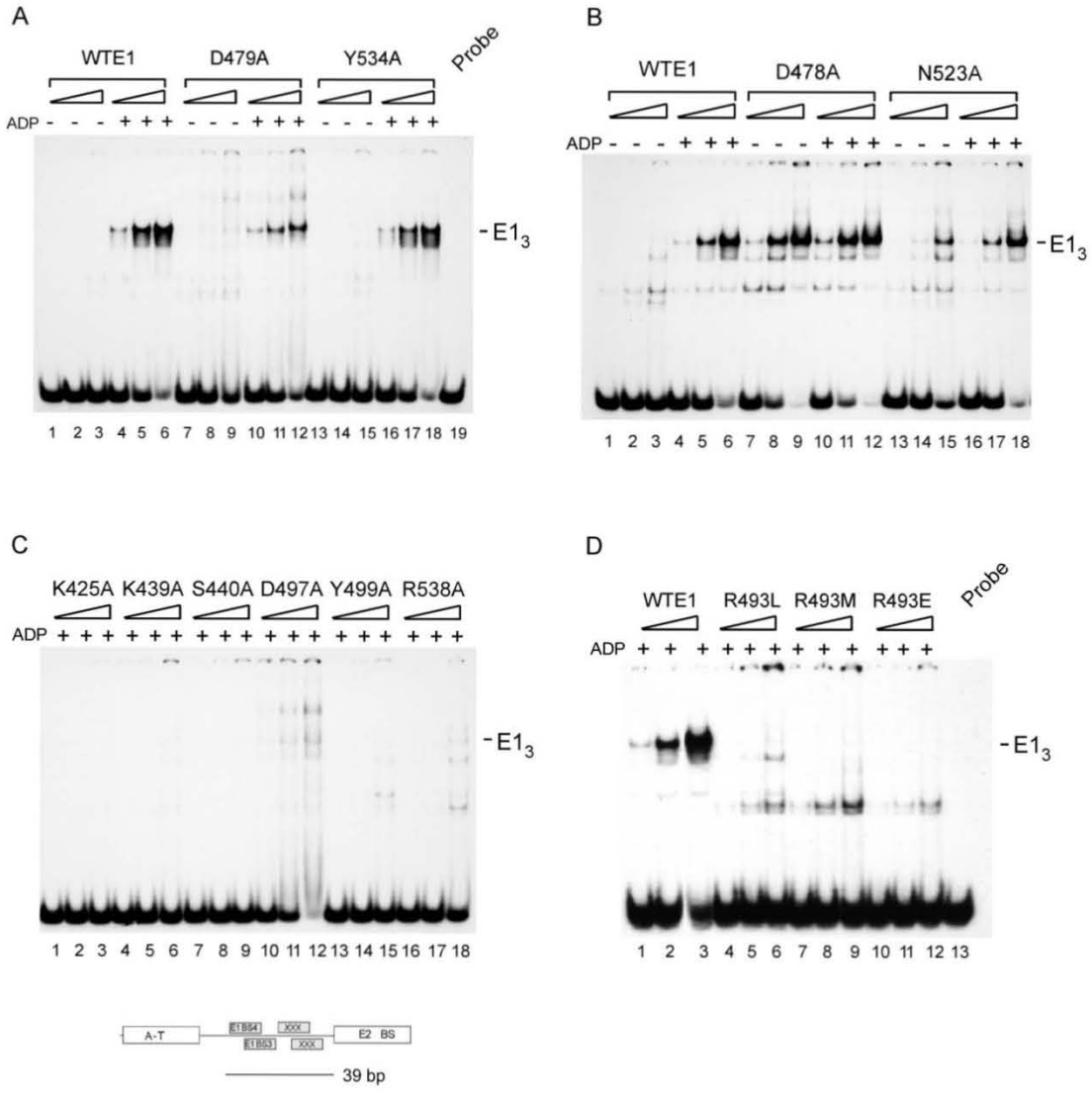
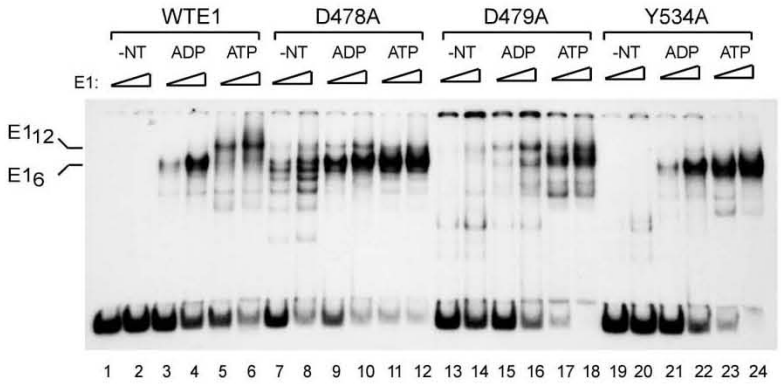

A
WTE1
D479A
Y534A
Probe
ADP
- - - + + + - - - + + + - - - + + +
$E1_3$
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
B
WTE1
D478A
N523A
ADP
- - - + + + - - - + + + - - - + + +
$E1_3$
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
C
K425A K439A S440A D497A Y499A R538A
ADP
+ + + + + + + + + + + + + + + + + +
$E1_3$
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
A-T
E1BS4
E1BS3
XXX
XXX
E2 BS
39 bp
D
WTE1
R493L
R493M
R493E
Probe
ADP
+ + + + + + + + + + + +
$E1_3$
1 2 3 4 5 6 7 8 9 10 11 12 13

Figure 4-3. E1 DT and DH formation of ATPase mutants

WT E1 and E1 point mutants were tested for DT and DH formation in EMSA using an 84 bp ori probe. Two quantities of E1 (6 and 12ng) were used in the absence of nucleotide, in the presence of 2 mM ADP or in the presence of 2 mM ATP as indicated above the lanes. The position of E1 DT ($E1_6$) and E1 DH ($E1_{12}$) are indicated. In panel **A**, WT E1, D478A, D479A and Y534A were tested; in panel **B**, WT E1 and N523A were tested.

A


WTE1
D478A
D479A
Y534A
-NT
ADP
ATP
E1:
E1$_{12}$
E1$_6$
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

B

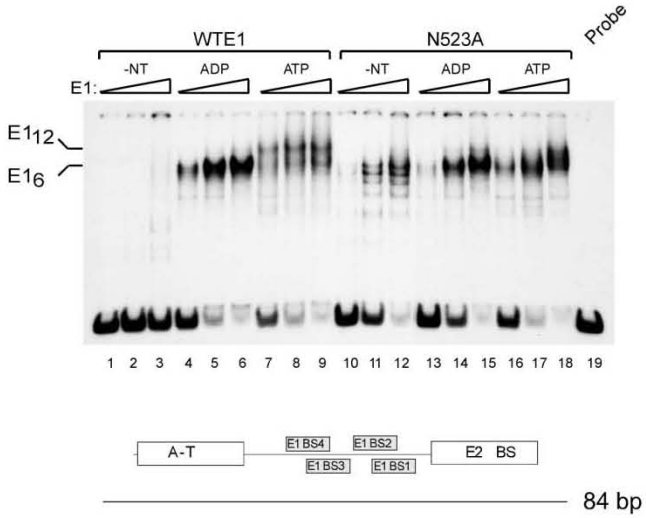
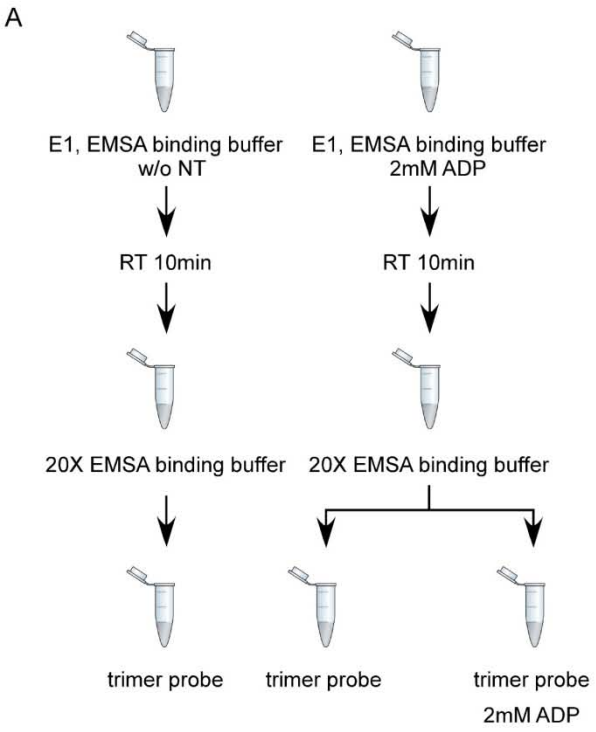
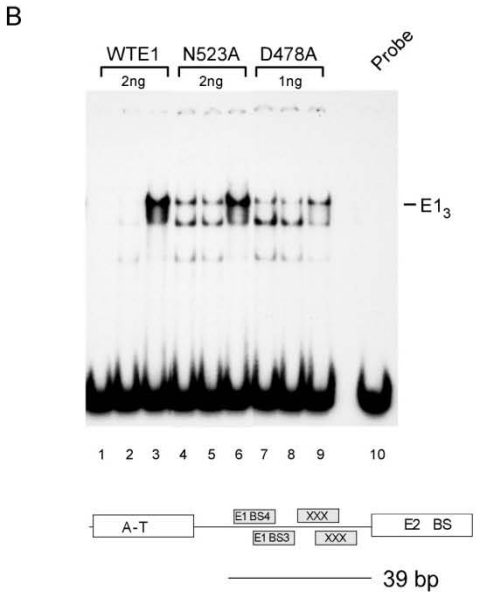

WTE1
N523A
Probe
-NT
ADP
ATP
E1:
E1$_{12}$
E1$_6$
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
A-T
E1 BS4
E1 BS2
E1 BS3
E1 BS1
E2 BS
84 bp

Figure 4-4. E1 binds ATP with a very short half-life. A. Schematic description of the experiment shown in Figure 4B. E1 was incubated in the absence or presence of nucleotide. The samples were then diluted 20-fold. Probe was then added, and the sample then loaded on an EMSA gel. In sample 1 in each set, the binding reaction without ADP was diluted without ADP resulting in a sample without ADP. In sample 2 in each set, the binding reaction was diluted without ADP resulting in a sample containing 0.1 mM ADP. In sample 3 in each set, the binding reaction was diluted with 2mM ADP resulting in a sample containing 2mM ADP. **B**. WT E1, N523A and D478A were incubated as described in the scheme in panel A and analyzed by EMSA using the 39bp trimer probe. Lane 10 contained probe alone.


A
E1, EMSA binding buffer w/o NT
E1, EMSA binding buffer 2mM ADP
RT 10min
RT 10min
20X EMSA binding buffer
20X EMSA binding buffer
trimer probe
trimer probe
trimer probe 2mM ADP


B
WTE1
N523A
D478A
Probe
2ng
2ng
1ng
E1$_3$
1 2 3 4 5 6 7 8 9 10
A-T
E1 BS4
E1 BS3
XXX
XXX
E2 BS
39 bp

Chapter 5 Discussion and Perspectives

The papillomavirus E1 protein is involved in multiple activities during viral genome replication. An important feature that distinguishes E1 from many other replication initiator proteins is that E1 can carry out both dsDNA melting and unwinding activity during the replication initiation process. Previous work has shown that E1 forms a complex, E1 DT, during initiation. E1 DT has been suggested to be responsible for dsDNA melting[17]. The assembly of DT requires cooperation of the two DNA binding activities in E1. The specific DNA binding activity of E1 DBD anchors E1 to E1 BS while the E1 oligomerization and helicase domains bind to the DNA sequences flanking the E1 BS with nonspecific activity.

5.1 DNA melting Vs. DNA unwinding

Melting of dsDNA and unwinding are two steps in the process of initiation of DNA replication. Melting is a crucial step that has to precede unwinding because all known helicases require an ssDNA template for assembly and to initiate dsDNA unwinding. In *E. coli*, the DNA melting activity is carried out by DnaA while DnaB unwinds dsDNA. Papillomavirus initiator E1 is responsible for both melting and unwinding for viral genome replication initiation. In Chapter 2, we showed that alanine substitution of K506, a conserved residue on the tip of the β-hairpin in the E1 helicase domain, results in a protein that lacks both melting and helicase activity. The mutation

H507A, in contrast, is defective only for melting. We propose that His507 can intercalate into the dsDNA and pry open the DNA. Our results demonstrate that although the E1 helicase domain carries out both melting and unwinding activities during replication initiation, these two activities can be separated.

5.1.1 DNA melting in other organisms

SV40 T-ag as well as replication initiators from all other polyomaviruses share conserved β-hairpin residues with E1. This indicates that all papovavirus initiators melt and unwind DNA using similar mechanisms as BPV1 E1. This conservation does not extend to other virus groups. The Rep protein from parvoviruses belongs to the SF3 helicase family as does SV40 T-ag and papillomavirus E1. However, while K506 is conserved in the Rep proteins, consistent with the function of Rep as a helicase, H507 is not. We believe that melting is not required because parvoviruses have ssDNA genomes. The helicase activity of Rep proteins, however, is required to unwind the double-stranded hairpin telomeres for viral genome replication[129].

In *E.coli*, which has a dsDNA genome, DNA melting and helicase activity are carried out by two replication factors, DnaA and DnaB, respectively. The bacterial initiator DnaA is suggested to use a mechanism that is dramatically different from E1 for dsDNA melting. Based on the DnaA monomer structure and biochemical data, it is proposed that DnaA forms a helical filament on the outside of dsDNA at oriC in the presence of ATP. This wrapping of DNA stabilizes positive DNA supercoils, and

negative writhe is generated in response to a positive torodial wrap [2, 12-14, 31]. Despite the different mechanism used by *E. coli* and papillomaviruses for DNA melting, A-T rich elements that can be distorted and readily melted are required for melting steps in both systems.

An activity that can melt DNA in preparation for DNA replication has not been identified in eukaryotes. Melting activity could reside in any of the proteins or protein complexes that are known to take part in the initiation of DNA replication (e.g., ORC, MCM 2–7, etc.) or in some unknown complex. A CDC45/MCM2-7/GINS complex has been shown to be able to unwind the dsDNA in *in vitro* helicase assays, and the MCM2-7 complex is believed to carry out the unwinding activity[23]. The eukaryotic initiator ORC is distantly related to bacterial initiator DnaA. However, eukaryotic origins lack the features present in the *E.coli oriC,* which contains multiple binding sites for DnaA. It is therefore unlikely that ORC assembles into a multi-subunit filament on the origin, as has been proposed for DnaA. Our study suggested an alternative way for local template melting. E1 DT is incorporated into the functional helicase, E1 DH, after the dsDNA melting in the assembly process. Because of the obvious facility of an arrangement in which local melting is an integral part of helicase assembly, it is conceivable that a similar strategy may be used in eukaryotes. For example, a subassembly of a multimeric helicase could contain the melting activity. Archaeal MCM proteins contain a β-hairpin structure, but it is unknown whether this β-hairpin functions in the helicase, in melting, or in both processes [132]. The precise feature that makes the E1 β-hairpin capable of melting DNA (the aromatic side chain at the tip) is not present in the β-hairpin of the

MCM protein [132]. However, the β-hairpin that is thought to intercalate into DNA and effect melting in the phage T7 RNA polymerase also lacks this particular feature.

5.1.2 DT in DNA melting

The E1 DT is an important intermediate in E1 DH formation and melts the DNA in preparation for the formation of the helicase [17]. In Chapter 2, we showed that H507 substitutions in E1 that fail to form DT also fail to support replication. This correlation demonstrates that DT plays an important role in DNA replication, most likely in dsDNA melting. To determine the precise role of the DT in melting, competitor DNA was added after DT was assembled and a time course of melting was performed. The results demonstrated that the DT is solely responsible for local template melting. Formation of the DH, which is exclusively responsible for unwinding, requires addition of E1 molecules to the E1 DT-melted-DNA-complex (Schuck and Stenlund, unpublished data).

5.1.3 A model of the transition from DNA melting to DNA unwinding

Well-studied hexameric DNA helicases such as DnaB and T7 helicase are believed to function by binding as hexamers to ssDNA, followed by translocation on ssDNA and displacement of the complementary strand [133-136]. The X-ray crystal structure of the helicase and oligomerization domains of E1 in the presence of ssDNA indicates that this is also the case for E1 protein[28]. A remaining question is how one of the DNA strands is displaced to the outside of the hexameric ring while one strand

remains inside the ring. The involvement of the β-hairpin and F464 in both melting and helicase activity provides some hints. The E1 DT, which forms on dsDNA, melts the dsDNA by interaction of the β-hairpin histidine with DNA, and this complex now becomes the substrate for DH formation. If, after melting, further addition of E1 molecules to the DT occurs specifically on one of the melted strands, the result would be exclusion of one strand and encirclement of the other strand by the hexamer, i.e., helicase loading (Figure 5-1). In such a model, the β-hairpin and the F464 residues can remain in contact with the DNA throughout the melting process and the transition to the DNA helicase, consistent with the dual function of the β-hairpin and F464.

5.2 Specific initiator loading Vs. Nonspecific DNA binding activity

A highly specific DNA binding activity is required for a replication initiator to identify a replication origin in the presence of large amounts of non-ori DNA. In *E. coli*, which uses a single replication origin *oriC* for replication, the bacterial initiator DnaA binds with high affinity to the DnaA boxes at *oriC* [2]. There is a much greater challenge for the initiation of papillomavirus replication due to the presence of large amounts of host genomic DNA. The DNA binding domain of E1 binds to the E1 BS in the ori with an intrinsically high specificity. However, in the full length E1 protein this specific binding is masked by the nonspecific DNA binding activity of its oligomerization and helicase domain. The auxiliary factor E2 is required to block the nonspecific DNA binding activity of E1. The nonspecific DNA binding activity obviously constitutes an obstacle to ori identification, which is one of main functions of the initiator protein. From

an evolutionary perspective, this seemingly inconvenient nonspecific dsDNA binding activity must be preserved for a special purpose, or the loading of E1 with high specificity binding without the help of E2 would be more efficient. This apparent paradox of the coexistence of two competing DNA binding activities in the initiator E1is resolved in Chapter 3, where we identified a large number of residues involved in the nonspecific DNA binding of E1. We show that these residues are important for viral replication initiation because they are required for E1 DT assembly and local ori melting.

The T-ag helicase domain shares a high degree of similarity with E1 helicase domain. As discussed above, T-ag uses the same mechanism as E1 for dsDNA melting. The essential histidine residue on the tip of the β-hairpin (H507 in E1 and H513 in T-ag) that is required for melting is also involved in the nonspecific dsDNA binding activity. This suggests that T-ag contains a similar nonspecific DNA binding activity as E1. This is supported by the finding that T-ag from polyomavirus (Py), which is very closely related to SV40, has a low selectivity for ori binding[137]. However, no viral auxiliary factor has been implicated in T-ag loading. The early palindrome (EP) sequence in the SV40 ori is in a similar position relative to the T-ag binding sites as the position that the E2 BS occupies relative to the E1 BS. This suggests that there might be an unidentified cellular factor that provides specificity for the binding of T-ag [8].

The Eukaryotic initiator complex ORC binds to replication origins and recruits other replication factors to initiate replication. In the budding yeast *S.cerevisiae*, ORC

binds autonomously replicating sequences (ARS) with modest specificity [11, 138]. In the fission yeast *S.pombe,* the binding sites for ORC are less well defined except that they are rich in A-T base pairs [139, 140]. Metazoan ORC binds to DNA with low sequence specificity in the loading process[141]. It is possible that analogous to the nonspecific binding activity of E1, the nonspecific binding activity of ORC provides some other unknown functions that are crucial for replication. Loading of ORC may require other facilitating factors or specific features of the chromatin such as epigenetic modifications.

5.3 ATP/Mg^{2+} in replication machinery assembly

The processes of ATP binding and hydrolysis constitute important regulatory tools during initiation of DNA replication. Generally, the nucleotide hydrolysis provides the energy for various biochemical processes including dsDNA melting and unwinding. E1 and many other initiator proteins are AAA+ (ATPases associated with various cellular activities) proteins with conserved motifs for ATP binding and hydrolysis.

ATP binding and hydrolysis by E1 has multiple functions during papillomavirus replication. First, ATP binding stimulates the DNA binding activity of E1 and promotes E1 complex assembly during replication initiation. The binding of nucleotide causes allosteric conformational change in E1 for it to form the DT [17, 28]. Both ADP and ATP can stimulate the assembly of the DT while ATP hydrolysis is required for the transition from DT to DH[17]. The results presented in Chapter 4 indicate that nucleotide binding has at least two functions in DT assembly. One is to activate DNA binding in a generic

way and the other is to specifically allow or stimulate DT formation. These two functions can be separated. By testing the residues involved in ATP binding and hydrolysis for DT assembly, we were able to distinguish those necessary for nucleotide binding from those required for ATP hydrolysis. Second, ori melting by E1 is dependent on ATP binding and hydrolysis. Third, ATP hydrolysis provides energy for dsDNA unwinding. Similar roles of ATP binding and hydrolysis in complex assembly, ori distortion and unwinding have also been demonstrated for SV40 T-ag[15, 21, 32, 142, 143]. The major role of ATP binding and hydrolysis of Rep is to promote the assembly of Rep helicase and provide the energy for dsDNA unwinding [129].

In *E.coli*, the initiator DnaA, helicase DnaB and helicase loader DnaC, three proteins that are crucial for replication initiation, also contain AAA+ modules for ATP binding and hydrolysis. In the presence of ATP, the initiator DnaA assembles into a multimeric complex on the replication origin and distorts it to generate the ssDNA template [12-14, 31]. ATP binding also promotes hexameric DnaB helicase assembly by stimulating the dimerization of DnaB trimers[144, 145]. Recent structural and biochemical data suggest that nucleotide binding of DnaC is required for its interaction with DnaA in the helicase loading process[146]. The replication initiation process of the eukaryotic cell is much more complicated than that of bacteria or viruses. Many more AAA+ proteins are involved. Orc1, Orc4, Orc5, CDC6 and MCM2-7 are all AAA+ family members. ATP binding is required for the ORC complex to bind to DNA [147, 148]. MCM complex loading is dependent on the ATP binding and hydrolysis by CDC6 [23, 149, 150]. Similar to papillomavirus, ATP binding and hydrolysis are important

ways of regulating replication initiation complex assembly and function in other systems. A similar strategy of studying ATP binding and hydrolysis pockets comparable to that used in Chapter 4 might help to dissect the functions of ATPases in replication initiation.

5.4 Perspectives

The results presented in this thesis provide insight into DNA melting and DT assembly during BPV1 replication initiation. However, there are still many questions that remain be elucidated.

1. Details of DT assembly and DNA melting

   We have identified many residues that are involved in the nonspecific DNA binding activities of E1 for E1 DT assembly. However, it is still not clear how these residues interact with dsDNA and which sequences they interact with. Which exact DNA sequences they interact with is likely determined by the distance between the DBD and the helicase/oligomerization domains. Because the linker regions between these domains are flexible, the distance could be significantly different than that in the hexamer structure of E1. Nucleotide binding can also cause motifs such as the β-hairpin and F464 to change their positions. Because the nonspecific binding of the E1 oligomerization and helicase domain to DNA sequences flanking the E1 BS is a weak interaction involving multiple residues, it is very difficult to elucidate how and where these residues bind to

DNA with biochemical methods. An X-ray crystal structure of E1 DT would be valuable to answer these questions.

An interesting aspect of the transition from DNA melting to unwinding is that the helical DT can transform itself into a ring structured DH by hydrolyzing ATP[17]. This suggests that the E1 subunits of DT change their relative positions at the same time as they modify the structure of the template DNA. Biophysical studies of the transition process might help to elucidate the molecular detail of this transition.

2. Cell cycle regulation of replication initiation complex assembly

E1 DT and DH have been demonstrated to be important complexes for ori fragment melting and unwinding *in vitro*. However, it is not clear how these complexes assemble in the host cells. Papillomavirus replication is coupled to host genome replication in that the viral replication occurs in S-phase. It is likely that the assembly or activity of E1 DT and DH is regulated by the cell cycle of the host cell. This will be a very interesting subject to be explored.

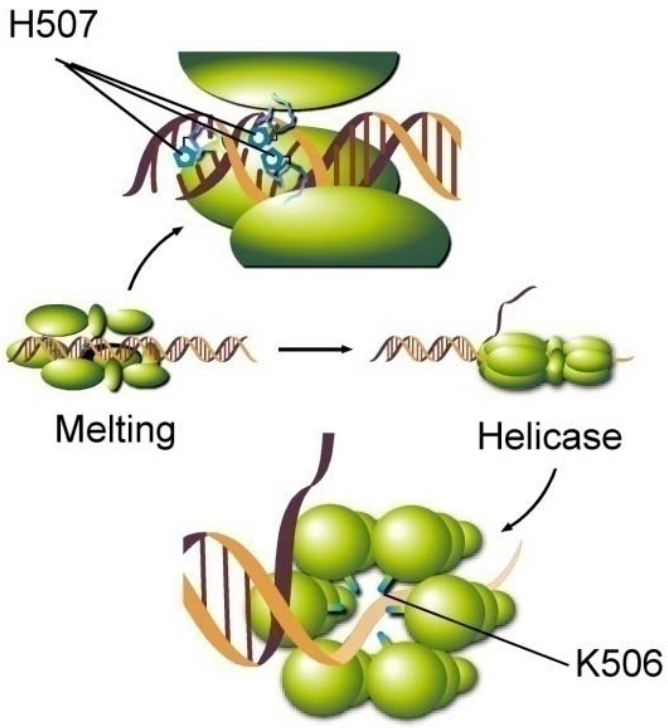
3. DNA melting in eukaryotic cells

No replication factor in eukaryotic cells has been demonstrated to be responsible for melting activity in replication initiation. Our knowledge of eukaryotic replication will be significantly improved if the protein or protein complex responsible for dsDNA melting in eukaryotic cells could be identified.

4. Potential application in medicine

My biochemical data confirmed the ATP binding and hydrolysis pocket of E1 predicted from the X-ray crystal structure[28]. Based on this structural and biochemical data, ATP analogs can be designed to specifically inhibit the activity of E1. This provides another possible treatment for cervical cancer caused by HPVs.

Figure 5-1. Model for the transition between the E1 DT complex, which melts DNA, and the E1 DH, which unwinds DNA. E1 forms a double trimer and H507 at the tip of the β-hairpin intercalates into the DNA in the minor groove. The DT DNA complex is the substrate for formation of the DH, which forms on one of the melted strands as the K506 at the tip of the hairpin interacts with ssDNA. For simplicity, only one trimer and one hexamer is shown.


H507
Melting
Helicase
K506

References

1. Stenlund, A., *Initiation of DNA replication: lessons from viral initiator proteins.* Nat Rev Mol Cell Biol, 2003. **4**(10): p. 777-85.
2. Mott, M.L. and J.M. Berger, *DNA replication initiation: mechanisms and regulation in bacteria.* Nat Rev Microbiol, 2007. **5**(5): p. 343-54.
3. Dutta, A. and S.P. Bell, *Initiation of DNA replication in eukaryotic cells.* Annu Rev Cell Dev Biol, 1997. **13**: p. 293-332.
4. Bell, S.P. and A. Dutta, *DNA replication in eukaryotic cells.* Annu Rev Biochem, 2002. **71**: p. 333-74.
5. Jacob, F. and S. Brenner, *[On the regulation of DNA synthesis in bacteria: the hypothesis of the replicon.]*. C R Hebd Seances Acad Sci, 1963. **256**: p. 298-300.
6. Weigel, C. and H. Seitz, *Bacteriophage replication modules.* FEMS Microbiol Rev, 2006. **30**(3): p. 321-81.
7. Stillman, B., *Initiation of chromosomal DNA replication in eukaryotes. Lessons from lambda.* J Biol Chem, 1994. **269**(10): p. 7047-50.
8. Deb, S., et al., *Domain structure of the simian virus 40 core origin of replication.* Mol Cell Biol, 1986. **6**(5): p. 1663-70.
9. Parsons, R., M.E. Anderson, and P. Tegtmeyer, *Three domains in the simian virus 40 core origin orchestrate the binding, melting, and DNA helicase activities of T antigen.* J Virol, 1990. **64**(2): p. 509-18.
10. Deb, S., et al., *The T-antigen-binding domain of the simian virus 40 core origin of replication.* J Virol, 1987. **61**(7): p. 2143-9.
11. Bell, S.P. and B. Stillman, *ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex.* Nature, 1992. **357**(6374): p. 128-34.

12. Bramhill, D. and A. Kornberg, *Duplex opening by dnaA protein at novel sequences in initiation of replication at the origin of the E. coli chromosome.* Cell, 1988. **52**(5): p. 743-55.

13. Sekimizu, K., D. Bramhill, and A. Kornberg, *ATP activates dnaA protein in initiating replication of plasmids bearing the origin of the E. coli chromosome.* Cell, 1987. **50**(2): p. 259-65.

14. Erzberger, J.P., M.L. Mott, and J.M. Berger, *Structural basis for ATP-dependent DnaA assembly and replication-origin remodeling.* Nat Struct Mol Biol, 2006. **13**(8): p. 676-83.

15. Borowiec, J.A. and J. Hurwitz, *Localized melting and structural changes in the SV40 origin of replication induced by T-antigen.* Embo J, 1988. **7**(10): p. 3149-58.

16. Borowiec, J.A., et al., *Binding and unwinding--how T antigen engages the SV40 origin of DNA replication.* Cell, 1990. **60**(2): p. 181-4.

17. Schuck, S. and A. Stenlund, *Assembly of a double hexameric helicase.* Mol Cell, 2005. **20**(3): p. 377-89.

18. Kelman, Z. and M. O'Donnell, *DNA replication: enzymology and mechanisms.* Curr Opin Genet Dev, 1994. **4**(2): p. 185-95.

19. Konieczny, I., *Strategies for helicase recruitment and loading in bacteria.* EMBO Rep, 2003. **4**(1): p. 37-41.

20. Stahl, H., P. Droge, and R. Knippers, *DNA helicase activity of SV40 large tumor antigen.* Embo J, 1986. **5**(8): p. 1939-44.

21. Mastrangelo, I.A., et al., *ATP-dependent assembly of double hexamers of SV40 T antigen at the viral origin of DNA replication.* Nature, 1989. **338**(6217): p. 658-62.

22. Dodson, M., et al., *Unwinding of duplex DNA from the SV40 origin of replication by T antigen.* Science, 1987. **238**(4829): p. 964-7.

23. Moyer, S.E., P.W. Lewis, and M.R. Botchan, *Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase.* Proc Natl Acad Sci U S A, 2006. **103**(27): p. 10236-41.

24. Waga, S., G. Bauer, and B. Stillman, *Reconstitution of complete SV40 DNA replication with purified replication factors.* J Biol Chem, 1994. **269**(14): p. 10923-34.

25. Waga, S. and B. Stillman, *Anatomy of a DNA replication fork revealed by reconstitution of SV40 DNA replication in vitro.* Nature, 1994. **369**(6477): p. 207-12.

26. Muller, F., Y.S. Seo, and J. Hurwitz, *Replication of bovine papillomavirus type 1 origin-containing DNA in crude extracts and with purified proteins.* J Biol Chem, 1994. **269**(25): p. 17086-94.

27. Enemark, E.J. and L. Joshua-Tor, *On helicases and other motor proteins.* Curr Opin Struct Biol, 2008. **18**(2): p. 243-57.

28. Enemark, E.J. and L. Joshua-Tor, *Mechanism of DNA translocation in a replicative hexameric helicase.* Nature, 2006. **442**(7100): p. 270-5.

29. Hopfner, K.P. and J. Michaelis, *Mechanisms of nucleic acid translocases: lessons from structural biology and single-molecule biophysics.* Curr Opin Struct Biol, 2007. **17**(1): p. 87-95.

30. Patel, S.S. and I. Donmez, *Mechanisms of helicases.* J Biol Chem, 2006. **281**(27): p. 18265-8.

31. Funnell, B.E., T.A. Baker, and A. Kornberg, *In vitro assembly of a prepriming complex at the origin of the Escherichia coli chromosome.* J Biol Chem, 1987. **262**(21): p. 10327-34.

32. Dean, F.B., et al., *The simian virus 40 T antigen double hexamer assembles around the DNA at the replication origin.* J Biol Chem, 1992. **267**(20): p. 14129-37.

33. McBride, A.A., *Replication and partitioning of papillomavirus genomes.* Adv Virus Res, 2008. **72**: p. 155-205.

34. zur Hausen, H., *Papillomaviruses in the causation of human cancers - a brief historical account.* Virology, 2009. **384**(2): p. 260-5.

35. Lambert, P.F., C.C. Baker, and P.M. Howley, *The genetics of bovine papillomavirus type 1.* Annu Rev Genet, 1988. **22**: p. 235-58.

36. Law, M.F., et al., *Mouse cells transformed by bovine papillomavirus contain only extrachromosomal viral DNA sequences.* Proc Natl Acad Sci U S A, 1981. **78**(5): p. 2727-31.

37. Gilbert, D.M. and S.N. Cohen, *Bovine papilloma virus plasmids replicate randomly in mouse fibroblasts throughout S phase of the cell cycle.* Cell, 1987. **50**(1): p. 59-68.

38. Mecsas, J. and B. Sugden, *Replication of plasmids derived from bovine papilloma virus type 1 and Epstein-Barr virus in cells in culture.* Annu Rev Cell Biol, 1987. **3**: p. 87-108.

39. Ustav, M. and A. Stenlund, *Transient replication of BPV-1 requires two viral polypeptides encoded by the E1 and E2 open reading frames.* Embo J, 1991. **10**(2): p. 449-57.

40. Ustav, M., et al., *Identification of the origin of replication of bovine papillomavirus and characterization of the viral origin recognition factor E1.* Embo J, 1991. **10**(13): p. 4321-9.

41. Yang, L., et al., *Activation of BPV-1 replication in vitro by the transcription factor E2.* Nature, 1991. **353**(6345): p. 628-32.

42. Chiang, C.M., et al., *Viral E1 and E2 proteins support replication of homologous and heterologous papillomaviral origins.* Proc Natl Acad Sci U S A, 1992. **89**(13): p. 5799-803.

43. Holt, S.E., G. Schuller, and V.G. Wilson, *DNA binding specificity of the bovine papillomavirus E1 protein is determined by sequences contained within an 18-base-pair inverted repeat element at the origin of replication.* J Virol, 1994. **68**(2): p. 1094-102.

44. Chen, G. and A. Stenlund, *The E1 initiator recognizes multiple overlapping sites in the papillomavirus origin of DNA replication.* J Virol, 2001. **75**(1): p. 292-302.

45. Androphy, E.J., D.R. Lowy, and J.T. Schiller, *Bovine papillomavirus E2 trans-activating gene product binds to specific sites in papillomavirus DNA.* Nature, 1987. **325**(6099): p. 70-3.

46. Bedrosian, C.L. and D. Bastia, *The DNA-binding domain of HPV-16 E2 protein interaction with the viral enhancer: protein-induced DNA bending and role of the nonconserved core sequence in binding site affinity.* Virology, 1990. **174**(2): p. 557-75.

47. Li, R., et al., *Specific recognition nucleotides and their DNA context determine the affinity of E2 protein for 17 binding sites in the BPV-1 genome.* Genes Dev, 1989. **3**(4): p. 510-26.

48. Newhouse, C.D. and S.J. Silverstein, *Orientation of a novel DNA binding site affects human papillomavirus-mediated transcription and replication.* J Virol, 2001. **75**(4): p. 1722-35.

49. Lu, J.Z., et al., *Two E2 binding sites (E2BS) alone or one E2BS plus an A/T-rich region are minimal requirements for the replication of the human papillomavirus type 11 origin.* J Virol, 1993. **67**(12): p. 7131-9.

50. Remm, M., R. Brain, and J.R. Jenkins, *The E2 binding sites determine the efficiency of replication for the origin of human papillomavirus type 18.* Nucleic Acids Res, 1992. **20**(22): p. 6015-21.

51. Russell, J. and M.R. Botchan, *cis-Acting components of human papillomavirus (HPV) DNA replication: linker substitution analysis of the HPV type 11 origin.* J Virol, 1995. **69**(2): p. 651-60.

52. Chiang, C.M., et al., *Control of human papillomavirus type 11 origin of replication by the E2 family of transcription regulatory proteins.* J Virol, 1992. **66**(9): p. 5224-31.

53. Sverdrup, F. and S.A. Khan, *Two E2 binding sites alone are sufficient to function as the minimal origin of replication of human papillomavirus type 18 DNA.* J Virol, 1995. **69**(2): p. 1319-23.

54. Schuck, S. and A. Stenlund, *ATP-dependent minor groove recognition of TA base pairs is required for template melting by the E1 initiator protein.* J Virol, 2007. **81**(7): p. 3293-302.

55. Yang, L., et al., *The E1 protein of bovine papilloma virus 1 is an ATP-dependent DNA helicase.* Proc Natl Acad Sci U S A, 1993. **90**(11): p. 5086-90.

56. Gai, D., et al., *Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen.* Cell, 2004. **119**(1): p. 47-60.

57. Enemark, E.J., et al., *Crystal structure of the DNA binding domain of the replication initiation protein E1 from papillomavirus.* Mol Cell, 2000. **6**(1): p. 149-58.

58. Li, D., et al., *Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen.* Nature, 2003. **423**(6939): p. 512-8.

59. Meinke, G., et al., *The crystal structure of the SV40 T-antigen origin binding domain in complex with DNA.* PLoS Biol, 2007. **5**(2): p. e23.

60. Meinke, G., P.A. Bullock, and A. Bohm, *Crystal structure of the simian virus 40 large T-antigen origin-binding domain.* J Virol, 2006. **80**(9): p. 4304-12.

61. Luo, X., et al., *Solution structure of the origin DNA-binding domain of SV40 T-antigen.* Nat Struct Biol, 1996. **3**(12): p. 1034-9.

62. Enemark, E.J., A. Stenlund, and L. Joshua-Tor, *Crystal structures of two intermediates in the assembly of the papillomavirus replication initiation complex.* Embo J, 2002. **21**(6): p. 1487-96.

63. Schuck, S. and A. Stenlund, *Role of papillomavirus E1 initiator dimerization in DNA replication.* J Virol, 2005. **79**(13): p. 8661-4.

64. Ferran, M.C. and A.A. McBride, *Transient viral DNA replication and repression of viral transcription are supported by the C-terminal domain of the bovine papillomavirus type 1 E1 protein.* J Virol, 1998. **72**(1): p. 796-801.

65. Sun, Y., H. Han, and D.J. McCance, *Active domains of human papillomavirus type 11 E1 protein for origin replication.* J Gen Virol, 1998. **79 (Pt 7)**(Pt 7): p. 1651-8.

66. Deng, W., et al., *Cyclin/CDK regulates the nucleocytoplasmic localization of the human papillomavirus E1 DNA helicase.* J Virol, 2004. **78**(24): p. 13954-65.

67. Lentz, M.R., et al., *The E1 replication protein of bovine papillomavirus type 1 contains an extended nuclear localization signal that includes a p34cdc2 phosphorylation site.* J Virol, 1993. **67**(3): p. 1414-23.

68. Hsu, C.Y., F. Mechali, and C. Bonne-Andrea, *Nucleocytoplasmic shuttling of bovine papillomavirus E1 helicase downregulates viral DNA replication in S phase.* J Virol, 2007. **81**(1): p. 384-94.

69. Ma, T., et al., *Interaction between cyclin-dependent kinases and human papillomavirus replication-initiation protein E1 is required for efficient viral replication.* Proc Natl Acad Sci U S A, 1999. **96**(2): p. 382-7.

70. Yu, J.H., et al., *Mitogen-activated protein kinases activate the nuclear localization sequence of human papillomavirus type 11 e1 DNA helicase to promote efficient nuclear import.* J Virol, 2007. **81**(10): p. 5066-78.

71. Cueille, N., et al., *Functional interaction between the bovine papillomavirus virus type 1 replicative helicase E1 and cyclin E-Cdk2.* J Virol, 1998. **72**(9): p. 7255-62.

72. Malcles, M.H., et al., *Regulation of bovine papillomavirus replicative helicase e1 by the ubiquitin-proteasome pathway.* J Virol, 2002. **76**(22): p. 11350-8.

73. Rangasamy, D. and V.G. Wilson, *Bovine papillomavirus E1 protein is sumoylated by the host cell Ubc9 protein.* J Biol Chem, 2000. **275**(39): p. 30487-95.

74. Lentz, M.R., et al., *A phosphorylation map of the bovine papillomavirus E1 helicase.* Virol J, 2006. **3**: p. 13.

75. Mechali, F., et al., *Bovine papillomavirus replicative helicase E1 is a target of the ubiquitin ligase APC.* J Virol, 2004. **78**(5): p. 2615-9.

76. Rangasamy, D., et al., *SUMO-1 modification of bovine papillomavirus E1 protein is required for intranuclear accumulation.* J Biol Chem, 2000. **275**(48): p. 37999-8004.

77. Lin, B.Y., et al., *Chaperone proteins abrogate inhibition of the human papillomavirus (HPV) E1 replicative helicase by the HPV E2 protein.* Mol Cell Biol, 2002. **22**(18): p. 6592-604.

78. Sedman, J. and A. Stenlund, *Co-operative interaction between the initiator E1 and the transcriptional activator E2 is required for replicator specific DNA replication of bovine papillomavirus in vivo and in vitro.* Embo J, 1995. **14**(24): p. 6218-28.

79. Seo, Y.S., et al., *Bovine papilloma virus (BPV)-encoded E1 protein contains multiple activities required for BPV DNA replication.* Proc Natl Acad Sci U S A, 1993. **90**(2): p. 702-6.

80. Blitz, I.L. and L.A. Laimins, *The 68-kilodalton E1 protein of bovine papillomavirus is a DNA binding phosphoprotein which associates with the E2 transcriptional activator in vitro.* J Virol, 1991. **65**(2): p. 649-56.

81. Hegde, R.S., et al., *Crystal structure at 1.7 A of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target.* Nature, 1992. **359**(6395): p. 505-12.

82. Antson, A.A., et al., *Structure of the intact transactivation domain of the human papillomavirus E2 protein.* Nature, 2000. **403**(6771): p. 805-9.

83. Harris, S.F. and M.R. Botchan, *Crystal structure of the human papillomavirus type 18 E2 activation domain.* Science, 1999. **284**(5420): p. 1673-7.

84. Lambert, P.F., et al., *Genetic assignment of multiple E2 gene products in bovine papillomavirus-transformed cells.* J Virol, 1989. **63**(7): p. 3151-4.

85. Doorbar, J., et al., *Detection of novel splicing patterns in a HPV16-containing keratinocyte cell line.* Virology, 1990. **178**(1): p. 254-62.

86. Hubbert, N.L., et al., *Bovine papilloma virus-transformed cells contain multiple E2 proteins.* Proc Natl Acad Sci U S A, 1988. **85**(16): p. 5864-8.

87. Choe, J., et al., *Bovine papillomavirus type 1 encodes two forms of a transcriptional repressor: structural and functional analysis of new viral cDNAs.* J Virol, 1989. **63**(4): p. 1743-55.

88. Lambert, P.F., B.A. Spalholz, and P.M. Howley, *A transcriptional repressor encoded by BPV-1 shares a common carboxy-terminal domain with the E2 transactivator.* Cell, 1987. **50**(1): p. 69-78.

89. Stubenrauch, F., et al., *The E8E2C protein, a negative regulator of viral transcription and replication, is required for extrachromosomal maintenance of human papillomavirus type 31 in keratinocytes.* J Virol, 2000. **74**(3): p. 1178-86.

90. Chiang, C.M., T.R. Broker, and L.T. Chow, *An E1M--E2C fusion protein encoded by human papillomavirus type 11 is asequence-specific transcription repressor.* J Virol, 1991. **65**(6): p. 3317-29.

91. Barsoum, J., et al., *Mechanism of action of the papillomavirus E2 repressor: repression in the absence of DNA binding.* J Virol, 1992. **66**(6): p. 3941-5.

92. Lim, D.A., et al., *Competition for DNA binding sites between the short and long forms of E2 dimers underlies repression in bovine papillomavirus type 1 DNA replication control.* J Virol, 1998. **72**(3): p. 1931-40.

93. Kurg, R., et al., *Bovine papillomavirus type 1 E2 protein heterodimer is functional in papillomavirus DNA replication in vivo.* Virology, 2009. **19**: p. 19.

94. Kurg, R., et al., *Characterization of the functional activities of the bovine papillomavirus type 1 E2 protein single-chain heterodimers.* J Virol, 2006. **80**(22): p. 11218-25.

95. Bonne-Andrea, C., S. Santucci, and P. Clertant, *Bovine papillomavirus E1 protein can, by itself, efficiently drive multiple rounds of DNA synthesis in vitro.* J Virol, 1995. **69**(5): p. 3201-5.

96. Bonne-Andrea, C., et al., *Bovine papillomavirus type 1 DNA replication: the transcriptional activator E2 acts in vitro as a specificity factor.* J Virol, 1997. **71**(9): p. 6805-15.

97. Sedman, T., J. Sedman, and A. Stenlund, *Binding of the E1 and E2 proteins to the origin of replication of bovine papillomavirus.* J Virol, 1997. **71**(4): p. 2887-96.

98. Lusky, M., J. Hurwitz, and Y.S. Seo, *The bovine papillomavirus E2 protein modulates the assembly of but is not stably maintained in a replication-competent multimeric E1-replication origin complex.* Proc Natl Acad Sci U S A, 1994. **91**(19): p. 8895-9.

99. Chen, G. and A. Stenlund, *Two patches of amino acids on the E2 DNA binding domain define the surface for interaction with E1.* J Virol, 2000. **74**(3): p. 1506-12.

100. Gillitzer, E., G. Chen, and A. Stenlund, *Separate domains in E1 and E2 proteins serve architectural and productive roles for cooperative DNA binding.* Embo J, 2000. **19**(12): p. 3069-79.

101. Berg, M. and A. Stenlund, *Functional interactions between papillomavirus E1 and E2 proteins.* J Virol, 1997. **71**(5): p. 3853-63.

102. Stenlund, A., *E1 initiator DNA binding specificity is unmasked by selective inhibition of non-specific DNA binding.* Embo J, 2003. **22**(4): p. 954-63.

103. Abbate, E.A., J.M. Berger, and M.R. Botchan, *The X-ray structure of the papillomavirus helicase in complex with its molecular matchmaker E2.* Genes Dev, 2004. **18**(16): p. 1981-96.

104. Sanders, C.M. and A. Stenlund, *Recruitment and loading of the E1 initiator protein: an ATP-dependent process catalysed by a transcription factor.* Embo J, 1998. **17**(23): p. 7044-55.

105. Chen, G. and A. Stenlund, *Characterization of the DNA-binding domain of the bovine papillomavirus replication initiator E1.* J Virol, 1998. **72**(4): p. 2567-76.

106. Chen, G. and A. Stenlund, *Sequential and ordered assembly of E1 initiator complexes on the papillomavirus origin of DNA replication generates progressive structural changes related to melting.* Mol Cell Biol, 2002. **22**(21): p. 7712-20.

107. Gillette, T.G., M. Lusky, and J.A. Borowiec, *Induction of structural changes in the bovine papillomavirus type 1 origin of replication by the viral E1 and E2 proteins.* Proc Natl Acad Sci U S A, 1994. **91**(19): p. 8846-50.

108. Borowiec, J.A., F.B. Dean, and J. Hurwitz, *Differential induction of structural changes in the simian virus 40 origin of replication by T antigen.* J Virol, 1991. **65**(3): p. 1228-35.

109. Shen, J., et al., *The roles of the residues on the channel beta-hairpin and loop structures of simian virus 40 hexameric helicase.* Proc Natl Acad Sci U S A, 2005. **102**(32): p. 11248-53.

110. Sedman, J. and A. Stenlund, *The papillomavirus E1 protein forms a DNA-dependent hexameric complex with ATPase and DNA helicase activities.* J Virol, 1998. **72**(8): p. 6893-7.

111. Dean, F.B., et al., *Simian virus 40 (SV40) DNA replication: SV40 large T antigen unwinds DNA containing the SV40 origin of replication.* Proc Natl Acad Sci U S A, 1987. **84**(1): p. 16-20.

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

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

114. Sanders, C.M. and A. Stenlund, *Transcription factor-dependent loading of the E1 initiator reveals modular assembly of the papillomavirus origin melting complex.* J Biol Chem, 2000. **275**(5): p. 3522-34.

115. Liu, X., S. Schuck, and A. Stenlund, *Adjacent residues in the E1 initiator beta-hairpin define different roles of the beta-hairpin in Ori melting, helicase loading, and helicase activity.* Mol Cell, 2007. **25**(6): p. 825-37.

116. von Hippel, P.H., *From "simple" DNA-protein interactions to the macromolecular machines of gene expression.* Annu Rev Biophys Biomol Struct, 2007. **36**: p. 79-105.

117. Sanders, C.M., *A DNA-binding activity in BPV initiator protein E1 required for melting duplex ori DNA but not processive helicase activity initiated on partially single-stranded DNA.* Nucleic Acids Res, 2008. **36**(6): p. 1891-9.

118. Castella, S., G. Bingham, and C.M. Sanders, *Common determinants in DNA melting and helicase-catalysed DNA unwinding by papillomavirus replication protein E1.* Nucleic Acids Res, 2006. **34**(10): p. 3008-19.

119. Iyer, L.M., et al., *Evolutionary history and higher order classification of AAA+ ATPases.* J Struct Biol, 2004. **146**(1-2): p. 11-31.

120. Gorbalenya, A.E., E.V. Koonin, and Y.I. Wolf, *A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses.* FEBS Lett, 1990. **262**(1): p. 145-8.

121. Hanson, P.I. and S.W. Whiteheart, *AAA+ proteins: have engine, will work.* Nat Rev Mol Cell Biol, 2005. **6**(7): p. 519-29.

122. Hickman, A.B. and F. Dyda, *Binding and unwinding: SF3 viral helicases.* Curr Opin Struct Biol, 2005. **15**(1): p. 77-85.

123. Sun, S., et al., *Identification of a 68-kilodalton nuclear ATP-binding phosphoprotein encoded by bovine papillomavirus type 1.* J Virol, 1990. **64**(10): p. 5093-105.

124. Titolo, S., et al., *Identification of domains of the human papillomavirus type 11 E1 helicase involved in oligomerization and binding to the viral origin.* J Virol, 2000. **74**(16): p. 7349-61.

125. Titolo, S., et al., *Role of the ATP-binding domain of the human papillomavirus type 11 E1 helicase in E2-dependent binding to the origin.* J Virol, 1999. **73**(7): p. 5282-93.

126. Yang, L., et al., *Activation of BPV-1 replication in vitro by the transcription factor E2.* Nature, 1991. **353**(6345): p. 628-32.

127. Titolo, S., et al., *Characterization of the minimal DNA binding domain of the human papillomavirus e1 helicase: fluorescence anisotropy studies and characterization of a dimerization-defective mutant protein.* J Virol, 2003. **77**(9): p. 5178-91.

128. James, J.A., et al., *Structure of adeno-associated virus type 2 Rep40-ADP complex: insight into nucleotide recognition and catalysis by superfamily 3 helicases.* Proc Natl Acad Sci U S A, 2004. **101**(34): p. 12455-60.

129. James, J.A., et al., *Crystal structure of the SF3 helicase from adeno-associated virus type 2.* Structure, 2003. **11**(8): p. 1025-35.

130. Seo, Y.S., et al., *Bovine papilloma virus (BPV)-encoded E2 protein enhances binding of E1 protein to the BPV replication origin.* Proc Natl Acad Sci U S A, 1993. **90**(7): p. 2865-9.

131. Yang, L., et al., *Transcription factor E2 regulates BPV-1 DNA replication in vitro by direct protein-protein interaction.* Cold Spring Harb Symp Quant Biol, 1991. **56**: p. 335-46.

132. Fletcher, R.J., et al., *The structure and function of MCM from archaeal M. Thermoautotrophicum.* Nat Struct Biol, 2003. **10**(3): p. 160-7.

133. Egelman, E.H., et al., *Bacteriophage T7 helicase/primase proteins form rings around single-stranded DNA that suggest a general structure for hexameric helicases.* Proc Natl Acad Sci U S A, 1995. **92**(9): p. 3869-73.

134. Jezewska, M.J., S. Rajendran, and W. Bujalowski, *Complex of Escherichia coli primary replicative helicase DnaB protein with a replication fork: recognition and structure.* Biochemistry, 1998. **37**(9): p. 3116-36.

135. Kaplan, D.L. and M. O'Donnell, *DnaB drives DNA branch migration and dislodges proteins while encircling two DNA strands.* Mol Cell, 2002. **10**(3): p. 647-57.

136. Yu, X., et al., *DNA is bound within the central hole to one or two of the six subunits of the T7 DNA helicase.* Nat Struct Biol, 1996. **3**(9): p. 740-3.

137. Lorimer, H.E., E.H. Wang, and C. Prives, *The DNA-binding properties of polyomavirus large T antigen are altered by ATP and other nucleotides.* J Virol, 1991. **65**(2): p. 687-99.

138. Marahrens, Y. and B. Stillman, *A yeast chromosomal origin of DNA replication defined by multiple functional elements.* Science, 1992. **255**(5046): p. 817-23.

139. Segurado, M., A. de Luis, and F. Antequera, *Genome-wide distribution of DNA replication origins at A+T-rich islands in Schizosaccharomyces pombe.* EMBO Rep, 2003. **4**(11): p. 1048-53.

140. Dai, J., R.Y. Chuang, and T.J. Kelly, *DNA replication origins in the Schizosaccharomyces pombe genome.* Proc Natl Acad Sci U S A, 2005. **102**(2): p. 337-42.

141. Cvetic, C. and J.C. Walter, *Eukaryotic origins of DNA replication: could you please be more specific?* Semin Cell Dev Biol, 2005. **16**(3): p. 343-53.

142. Greenleaf, W.B., et al., *Systematic study of the functions for the residues around the nucleotide pocket in simian virus 40 AAA+ hexameric helicase.* J Virol, 2008. **82**(12): p. 6017-23.

143. Wessel, R., J. Schweizer, and H. Stahl, *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, 1992. **66**(2): p. 804-15.

144. Bujalowski, W., M.M. Klonowska, and M.J. Jezewska, *Oligomeric structure of Escherichia coli primary replicative helicase DnaB protein.* J Biol Chem, 1994. **269**(50): p. 31350-8.

145. San Martin, M.C., et al., *A structural model for the Escherichia coli DnaB helicase based on electron microscopy data.* J Struct Biol, 1995. **114**(3): p. 167-76.

146. Mott, M.L., et al., *Structural synergy and molecular crosstalk between bacterial helicase loaders and replication initiators.* Cell, 2008. **135**(4): p. 623-34.

147. Speck, C., et al., *ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA.* Nat Struct Mol Biol, 2005. **12**(11): p. 965-71.

148. Siddiqui, K. and B. Stillman, *ATP-dependent assembly of the human origin recognition complex.* J Biol Chem, 2007. **282**(44): p. 32370-83.

149. Randell, J.C., et al., *Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase.* Mol Cell, 2006. **21**(1): p. 29-39.

150. Bowers, J.L., et al., *ATP hydrolysis by ORC catalyzes reiterative Mcm2-7 assembly at a defined origin of replication.* Mol Cell, 2004. **16**(6): p. 967-78.