

Rowan University

## Rowan Digital Works

---

Graduate School of Biomedical Sciences  
Theses and Dissertations

Rowan-Virtua Graduate School of Biomedical  
Sciences

---

8-2017

# Molecular Mechanisms of DNA Replication Initiation in HPVs with Genetic Variations Leading to Cellular Carcinogenesis

Gulden Yilmaz  
*Rowan University*

Follow this and additional works at: [https://rdw.rowan.edu/gsbs\\_etd](https://rdw.rowan.edu/gsbs_etd)



Part of the [Cancer Biology Commons](#), [Cell Biology Commons](#), [Cellular and Molecular Physiology Commons](#), [Laboratory and Basic Science Research Commons](#), [Molecular Biology Commons](#), [Molecular Genetics Commons](#), [Neoplasms Commons](#), [Virology Commons](#), and the [Virus Diseases Commons](#)

---

### Recommended Citation

Yilmaz, Gulden, "Molecular Mechanisms of DNA Replication Initiation in HPVs with Genetic Variations Leading to Cellular Carcinogenesis" (2017). *Graduate School of Biomedical Sciences Theses and Dissertations*. 18.

[https://rdw.rowan.edu/gsbs\\_etd/18](https://rdw.rowan.edu/gsbs_etd/18)

This Dissertation is brought to you for free and open access by the Rowan-Virtua Graduate School of Biomedical Sciences at Rowan Digital Works. It has been accepted for inclusion in Graduate School of Biomedical Sciences Theses and Dissertations by an authorized administrator of Rowan Digital Works.

**MOLECULAR MECHANISMS OF DNA REPLICATION  
INITIATION IN HPV<sub>s</sub> WITH GENETIC VARIATIONS LEADING  
TO CELLULAR CARCINOGENESIS**

Gulden Yilmaz, B.S.

A Dissertation submitted to the Graduate School of Biomedical Sciences, Rowan  
University in partial fulfillment of the requirements of the Ph.D. Degree.

Stratford, NJ 08084

August 2017

## TABLE OF CONTENTS

	PAGE
TABLE OF CONTENTS	2
ACKNOWLEDGEMENTS	7
ABSTRACT	8
INTRODUCTION	10
RATIONALE	17
MATERIALS AND METHODS	20
EXPERIMENTAL RESULTS	33
SECTION I: Mechanism of interaction of the Human Papillomavirus E2 initiator protein with the origin of DNA replication	33
A. Purification of HPV-11 E2 protein and its structure in solution	33
1. Purification of full-length low-risk protein	33
2. Oligomeric structure of purified E2 protein in solution	34
B. E2 protein DNA binding preference	36
1. E2 activity analysis with specific and non-specific DNA	36
2. E2 DNA binding preference with double- and single-stranded DNA	36
C. E2 binds differentially to its DNA binding sites	38
1. Analysis of E2 with cognate E2 binding sites	38

2. Challenge assay with E2 and E2 binding sites	41
D. The composition and length of the spacer affected E2 binding affinity	43
1. Analysis of spacer length on E2 binding activity	43
2. Analysis of spacer sequence on E2 binding affinity	43
E. Effect of flanking sequences of the E2 binding sites on E2 binding affinity	46
F. Cooperativity and higher order complex formation with multiple binding sites	49
G. Modulation of DNA binding ability of E2 by ionic strength and temperature	53
1. Physical nature of E2-DNA interaction	53
2. Thermodynamic analysis of E2-DNA binding	55
<b>SECTION II: Mutations in the Human Papillomavirus replication origin may mediate viral oncogenesis</b>	<b>58</b>
A. Purification and analysis of full-length recombinant E2 from HPV-16	58
1. Purification of full-length high-risk E2 protein	58
2. Analysis of specific DNA binding activity of E2	60
3. Analysis of high-risk E2 with cognate E2 binding sites	62
B. Nucleotide variation in the binding site consensus sequence leads to decreased E2 binding affinity	64
1. Analysis of E2 binding to HPV-16 SNV	64
2. Analysis of E2 binding to other high-risk SNVs	66
C. Multi-protein E2-DNA complex formation with DNA containing multiple E2 binding sites	69

<b>D. E2 binding site analysis identifies common variant among high-risk genotypes</b>	<b>73</b>
1. Bioinformatics analysis of low-risk E2 binding sites	73
2. Bioinformatics analysis of high-risk E2 binding sites	73
3. Bioinformatics analysis of probable high-risk E2 binding sites	74
<b>E. Low-risk E2 protein cannot bind E2 binding site containing high-risk SNV</b>	<b>77</b>
<b>F. Structural analysis of E2 on HPV origin by Atomic Force Microscopy (AFM)</b>	<b>80</b>
1. Design of low-risk and high-risk origin constructs for AFM	80
2. Structural analysis of low-risk E2 protein on low-risk HPV origin	81
3. Structural analysis of high-risk E2 protein on high-risk HPV origin	82
<b>SECTION III: Characterization of E1 helicase and E2 interaction at the origin</b>	<b>85</b>
A. Cloning of E1 helicase	85
B. Expression of E1 helicase	86
C. Purification of E1 helicase	86
D. Analysis of E1 helicase activity	90
E. E2 had no effect on E1 DNA helicase activity	92
F. Characterization of E1-E2-DNA complex formation	94
1. E1 DNA helicase bound to E2-DNA complex but not free double-stranded DNA	94
2. E1-E2-DNA complex formation does not require E1 binding site	97

3. E1-E2-DNA complex formation requires at least two E2 binding sites	97
G. ssDNA had no effect on E1-E2-DNA complex formation	101
<b>DISCUSSION</b>	<b>103</b>
<b>SECTION I: Structural and functional studies of the full-length low-risk HPV E2 protein</b>	<b>103</b>
A. Purification and oligomeric structures of E2	103
B. Mechanism of E2-DNA interaction in the HPV origin	105
C. Effect of spacer and flanking regions on E2 recognition	106
D. Cooperativity among E2 binding sites	107
E. Ionic and hydrophobic interactions in E2-DNA complex formation	109
<b>SECTION II: Identification of SNV in high-risk HPV origin and its effects on high-risk E2 function</b>	<b>110</b>
A. Hierarchy of high-risk E2 binding to its cognate E2 binding sites	110
B. Effect of SNV on E2 DNA binding affinity	111
C. Bioinformatics analysis of E2 binding sites	112
D. Effect of E2 multimer formation on the origin with one or more SNVs	113
E. Low-risk E2 protein cannot bind E2 binding site containing high-risk SNV	114
F. Structural analysis of E2 on HPV origin by Atomic Force Microscopy (AFM)	115

<b>G. A putative model for HPV E2 DNA binding and activation of replication at the low-risk and high-risk origins</b>	<b>116</b>
<b>SECTION III: HPV E1 helicase activity and E2 interaction</b>	<b>121</b>
<b>A. Purification of HPV E1 DNA helicase</b>	<b>121</b>
<b>B. DNA unwinding activity of E1 helicase</b>	<b>122</b>
<b>C. Effect of E2 on E1 helicase activity</b>	<b>122</b>
<b>D. Characterization of E1-E2-DNA complex formation</b>	<b>123</b>
<b>1. E1 DNA helicase bound to E2-DNA complex but not free double-stranded DNA</b>	<b>123</b>
<b>2. E1-E2-DNA complex formation does not require E1 binding site</b>	<b>124</b>
<b>3. E1-E2-DNA complex formation requires at least two E2 binding sites</b>	<b>124</b>
<b>E. ssDNA had no effect on E1-E2-DNA complex formation</b>	<b>125</b>
<b>F. Putative model of E1-E2 interaction at the origin for initiation of HPV replication</b>	<b>126</b>
<b>SUMMARY AND CONCLUSIONS</b>	<b>129</b>
<b>REFERENCES</b>	<b>131</b>
<b>ABBREVIATIONS</b>	<b>143</b>
<b>ATTRIBUTES</b>	<b>145</b>
<b>LIST OF FIGURES</b>	<b>146</b>
<b>LIST OF TABLES</b>	<b>149</b>

## **ACKNOWLEDGEMENTS**

I would like to convey my utmost appreciation to Dr. Subhasis Biswas, my thesis advisor, for his endless guidance, motivation, and patience throughout the course of my studies. I wish to express my infinite gratitude to Dr. Esther Biswas-Fiss for her encouragement, support and direction throughout my research studies. I wish to thank Drs. Salvatore Caradonna, Jennifer Fischer, and Susan Muller-Weeks for serving as my thesis committee. I am thankful to my colleagues for insightful discussions, especially Ms. Meera J. Patel and Mr. Lavesh Bhatia. I am very grateful for all the help from the support staff of the Graduate School and Departments of Cell and Molecular Biology during the course of my Ph.D. program. Last but not least, this work would not have been possible without the assistance and encouragement of my incredible family. I would like to thank my remarkable husband, Yaşar Yılmaz, whom has always inspired and supported my ambitious goals. My brilliant sisters, Gülüstan, Gülcan, and Gülhatun, have always been a limitless source of guidance, motivation, and positivity. This thesis is dedicated to my outstanding parents Mrs. Perihan Kaplan and Mr. (late) Ali Kaplan, whom are my leading inspirations to achieve success in endeavors throughout my lifetime.

## **ABSTRACT**

Human papillomaviruses are a vast family of double-stranded DNA viruses containing non-carcinogenic and carcinogenic types, whose crucial differences remain unknown, except for the difference in the frequency of DNA replication. The human papillomavirus (HPV) E2 protein regulates the initiation of viral DNA replication and transcription. Its recognition and binding to four 12 bp palindromic sequences in the viral origin is essential for its function. Little is known about the DNA binding mechanism of the E2 protein found in HPV types that have low risk for oncogenicity (low-risk) as well as the roles of various elements of the individual binding sites. The binding sites in the origin of all HPVs are separated by variable spacer and flanking regions; however, their importance in E2-DNA recognition and regulation remains unclear. Analysis of low-risk E2 DNA binding affinity unraveled multiple sequence elements that appeared to influence target discrimination including the sequence of spacer region, flanking sequences, and proximity of E2 binding sites. The results for the low-risk E2 were compared to those of the E2 encoded by HPVs that have high risk for oncogenicity (high-risk). Single nucleotide variations in the high-risk E2 binding sites were identified by sequence analysis of carcinogenic and non-carcinogenic HPVs. Sequence variations in binding sites of carcinogenic HPVs were correlated to attenuated formation of the E2-DNA complex. Differences in E2 binding in the origin of carcinogenic and non-carcinogenic HPVs were observed. Further analysis of viral replication initiation included the assessment of E2 in the presence of the HPV E1 helicase, also required in this process. The fundamental biochemical properties of the E1-E2-DNA complex formation were evaluated. E1

was localized to the DNA by E2 bound to the DNA, through E1-E2 protein interactions. The data presented here suggests the mechanism of E1-E2 interaction at the origin in HPV DNA replication initiation. E2 binding to the origin is tightly linked to the activation of the DNA replication origin as well as initiation of DNA replication. Together, these results allow us to elucidate a model for replication initiation in carcinogenic and non-carcinogenic HPVs and propose a correlation to HPV-mediated cellular carcinogenesis.

## **INTRODUCTION**

Human papillomavirus (HPV) is the most common sexually transmitted infection in the United States, with 80% of all women infected by the age of 50 [1]. The largest group of HPVs consists of the Alpha papillomaviruses, which are further classified as low-risk and high-risk based on their ability to induce cancer [2]. Low-risk or non-oncogenic HPVs, such as HPV-6 and HPV-11, are correlated with benign lesions, such as low-grade cervical abnormalities, and 90% of all anogenital wart cases [3]. These HPV types commonly occur in recurrent respiratory papillomatosis (RRP), which is a life-threatening disease caused by the recurrence of warts or papillomas present in the upper respiratory tract [4]. Juvenile-onset recurrent respiratory papillomatosis (JORRP) results from the vertical transmission of HPV from mother to infant during delivery [5]. JORRP correlates with extensive morbidity and the requirement of an average of 13 lifetime surgeries to remove papillomas and open the airway [5].

Unlike the low-risk types, high-risk HPV types are correlated with various forms of cancers [6]. The International Agency for Research on Cancer classifies HPV types as non-carcinogenic, high-risk/carcinogenic, and probable high-risk/carcinogenic based on HPV type-specific prevalence in invasive cervical cancer compared to their respective frequency in normal cytology [7-9]. High-risk HPVs have been linked to 90% of cervical and anal cancers, 70% of vaginal and vulvar cancers, and 60% of penile cancers [10]. More recently, HPV has been associated with head and neck squamous cell carcinoma [11]. Since approximately 200 HPV

variants have been identified by DNA sequence analysis, it is not possible to protect against all HPV strains simply by vaccination [12, 13]. More importantly, there is no treatment to aid those individuals already infected with the virus.

Human papillomaviruses are approximately 7.9 kb circular double-stranded DNA tumor viruses that explicitly target the basal cells of the epithelial mucosa [14]. Infection by papillomaviruses requires that virus particles access the epithelial basal layer and enter the dividing basal cells [15]. Papillomavirus particles disassemble in endosomes, and capsid protein L2 transfers viral DNA to the nucleus [15]. The viral genome establishes itself as a stable episome in the basal cells [15]. The HPV genome consists of early (E) and late (L) genes, as well as a non-coding region. The late expressing open reading frames (ORFs) encode the L1 and L2 capsid proteins. The early expressing ORFs encode regulatory proteins, such as the E1, E2, E6, and E7 proteins. E6 and E7 proteins are oncoproteins and known to be involved in carcinogenesis [16]. The primary viral proteins expressed in the basal cells are E1 and E2, both of which are essential for initiation of papillomavirus DNA replication *in vivo* [17]. The virus replicates as multi-copy episomal nuclear plasmids, requiring the viral DNA replication origin. The origin-binding protein, E2, and the replicative DNA helicase, E1, are the only two virus-encoded proteins that are required for this process [18-20]. All other required replication factors, such as DNA polymerase and ssDNA-binding protein (RPA), are appropriated by the virus from the host cellular DNA replication machinery [21, 22]. E2 is vital in the activation of the viral DNA

replication origin [17, 23, 24]. Specifically, E2 is important in the preinitiation complex formation to block non-specific E1-DNA interactions [23].

During initiation of DNA replication, E2 binds to specific sequence motifs within the viral replication origin [25]. All HPV genomes contain four E2 binding sites in a short conserved spatial arrangement in the upstream regulatory region of the *E6* and *E7* oncogenes, which constitutes the origin of DNA replication [26]. E2 binding sites are numbered as binding site 1 through 4 based on their proximity to the early *E6/E7* promoter. For example, E2 binding site 1 (BS1) and binding site 2 (BS2), separated by 3-5 bp, are closest to the early promoter for *E6* and *E7* viral oncogenes [26]. The separation between BS2 and binding site 3 (BS3) is 64-73 bp, and the spacing between BS3 and E2 binding site 4 (BS4) is 279-407 bp [26]. Binding site 3 is required for optimal origin activation, in addition to binding sites 1 and 2 [20, 27]. These sites represent a conserved sequence element in all HPV types, whereas non-conserved sequences exist in other parts of the viral long control region [28]. The consensus sequence of all four binding sites is comprised of a 12 bp palindromic sequence, ACCG(N)<sub>4</sub>CGGT, where N is any nucleotide in the “spacer” region [25, 29]. The potential involvement of these sites in HPV-associated carcinogenesis remains unknown. The spatial arrangement and multiple E2 binding sites in the viral origin are highly conserved among HPVs and are likely significant in the activation of DNA replication initiation. E2 binding to these multiple sites in the origin may lead to the formation of multi-protein looped protein-DNA complexes, a process reminiscent of the *E. coli* replication initiator protein, DnaA, binding to its origin and forming a higher-order complex, during the initiation of DNA replication

as demonstrated by Kornberg and coworkers [30-33]. Incidentally, the lambda bacteriophage O protein and eukaryotic origin recognition complex (ORC) also recognize and remodel their respective origins in a similar fashion [34-36]. Early studies have utilized bovine papillomavirus (BPV) or truncated human papillomavirus E2 to characterize the role of E2 [37, 38]. However, further studies are necessary to elucidate the activity of the full-length HPV E2 protein and the HPV origin.

E2 also regulates transcription of *E6* and *E7* genes and presumably requires these four binding sites [20, 39-41]. Binding sites 1 and 2 are involved in repressing expression of viral oncogenes *E6* and *E7*. The binding of E2 at these proximal elements in the viral genome hinders the binding of cellular factors Sp1 and TFIID that are needed for RNA polymerase-induced transcription of these genes [42-45]. Binding site 4 has also been implicated in transcriptional activation [46]. Hence, E2 can control or in this case suppress oncogene transcription, depending on the context of the binding sites and associated cellular factors.

The replication of low-risk HPV types is known to be more facile in comparison to high-risk HPV types [20]. Specifically, the initial amplification of the low-risk HPV genome is drastically more effective than that of the high-risk types [47]. It has been reported that the most carcinogenic type, HPV-16, possessed the least efficient replication system of those high-risk types that were tested [47]. Hence, the molecular basis and significance of this difference requires an in-depth investigation to understand possible connection(s) to carcinogenesis.

In HPV-mediated cancers, the viral genomes are frequently found to be integrated into the host cellular genome [48-50]. The level of HPV integration is positively correlated with cervical intraepithelial neoplasia grades and has been proposed as a marker of disease progression [51-54]. The E2 ORF in HPV is commonly disrupted or deleted during its integration into the host genome [55]. Since the E2 protein is known to repress *E6* and *E7* expression, the loss of E2 during integration subsequently results in the increased expression of E6 and E7 oncoproteins [56, 57]. E6 and E7 serve to inactivate and accelerate the degradation of tumor suppressor proteins p53 and pRb, activating the neoplastic progression [58, 59].

The E2 binding sites are located adjacently to the E1-binding loci at the origin [60]. The E1 and E2 proteins function by binding to the viral origin of replication, recruiting cellular proteins that mediate viral DNA replication, and regulating viral gene expression [61]. E1 is a member of the superfamily 3 of NTP-binding proteins [62]. The E1 protein consists of a DNA-binding domain (DBD) and an ATPase/helicase domain [63]. An 18 bp inverted-repeat element within the origin is referred to as the E1 binding site [64]. The C-terminal domain of the protein includes the ATP-binding domain and is necessary for oligomerization into hexamers for functional ATPase and helicase activities [63, 65]. This E1 domain is also involved in interaction with the host polymerase alpha primase, as well interaction with the transactivation domain of E2, in a mutually exclusive manner [21, 66]. E1 is involved in many steps of replication initiation, such as DNA melting, and subsequent unwinding of DNA [67].

Most of our current knowledge of the mechanism of activation by E1 is based on studies of BPV or truncated HPV E1 polypeptides [68]. Although BPV and HPV E1 proteins may possess similar functions, significant protein sequence differences exist, suggesting potential characteristic distinctions [69]. For example, the HPV E1 region corresponding to the minimal DNA-binding domain of BPV E1 was not sufficient for binding to the origin [63, 70]. These studies reported that HPV E1 also required the C-terminal oligomerization/enzymatic domain to interact with the origin [63, 70]. HPV E1 exhibits less sequence specificity than BPV E1 [71]. Unlike HPV E1, BPV E1 and E2 proteins interact using DNA binding domains of both proteins [72]. Therefore, it is apparent that BPV E1 properties are not directly comparable to HPV E1. Further studies are required to determine HPV E1 activities and its interaction with E2 and the origin of viral DNA replication.

E1 is the only viral enzyme produced by HPV, making it an essential target for anti-viral drug discovery. The key role of E1 in HPV replication, its high degree of conservation among various HPV types, and its enzymatic activity are attributes which make E1 an attractive target for therapeutic intervention. Inhibition of E1 would provide an opportunity to inhibit the early stages of the viral life cycle. The high degree of conservation of E1 would permit a ubiquitous drug, targeting several HPV types. However, the major hindrance of this development has been expression and purification of the full-length HPV E1 protein, which could be used to develop high throughput screens of potential therapeutic agents.

In the study presented in this thesis, the expression and purification of sufficient quantities of E1 and E2 protein has enabled the characterization of the activities of these proteins. In order to understand the relationship between HPV types and the development of cancer, a systematic functional analysis of full-length E2 proteins from both low-risk and high-risk genotypes was undertaken to identify the mechanisms involved in the origin recognition in each type. A comprehensive analysis is presented to elucidate the structural features and functional properties of low-risk E2 protein and its mechanism of DNA binding and target discrimination. The mechanism of interaction of the high-risk E2 protein with its cognate binding sites was also presented and the results were contrasted with those of the low-risk E2 protein. A combination of biochemical analysis of E2-DNA binding and bioinformatics analysis of benign, high-risk, and probable high-risk HPV DNA sequences was conducted. The results are depicted using a model of origin activation of both low-risk and high-risk HPVs. This study has further analyzed HPV replication initiation by E2 in the presence of the E1 helicase with the HPV origin. The binding of E2 dimers at the E2 binding sites located in the origin is required for E1 interaction. More specifically, the binding of at least two E2 binding sites are required for the recruitment of the E1 helicase. Together, these data suggested that E1 does not directly bind the DNA, but instead is bound to the E2 proteins in the E2-DNA complex. The data presented here suggests an essential E1-E2 interaction at the origin in HPV DNA replication initiation. In summary, E2 first binds to multiple sites in the origin, which then attracts E1 to join the E2-DNA complex, creating a larger E1-E2-DNA complex for initiation of viral replication.

## RATIONALE

Persistent human papillomavirus (HPV) infection is a known risk factor for several epithelial squamous cell carcinomas. Despite the high incidence of HPV infection and its associated malignancies, there are currently no effective anti-viral agents available for therapy. HPV is classified into high-risk and low-risk types; low-risk types are correlated with benign lesions such as genital warts and laryngeal papillomatosis, whereas high-risk types are associated with the development of cancer. Two important virally encoded proteins expressed in basal cells are E1 and E2, which are both essential for initiation of papillomavirus DNA replication *in vivo*.

The E2 protein, an origin-binding protein, is required for the initiation of episomal HPV replication. In addition, the disruption of the viral *E2* gene sequence occurs during genomic integration of HPV DNA. The disruption of the E2 open reading frame leads to the release of HPV *E6* and *E7* oncogene from E2-mediated repression, overexpression of oncoproteins E6 and E7, and suppression of key tumor suppressor pathways.

Recognition of the four E2 DNA binding site sequences is the first step in regulation of the virus by E2. Although individual binding sites are associated with separate E2 processes, the mechanism by which E2 distinguishes between its four binding sites remains poorly understood. Early studies have utilized BPV-1 or truncated human papillomavirus E2 consisting of the 84 amino acid DNA-binding domain to characterize the role of E2 [37, 38]. However, the full-length protein (367

amino acids) may function differently than its partial or truncated counterparts. Differences in E2-DNA binding affinities were reported when analyzing native and truncated (DNA-binding domain only) BPV-1 E2 proteins [73]. This same study found that regions extrinsic to the E2 DNA-binding domain are necessary for optimal cooperative DNA binding [73]. Thus, the E2 polypeptide sequences outside of the defined E2 DNA-binding domain may influence DNA binding. The central question remains of how E2 confers multi-functional viral regulation through these conserved binding sites. A comprehensive study analyzing E2 in the context of its cognate binding sites, and potential other factors, may contribute to understanding how this virus may be controlled.

The E1 protein is involved in many stages of DNA replication initiation, such as DNA melting, and unwinding of DNA during elongation phase. However, most of our current knowledge of the mechanism of action of E1 is based on studies of full-length BPV or truncated HPV E1 polypeptides [68]. BPV and HPV E1 proteins possess significant protein sequence differences, suggesting potential functional differences [69]. Unlike BPV E1, HPV E1 exhibits little sequence specificity [71] and E1 and E2 do not interact using their respective DNA binding domains [72]. Therefore, it appears that BPV E1 and HPV E1 may possess diverse activities. Truncated forms of HPV-11 E1 that were evaluated did not support efficient viral DNA replication [63, 74]. Further studies are required to determine the full-length HPV E1 activities and its interaction with E2 at the HPV origin of replication.

HPV encompasses a large family of viruses that range from benign to highly carcinogenic, associated with genital, oropharyngeal and a variety of other cancer forms [6, 10, 11]. Sequences of HPV viral genomes, regardless of the carcinogenicity, are highly homologous. Consequently, the ability of some HPV types to cause carcinogenic transformation of infected cells remains a paradox. Earlier studies indicate that the benign and carcinogenic forms of HPV differ in DNA replication efficiency of the episomal viral DNA [20, 47]. However, the basis of this difference also remains unknown.

In summary, this thesis addressed the core hypothesis that altered viral DNA replication facilitates the pathology of carcinogenesis. The goal was to understand how high-risk and low-risk HPV genotypes affect the biochemistry of DNA replication. It was of importance to produce mechanistic insights into the structural and functional roles of the low- and high-risk HPV proteins involved in DNA replication. More specifically, the aim was to comprehend how high- and low-risk HPV genotypes influence E2 function. The correlation between HPV DNA replication, HPV genotype, and associated risk of carcinogenesis was examined. A second aim of this research was to characterize the interaction of E1 and E2 at the origin. To accomplish this, a combination of biochemical and biophysical *in vitro* analyses of HPV E2 and E1 recombinant proteins were used. A detailed understanding of the structure and function of these proteins provided insight into HPV-related diseases and the molecular mechanisms involved in the development of targeted therapies.

## **MATERIALS & METHODS**

### **1. Reagents**

ACS reagent grade chemicals from Fischer Chemical Company (Pittsburgh, PA) were used to prepare all buffers and solutions. Synthetic oligonucleotides were obtained from Sigma Genosys Inc. (Woodlands, TX) and sequences are as shown in Tables 1-7. All recombinant plasmids were sequenced completely at the DNA sequencing facility of Eurofins MWG Operon LLC (Huntsville, AL). Protein sequencing and analysis was performed by the University of Florida (Gainesville, FL) Department of Chemistry Mass spectrometry services.

### **2. Buffers**

Buffer A contained 25 mM Tris-HCl (pH 8.0), 0.1 M NaCl, and 10% sucrose. Buffer B comprised of 10 mM imidazole, 20 mM Tris-HCl (pH 8.0), 500 mM NaCl, and 2 mM DTT. Buffer C consisted of 25 mM imidazole, 20 mM Tris-HCl (pH 8.0), 500 mM NaCl, and 2 mM DTT. Buffer D contained 50 mM imidazole, 20 mM Tris-HCl (pH 8.0), 500 mM NaCl, and 2 mM DTT. Buffer E was 200 mM imidazole, 20 mM Tris-HCl (pH 8.0), 500 mM NaCl, and 2 mM DTT.

### **3. Cloning, expression and purification of low-risk HPV-11 E2**

The HPV-11 *E2* gene was amplified from purified plasmid HPV-11 DNA, GenBank number M14119 (ATCC, Rockland, MA), using primers referred to as “Primer 1” and “Primer 2” (Table 1). The PCR product was cloned into a T7 expression vector, pET28a, engineered to have an N-terminal histidine tag (EMD

Millipore, Billerica, MA), using standard recombinant DNA technology [75]. The validity of cloned *E2* gene, in the pET28a-E2 plasmid, was verified by DNA sequencing. *E. coli* Lemo21(DE3) strain (New England Biolabs, Ipswich, MA) was transformed with the pET28a-E2 plasmid for subsequent expression of the protein. *E. coli* cells were grown with 1 mM L-rhamnose and shaking at 37°C to OD<sub>600</sub> of 0.4. L-rhamnose was added to control the tunable expression system in Lemo21(DE3) cells, further discussed in the results section. To promote expression of the recombinant protein, isopropyl-β-D-thiogalactopyranoside (IPTG) was then added to a final concentration of 0.4 mM and incubation at 37°C was continued for 15 hours. The cells were harvested by centrifugation and resuspended in buffer A containing 1 μg/ml of each protease inhibitor including pepstatin, leupeptin, chymostatin, and aprotinin. Induced cells were extracted with the addition of lysozyme following standard protocol [76]. The extract was loaded onto cOmplete His-tag purification resin (Roche, Pleasanton, CA), equilibrated with buffer B. The column was washed thoroughly with buffer C, followed by a brief wash with buffer D. The protein was eluted using buffer E containing 10% glycerol. Fractions containing E2 were identified by SDS-PAGE. These fractions were pooled and subjected to reverse phase chromatography. The protein eluted using a reverse gradient from high salt (2 M NaCl) to lower salt (0.5 M NaCl). E2 eluted near 0.5 M NaCl concentration in the fully soluble form. Protein homogeneity was determined by SDS-PAGE (Figure 2a) and verified by mass spectrometry conducted at University of Florida (Gainesville, FL) Department of Chemistry Mass spectrometry services.

#### **4. Cloning, expression and purification of high-risk HPV-16 E2**

The HPV-16 *E2* gene was amplified from purified plasmid HPV-16 DNA, GenBank number NC\_001526 (Addgene, Inc., Cambridge, MA), using primers referred to as “Primer 3” and “Primer 4” (Table 1). Using standard recombinant DNA technology, the PCR product was cloned into a T7 expression vector, pET28a (EMD Millipore, Billerica, MA), and contained a His<sub>6</sub> purification tag at the N-terminus. The *E2* gene present in the pET28a-E2 plasmid was fully sequenced prior to further studies. Overexpression of the protein in the *E. coli* cells was performed as mentioned above for low-risk E2.

Purification of E2 was performed by His-tag affinity chromatography as described above for low-risk E2. These fractions were pooled and subjected to hydroxyapatite chromatography. The protein eluted using a linear gradient from low phosphate (10 mM KPO<sub>4</sub>, pH 7.5, 500 mM NaCl, 10% (v/v) glycerol, 2 mM DTT) to high phosphate (75 mM KPO<sub>4</sub>, pH 7.5, 500 mM NaCl, 10% (v/v) glycerol, 2 mM DTT). Protein homogeneity was determined by SDS-PAGE (Figure10a).

#### **5. Mass spectrometry**

Purified HPV-11 and HPV-16 E2 proteins (2 µg each) were loaded separately onto a 10% polyacrylamide gel. Following electrophoresis, the protein gel was fixed for 15 minutes to prevent the diffusion of proteins. Fixation was followed by a rinse cycle and subsequently stained with Coomassie Blue. The SDS-PAGE gel was destained with fixing solution. The clearly visible band was extracted from the rest of

the gel. Protein sequencing and analysis was performed by the University of Florida (Gainesville, FL) Department of Chemistry Mass spectrometry services.

#### **6. Dynamic Light Scattering of purified E2 protein**

Dynamic Light Scattering experiments were performed using DynaPro NanoStar instrument and DYNAMICS 7.1 software (Wyatt Technology, Goleta, CA). Sample parameters were set, including solvent as 10% glycerol. Mw-R model used was “branched polymers”. The measurement time period was 400 seconds, consisting of 20 acquisitions, at 20 seconds each. Three separate measurements were performed.

#### **7. Full length protein structure modeling**

Structure of the full-length HPV-11 E2 protein was modeled using Robetta protein structure prediction software (<http://www.robetta.org/>) at the University of Washington, Seattle, WA. Robetta methodology uses a combination of Rosetta homology modeling software and ab initio fragment assembly with GinzU domain prediction software that allows modeling of full length protein with partial or no available crystal structure(s). This methodology was chosen for creating the three dimensional model as crystal structures of the individual domains of E2 are available but none are available for full-length E2 protein.

## **8. Electrophoretic mobility shift assay (EMSA)**

Synthetic oligonucleotides were used for end-labeling reactions with T4 polynucleotide kinase and ( $\gamma$ - $^{32}\text{P}$ ) ATP (Tables 2-7). DNA probes were purified by Bio-Gel P-6 column (Bio-Rad, Hercules, CA).

The binding reactions were carried out in a 10- $\mu\text{l}$  final volume containing the indicated amount of protein and 12.5 femtograms of  $\gamma$   $^{32}\text{P}$ -labeled DNA probe in 10 mM HEPES (pH 7.9), 4 mM  $\text{MgCl}_2$ , 10% glycerol, 5 mM dithiothreitol, 0.2 mM EDTA, 0.1 mg/ml bovine serum albumin, 10 ng/ $\mu\text{l}$  poly(dI-dC), and 50 mM NaCl. The binding reactions were allowed to proceed for 10 minutes at 30° C instead of 37°C to minimize the effect of any possible trace nuclease, DNase, or protease contaminants. The mixtures were immediately loaded on 5% polyacrylamide gel in 0.5X TBE and run at 170 V for 45 minutes at 4 °C, except for the cooperativity assay which was run for 2 hours at 170 V. Following electrophoresis, the gels were dried and exposed to X-ray films. Quantification was accomplished by PhosphorImager analysis using Typhoon 9410 Variable Mode Imager (GE Healthcare Life Sciences, Rahway, NJ). In competition experiments, reactions were performed as indicated above, with the addition of non-radio-labeled/competitor oligonucleotides in increasing amounts. The competitor DNA was added simultaneously with the labeled probe. Each experiment was quantified independently by PhosphorImager analysis. The probe (BS1+2+E1BS) used in the cooperativity EMSA was generated by PCR

amplification of the HPV-11 upstream regulatory region spanning HPV-11 genome nucleotide 7092-7933/1-92 using Primer 5 and Primer 6 (Table 1).

### **9. Calculation of equilibrium binding constants**

The free probe in each lane after EMSA was quantified using Image Quant 5.2 (GE Healthcare Life Sciences, Rahway, NJ) without (x) and with (y) an increasing amount of HPV-11 E2 protein. The % bound probe (B) was determined using Equation 1.

Equation 1:  $B=100\% (1-y/x)$

Non-linear regression analysis of the fractional occupancy (bound probe vs.  $\log ([E2])$ ) was carried out using Prism 6.0 software (Graph Pad Software Inc.) utilizing a variable slope sigmoidal dose-response model.

### **10. Cloning of HPV-11 and HPV-16 origin and Atomic Force Microscopy of E2-DNA complexes**

The HPV-11 origin (GenBank number M14119 from ATCC) could not be amplified directly from purified plasmid HPV-11 DNA, due to the interruption of the HPV-11 origin by pBR322 vector DNA. As a result, the HPV-11 DNA was excised from the vector and recombined. Briefly, the HPV-11 plasmid was digested with restriction enzyme BamH1, and separated on a 1% agarose gel. The correct DNA band was excised from the gel and purified using GeneClean Kit (MP Biomedicals, Santa Ana, CA), followed by ligation using Clonables Ligation Premix (EMD Millipore, Billerica, MA). Following ligation of the DNA, the HPV-11 origin

containing all four E2BSs was generated by PCR using Primer 7 and Primer 8 (Table 1). The primers for this substrate were designed for regions in HPV-11 DNA. The HPV-16 DNA substrate containing all four E2BSs was generated by PCR using Primer 9 and Primer 10 (Table 1) from purified plasmid HPV-16 DNA, with GenBank number NC\_001526 (Addgene, Inc., Cambridge, MA). Both HPV-11 and HPV-16 PCR products were purified using GeneJet PCR Purification Kit (Thermo Scientific, Waltham, MA). The correct DNA band from the purified HPV-11 PCR product was separated on a 0.8% agarose gel, excised from the gel, and purified using GeneClean Kit (MP Biomedicals, Santa Ana, CA). Using standard recombinant DNA technology, the PCR products were cloned into a M13KS vector (Stratagene, La Jolla, CA) for sequencing purposes. Both HPV-11 and HPV-16 origins were fully sequenced prior to further studies.

## **11. Binding reactions for AFM**

The binding reactions were carried out in a 20- $\mu$ l final volume containing 100 ng of linear DNA and 150 ng of E2 protein in 20 mM HEPES (pH 7.9) and 2 mM ATP, modified from previously published method [33]. The binding reactions were allowed to proceed for 20 minutes at 37°C and then treated with glutaraldehyde for 5 minutes at room temperature. E2-DNA complexes were brought to a final volume of 50  $\mu$ l with 20 mM HEPES (pH 7.9). These complexes were then separated by gel filtration through 0.5 ml columns of 6% crosslinked agarose beads (Agarose Bead Technologies Inc., Miami, FL) equilibrated with 20 mM HEPES (pH 7.9). Atomic Force Microscopy (AFM) imaging was performed at Delaware Biotechnology

Institute (University of Delaware, Newark, DE) using a multimode AFM in peak force tapping mode. The E2-DNA complex solution (5  $\mu$ l) and 5  $\mu$ l 2M NiCl<sub>2</sub> in HEPES was deposited on freshly cleaved mica and incubated at room temperature for 5 minutes for binding, followed by four rinses with deionized water. The samples were then allowed to air dry and positioned for imaging. The AFM images were analyzed by using NanoScope Analysis software (Bruker, Billerica, MA).

## **12. Cloning, expression, and purification of E1**

The HPV-11 *E1* gene was amplified from purified plasmid HPV-11 DNA, GenBank number M14119, obtained from American Type Culture Collection (ATCC) using Primers #11 and #12 (Table 1). The PCR product was cloned into a T7 expression vector, pET28a, engineered to have an N-terminal histidine tag (EMD Millipore, Billerica, MA), using standard recombinant DNA technology [75]. The validity of cloned *E1* gene, in the pET28a-E1 plasmid, was verified by DNA sequencing. *E. coli* Lemo21 (DE3) strain (New England Biolabs, Ipswich, MA) was then transformed with the pET28a-E1 plasmid for subsequent expression of the protein. *E. coli* cells were grown with 1 mM L-rhamnose and shaking at 37°C to OD<sub>600</sub> of 0.4. Isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG) was then added to a final concentration of 0.4 mM and incubation at 37°C was continued for 15 hours. The cells were harvested by centrifugation and resuspended in buffer A containing 1  $\mu$ g/ml of each protease inhibitor including pepstatin, leupeptin, chymostatin, and aprotinin. Induced cells were extracted with the addition of lysozyme following

standard protocol [76]. Low expression and reduced solubility of recombinant low-risk E1 protein was observed.

Custom gene synthesis was utilized to remove GC-rich sequences as well as to remove codons not suitable in *E. coli* to increase the level of expression and solubility of protein expression in *E. coli* host (GenScript, Piscataway, NJ). This optimized synthesized DNA sequence was cloned into an *E. coli* expression vector, engineered with an N-terminal S-tag and poly-histidine tag. In order to increase the binding potential of poly-histidine-tagged E1 for this column, the E1 optimized sequence was later sub-cloned without the S-tag, which could have been potentially hindering the full binding to cobalt by the His-tagged E1. The extract was loaded onto HisPur Cobalt purification resin (Thermo Fischer Scientific, Waltham, MA), equilibrated with buffer B. The column was washed thoroughly with buffer B, followed by a brief wash with buffer D. The protein was eluted using buffer E containing 10% glycerol. Fractions containing E1 were identified by SDS-PAGE. Protein homogeneity was determined by SDS-PAGE (Figure 21a) and verified by mass spectrometry conducted at University of Florida Department of Chemistry Mass spectrometry services (Gainesville, FL) (Figure 21b).

### **13. Anti-His tag antibody Western analysis**

Elutions from HisPur Cobalt purification resin were resolved in 10% SDS-PAGE. Following electrophoresis, the gel was transferred to Protran Premium NC Membrane (Genesee Scientific, San Diego, CA). The nitrocellulose membrane was incubated in 1XTBST buffer containing 3% BSA and 1% gelatin for 30 minutes at

room temperature. The membrane was then incubated overnight at 4°C with His-probe Antibody (sc-804) diluted 1:500 (Santa Cruz Biotechnology, Dallas, TX). After washing the membrane, it was incubated with horseradish peroxidase-coupled goat anti-rabbit IgG from the Pierce Fast Western Blotting Kit, Super Signal West Pico, Rabbit (Thermo Scientific, Waltham, MA). Subsequent steps were followed as described in the instruction manual for this kit.

#### **14. Anti-DnaB antibody Western analysis**

Anti-DnaB antisera was prepared from purified *E. coli* DnaB helicase in mouse following standard antibody generation protocol [77]. Western blot analysis was carried out following standard procedures [75] using *E. coli* anti-DnaB antisera. Briefly, proteins were resolved in 10% SDS-PAGE. Following electrophoresis, the gel was transferred to Protran Premium NC Membrane (Genesee Scientific, San Diego, CA). The nitrocellulose membrane was incubated in 1XTBST buffer containing 3% BSA and 1% gelatin for 30 minutes at room temperature. The membrane was then incubated overnight at 4°C with *E. coli* anti-DnaB antisera. After washing the membrane, it was incubated with alkaline phosphate-coupled anti-mouse IgG (Promega, Madison, WI). The membrane was then washed and incubated with 5-Bromo-4-chloro-3-indolyl phosphate (BCIP) (Sigma-Aldrich, St. Louis, MO) and nitro blue tetrazolium (NBT) (Sigma-Aldrich, St. Louis, MO) prepared according to manufacturer's instructions for colorimetric detection.

#### **15. Helicase assay**

The helicase assays were based on methods described previously [78]. A synthetic oligonucleotide, 60 bp in length, referred to as Primer 13 in Table 1, was designed to be complementary to a 50 bp sequence between nucleotides 6268-6317 of M13mp19 ssDNA, radio-labeled at the 5' end using T4 polynucleotide kinase. The radio-labeled oligonucleotide was hybridized to M13mp19 ssDNA as previously reported [78]. The hybridized oligomer contained 5 nucleotide tails (nonhomologous regions) on both 5' and 3' termini. Excess unhybridized radio-labeled oligonucleotides were purified by Bio-Gel P-6 column (Bio-Rad, Hercules, CA).

The helicase reactions were carried out in a 10- $\mu$ l final volume containing mM Tris-HCl (pH 7.5), 10% glycerol, 0.1 mg/ml BSA, and 5 mM DTT, 10 mM MgCl<sub>2</sub>, and enzyme as indicated. The reaction were incubated at 37 °C for 2 hours and terminated by the addition of 2  $\mu$ l of 2.5% SDS, 60 mM EDTA, and 1% bromophenol blue. Each reaction was immediately loaded on 8% polyacrylamide gel in 1X TBE and 0.1% SDS run at 180 V for 1 hour. Following electrophoresis, the gels were dried and exposed to X-ray films. Quantification was accomplished by PhosphorImager analysis using Typhoon 9410 Variable Mode Imager (GE Healthcare Life Sciences, Rahway, NJ).

## **16. EMSA with E1 and E2**

Synthetic oligonucleotides were used for end-labeling reactions with T4 polynucleotide kinase and ( $\gamma$ -<sup>32</sup>P) ATP. DNA probes were purified by Bio-Gel P-6 column (Bio-Rad, Hercules, CA).

The binding reactions were carried out in a 10- $\mu$ l final volume containing the indicated amounts of E1 and E2 proteins and 70 pmol of  $\gamma$  <sup>32</sup>P-labeled DNA probe in 10 mM HEPES (pH 7.9), 4 mM MgCl<sub>2</sub>, 10% glycerol, 5 mM dithiothreitol, 0.2 mM EDTA, 0.1 mg/ml bovine serum albumin, 1.4 mM ATP, and 50 mM NaCl. The binding reactions were allowed to proceed for 10 minutes at 30° C instead of 37°C to minimize the effect of any possible trace nuclease, DNase, or protease contaminants. The mixtures were immediately loaded on 5% polyacrylamide gel in 0.5X TBE and run at 170 V for 1 hour 15 minutes at 4 °C. Following electrophoresis, the gels were dried and exposed to X-ray films. Quantification was accomplished by PhosphorImager analysis using Typhoon 9410 Variable Mode Imager (GE Healthcare Life Sciences, Rahway, NJ). In competition experiments, reactions were performed as indicated above, with the addition of non-radio-labeled/competitor oligonucleotides in increasing amounts. Each experiment was quantified independently by PhosphorImager analysis. The probe (BS1+2+E1BS) used in the EMSA was generated by PCR amplification of the HPV-11 upstream regulatory region spanning HPV-11 genome nucleotide 7092-7933/1-92 using Primer 5 and Primer 6 (Table 1).

Oligonucleotide	Sequence	Assay Component
Primer 1	5'-CTCCTCCATATGGAAGCAATAGCCAAGC-3'	Forward primer for PCR amplification of HPV-11 <i>E2</i> gene from purified plasmid HPV-11 DNA
Primer 2	5'-TCTTCTCTCGAGGTAGTTGTTGCTGCAGC-3'	Reverse primer for PCR amplification of HPV-11 <i>E2</i> gene from purified plasmid HPV-11 DNA
Primer 3	5'-CTCCTCCCATGGAGACTCTTTGCCAACG-3'	Forward primer for PCR amplification of HPV-16 <i>E2</i> gene from purified plasmid HPV-16 DNA
Primer 4	5'-TCTTCTGGATCCAGACAAAAGCAGCGGACGTAT-3'	Reverse primer for PCR amplification of HPV-16 <i>E2</i> gene from purified plasmid HPV-16 DNA
Primer 5	5'-GTAACCCACACCCTACATATTTTCCTTC-3'	Forward Primer PCR amplification of the HPV-11 upstream regulatory region for generation of "BS1+2+E1BS" probe
Primer 6	5'-TCTGCTAATTTTTTGGGCTGGTTTATAT-3'	Reverse Primer PCR amplification of the HPV-11 upstream regulatory region for generation of "BS1+2+E1BS" probe
Primer 7	5'-CTCCTCGGATCCTGGAGGACTGGAAC TTTGGT-3'	Forward primer for PCR amplification of HPV-11 origin for AFM studies
Primer 8	5'-TCTTCTAAGCTTTATCTCTGCGGTGGTCAGTG-3'	Reverse primer for PCR amplification of HPV-11 origin for AFM studies
Primer 9	5'-CTCCTCGGATCC CCA CTA TTT TGG AGG ACT GGA ATT TTG GTC-3'	Forward primer for PCR amplification of HPV-16 origin for AFM studies
Primer 10	5'-TCTTCTAAGCTTCCCGAAAAGCAAAGTCATATA CCTCACGTCGC-3'	Reverse primer for PCR amplification of HPV-16 origin for AFM studies
Primer 11	5'-CTCCTCCCATGGCGGACGATTCAGGTAC-3'	Forward primer for PCR amplification of HPV-11 <i>E1</i> gene from purified plasmid HPV-11 DNA
Primer 12	5'-TCTTCTGAATTCGAGTCGTTGTCCTGCAGG-3'	Reverse primer for PCR amplification of HPV-11 <i>E1</i> gene from purified plasmid HPV-11 DNA
Primer 13	5'-GGGTCTCACGACGTTGTA AACGACGGCCAGTG AATTCGAGCTCGGTACCCGGGGTAGGA-3'	Complementary to a 50 bp sequence M13mp19 ssDNA, used in helicase assay as substrate

**Table 1. Oligonucleotide sequences used in this study.** Shown are the synthetic oligonucleotides used in this study.

## **EXPERIMENTAL RESULTS**

### **SECTION I: Mechanism of interaction of the Human Papillomavirus E2 initiator protein with the origin of DNA replication**

#### **A. Purification of HPV-11 E2 protein and its structure in solution**

Earlier studies have used truncated low-risk E2 proteins for mechanistic studies, possibly due to the low expression and solubility of the full-length E2 protein [79, 80]. The optimization of the expression and purification of E2 has enabled us to conduct mechanistic, structural, and functional studies for the full-length E2 protein (Figure 1a) with its DNA binding sites from the HPV-11 origin (Figure 1b). No crystal structures are currently available for the full-length HPV E2 proteins. Therefore, homology-based modeling of HPV E2 was utilized to demonstrate the structure and spatial arrangement of the full-length low-risk HPV E2 protein (Figure 1c).

#### **1. Purification of full-length low-risk E2 protein**

Optimal conditions were investigated for both E2 expression and solubility. Lemo21(DE3) cells (New England Biolabs, Ipswich, MA), with a tunable expression ability, was analyzed. Lemo21(DE3) is a derivative the BL21(DE3) *E. coli* cell line, containing an additional plasmid, pLEMO, encoding the natural inhibitor of T7 RNA polymerase, T7 lysozyme. The gene encoding the T7 lysozyme is modulated by the rhamnose promoter. The addition of rhamnose to the culture stimulates T7 lysozyme synthesis, inhibiting T7 RNA polymerase. This allows for tighter control of

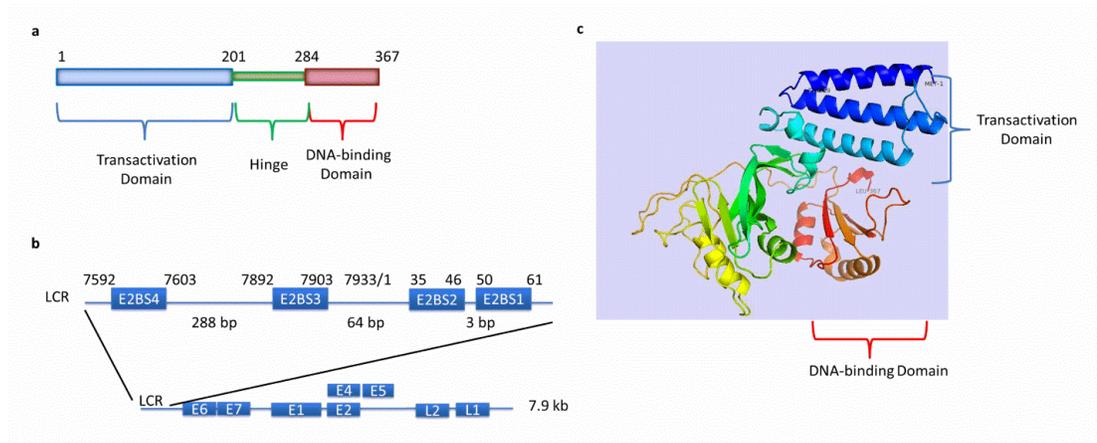
expression. When Lemo21(DE3) is grown in the absence of rhamnose, the strain exhibits features of the pLysS-containing strain, lowering background expression of target genes. The addition of rhamnose to the bacterial culture yielded soluble E2 protein.

E2 protein was soluble in solution containing at least 500 mM NaCl and appeared to aggregate at lower salt concentrations. Thus, purification of soluble E2 required chromatographic steps that can be carried out at high salt conditions. Reverse phase chromatography and the histidine-affinity chromatography were ideal, as these are normally performed at high salt (>500 mM). E2 protein was first purified by the histidine-affinity chromatography and was analyzed by SDS-PAGE. The fractions enriched in E2 were pooled, subjected to reverse phase chromatography, and eluted with a reverse salt gradient (Figure 2a). E2 was purified to homogeneity. Trypsin digestion and protein sequencing by mass spectrometry (MS) confirmed that low-risk E2 was indeed purified (Figure 2c).

## **2. Oligomeric structure of purified E2 protein in solution**

The oligomeric structure of purified E2 protein in solution was determined by Dynamic Light Scattering (DLS). DLS measures the hydrodynamic radius of macromolecules in solution. The generated histogram contained a single peak, which is referred to as mono-modal size distribution, suggesting a single species population (Figure 2b). The molar mass of  $79 \pm 7$  kDa was estimated from the measured radius of  $5.05 \pm 0.18$  nm, indicative of a dimer (Figure 2b). These results showed that the

E2 peak comprised 100% intensity and 100% mass of the sample measured, confirming the homogeneity of the purified E2 protein (Figure 2b).



**Figure 1. Structure of E2 protein, HPV-11 genome and its predicted crystal structure.** (a) Structure of the HPV-11 full-length E2 protein indicating the N-terminal transactivation, hinge and C-terminal DNA-binding domains (drawn to scale). (b) The location and structure of the viral DNA replication origin in the HPV-11 genome is shown. The enlargement of the LCR indicates the location of the four E2 binding sites (E2BS1-4) (not drawn to scale). (c) A homology-based model of full-length E2 protein from Robetta Structure Prediction server at the University of Washington indicating structures of the transactivation and DNA-binding domains. The first and last amino acids are annotated by MET-1 and LEU-367, respectively.

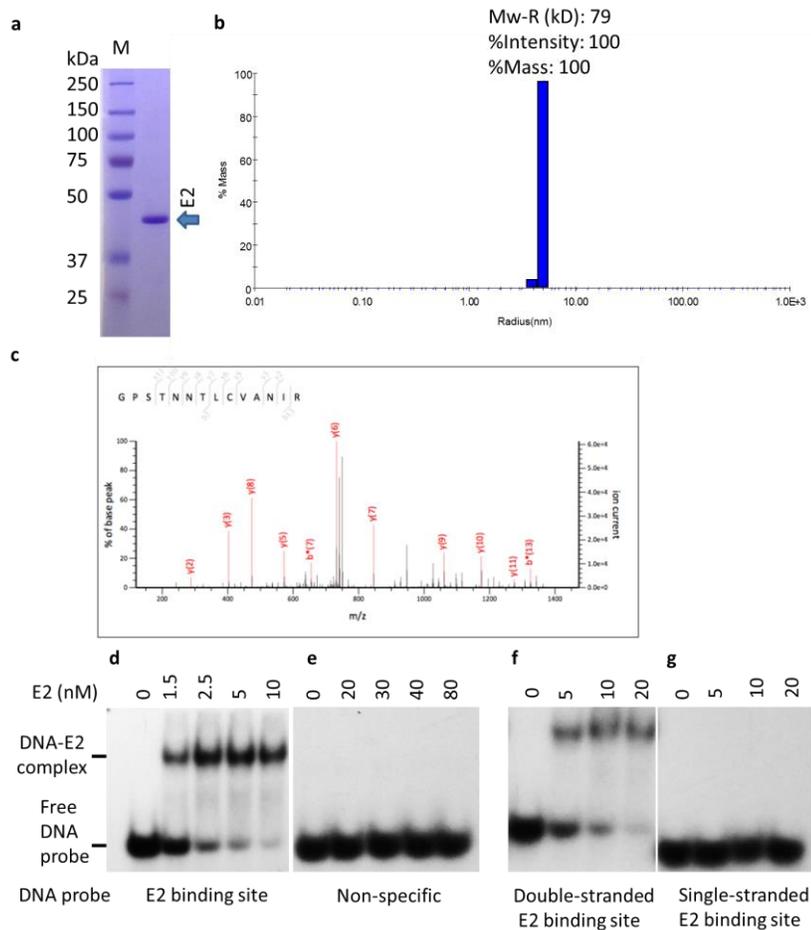
## **B. E2 protein DNA binding preference**

### **1. E2 activity analysis with specific and non-specific DNA**

The specificity of binding to its four recognition sequences as well as non-specific and single-stranded DNA was analyzed using electrophoretic mobility shift assay (EMSA). In general, <sup>32</sup>P-labeled duplex oligonucleotides containing the E2 recognition sequences (Table 2), as well as <sup>32</sup>P-labeled non-specific DNA (Figure 2e, Table 2) were individually incubated with increasing concentrations of E2 protein, in the presence of poly(dI-dC). The incubation of E2 with a specific oligonucleotide, E2 binding site 4, resulted in the exponential increase of a slower migrating band, indicating the formation of the E2-DNA complex (Figure 2d). However, E2 incubation with non-specific DNA did not result in complex formation, even at 80 nM concentration of E2 (Figure 2e), indicating that the E2 binding to cognate binding site DNA sequences appeared highly specific.

### **2. E2 DNA binding preference with double- and single-stranded DNA**

Similarly, the binding of E2 to the single-stranded E2 binding site was investigated. E2 was able to bind only the double-stranded form (Figure 2f), but not the single-stranded form of E2 binding site 4 (Figure 2g). Together, these data suggested that E2 bound specifically to its recognition sequence in the double-stranded structure only.



**Figure 2. E2 protein purity, oligomeric structure, mass spectrometric data, and DNA preference.** (a) Coomassie Blue-stained SDS-PAGE shows purified E2 protein at predicted size of 43 kDa on 10% polyacrylamide separating gel. “M” represents Precision Plus Protein Standards (Bio-Rad). (b) DLS analysis of E2 protein sample, using 3 separate measurements. A representative regularization graph of E2 showing a single mono-modal peak with a mass of 79 kDa for E2 protein is presented. (c) The displayed MS/MS (product-ion) spectra is of a 20 amino acid peptide unique to HPV-11 E2, generated after the fragmentation of the purified recombinant HPV-11 E2 used in this study. (d) DNA binding by E2 protein was analyzed by EMSA using a <sup>32</sup>P-labeled E2 binding site 4 or (e) non-specific DNA probe (Table 2) as indicated. (f) E2 protein was titrated with either double-stranded or (g) single-stranded E2 binding site. The E2 protein concentrations were as indicated.

## **C. E2 binds differentially to its DNA binding sites**

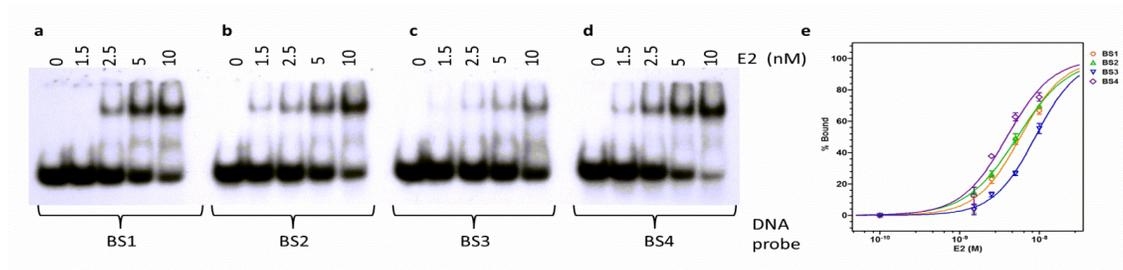
### **1. Analysis of E2 with cognate E2 binding sites**

The potential differences in E2 binding affinity for its four E2 binding sites were explored in a quantitative manner by EMSA. E2 protein was titrated in increasing amounts with labeled oligonucleotides containing E2 binding sites 1-4 (Figure 3). The free and bound DNA quantities were determined using PhosphorImager and ImageQuant software. The percentage of bound DNA for each binding site was plotted in Figure 3e. The binding affinities were determined using non-linear regression analysis of the plots in Figure 3e (% bound vs. log [E2]). The dissociation constant ( $K_D$ ) was computed using a variable slope sigmoidal dose-response model in GraphPad PRISM software and these values are shown in Table 2. This experiment was repeated three times. Highest binding affinity was observed for binding site 4 (BS4), with a  $K_D$  of  $3.9 \pm 0.5$  nM (Figure 3d). Intermediate binding affinity was observed with binding site 1 (BS1) and binding site 2 (BS2), with  $K_D$  of  $5.8 \pm 0.7$  nM and  $5.3 \pm 0.4$  nM respectively (Figure 3a and 3b). Slightly lower binding affinity was observed with binding site 3 (BS3), corresponding to a  $K_D$  of  $8.9 \pm 0.9$  nM (Figure 3c). Thus, the order of binding was as follows: BS4>BS2~BS1>BS3.

**Table 2. Sequences for oligonucleotides used as low-risk E2 binding sites.** Sequences for only one strand of each duplex oligonucleotide are shown. E2 consensus sequences are underlined. Each  $K_D$  was calculated and averaged from the experiment performed in triplicate.

Oligonucleotide	Sequence	# $K_D$ (nM)
Non-specific	TGGCGGACCCACCCAGCCCCACGCGCA	ND*
BS1	GGTTCA <u>ACCGAAAACGGT</u> TATATG	$5.8 \pm 0.7$
BS2	GGAGGG <u>ACCGAAAACGGT</u> TCAACC	$5.3 \pm 0.4$
BS3	CCTGCA <u>ACCGGTTTCGGT</u> TACCCA	$8.9 \pm 0.9$
BS4	CTTGCA <u>ACCGTTTTTCGGT</u> TGCCCT	$3.9 \pm 0.5$

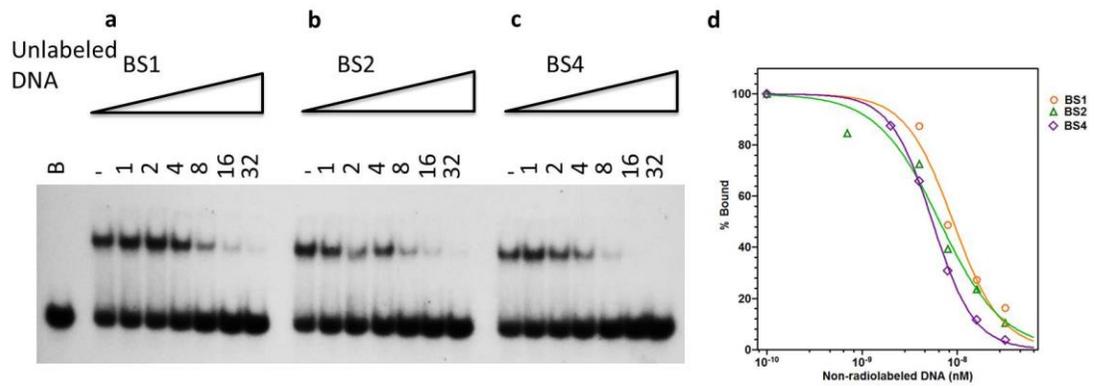
#  $K_D$  values represent the average of three independent measurements. Standard deviations were incorporated into  $K_D$ , accordingly. \*ND stands for “Not Determined”.



**Figure 3. Differential binding of E2 protein to its four binding sites.** (a-d) E2 protein was titrated to each of its four binding site probes (Table 2) labeled with  $^{32}\text{P}$  as indicated. The E2 protein concentrations were as indicated. A representative EMSA is shown, since these experiments were repeated three times. (e) Quantitative analysis of E2 binding to BS1, BS2, BS3, and BS4 as shown in Figure 3a-d by Prism 5 software from GraphPad Inc. Data represent the average of three experiments.

## 2. Challenge assay with E2 and E2 binding sites

In order to verify this hierarchy of E2 binding, a parallel EMSA challenge assay was performed. In these experiments, standard amounts of E2 protein concentration and  $^{32}\text{P}$ -labeled oligonucleotide (BS1) were titrated with unlabeled oligonucleotide. Essentially, the addition of unlabeled oligonucleotide would compete with the  $^{32}\text{P}$ -labeled oligonucleotide based on its sequence and decrease E2-DNA complex formation. BS4 was the most effective competitor, followed by BS1 and BS2 in effectiveness with  $K_D$  of  $4.6 \pm 0.4$ ,  $6.2 \pm 1.4$ , and  $6.0 \pm 0.6$  nM respectively (Figure 3a-c). Plot of this data indicated a decrease in bound  $^{32}\text{P}$ -labeled DNA (Figure 4d). The affinity of E2 for the competing unlabeled oligonucleotide could be calculated from these plots. Competitor BS3 required a higher concentration of unlabeled oligonucleotide to inhibit binding to  $^{32}\text{P}$ -labeled BS1 (data not shown). The non-specific DNA was not able to inhibit complex formation, even at much higher concentrations (data not shown). Together, these data confirm the unique hierarchy and binding affinities of E2 to its binding sequences.



**Figure 4. EMSA challenge assay verifies E2 hierarchy of binding.** (a-c) Increasing amounts of non-radio-labeled oligonucleotides containing individual E2 binding sites, as indicated, were titrated to challenge the E2 + <sup>32</sup>P-BS1 complex formation. The E2 protein concentration remained constant (5 nM). The lane that contained no E2 is marked as B. The lane with E2 but no added competitor was represented by (-). (d) Quantitative analysis of challenge assay as shown in Figure 4a-c using PhosphorImager analysis.

## **D. The composition and length of the spacer affected E2 binding affinity**

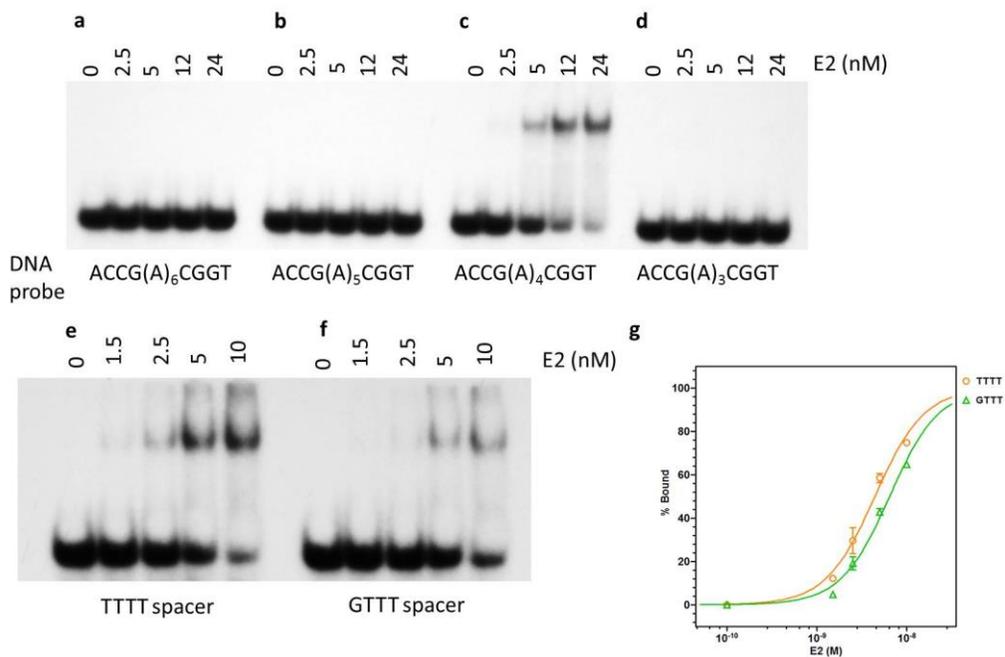
### **1. Analysis of spacer length on E2 binding activity**

The high degree of length conservation of the spacer region, between the conserved E2 half-sites, suggests physiological significance; however, it remained untested. This was the basis for the study. Oligonucleotides were designed based on wild-type E2 BS1 (ACCGAAAACGGT or ACCG(A)<sub>4</sub>CGGT). Spacer lengths were altered as follows: ACCG(A)<sub>6</sub>CGGT, ACCG(A)<sub>5</sub>CGGT, and ACCG(A)<sub>3</sub>CGGT all in duplex form. E2 protein was only able to form a DNA-protein complex with wild-type E2 BS1 containing the spacer of conserved four nucleotide length (ACCG(A)<sub>4</sub>CGGT) as seen in Figure 5c. No complex formation was observed with altered spacers (Figure 5a, 5b, and 5d). Oligonucleotides ACCG(A)<sub>2</sub>CGGT and ACCG(A)<sub>1</sub>CGGT were also tested; however, no bound E2 was observed (data not shown). These data suggested that the conserved length of four nucleotides in the spacer was essential for E2-DNA recognition.

### **2. Analysis of spacer sequence on E2 binding affinity**

E2 protein bound differentially to its various binding sites, even though the core palindromic sequence remained constant (Figure 3, Table 2). Thus, the conserved E2 half-sites (ACCG or CGGT) did not account for the difference in affinities. After a sequence analysis of the E2 binding sites, a trend in the spacer region amongst other low-risk HPV genotypes was observed. All spacers of BS3 contain a guanine, whereas the other binding sites contain primarily AT-rich sequences. Since in nature BS3 has a different spacer sequence, the ramification of

E2 recognition in relation to spacer composition was evaluated. The experiments reported here quantitatively analyzed the binding of HPV-11 E2 to binding sites that differed only in their spacer sequences. The highest affinity site (BS4) contains a TTTT spacer, whereas the lowest affinity site (BS3) contains a GTTT spacer. Here, BS4 was altered by a single nucleotide in the spacer region to represent that of BS3. HPV-11 E2 had higher preference to BS4 with wild-type spacer TTTT (Figure 5e). The binding affinity for wild-type BS4 was  $4.4 \pm 0.4$  nM, consistent with my studies mentioned previously. Decreased binding was evident when the thymine in the spacer was replaced with a guanine, as seen in BS3. This BS4 mutant, containing GTTT spacer, yielded a binding affinity of  $6.4 \pm 0.6$  nM (Figure 5f), shown quantitatively in Figure 5g. This may partly explain the lowered affinity for BS3, although wild-type BS3 possessed a  $K_D$  of  $8.9 \pm 0.9$  nM. This implies that the spacer composition does play a role in E2 binding preference; however, other elements may also influence the binding affinity.



**Figure 5. Binding of E2 to DNA binding site with altered spacer lengths and sequence.** (a-d) Synthetic oligonucleotides containing E2 BS1 with varying lengths of spacer, (A)<sub>x</sub>, where X refers to the number of adenine residues used to examine the roles of spacer length. EMSA was performed with titration of E2 protein, concentrations were as indicated. (e) Oligonucleotides contained HPV-11 E2 BS4 with either its native TTTT spacer or (f) spacer of BS3 (GTTT). E2 protein was titrated with each probe, with E2 protein concentrations as indicated. (g) Quantitative analysis of Figure 5e and Figure 5f using PhosphorImager analysis.

### **E. Effect of flanking sequences of the E2 binding sites on E2 binding affinity**

In wild-type E2 recognition sites, each consensus half-site sequence remains constant but the spacer and flanking sequences are different. The binding of E2 to a series of binding sites that differed only in their flanking sequences was quantitatively analyzed (Figure 6). In order to evaluate the effects of the flanking sequences on E2 binding affinities, duplex oligonucleotides were created using a constant consensus and spacer sequence (ACCGAAAACGGT), with different 6 base pair flanking sequences obtained from the native HPV-11 origin (Table 3). BS3 flanking sequences were used to determine their role in the decreased E2 affinity in Figure 3c.

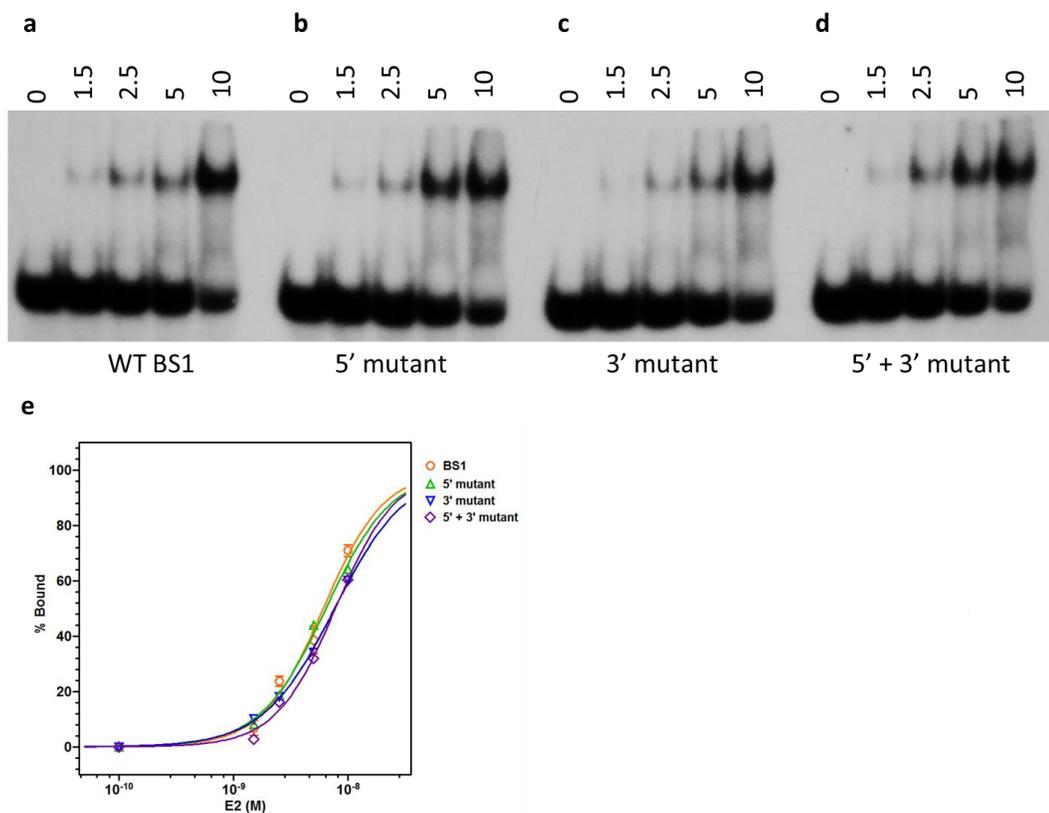
E2 was able to form a DNA-protein complex with each oligonucleotide (Figure 6). When the BS1 5' flanking sequence was replaced with the BS3 5' flanking sequence (5' mutant), the binding constant remained unaffected (Figure 6b). The resulting binding affinities of wild-type (WT) BS1 and 5' mutant were  $6.0 \pm 0.7$  nM and  $6.4 \pm 0.6$  nM, respectively (Figure 6a and Figure 6b). However, the replacement of the BS1 3' flanking sequence with that of BS3 (3' mutant) had an effect on the binding affinity of E2 to this site, as suggested by a higher  $K_D$  of  $7.6 \pm 0.6$  nM (Figure 6c, Table 3). The double mutant (5'+3' mutant) was replaced with both the 5' and 3' flanking sequences from BS3. This resulted in similar binding affinity as 3' mutant, with  $K_D$  of  $7.7 \pm 1.7$  nM (Figure 6d, Table 3). In summary, E2 bound wild-type BS1 and 5' mutant similarly, and slightly less to 3' mutant and double 5'+3' mutant (Table 3). Together, this data suggested that the flanking sequences affect E2 binding.

**Table 3. Sequences for flanking sequence oligonucleotides used as E2 binding sites.** Sequences for only one strand of each duplex oligonucleotide are shown. E2 consensus sequences are underlined. Mutant flanking sequences from BS3 are bolded. Each  $K_D$  was calculated and averaged from the experiment performed in

Oligonucleotide	Sequence	# $K_D$ (nM)
WT BS1	5' GGTTCA <u>ACCGAAAACGGT</u> TATATG 3'	6.0 ± 0.7
5' mutant	5' <b>CCTGCA</b> <u>ACCGAAAACGGT</u> TATATG 3'	6.4 ± 0.6
3' mutant	5' GGTTCA <u>ACCGAAAACGGT</u> <b>TACCCA</b> 3'	7.6 ± 0.6
5' + 3' mutant	5' <b>CCTGCA</b> <u>ACCGAAAACGGT</u> <b>TACCCA</b> 3'	7.7 ± 1.7

triplicate.

#  $K_D$  values represent the average of three independent measurements.



**Figure 6. Effect of flanking sequences on DNA binding by E2 protein.** Four different oligonucleotides were analyzed for E2 binding affinity by EMSA. Sequences for all oligonucleotides used are represented in Table 3. (a) One oligonucleotide contained native flanking sequences of E2 BS1 (WT BS1). All the other three oligonucleotides possessed the flanking sequences of BS3. (b) The BS1 5' flanking sequence was replaced with the BS3 5' flanking sequence (5' mutant). (c) BS1 3' flanking sequence was substituted with that of BS3 (3' mutant). (d) The double mutant (5'+3' mutant) was replaced with both the 5' and 3' flanking sequences from BS3. The E2 protein concentrations were as indicated. (e) Quantitative analysis of flanking sequences on DNA binding by E2 protein using PhosphorImager analysis.

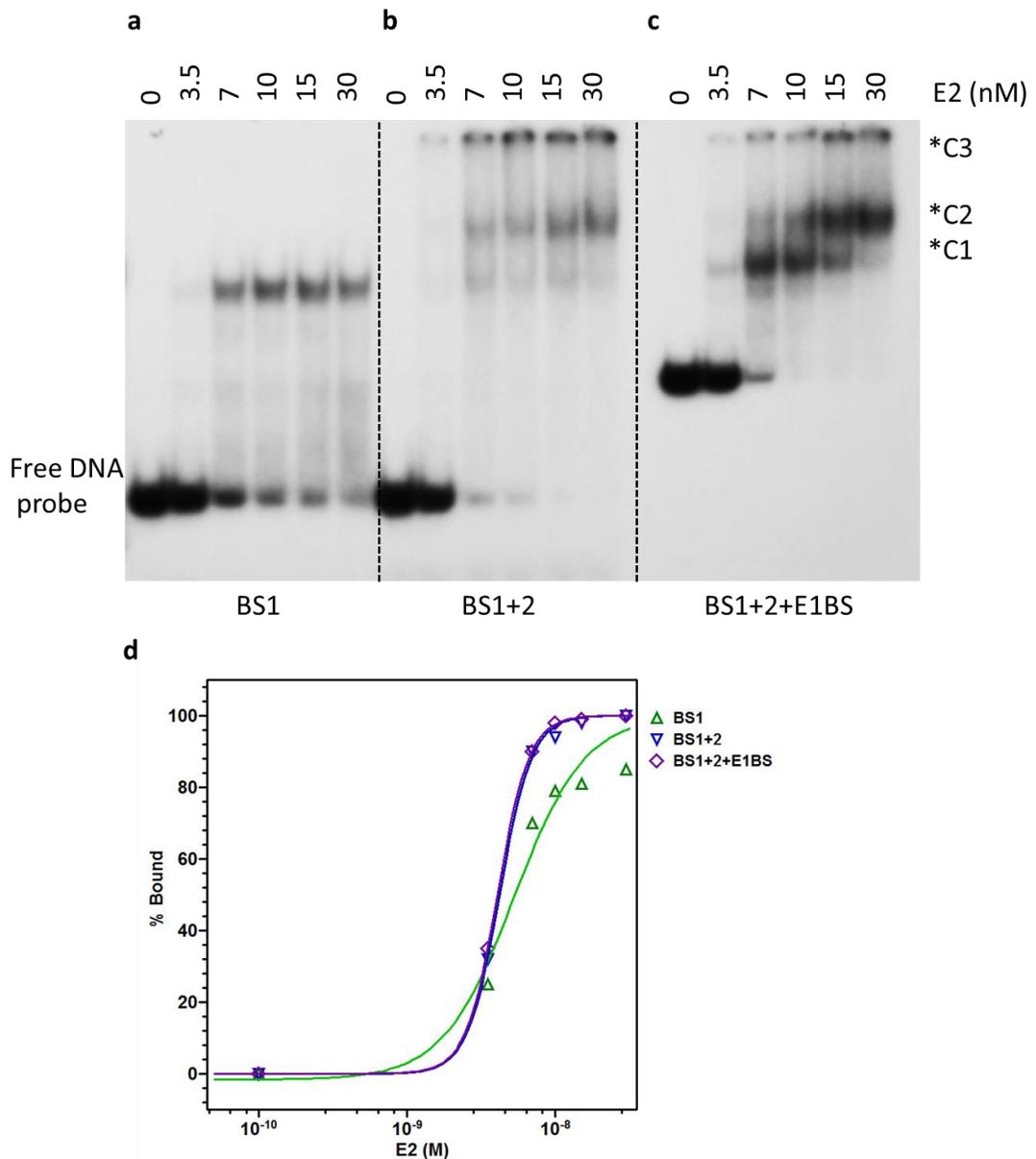
## **F. Cooperativity and higher order complex formation with multiple binding sites**

E2 BS1 and BS2 are in close proximity in all HPV origins separated by a few nucleotides. Some previous studies have supported E2 DNA binding cooperativity, while others have disagreed on its involvement [33, 73, 81, 82]. Therefore, when multiple sites are present, it was of interest to examine whether cooperativity plays a role in the overall binding affinity. To address this, oligonucleotides were synthesized that contained either one E2 binding site (BS1), or HPV-11 native spatial arrangement of two E2 binding sites (BS1+2). The third probe (BS1+2+E1BS) used in the cooperativity EMSA was generated by PCR amplification of the HPV-11 upstream regulatory region spanning HPV-11 genome nucleotide 7092-7933/1-92. This oligonucleotide incorporated the E1 binding site (E1BS), which is important for replication of HPV. All three oligonucleotide sequences are listed in Table 4. By EMSA analysis, a single E2-DNA complex (C1) was observed with one binding site and yielded a  $K_D$  of  $5.3 \pm 0.6$  nM (Figure 7a), consistent with my data mentioned previously for BS1. The incubation of E2 with oligonucleotide containing two binding sites (BS1+2) resulted in a higher binding affinity of  $4.2 \pm 0.3$  nM (Figure 7b). Notably, with the addition of this second binding site, higher order complex formation was also observed (C2) (Figure 7b). In addition, as the formation of higher order complexes increased, the appearance of the lower order complexes concomitantly decreased. This pattern is more visibly evident in the oligonucleotide containing two E2 binding sites as well as an E1 binding site (BS1+2+E1BS), which

was an extended version of the HPV-11 origin (Figure 7c). Experiments with BS1+2 and BS1+2+E1BS yielded similar binding patterns as well as binding affinities as shown in Figure 7d. Together, these data suggested that E2 may engage in cooperative DNA binding. In other words, the binding of the first E2 site led to more efficient binding to the second site, as indicated by the higher binding affinity. The higher order complex was seen only with E2 binding to two E2 binding sites (oligonucleotides BS1+2 and BS1+2+E1BS), likely two E2 dimers binding on the DNA. Another apparent feature, observed only when two E2 binding sites are present, is the formation of a third complex (C3), indicative of a much larger, multimeric E2-DNA complex that perhaps resembles a structure observed by electron microscopy a few years earlier [33].

**Table 4. Sequences of oligonucleotides utilized in cooperative DNA binding assays.** Sequences for only one strand of each duplex oligonucleotide are shown. E2 binding sites are bold and underlined. The E1 binding site is italicized. The sizes of each respective oligonucleotide are presented in ( ).

<b>Oligonucleotide</b>	<b>Sequence</b>
BS1 (50 bp)	GGAGGGATTGAAAAC <b><u>TTTTCAACCGAAAACGGTTAT</u></b> ATATAAACCAGCCC
BS1+2 (50 bp)	GGCGGG <b><u>ACCGAAAACGGTTCAACCGAAAACGGTTA</u></b> TATATAAACCAGCCC
BS1+2+E1BS (124 bp)	GTAACCCACACCCTACATATTT <b><u>CCTTCTTATACTTAATA</u></b> <i>ACAATCTTAGTTTAAAAAAGAGGAGGG</i> <b><u>ACCGAAAAC</u></b> <b><u>GGTTCAACCGAAAACGGTT</u></b> <i>TATATATAAACCAGCCCA</i> AAAAATTAGCAGA



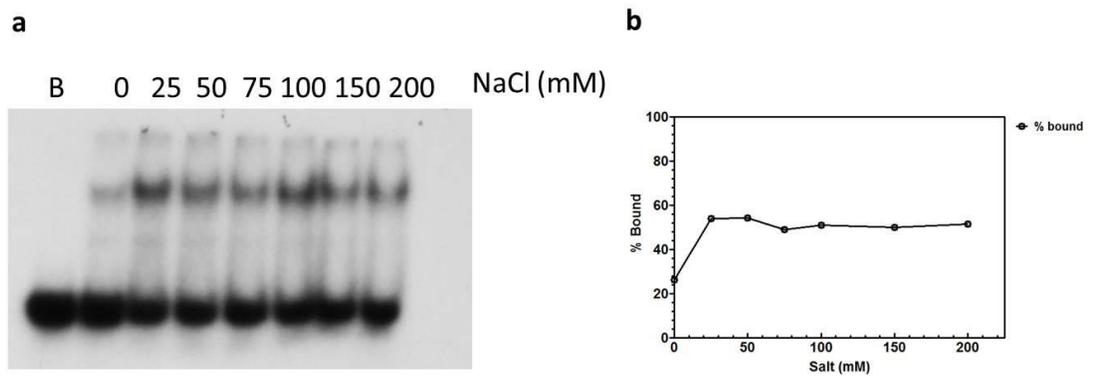
**Figure 7. Analysis of cooperative DNA binding of E2 protein.** EMSA was performed in the presence of increasing amounts of E2 protein as indicated above each lane. C1 indicates complex with one E2 dimer, C2 indicates complex with two E2 dimers and C3 refers to complex containing more than two E2 dimers. (a) Oligonucleotides contained E2 BS1 alone (BS1), (b) combination of BS1 and BS2 (BS1+2), or (c) BS1, BS2 and E1 binding site (BS1+2+E1BS). (d) Quantitative analysis of E2 cooperative DNA binding.

## **G. Modulation of DNA binding ability of E2 by ionic strength and temperature**

The nature of the E2-DNA interaction using thermodynamic analyses may help us characterize E2-mediated regulation. X-ray crystal structure of the truncated HPV E2 binding site indicated that increased salt concentration increased the angle of DNA curvature, which may point to structural flexibility [83]. It was suggested that this curvature may provide the appropriate three-dimensional structure for protein binding [83]. It was of interest to examine the effect of ionic strength and temperature on E2 binding affinity.

### **1. Physical nature of E2-DNA interaction**

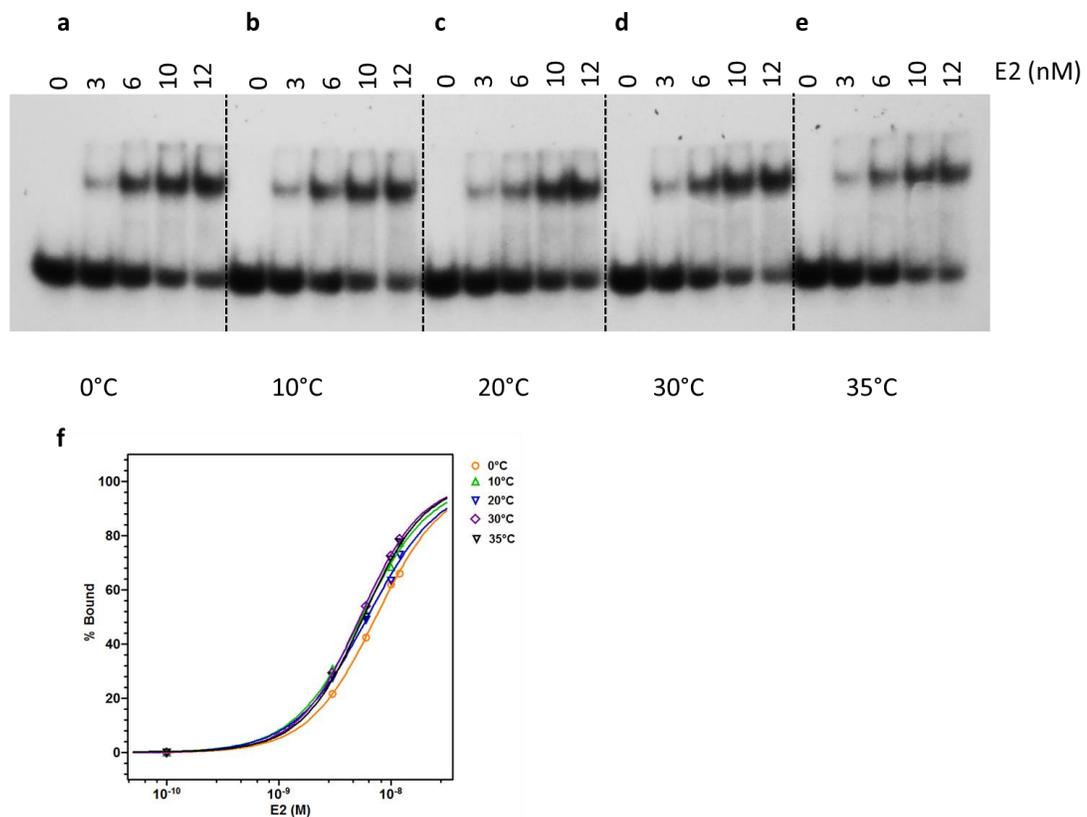
The physical nature of the DNA-protein interaction was analyzed. E2-DNA complex formation was examined as a function of NaCl concentration by EMSA. In this experiment, standard concentrations of E2 protein (6 nM) and <sup>32</sup>P-labeled oligonucleotide (BS1) were used. The best binding was observed at or above 25 mM NaCl (Figure 8a). This binding remained constant even at higher salt concentrations (Figure 8b). Salt enhanced DNA binding by E2, and the complex was stable at high salt concentrations. The observed salt dependence suggests that DNA binding of E2 may be mostly entropy-driven or hydrophobic in nature.



**Figure 8. Analysis of ionic strength on E2-DNA complex formation. (a)** EMSA was performed with standard concentration of E2 (6 nM) in the absence or presence of increasing amounts of salt (in mM) as indicated above each lane. “B” represented  $^{32}\text{P}$ -labeled E2 binding site 1, in absence of E2 protein. **(b)** Quantification of bound DNA at various salt concentrations reveals effect of ionic strength.

## **2. Thermodynamic analysis of E2-DNA binding**

In order to understand the thermodynamics of E2-DNA binding interactions, E2 binding to BS1 was studied at different temperatures (Figure 9a-e). Although similar binding affinities were seen at most temperatures, slightly lower binding affinity was observed at 0°C (Figure 9a, Table 5). Thus, DNA binding was not significantly influenced by temperature (Figure 9f). E2 is capable of binding at a wide range of temperatures. Together, these studies suggest the E2-DNA interaction is likely driven by hydrophobic interactions and less by ionic interactions.



**Figure 9. Thermodynamic parameters of E2-DNA complex formation. (a-e)** Temperature dependence of DNA binding of E2 was measured at indicated temperatures. E2 protein was incubated with <sup>32</sup>P-labeled E2 binding site 1 with increasing concentrations of E2. Five different temperatures were analyzed: 0°C, 10°C, 20°C, 30°C, and 35°C, as shown. **(f)** Quantification of bound DNA at various temperatures reveals binding isotherm.

**Table 5. Thermodynamic analysis of binding affinities for E2 protein.** Each  $K_D$  was calculated for the temperatures as indicated.

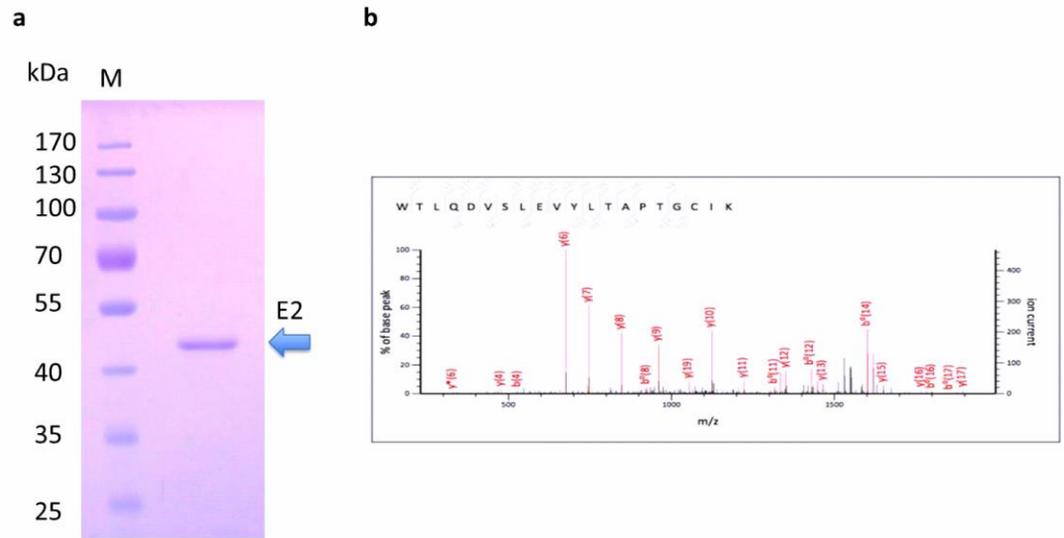
<b>Temperature</b>	<b><math>K_D</math> (nM)</b>
0°C	$7.3 \pm 0.7$
10°C	$5.5 \pm 1.5$
20°C	$6.2 \pm 1.4$
30°C	$5.3 \pm 0.4$
35°C	$5.7 \pm 1.1$

## **Section II: Mutations in the Human Papillomavirus replication origin may mediate viral oncogenesis**

### **A. Purification and analysis of full-length recombinant E2 from HPV-16**

#### **1. Purification of full-length high-risk E2 protein**

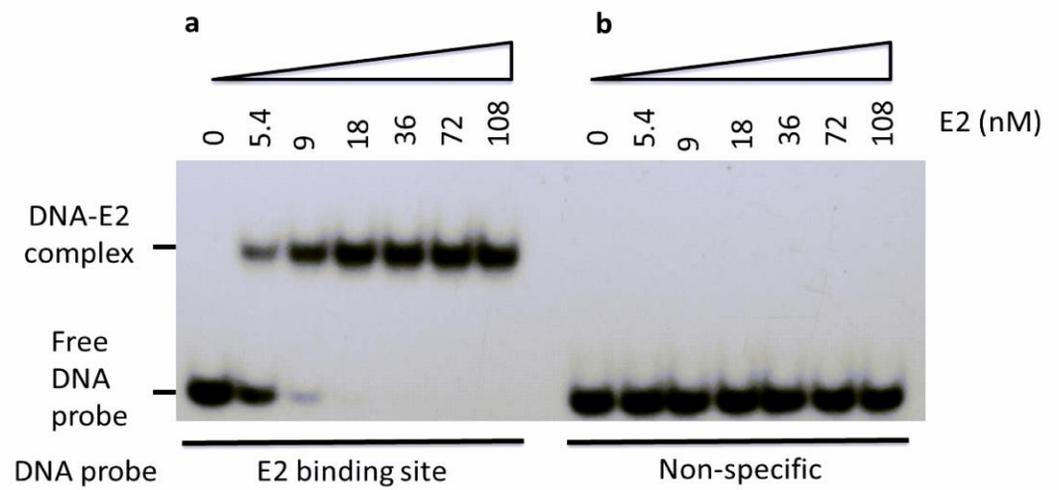
The full-length E2 protein from high-risk HPV-16 was cloned, expressed, and purified in order to study its role in replication origin activation. A methodology for the purification of the full-length high-risk E2 proteins was developed that exploited its unique physico-chemical properties. As observed with low-risk E2 protein, high-risk E2 required high salt concentrations ( $\geq 500$  mM) for solubility and appeared to aggregate at lower salt concentrations. Using a combination of nickel-affinity chromatography and hydroxyapatite chromatography, high-risk E2 was purified to homogeneity (Figure 10a). Its purity was confirmed by LC-MS/MS (Figure 10b).



**Figure 10. SDS-PAGE analysis and mass spectrometric data of high-risk E2 protein used in this study.** (a) 10% SDS-PAGE, visualized with Coomassie Blue, shows purified high-risk E2 protein at predicted size of 43 kDa on 10% polyacrylamide separating gel. “M” refers to PageRuler Prestained Protein Ladder (Thermo Scientific). (b) The displayed MS/MS (product-ion) spectra is of a 20 amino acid peptide unique to HPV-16 E2, generated after the fragmentation of the purified recombinant HPV-16 E2 used in this study.

## **2. Analysis of specific DNA binding activity of high-risk E2**

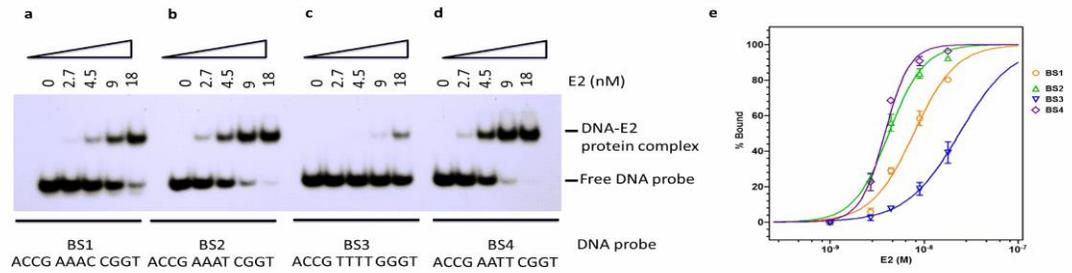
DNA binding activity and specificity of recombinant high-risk E2 protein was examined by electrophoretic mobility shift assay (EMSA) using its cognate binding sequences as well as non-specific DNA (Figure 11). A single DNA-protein complex was observed with an oligonucleotide probe containing HPV-16 E2 binding site 4, which increased with higher concentrations of E2 protein (Figure 11a). However, no detectable complex formation was observed with the non-specific DNA probe (sequence shown in Table 2), even at high concentrations of E2, indicating specificity (Figure 11b).



**Figure 11. EMSA analysis of high-risk E2 DNA binding activity and specificity.** (a) DNA binding by recombinant E2 was analyzed by EMSA using a  $^{32}\text{P}$ -labeled E2 binding site 4. (b) E2 protein was titrated with non-specific DNA probe. The E2 protein concentrations were as indicated.

### 3. Analysis of high-risk E2 with cognate E2 binding sites

To evaluate E2 binding preferences within the high-risk HPV origin, E2 binding to its four cognate DNA binding sites (BS1 to BS4) was examined (Figure 12). These oligonucleotides were derived directly from the HPV-16 genome sequence, and contained consensus binding sites as well as flanking sequences, as shown in Table 6. E2 bound BS4 most efficiently as shown in Figure 12d. E2 bound BS1 and BS2 efficiently (Figure 12a and 12b). However, E2 bound BS3 with very low affinity (Figure 12c). Quantitative analysis of EMSA demonstrated that E2 bound BS4 with a  $K_D$  of  $3.8 \pm 0.4$  nM (Fig 12e). With BS1 and BS2,  $K_D$  values were  $7.7 \pm 0.7$  nM and  $4.2 \pm 0.4$  nM, respectively. Overall, E2 bound BS1, BS2, and BS4 with high affinity, while a significantly lower binding affinity,  $> 18$  nM, was observed with BS3 as shown in Figure 12. Thus, the order of binding was as follows: BS4>BS2>BS1>>BS3



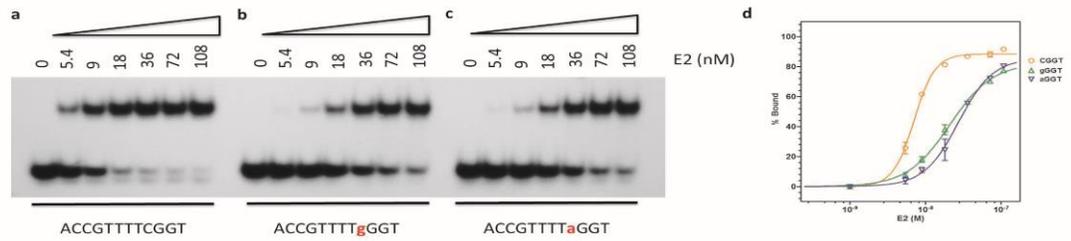
**Figure 12. High-risk E2 binds differentially to its four binding sites in the origin.**

(a-d) High-risk HPV-16 E2 protein was titrated with each of its four cognate E2 binding site probes (Table 6) labeled with <sup>32</sup>P as indicated. The E2 protein concentrations were as indicated. The lower band represents the free DNA probe. The higher band indicates the formation of the DNA-E2 complex. A representative EMSA of the triplicate is shown. (e) Quantitative analysis of E2 binding to these four individual binding sites is shown graphically.

## **B. Nucleotide variation in the binding site consensus sequence correlated to decreased E2 binding affinity**

### **1. Analysis of E2 binding to HPV-16 SNV**

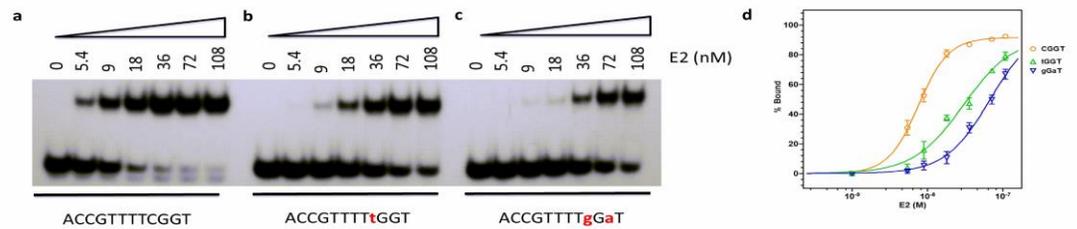
Sequence analysis of HPV-16 BS3, with low E2 binding affinity, revealed a single nucleotide variation (SNV), with a substitution of a conserved cytosine to guanine in the half-site of the consensus sequence (CGGT), while the other three binding sites (BS1, BS2, and BS4) all contained the consensus sequences. The SNV in the conserved half-site was further examined to determine whether it was responsible for the impaired binding of BS3 by E2. Variation in the consensus sequence is indicated by the lowercase bold lettering throughout this study. A wild-type oligonucleotide probe was constructed from HPV-16 BS3 sequence where the SNV (**g**GGT) was reversed by introducing a cytosine for the guanine in the consensus half-site sequence. High binding affinity was restored upon reverse mutation of the SNV in the HPV-16 BS3 sequence and the corresponding  $K_D$  was  $7 \pm 0.3$  nM (Figure 13a), which is comparable to other high affinity binding sites, with no SNV, examined in this study (Figure 12a). Whereas, HPV-16 BS3 with ACCGTTTT**g**GGT variation resulted in much lower binding affinity with a  $K_D$  of  $21 \pm 3$  nM (Figure 13b). This suggests that a substitution of cytosine in the CGGT sequence with a guanine in the consensus half-site (**g**GGT) of high-risk HPV-16 BS3 led to the observed attenuation of the binding affinity.



**Figure 13. Single nucleotide variation (SNV) in E2 binding site results in attenuated binding affinity.** (a) Synthetic oligonucleotides containing E2 BS3 with high-risk SNVs were used to examine the role of the SNV on protein-DNA complex formation. EMSA was performed with titration of E2 protein, concentrations were as indicated. Oligonucleotides were based on wild-type consensus sequence ACCGTTTTTCGGT. This oligonucleotide was compared to altered forms of this sequence, specifically the half-site (CGGT) with the alteration shown in bold and red lowercase letters. These represented SNVs found in high-risk types. Each SNV studied is found in at least one high-risk type. (b) One altered oligonucleotide included the sequence ACCGTTTTgGGT, which represented BS3 of the most carcinogenic HPV type, HPV-16. This oligonucleotide was also used in Figure 12 (BS3). (c) The oligonucleotide analyzed for E2 binding affinity possessed the ACCGTTTTaGGT sequence. This represented a SNV observed in several HPV types, such as HPV-18. A representative EMSA of the triplicate is shown. (d) Quantitative analysis of E2 binding to these sites is shown graphically.

## 2. Analysis of E2 binding to other high-risk SNVs

It was further investigated whether a lowered affinity to high-risk BS3 could be a common property among high-risk types. High-risk E2 binding to BS3, or other E2 sites with SNVs, from a number of high-risk HPVs, including HPV-18, HPV-31, and HPV-35 was examined. The binding of E2 to sites that differed only in the E2 half-site was quantitatively analyzed (Figure 13d, Figure 14d). Oligonucleotides were designed based on wild-type sequence ACCGTTTTCGGT, mentioned previously (Fig. 13a, Fig. 14a). E2 consensus sites were altered as follows: ACCGTTTTaGGT (present in BS3 of HPV-18), ACCGTTTTtGGT (present in BS2 of HPV-31), and ACCGTTTTgGaT (present in BS3 of HPV-35) all in duplex form to ensure the most physiologically relevant comparison of E2 binding (Table 6). Each nucleotide variation is representative of at least one high-risk type. E2 formed DNA-protein complexes with all sequences but with significantly lowered affinities. As observed in Fig. 13c, the oligonucleotide containing the sequence ACCGTTTTaGGT also resulted in lower binding than wild-type (ACCGTTTTCGGT); with a  $K_D$  of 27 nM, comparable but somewhat lower than HPV-16 BS3 (ACCGTTTTgGGT) (Fig. 13b). The lowest binding affinity was observed with ACCGTTTTtGGT and ACCGTTTTgGaT, with  $K_D$  values of 30 nM and 67 nM, shown in Fig. 14b and 14c, respectively. The nucleotide variations chosen for evaluation of E2 binding affinity represented common SNVs found in BS3 (or BS2) in many high-risk HPVs, and led in each case to significantly decreased E2 binding affinity.



**Figure. 14. SNV leads to decreased E2 DNA binding.** (a) Synthetic oligonucleotides containing E2 BS3 with high-risk SNVs were used to examine the role of the SNV on protein-DNA complex formation. EMSA was performed with titration of E2 protein, concentrations were as indicated. Oligonucleotides were based on wild-type consensus sequence ACCGTTTTTCGGT, also shown in Figure 13a. This oligonucleotide was compared to altered forms of this sequence, specifically the half-site (CGGT) with the alteration shown in bold and red lowercase letters. These represented SNVs found in high-risk types. Each SNV studied is found in at least one high-risk type. (b) A less common SNV was analyzed with an oligonucleotide sequence of ACCGTTTTtGGT. This SNV was only present in one HPV type, in the BS2 of high-risk HPV-31. (c) Sequence ACCGTTTTgGaT represented high-risk HPV-35 BS3 containing two alterations in the same consensus half-site. A representative EMSA of the triplicate is shown. (d) Quantitative analysis of high-risk E2 binding to these sites is shown graphically.

**Table 6. Sequences for oligonucleotides used as high-risk E2 binding sites.** Sequences for only one strand of each duplex oligonucleotide are shown. E2 half-sites are underlined. Variation in the consensus sequence is indicated by the lowercase bold lettering throughout this study. Each  $K_D$  was calculated and averaged from the experiment performed in triplicate.

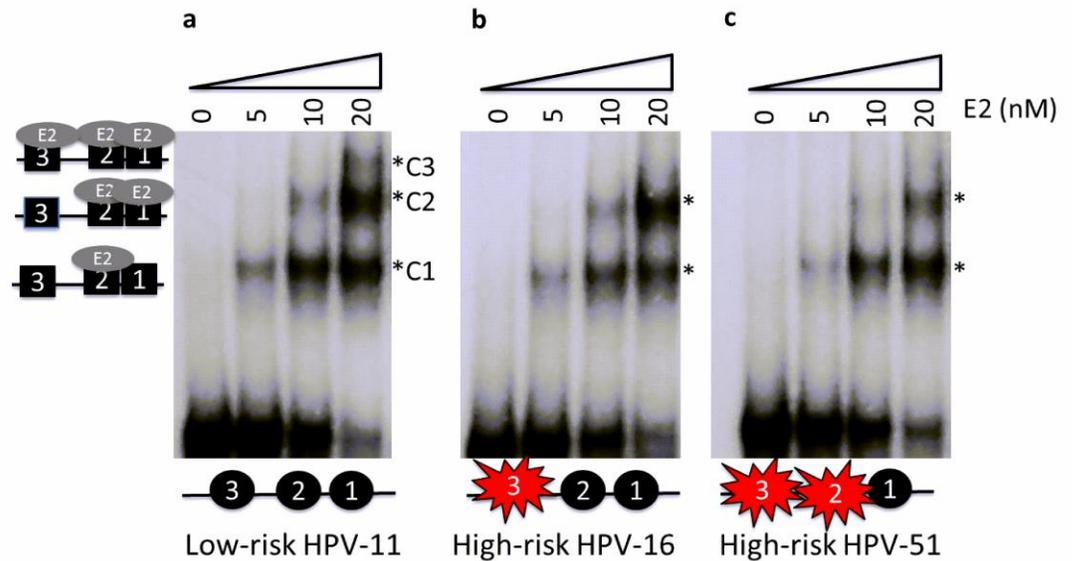
Oligonucleotide	Sequence	$K_D$ (nM)
<b>BS1</b>	GGCCGGTTGA <u>ACCGAAACCGTT</u> AGGGC	7.7
<b>BS2</b>	GCCGGGCGTA <u>ACCGAAATCGGT</u> TGGGCG	4.2
<b>BS4</b>	GGGCTTCA <u>ACCGAATTCGGT</u> TGCGGGCG	3.8
<b>HPV-16 BS3</b>	GGTGTGTGCAA <u>ACCGTTTT</u> <b>g</b> GGTTACGG	21
<b>HPV-18 BS3</b>	GGTGTGTGCAA <u>ACCGTTTT</u> <b>a</b> GGTTACGG	27
<b>HPV-31 BS2</b>	GGTGTGTGCAA <u>ACCGTTTT</u> <b>t</b> GGTTACGG	30
<b>HPV-35 BS3</b>	GGTGTGTGCAA <u>ACCGTTTT</u> <b>gGa</b> TTACGG	67

### **C. Multi-protein E2-DNA complex formation with DNA containing multiple**

#### **E2 binding sites**

Multi-protein E2-DNA complex formation was examined with DNA templates that contained BS1, BS2, and BS3 sequences from HPV-16 as well as other high-risk HPVs. Several high-risk HPVs were found to contain SNVs in multiple E2 binding sites. For example, HPV-51 contains SNVs in the BS3 half-site (**gtGT**), but also contains a SNV in its BS2 sequence (**gGGT**). In all HPVs, BS1, BS2, and BS3 are in close proximity in the origin. Therefore, multi-protein-DNA complex formation was examined with a template containing all three sites, similar to the complex that may form *in vivo* on the origin DNA containing these binding sites. Oligonucleotides used in this study are listed in Table 7. The wild-type oligonucleotide, spanning nucleotides 7889-7933/1-65, contained three consensus binding sites (BS1, BS2 and BS3) derived from the low-risk HPV-11 origin (Figure 15a). The second oligonucleotide in this study was designed to contain two wild-type binding sites (BS1 and BS2) and one binding site (BS3) with a SNV from high-risk HPV-16, representing the HPV-16 origin (Figure 15b). The third oligonucleotide contained two nucleotide variations, one present in BS2 (**gGGT**) and other in BS3 (**gtGT**), as observed in high-risk HPV-51 (and probable high-risk HPV-82) (Figure 15c). By EMSA analysis, three protein-DNA complexes were observed particularly at higher E2 concentration with the low-risk origin. E2-DNA complex 1 (C1) was observed at the lowest concentration of E2 (Figure 15a). Subsequently, complex 2 (C2) was observed with higher E2 levels, followed by the formation of complex 3

(C3) (Figure 15a). C1 was formed by one E2 dimer binding presumably to the highest affinity site. Similarly, the C2 represented two E2 dimers binding to the two higher affinity sites. C3 was likely formed by three E2 dimers bound at all three sites, including BS3 with lowest affinity. With the high-risk HPV-16 origin, C1 formation was observed, but predominantly C2 was observed at higher concentrations of E2 (Figure 15b). The C3 that was observed in the low-risk HPV-11 origin is not evident with the high-risk HPV-16 origin. Similarly, C3 is also significantly attenuated in high-risk origin with two SNVs in the E2 binding sites of HPV-51 (Figure 15c). Since C3 is not observed in both high-risk types HPV-16 and HPV-51, this may suggest that the single nucleotide variations of BS3, present in both these origins, may be responsible. In the HPV-51 origin with two SNVs, C2 formation observed, but predominantly C1 formation with the highest concentration of E2 (Figure 15c). These data support the notion that the second nucleotide variation is accountable for the unique E2 binding pattern. Based on the different distributions of E2 binding in the various origins as shown here, SNVs in E2 binding sites could hinder multi-protein complex formation and remodeling of the E2 origin, a prerequisite for the origin activation.



**Figure 15. Multi-protein E2-DNA complex formation in the HPV origin containing three E2 binding sites.** (a) Low-risk HPV-11 origin contained native, intact E2 BS1, BS2, and BS3 in a single oligonucleotide. As depicted in the schematic on the left, C1 indicates complex with one E2 dimer. C2 represents complex of two dimers with the origin, as shown in the schematic on the left. C3 refers to a complex containing three dimers on the DNA, as illustrated by the schematic. (b) The second oligonucleotide for this study required the alteration of low-risk HPV-11 origin to incorporate only the high-risk HPV-16 BS3 nucleotide variation, representing the HPV-16 origin. This oligonucleotide kept all other sites constant, as in the low-risk origin sequence. (c) The third oligonucleotide contained two nucleotide variations, one present in BS2 (gGGT) and other in BS3 (gtGT) as seen in high-risk HPV-51 (and HPV-82 in probable high-risk group). A representative EMSA of the triplicate is shown. Sequences used are represented in Table 7.

**Table 7. Sequences for oligonucleotides used in multi-protein E2-DNA complex formation.** Sequences for only one strand of each duplex oligonucleotide are shown. Variation in the consensus sequence is indicated by the lowercase bold lettering throughout this study.

Oligonucleotide	Sequence
<b>Low-risk HPV-11 Origin</b>	GCAACCGGTTTCGGTTACCCACACCCTACATATTCCTTCTTATA CTTAATAACAATCTTAGTTTAAAAAAGAGGAGGGACCGAAAAC GGTTCAACCGAAAACGGTTATA
<b>High-risk HPV-16 Origin</b>	GCAACCGGTTT <b>g</b> GGTTACCCACACCCTACATATTCCTTCTTATA CTTAATAACAATCTTAGTTTAAAAAAGAGGAGGGACCGAAAAC GGTTCAACCGAAAACGGTTATA
<b>High-risk HPV-51 origin</b>	GCAACCGGTTT <b>gt</b> GTTACCCACACCCTACATATTCCTTCTTATA CTTAATAACAATCTTAGTTTAAAAAAGAGGAGGGACCGAAA <b>g</b> GGTTCAACCGAAAACGGTTATA

#### **D. E2 binding site analysis identifies common variants among high-risk genotypes**

Based on the findings that SNVs in the E2 binding sites influence origin binding and therefore may influence frequency of origin activation and replication initiation, a comprehensive bioinformatics analysis of the sequences of E2 binding sites from all known low-risk and high-risk subtypes was performed. The upstream regulatory region sequences of the Alpha papillomaviruses were obtained from the Papillomavirus Genome (PaVE) Database. E2 binding sites were identified based on their homology to the conserved consensus sequence ACCG(N)<sub>4</sub>CGGT and relative positions of those sites within the origin.

##### **1. Bioinformatics analysis of low-risk E2 binding sites**

Multiple sequence alignment of E2 binding sites were generated with genotypes of similar cancer risk characterized by previous studies [2, 84]. Binding sites from all low-risk genotypes display complete half-site consensus sequences in all four binding sites as shown in Figure 16a. No variation was observed in the consensus sequence conforming to ACCG(N)<sub>4</sub>CGGT.

##### **2. Bioinformatics analysis of high-risk E2 binding sites**

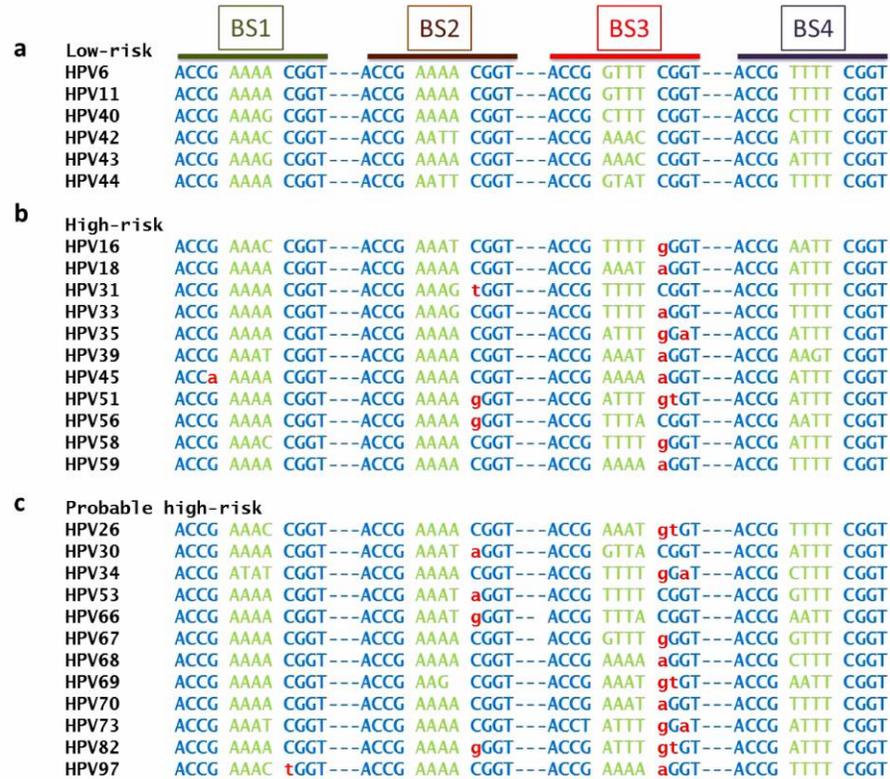
Multiple sequence alignment of 11 established high-risk genotypes revealed variation in the consensus half-site of at least one of the E2 binding sites (Figure 16b). For example, the sequence for HPV-16 BS3 was ACCGTTTTgGGT, substituting the well-conserved cytosine in the CGGT sequence with guanine. Nine of the 11 high-

risk genotypes displayed nucleotide variation in BS3 at the same position, with HPV-35 and HPV-51 containing a second SNV in the same half-site of BS3 altered (Figure 16b). The other two high-risk types HPV-31 and HPV-56 contained a SNV in BS2, as shown in Figure 16b. Two high-risk types had SNVs in both BS2 and BS3 (HPV-51) or BS1 and BS3 (HPV-45), as shown in Figure 16b. Thus, the presence of an altered E2 binding site, primarily BS3 and/or BS2, correlated well with higher cancer risk.

### **3. Bioinformatics analysis of probable high-risk E2 binding sites**

The International Agency for Research on Cancer classified Group 2A/B as “probably/possibly carcinogenic to humans” (probable high-risk) due to limited evidence in cancer [84]. Similar to the high-risk bioinformatics analysis, 12 probable high-risk genotypes contained at least 1 binding site with a SNV in the E2 consensus half-site, primarily in BS3. The most frequent alteration of BS3 was a single nucleotide variation of the conserved cytosine in the CGGT half-site sequence, as seen in 9 of the 12 probable high-risk types (Figure 16c). Five probable high-risk types contained a second position of the same half-site altered within BS3, including HPV-26, HPV-34, HPV-69, HPV-73, and HPV-82 (Figure 16c). HPV-30, HPV-53, and HPV-66 contained a SNV in only BS2 (Figure 16c). In addition to BS3, HPV-82 and HPV-97 had an SNV in BS2 or BS1 respectively (Figure 16c). HPV-69 BS2 had a shortened spacer, only three nucleotides in length. HPV-51 and HPV-82 contain similar nucleotide variations in E2 binding sites, and also possess similar carcinogenic potential. The E2 binding site nucleotide variation, predominantly in

BS3, present in high-risk and probable high-risk groups correlated well with the high cancer risk assessed by previous epidemiological studies [85]. Probable high-risk HPV-85 and high-risk HPV-52 were the only types that did not contain any nucleotide variations in the consensus sequence and misclassification could not be ruled out.



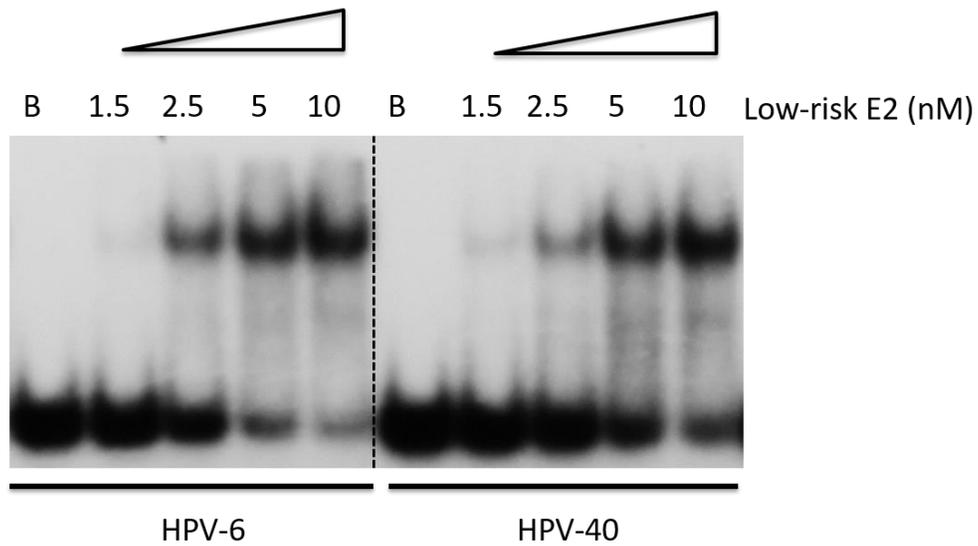
**Figure 16. Bioinformatics analysis of E2 binding sites and detection of SNVs in the consensus sequence.** (a) Sequence alignment of E2 binding sites 1-4 from HPV genomes, as specified. Genomes were obtained for low-risk, high-risk, and probable high-risk from PaVE and analyzed only for sites conforming to the E2 consensus binding sequence ACCG(N)<sub>4</sub>CGGT. The color codes are as follows: blue: consensus half-site, green: spacer, red: SNV. (b) Presented is the alignment of E2 binding sites 1-4 from each high-risk type indicated. SNVs were identified and presented in red lowercase letters, specifically in the E2 consensus half-site. (c) The HPV types belonging to the International Agency for Research on Cancer probable high-risk group are listed. Each genome was analyzed for E2 binding sites 1-4, as shown.

## **E. Low-risk E2 protein cannot bind E2 binding site containing high-risk SNV**

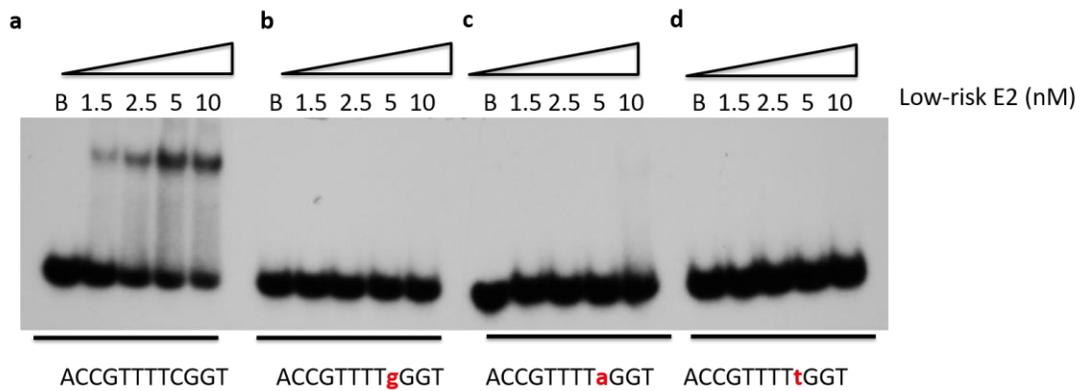
Binding sites from all low-risk genotypes display complete half-site consensus sequences in all four binding sites as shown in Fig. 16a. No variation was observed in the consensus sequence conforming to ACCG(N)<sub>4</sub>CGGT. Analysis of 11 established high-risk genotypes revealed variation in the consensus half-site of at least one of the E2 binding sites (Fig. 16b). This study further examined whether the SNV in the conserved half-site could also have an effect on DNA binding by low-risk E2 protein. As a control, low-risk HPV-11 E2 protein was incubated with BS3 sites from other low-risk types, including HPV-6 and HPV-40 (Figure 17). As expected, HPV-11 E2 was able to bind both HPV-6 BS3 and HPV-40 BS3 efficiently, as indicated by  $K_D$  of  $5.7 \pm 2.1$  nM and  $6.8 \pm 1.7$  nM, respectively (Figure 17). This suggests that low-risk HPV-11 E2 can bind to other low-risk BS3.

Variation in the consensus sequence is indicated by the lowercase bold lettering throughout this study. As mentioned previously, wild-type oligonucleotide probe was constructed from HPV-16 BS3 sequence where the SNV (**g**GGT) was removed by introducing a cytosine for the guanine in the consensus half-site sequence. High binding affinity was seen with the wild-type oligonucleotide with low-risk E2, which is comparable to other high affinity binding sites, with no SNV, examined (Figure 18a). Whereas, HPV-16 BS3 with ACCGTTTT**g**GGT variation resulted in no binding with low-risk E2 protein (Figure 18b). Similarly, no DNA binding by low-risk E2 protein was observed with oligonucleotides ACCGTTTT**a**GGT and ACCGTTTT**t**GGT, shown in Figure 18c and Figure 18d,

respectively. This suggests that a substitution of cytosine in the CCGT sequence, observed in high-risk BS3, led to dramatic attenuation of the binding affinity in low-risk E2.



**Figure 17. Low-risk HPV-11 E2 can bind to other low-risk BS3.** DNA binding by HPV-11 E2 was analyzed by EMSA using a  $^{32}\text{P}$ -labeled E2 binding site 3 from low-risk types HPV-6 and HPV-40. The E2 protein concentrations were as indicated. The lower band represents the free DNA probe. The higher band indicates the formation of the DNA-E2 complex.



**Figure 18. Single nucleotide variation (SNV) in E2 binding site results in attenuated binding affinity of low-risk E2.** (a) Synthetic oligonucleotides containing E2 BS3 with high-risk SNVs were used to examine the role of the SNV on low-risk protein-DNA complex formation. EMSA was performed with titration of low-risk E2 protein, concentrations were as indicated. Oligonucleotides were based on wild-type consensus sequence ACCGTTTTCGGT. This oligonucleotide was compared to altered forms of this sequence, specifically the half-site (CGGT) with the alteration shown in bold and red lowercase letters. These represented SNVs found in high-risk types. Each SNV studied is found in at least one high-risk type. (b) One altered oligonucleotide included the sequence ACCGTTTTgGGT, which represented BS3 of the most carcinogenic HPV type, HPV-16. (c) The oligonucleotide analyzed for E2 binding affinity possessed the ACCGTTTTaGGT sequence. This represented a SNV observed in several HPV types, such as HPV-18. (d) A less common SNV was analyzed with an oligonucleotide sequence of ACCGTTTTtGGT. This SNV was only present in one HPV type, in the BS2 of high-risk HPV-31.

## **F. Structural analysis of E2 on HPV origin by Atomic Force Microscopy (AFM)**

Previous studies have visualized multi-protein E2-DNA complex formation with low-risk E2 protein and its respective HPV origin DNA, but lacked a direct comparative analysis of the low-risk and high-risk E2-DNA structures [33, 86, 87]. This was the goal of this study and was investigated using Atomic Force Microscopy (AFM). AFM was an ideal method to observe structures of protein or protein-DNA complexes. Unlike electron microscopy, AFM does not require heavy metal staining and excessive drying, making the analysis possible under physiological conditions.

### **1. Design of low-risk and high-risk origin constructs for AFM**

Two linear DNA substrates containing either HPV-11 or HPV-16 origin regions were generated by PCR amplification of HPV-11 DNA spanning nucleotides 7076-7933/1-230 or HPV-16 DNA spanning nucleotides 6824-7905/1-248, respectively. These two DNA fragments were used as substrates for E2 binding (Figure 19). HPV-16 and HPV-11 DNA fragments were approximately 1088 bp and 1329 bp, respectively. HPV-16 DNA was slightly longer due to the inherent differences in spatial distance between E2 binding sites (Figure 19b), as well as optimal placement of PCR primers. BS4 was located in the middle of each DNA fragment with 516 bp on one side and 560 bp on the other side in the HPV-11 origin substrate (Figure 19). HPV-16 BS4 was located 627 bp from one end and 690 bp from the other end in the HPV-16 origin substrate (Figure 19). The other binding sites, BS1, BS2, and BS3 were generally clustered at the end of the DNA fragment

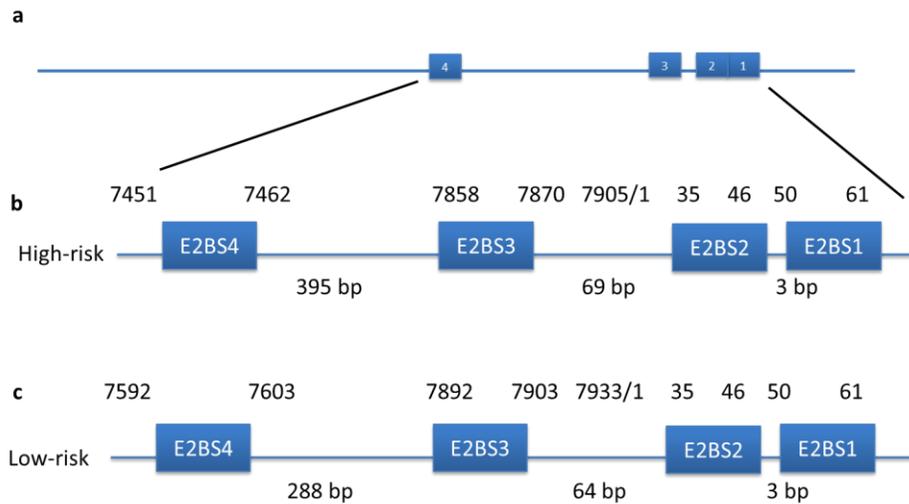
(Figure 19a). The HPV-11 BS1, BS2, and BS3 cluster was located 985 bp from one end and 169 bp from the other end (Figure 19a). Similarly, the HPV-16 BS1, BS2, and BS3 cluster was located 1034 bp from one end and 187 bp from the other end (Figure 19a). Each DNA template contained all four E2 binding sites in their native spatial arrangement obtained from their respective HPV DNA (Figure 19b, Figure 19c).

## **2. Structural analysis of low-risk E2 protein on low-risk HPV origin**

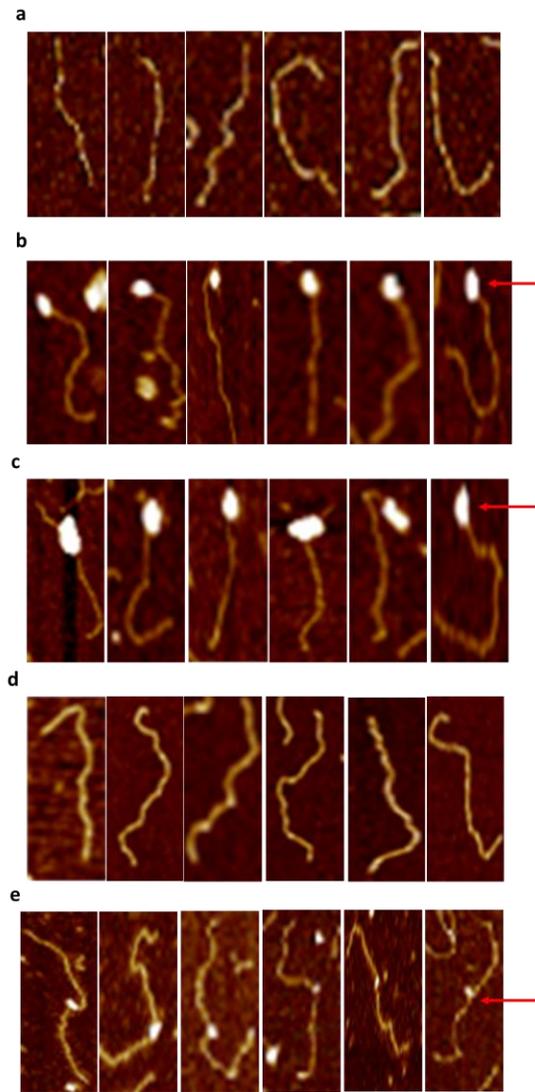
The DNA fragments were individually incubated with their respective E2 proteins. After incubation, E2-DNA complexes were subjected to fixation with glutaraldehyde. Low-risk DNA in the absence of protein was also visualized (Figure 20a). Approximately 200 DNA molecules were analyzed for each protein. E2 binding to the DNA was visualized as the white globule formation along the DNA, indicated with the red arrow in Figure 20b-d. Examination of low-risk E2-DNA complexes revealed several globule structures with E2 bound mostly at the end of the DNA molecule (Figure 20b, Figure 20c). The observed position of the low-risk E2 was consistent with the known locations of the BS1, BS2, and BS3 cluster (Figure 19a). This complex was most likely a particle containing three E2 dimers on the DNA. Further inspection of the low-risk E2-DNA particles revealed a range of sizes (Figure 20b, Figure 20c). Some low-risk E2 was observed as large masses bound to the DNA (Figure 20c). It is possible that this complex contained multiple E2 dimers.

### **3. Structural analysis of high-risk E2 protein on high-risk HPV origin DNA**

High-risk DNA in the absence of protein was also visualized using AFM (Figure 20d). Unlike the low-risk protein, high-risk E2 seemed to form a smaller complex on the DNA (Figure 20e). Also, the location of bound high-risk E2 on the DNA was different than the position of the low-risk E2 on the DNA (Figure 20e). More specifically, high-risk E2-DNA complexes were frequently observed in the center of the DNA substrate (Figure 20e). This position was consistent with the known location of BS4 on this DNA substrate (Figure 19a). This high-risk E2-DNA complex was likely a single E2 dimer binding to BS4. In rare cases, high-risk E2 was observed toward the end of the DNA molecule, and low-risk E2 was infrequently bound in the center of the DNA. Overall, these data demonstrated that low-risk E2 bound to the DNA and formed a large complex. On the other hand, high-risk E2 bound to its DNA in a linear fashion was not observed to form large complexes.



**Figure 19. DNA substrates for E2-DNA structural analysis using AFM.** (a) Linear DNA substrates containing either HPV-11 (low-risk) or HPV-16 (high-risk) long control regions were generated by PCR amplification of HPV-11 DNA spanning nucleotides 7076-7933/1-230 or HPV-16 DNA spanning nucleotides 6824-7905/1-248, respectively. The distances and locations of the E2 binding sites on the DNA are shown, drawn to scale. These DNA fragments were used as substrates for (b) high-risk and (c) low-risk E2 protein-DNA visualization with AFM. Each DNA template contained all four E2 binding sites in their native spatial arrangement from their respective HPV DNA.



**Figure 20. Structural analysis of low-risk and high-risk E2-DNA complex using AFM.** (a) Low-risk HPV DNA containing all four E2 binding sites was visualized using AFM in the absence of E2 protein. (b-c) Low-risk E2 protein was bound to low-risk HPV DNA fragment, forming a large E2-DNA complex, indicated by the red arrow. (d) High-risk HPV DNA containing all four E2 binding sites was visualized using AFM, in the absence of E2 protein. (e) High-risk E2 protein was observed on high-risk HPV DNA substrate, forming a small E2-DNA complex, indicated by the red arrow.

### **Section III: E1 helicase and E2 interaction at the origin**

Previous studies have used truncated low-risk E1 proteins for mechanistic studies, possibly due to the low expression and solubility of the full-length E1 protein [88, 89]. The optimization of the expression and purification of E1 has enabled us to conduct mechanistic, structural, and functional studies for the full-length E1 protein.

#### **A. Cloning of E1 helicase**

To further explore the mechanism of HPV DNA replication, the E1 helicase was cloned, expressed and purified using standard recombinant DNA technology [75]. The gene encoding E1 was amplified from HPV-11 genomic DNA obtained from ATCC (Materials and Methods). The PCR product was then cloned into a T7 expression vector for subsequent production in *E. coli*. The ORF was sequenced to verify the desired insert and to rule out the presence of fortuitous mutations introduced through PCR. Low expression and reduced solubility of recombinant low-risk E1 protein was observed. Low expression may have been due to the complex, GC-rich nature of the *E1* gene. In order to increase the level of expression, custom gene synthesis was utilized, which is used for optimizing DNA sequences by removing GC-rich sequences as well as heterologous codons to significantly increase protein expression in the *E. coli* host (GenScript). This optimized DNA sequence was cloned into an *E. coli* expression vector, engineered with an N-terminal S-tag and poly-histidine tag.

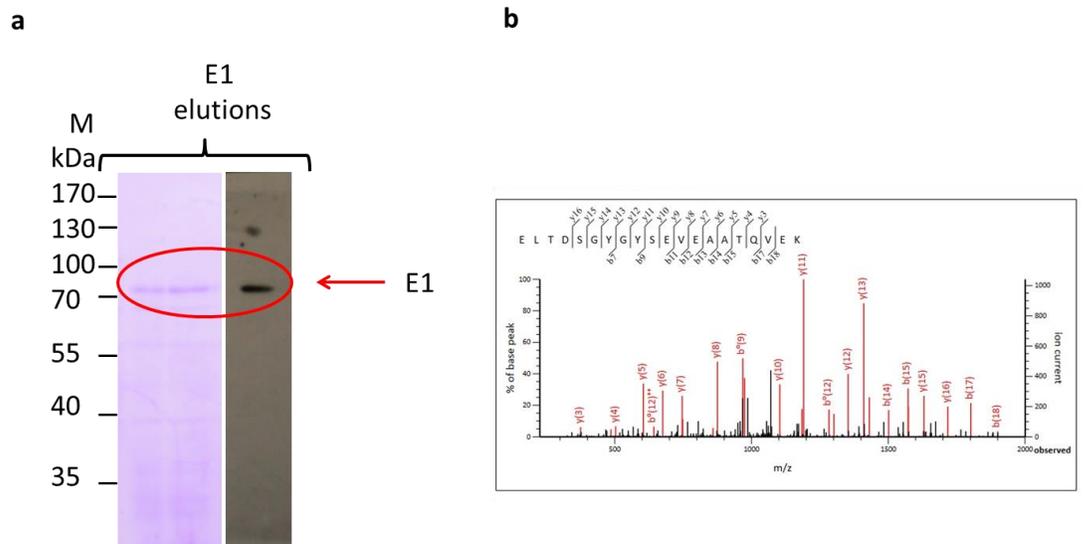
## **B. Expression of E1 helicase in *E. coli***

Optimal conditions were investigated for both E1 expression and solubility. Lemo21(DE3) cells (New England Biolabs, Ipswich, MA), with the tunable expression ability, were analyzed. Previously, the Lemo21(DE3) cell line was successful in facilitating the solubility of E2 protein, thus this approach was also utilized for E1 helicase. Similar to the approach taken with E2 protein, a range of IPTG, temperature, duration and rhamnose were explored. The optimal conditions in terms of expression level and solubility were found to be 16-hour incubation at 37°C, in the presence of 0.4 mM IPTG and 1 mM rhamnose. Extraction of soluble E1 protein was carried out as described for E2 protein (Materials and Methods). Pure E1 could be isolated following high salt extraction (500 mM NaCl). The solubility of E1 also required high salt and the protein appeared to aggregate at lower salt concentrations (< 500 mM NaCl), similar to that observed with E2 protein.

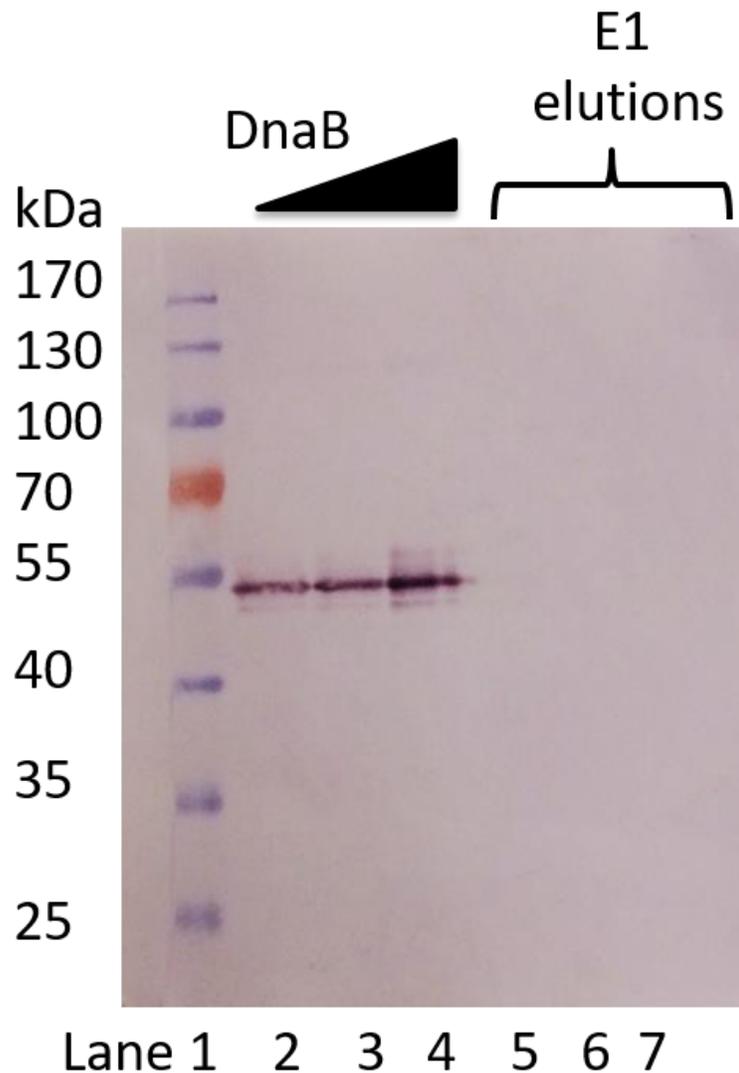
## **C. Purification of E1 protein**

Several protein purification approaches were examined for developing a suitable purification scheme, including S protein agarose, ion exchange chromatography, ssDNA cellulose, and immobilized metal affinity chromatography (IMAC) with either Nickel or cobalt as the affinity metal. Best results were obtained using the IMAC utilizing cobalt as the affinity metal. In order to increase the binding potential of poly-histidine-tagged E1 for this column, the E1 optimized sequence was sub-

cloned without the S-tag. This approach facilitated efficient purification by metal affinity chromatography as shown by SDS-PAGE analysis and confirmed by anti-His tag western in Figure 21. Purification experiments were carried out in buffer containing 500 mM NaCl, which reduced E1 protein aggregation. In order to address whether the E1 preparation contained the *E. coli* helicase, DnaB, a western blot using polyclonal mouse anti-DnaB antibody was also performed. The presence of DnaB in the protein preparation could potentially interfere with the later assessment of the E1 protein functionality, since it is also a helicase. Control samples of DnaB protein were used, which provided signal with the anti-DnaB antibody (Figure 22, Lanes 2-4). However, the E1 protein samples provided no observable signal with the anti-DnaB antibody, confirming the absence of DnaB helicase in the E1 protein sample (Figure 22, Lanes 5-7). Studies were conducted with homogenous E1 preparation. LC/MS-MS confirmed the homogeneity of the purified E1 (Figure 21b).



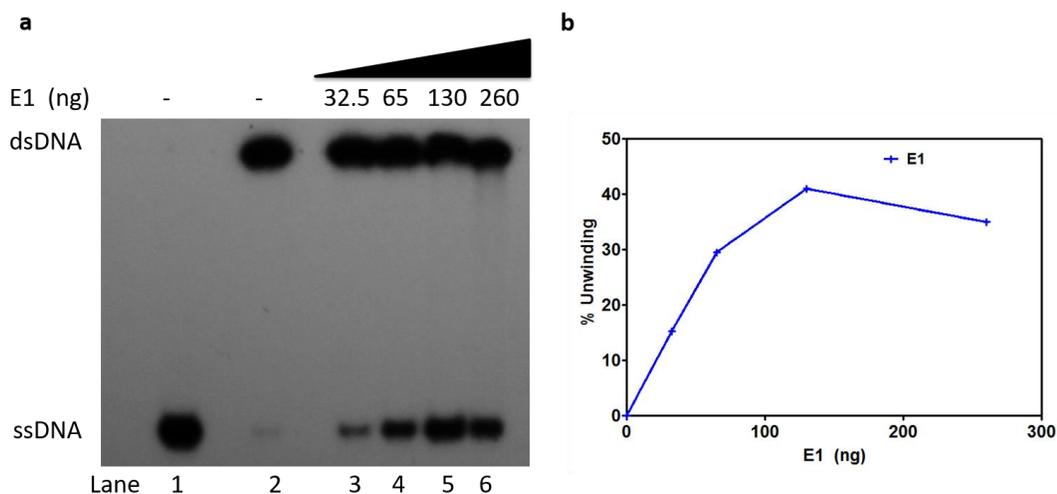
**Figure 21. E1 protein purity and mass spectrometric data.** (a) E1-containing elutions from cobalt resin in Coomassie Blue-stained SDS-PAGE (left) and western using His-probe antibody (right) are shown. E1 protein is evident at predicted size of 75 kDa. “M” represents PageRuler Prestained Protein Ladder (Thermo). (b) The displayed MS/MS (product-ion) spectra is of a 20 amino acid peptide unique to HPV-11 E1, generated after the fragmentation of the purified recombinant HPV-11 E1 used in this study.



**Figure 22. E1 protein preparations did not contain DnaB helicase.** DnaB helicase from *E.coli* was titrated as a positive control in the western using anti-DnaB probe. DnaB was evident at predicted size of 54 kDa (Lanes 2-4). E1-containing elutions from cobalt resin in western using DnaB-probe antibody were also included in DnaB western (Lanes 5-7). Lane 1 represents PageRuler Prestained Protein Ladder (Thermo). E1 elutions from cobalt column were negative for DnaB helicase (Lanes 5-7), confirming the absence of DnaB helicase in the E1 protein sample.

#### **D. Analysis of E1 helicase activity**

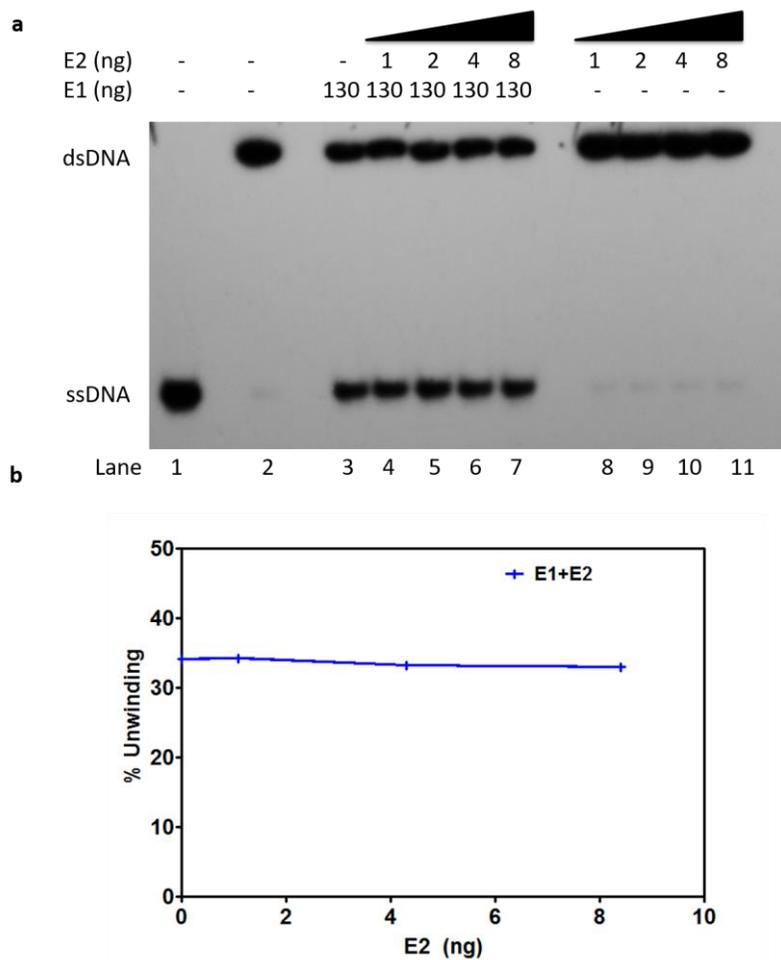
DNA replication requires DNA helicase activity during the elongation phase [89]. Consequently, the DNA helicase activity of E1 was analyzed using a partial duplex DNA substrate containing a 60 bp <sup>32</sup>P-radio-labeled oligonucleotide complementary to a 50 bp sequence of M13mp19, as described in the Materials and Methods. Controls included substrate, containing no E1 enzyme, in heat-denatured form (ssDNA) and in native double-stranded form (Figure 23a, Lanes 1-2). The increase in free displaced single-stranded oligonucleotide with increasing E1 protein titration represented unwinding of the dsDNA (Figure 23a). E1 was able to unwind M13-based substrates efficiently; DNA unwinding was quantified using PhosphorImager analysis (Figure 23b). The results presented in Figure 23 clearly demonstrated that E1 had measurable DNA helicase activity. Inhibition was observed at the highest titration point (260 ng), possibly due salt inhibition (Figure 23b, Lane 6). Together, these results indicated that the purified E1 was enzymatically active.



**Figure 23. DNA helicase activity of E1 protein.** Protein titration of DNA unwinding activity on a 60 bp partial duplex substrate. Details of the assay are given in Materials and Methods. **(a)** Lane 1: heat denatured substrate, no E1 enzyme added. Lane 2: native substrate, no E1 enzyme added. Lanes 3-6: E1 protein concentrations as indicated. Increase in lower single-stranded DNA with E1 titration represents unwinding of the dsDNA; therefore, confirming helicase activity. **(b)** Quantification of E1 helicase activity by PhosphorImager analysis.

### **E. E2 had no effect on E1 DNA helicase activity**

As E1 showed helicase activity, any involvement of E2 in the helicase function of E1 was investigated. The goal of this set of experiments was to explore collaborative activities of these two proteins, since they are both required for initiation of HPV DNA replication. E1 was able to unwind M13-based substrates efficiently, in the absence of E2 (Figure 24, Lane 3). However, even at the highest concentrations of E2 protein evaluated, it had no effect on E1 helicase activity (Figure 24b). As a control, E2 alone was also analyzed for helicase activity (Figure 24a, Lanes 8-11). No helicase activity was observed for E2 in the absence of E1 (Figure 24a, Lanes 8-11). The results presented in Figure 24 clearly demonstrated that E2 had no effect on E1 DNA helicase activity under the conditions investigated. Mechanistically, this suggests that E2 is not required for E1 helicase activity. It is also possible that E2 protein did not physically or functionally appear to interact with E1.



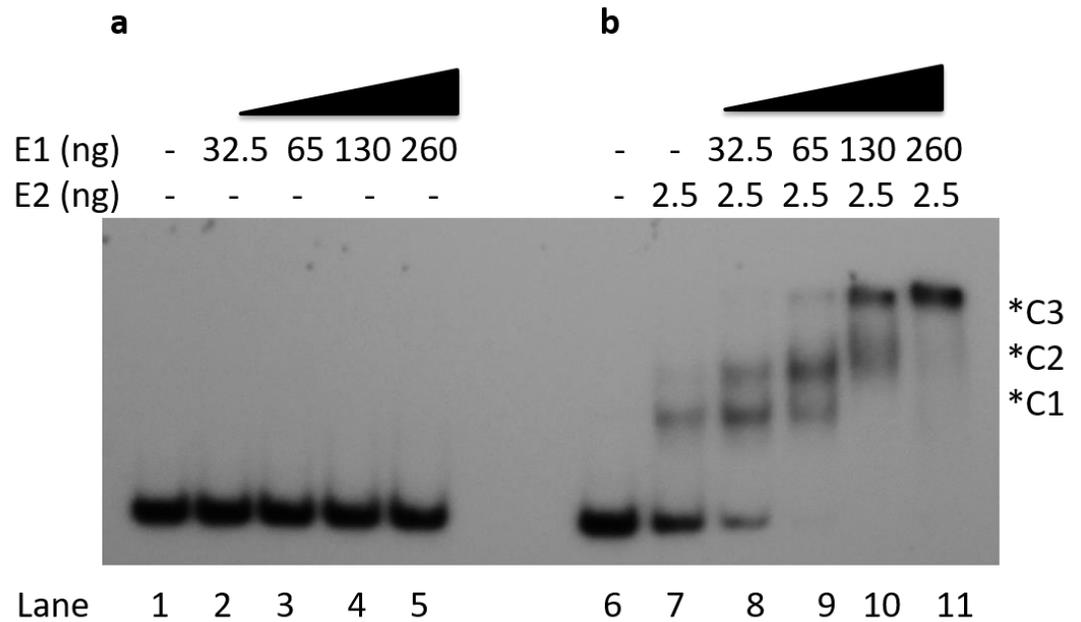
**Figure 24. E2 had no effect on DNA helicase activity of E1.** A standard helicase assay was carried out using increasing amounts of E2, in the presence of 130 ng of E1 protein. **(a)** Lane 1: heat denatured substrate, no E1 enzyme added. Lane 2: native substrate, no E1 enzyme added. Lanes 3-7: E1 protein (130 ng). E1 with titration of E2 in helicase assay amounts, including 1 ng (Lane 4), 2 ng (Lane 5), 4 ng (Lane 6), and 8 ng (Lane 7) of E2 protein. As a control, E2 alone was tested for helicase activity using increasing E2 concentrations, including 1 ng (Lane 8), 2 ng (Lane 9), 4 ng (Lane 10), and 8 ng (Lane 11). **(b)** Quantification of E1 helicase activity in the presence of E2 using PhosphorImager analysis.

## **F. Characterization of E1-E2-DNA complex formation**

### **1. E1 DNA helicase bound to E2-DNA complex but not free double-stranded DNA**

Sequence-specific DNA binding is an important function of replication proteins. E2 and E1 are both required for efficient viral DNA replication. Previous foot-printing studies indicated a possible localized E1 binding to a region of the origin DNA proximal to E2 binding sites 1 and 2 [64]. The binding of E1 to the origin was evaluated using EMSA of DNA fragment containing the native E1 binding site and two E2 binding sites as they appear in the HPV origin, depicted above as E2BS1+2+E1BS (Table 4). The titration of E1 with this oligonucleotide did not result in a slower migrating band; therefore, no DNA binding was observed (Figure 25a, Lanes 2-5). However, E2 was observed to form a complex with the same DNA, as seen in my previous experiments (Figure 25b, Lane 7). As expected, E2 alone produced two complexes (Figure 25b, Lane 7), as mentioned earlier with experiments involving cooperative E2-DNA binding (Figure 7). Complex 1 (C1) is likely an E2 dimer binding to one E2 binding site. Complex 2 (C2) is two E2 dimers bound to both to E2 binding sites. At higher E1 concentrations in the reaction, containing E2 and DNA, formation of a much larger and higher order complex (C3) was observed (Figure 25b, Lanes 8-11). At low concentrations of E1 this complex was not observed (Figure 25b, Lanes 8-9). However, low concentrations of E1 did enhance E2 DNA binding affinity, and reinforced formation of C2 (Figure 25b, Lanes 8-9). Higher E1 protein

concentrations resulted in the super-shifted C3 band, representative of the E1-E2-DNA complex (C3) (Figure 25b, Lanes 10-11). Since DNA binding was not observed with E1 alone, the association of E1 with this DNA is likely by due to its interaction with the E2 protein and not DNA. E1 also enhanced E2 binding to the DNA, suggesting cooperativity of these proteins at the origin. Together, these data demonstrate that the association of E1 with the protein-DNA complex was only observed in the presence of E2.



**Figure 25. E1 binds DNA in the presence of E2.** (a) Assays contained titration of E1 protein alone or (b) combination of standard E2 protein concentration (2.5 ng) with titration of E1 protein using radio-labeled oligonucleotide containing BS1, BS2 and E1 binding site (BS1+2+E1BS). EMSA was performed in the presence of increasing amounts of E1 protein including 32.5 ng, 65 ng, 130 ng, and 260 ng. C1 indicates complex with one E2 dimer, C2 indicates complex with two E2 dimers and C3 refers to complex containing two E2 dimers and E1 protein.

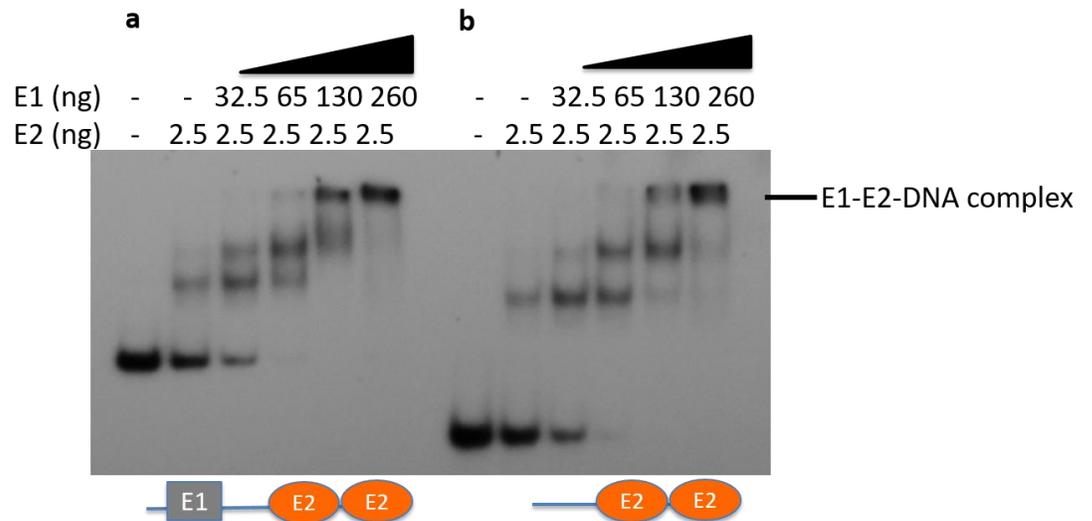
## **2. E1-E2-DNA complex formation does not require the E1 binding site**

Since the results suggested that E2 is required for E1 to bind to the E2-DNA complex, the requirement of the E1 binding site in the formation of the E1-E2-DNA complex was explored. As shown earlier, the E1-E2-DNA complex can form with oligonucleotide BS1+2+E1BS, containing the E1 binding site along with two E2 binding sites (Figure 26a). A similar experiment was performed with oligonucleotide BS1+2 (Table 4), possessing only two E2 binding sites (Figure 26b). The difference in migration of the free probes is due to the difference in radio-labeled substrate size. DNA probe BS1+2+E1BS is 124 bp in length, whereas BS1+2 is 50 bp long. The titration of E1 with E2 and BS1+2 DNA, with no E1 binding site, yielded similar results to the DNA containing the E1 binding site (BS1+2+E1BS). In other words, the E1-E2-DNA complex formation was not dependent on the presence of the E1 binding site (Figure 26).

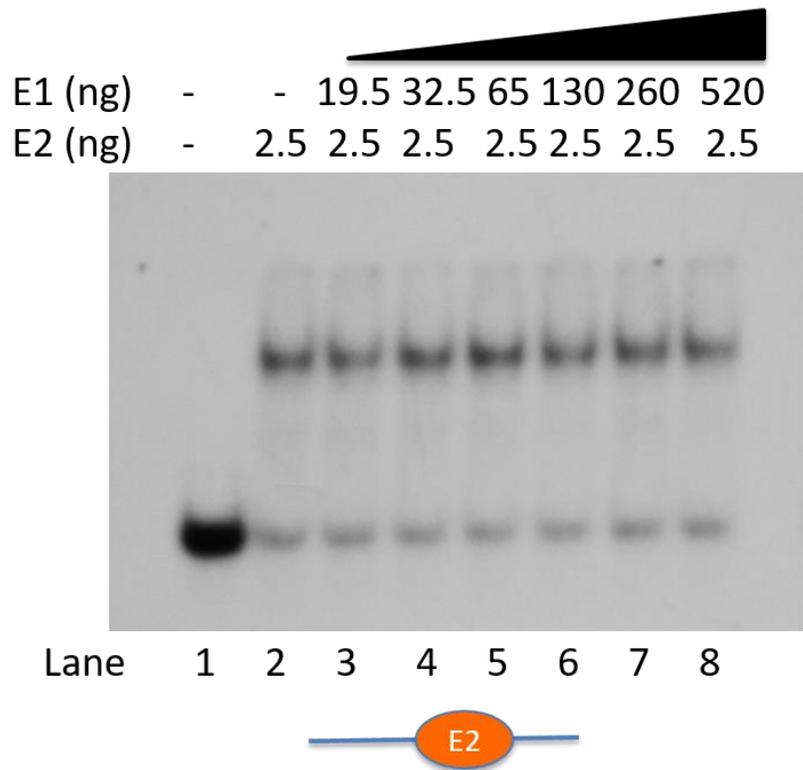
## **3. E1-E2-DNA complex formation required at least two E2 binding sites**

These results suggest that E1 binding to the origin is dependent upon E2 binding to the sites in the origin. This required the evaluation of the roles of each E2 binding site in the E1-E2-DNA complex formation. Hence, E1-E2-DNA complex formation was evaluated with DNA containing a single E2 binding site (Figure 27). E2 formed a single complex with the DNA that contained only one E2 binding site (Figure 27, Lane 2). Titration with E1, even at high concentrations, did not affect E2 binding (Figure 27, Lanes 3-8). In addition, no higher order complex was formed with increasing concentrations of E1 (Figure 27, Lanes 3-8), unlike that observed in

the presence of two E2 binding sites. This suggests that the E1-E2-DNA complex requires at least two E2 binding sites or binding of two E2 dimers. Together, these data suggested that upon binding to two binding sites in the origin DNA, E2 recruits E1 to the origin and provides a platform for E1 binding. In addition, E1 enhances E2 binding to its respective sites in the viral origin. This DNA binding pattern is indicative of cooperative DNA binding of E1 and E2 proteins at the origin. This is in agreement with reports that a minimum of two E2 binding sites are required for HPV replication [90, 91].



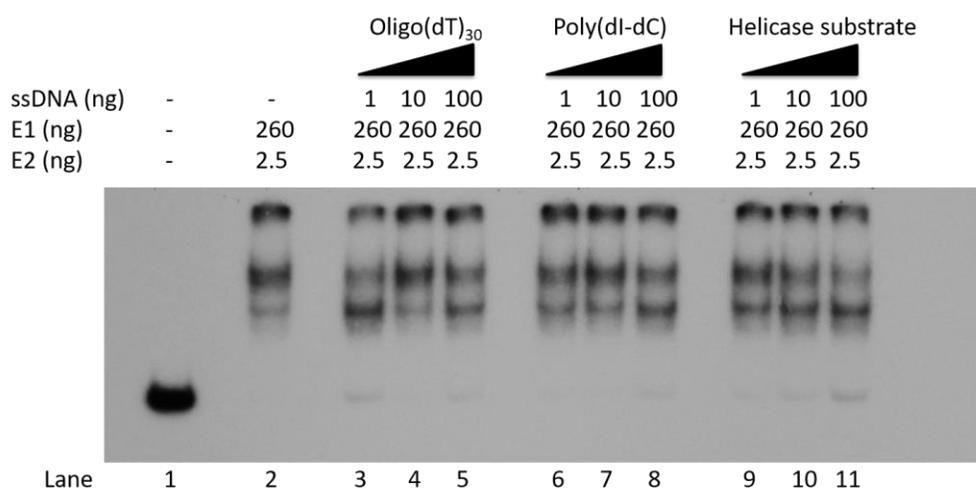
**Figure 26. E1-E2-DNA complex formation does not require the E1 binding site.** Assays contained titration combination of standard E2 protein concentration (2.5 ng) with titration of E1 protein using radio-labeled **(a)** oligonucleotide containing E2 DNA binding sites BS1, BS2 and E1 binding site (BS1+2+E1BS) or **(b)** oligonucleotide containing only BS1 and BS2 (BS1+2). EMSA was performed in the presence of increasing amounts of E1 protein including 32.5 ng, 65 ng, 130 ng, and 260 ng.



**Figure 27. E1-E2-DNA complex formation requires at least two E2 binding sites.** Assays contained titration combination of standard E2 protein concentration (2.5 ng) with titration of E1 protein using radio-labeled oligonucleotide containing only E2 DNA binding site BS4. EMSA was performed in the presence of increasing amounts of E1 protein including 19.5 ng (Lane 3), 32.5 ng (Lane 4), 65 ng (Lane 5), 130 ng (Lane 6), 260 ng (Lane 7) and 520 ng (Lane 8).

### **G. ssDNA had no effect on E1-E2-DNA complex formation**

Since E1 can bind ssDNA, it was addressed whether E1 could bind to ssDNA and E2 simultaneously. Competition experiments were performed to determine if ssDNA inhibited E1-E2-DNA complex formation (Figure 28). DNA substrate specificity of purified E1 was also examined by using multiple ssDNA substrates, including oligo(dT)<sub>30</sub>, poly(dI-dC), and 60 bp oligonucleotide (Primer 13, Table 1) (Figure 28). As seen previously, E2 and E1 formed a complex with DNA, containing no competitor ssDNA (Figure 28, Lane 2). Oligo(dT)<sub>30</sub> ssDNA had very little effect on E1-E2-DNA complex formation, even with the highest concentration tested (Figure 28, Lanes 3-5). Poly(dI-dC) had no effect on the formation of the E1-E2-DNA complex (Figure 28, Lanes 6-8). The helicase substrate, used previously in the helicase assay mentioned above, had a modest effect on E1-E2-DNA formation only at the highest concentration tested (100 ng) (Figure 28, Lanes 9-11). Overall, however, very little inhibition was observed if non-radio-labeled ssDNA was added before mixing E1 and E2 with the radio-labeled probe (Figure 28). Similar results were found when the non-radio-labeled ssDNA was added after or during E1-E2-DNA complex formation. E1 did not show any significant preference for ssDNA substrates. In other words, once E1 bound to E2-DNA complex, it either changed to a conformation that was unfavorable for binding with ssDNA or ssDNA binding did not change its interaction with E2 dimers.



**Figure 28. Effect of ssDNA on E1-E2-DNA complex formation.** Assays contained combination of standard protein concentrations of E2 (2.5 ng) and E1 (260 ng) using radio-labeled oligonucleotide containing BS1, BS2 and E1 binding site (BS1+2+E1BS). EMSA was performed in the absence (Lane 2) or presence of non-radio-labeled competitor ssDNA including Oligo(dT)<sub>30</sub> (Lanes 3-5), poly(dI-dC) (Lanes 6-8), and helicase substrate (Lanes 9-11). Competitor ssDNA were titrated in increasing amounts of 1 ng (Lanes 3, 6, 9), 10 ng (Lanes 4, 7, 10), and 100 ng (Lanes 5, 8, 11).

## **DISCUSSION**

### **Section I: Structural and functional studies on the full-length low-risk HPV E2 protein**

HPV-encoded E2 protein recognizes and binds to specific recognition sequences in the replication origin and initiates viral DNA replication. Most of our current knowledge of the mechanism of activation by E2 is based on studies of full-length bovine papillomavirus or a ~10 kDa HPV E2 DNA-binding domain polypeptide [37, 38, 92]. Although E2 is the modulator of nucleic acid transactions in both BPV and HPV, BPV has more than four E2-binding sites in the origin, suggesting diverse regulatory mechanisms. Multiple sequence alignment of human and bovine E2 proteins demonstrates limited homology between the two proteins. Thus, the contributions of the large N-terminal region of E2 (~31 kDa), representing more than 75% of the protein, on DNA binding could be very significant. Therefore, the native full-length low-risk HPV-11 E2 protein was cloned, expressed, purified and evaluated for its interaction with its binding sites in the HPV origin.

#### **A. Purification and oligomeric structures of E2**

Many mechanistic studies of HPV E2 protein were carried out with the 10 kDa DNA-binding domain of the protein, which is highly soluble, whereas the purification of the full-length E2 protein is very difficult. Thus, development of an optimized expression and purification scheme for HPV E2 was needed to perform mechanistic, structural, and functional studies for the full-length protein. My initial

studies with the purification of E2 protein indicated that E2 is soluble at low concentration ( $\leq 100$   $\mu\text{g/ml}$ ); however, it tends to precipitate with increase in concentration unless the salt concentration is higher. This made it difficult to purify this protein in sufficient quantities. A purification scheme that allows purification of E2 at high salt conditions was developed. Interestingly, DnaA, the *E. coli* origin-binding protein, also requires high salt conditions to maintain solubility in aqueous solution [93].

Our knowledge of HPV E2 oligomeric structure is based primarily on studies involving individual transactivation and/or DNA-binding domain polypeptides [94-96]. Based on crystallographic data derived from both the transactivation domain fragment of high-risk HPV-16 E2 and the DNA-binding domain fragment of HPV-6 E2, they appear to exist as dimers [94-96]. However, it is not clear whether this is unique to their physical states in crystal form. At present, the oligomeric structure of full-length HPV E2 in solution remains unknown. Since both the transactivation and DNA-binding domain fragments form dimer in crystal, E2 may potentially form dimers to multimers, which remains to be investigated. Using DLS, E2 appeared to exist as a dimer in solution at low ( $\leq 0.1$   $\text{mg/ml}$ ) protein concentration (Figure 2b). Higher concentrations of the protein resulted in a mixed oligomeric population, suggesting concentration-dependent self-aggregation of E2. *E. coli* DnaA portrays similar self-aggregation, which may be an inherent property of origin-binding proteins and may be necessary for DNA looping [97].

## **B. Mechanism of E2-DNA interaction in the HPV origin**

E2 protein modulates HPV DNA replication and transcription through its binding to four DNA recognition sequences in the origin (Figure 1b). Recombinant full-length low-risk E2 protein bound all four DNA binding sites with reasonably high affinity in the EMSA analysis (Figure 3). The results indicated that the E2 protein bound with highest affinity to BS4 ( $K_D = 3.9$  nM). The binding to all four sites observed with binding affinity hierarchy was as follows: BS4>BS1≈BS2>BS3 with  $K_D$  values ranging from 3.9 to 8.9 nM (Table 2). The higher affinity binding of E2 to BS4 implies that this site will be predominantly occupied even at low concentration of E2 protein. Subsequently, higher concentrations will allow E2 to bind BS1 and BS2, sites involved in the repression of transcription. This is consistent with previous studies demonstrating transcriptional repression by increased E2 expression [98]. My studies demonstrated that E2 bound BS3 with lowest affinity. It was earlier shown that binding to BS3 is required for DNA replication [20], viewed in the context of my results, this suggests that E2 binding to BS4, BS1, BS2 and finally BS3 will lead to the activation of viral DNA replication origin. The order of binding affinity may be important for preferential and sequential occupancy of DNA binding sites by E2 especially at low E2 expression levels in the host cells [99]. Since regulation of the cellular functions by E2 is dependent on its ability to bind DNA, the binding affinities to these four sites appear critical in controlling viral processes, such as transcription and replication. Similarly, *E. coli* DnaA protein binds DNA sequences located in prokaryotic origins with varied affinities and may be a common

theme amongst origin-binding proteins [30, 100]. Incidentally, my results with full-length E2 protein are strikingly different than that observed with studies carried out with the ~ 10 kDa HPV E2 DNA-binding domain fragment. The highest binding affinity was observed with BS3, intermediate binding with BS1 and BS2, and weak binding with BS4 [92]. With the full-length low-risk E2, highest binding to BS4 and least binding to BS3 was observed. Thus, together these results suggest that the 31 kDa transactivation/hinge domain may play a significant role in modulation of E2 DNA binding and binding affinity. I also found that, because of the association properties of E2 protein, measurement of DNA binding affinity using fluorescence anisotropy, that my laboratory has used previously, could be very difficult [100]. The self-association of the HPV E2 automatically increases the anisotropy generating aberrant  $K_D$  values. Therefore, EMSA and its quantitative analysis for determining DNA binding constants for E2 were used. This is the first study to evaluate the full-length low-risk HPV-11 E2 protein and the DNA binding affinities for its four E2 binding sites.

### **C. Effect of spacer and flanking regions on E2 recognition**

Sequence analysis of the four E2 DNA binding sites, as shown in Table 2, indicates that there are two sequences (ACCG and CGGT) in each E2 half-site of the palindrome which remain highly conserved amongst the four binding sites. However, the spacer sequence appears to change in BS3. My studies with the full-length E2 indicated that the E2 binding affinity is lower for BS3. Thus, this study examined whether spacer sequence variations could potentially account for the E2-DNA

binding affinities. A common feature amongst the low-risk HPV genotypes was the presence of a guanine in BS3 spacer sequences, whereas other E2 binding sites contained only adenine or thymine. EMSA analysis indicated that substitution of thymine with guanine led to a decrease in the E2-DNA affinity (Figure 5f). This may partially explain the lower affinity of E2 for BS3, containing the least A/T-rich spacer of all the other sites. This is consistent with the proposal that the intrinsic shape or flexibility of the N<sub>4</sub> non-contacted spacer within the E2 binding site plays an indirect role in determining the E2 binding affinity [101].

The sequences flanking the consensus sequence were also evaluated for their effect on E2 binding. The results indicated that the flanking sequences did have an effect on E2 binding affinity in the EMSA analysis (Figure 6). When the flanking sequences of BS3 substituted those of BS1, there was a decrease in binding affinity. This may partially explain the lowered affinity of E2 for BS3. The flanking sequence of each E2 binding site is different, which may have a role in the regulation of E2 binding and affinity. Together, results of these studies appeared to support the notion that the unique regions within and surrounding individual, naturally occurring E2 DNA binding sites promote differential E2 binding affinities, which may be a strategic feature of E2-mediated regulation. This is the first study to examine the effect of these regions on the full-length low-risk E2 binding affinity.

#### **D. Cooperativity among E2 binding sites**

The spatial arrangement of E2 binding sites in the viral origin is highly conserved among papillomaviruses. The importance of this consistent layout in the

context of viral regulation remains unknown. The short distance between two binding sites could permit cooperative interaction between two E2 dimers, and subsequent higher levels of regulation at low concentrations of full-length E2 proteins.

Cooperativity in DNA binding was observed with low-risk E2 when there were two DNA binding sites in close proximity (Figure 7). Similarly, BPV-1 E2 protein and *E. coli* DnaA also display cooperative binding to two adjacent DNA boxes [102, 103]. The higher order complex with HPV E2 protein was seen only with two E2 dimers binding to adjacent E2 binding sites. It was assumed that this data represents two E2 dimers binding side-by-side with a possible cooperative interaction. These two E2 binding sites also resulted in the formation of a third complex (C3), indicative of a much larger, multimeric E2-DNA complex. This suggests possible DNA-induced E2 trimer or multimer formation. Due to the high specificity of the E2 protein, this cooperativity is not likely due to the size of the DNA. Oligonucleotides “BS1” and “BS1+2”, were the same length but had different binding patterns, indicating that the additional E2 site in “BS1+2” was responsible. Given the cooperative nature of E2-DNA binding and the existence of a consensus binding sequence, the initial event may be the interaction with this sequence, followed by the addition of dimers in both directions to form a larger complex. Interestingly, an electron microscopy study provided evidence that HPV E2 binding to the origin leads to the formation of protein multimers that create DNA loops in the presence of all four E2 binding sites [33]. This is known to be an evolutionarily

conserved essential first step in origin activation and the initiation of DNA replication [30]. Hence, oligomerization and cooperativity of full-length origin-binding proteins may be a necessary pre-requisite for DNA looping and the activation of the replication origin.

### **E. Ionic and hydrophobic interactions in E2-DNA complex formation**

The nature of the E2-DNA interaction using thermodynamic analyses helped us characterize E2-mediated regulation. X-ray crystal structure of the truncated HPV E2 binding site indicated that increased salt concentration increased the angle of DNA curvature, which may point to structural flexibility [83]. It was suggested that this curvature may provide the appropriate three-dimensional structure for protein binding [83]. This is consistent with our results that show an increase in binding with the addition of salt (Figure 8). The observed salt dependence suggests that DNA binding of E2 may be mostly entropy-driven or hydrophobic in nature. In order to understand the thermodynamics of E2-DNA binding interactions, E2 binding to BS1 was studied at different temperatures (Figure 9a-e). These results indicate that DNA binding was not significantly influenced by temperature (Figure 9f). E2 is capable of binding at a wide range of temperatures. Together, these studies suggest the E2-DNA interaction is likely driven by hydrophobic interactions and less by ionic interactions.

## **Section II: Identification of SNV in high-risk HPV origin and its effects on high-risk E2 function**

HPV encompasses a large family of viruses that range from benign to highly carcinogenic, associated with genital, oropharyngeal and a variety of other cancer forms [6, 10, 11]. Sequences of HPV viral genomes, regardless of the carcinogenicity, are highly homologous. Consequently, the ability of some HPV types to cause carcinogenic transformation of infected cells remains a paradox. Earlier studies indicate that the benign and carcinogenic forms of HPV differ in DNA replication efficiency of the episomal viral DNA [20, 47]. However, the basis of this difference also remains unknown.

### **A. Hierarchy of high-risk E2 binding to its cognate E2 binding sites**

HPV-encoded E2 protein is pivotal in DNA replication as it initiates HPV DNA replication through binding to four specific DNA binding motifs in the replication origin. My results demonstrated that the high-risk (HPV-16) E2 protein bound to three of its four binding sites efficiently. It bound BS4 with highest affinity ( $K_D = 3.8$  nM) and the binding hierarchy was found to be as follows: BS4>BS2>BS1>>BS3 with  $K_D$  values ranging from 3.8 to 21 nM (Figure 12). It bound BS3 with at least ~5.5 fold lower affinity than BS4. More uniform binding to all four binding sites was observed by low-risk E2 protein as mentioned above (Figure 3).

## B. Effect of SNV on E2 DNA binding affinity

In order to determine whether the observed origin binding characteristics of E2 protein are unique to HPV-16 or a common feature for all high-risk HPVs, a thorough analysis of E2 binding to BS3 from several well-known high-risk HPVs was carried out. The sequences of E2 binding sites from different high-risk types were also evaluated for their effect on E2 binding. The results with HPV-18 BS3 demonstrated a seven-fold decrease ( $K_D = 27$  nM) in E2 binding affinity compared to E2 affinity for BS4 (Figure 13c). A substitution of cytosine in the conserved E2 half-site (CGGT) with adenine aGGT in HPV-18 BS3 was also observed. A critical evaluation of DNA sequences of HPV origins indicated that variation (aGGT) in BS3 is common amongst several high-risk types, such as HPV-18, HPV-33, HPV-39, HPV-45, and HPV-59. DNA binding by E2 to sequences in other high-risk group members was also examined, such as HPV-31(BS2) and HPV-35 (BS3), as shown in Fig. 14b and 14c, respectively. Each of these variant binding sites demonstrated much weaker binding to E2, even lower than HPV-16 BS3. HPV-31 BS2 has a tGGT variation whereas HPV-35 BS3 has a gGaT variation, both of which have reduced binding affinities to E2. Together, results of these studies appeared to support the notion that a nucleotide variation in the consensus sequence, as observed in these HPVs, appeared to result in significantly decreased E2 binding affinity, which may be a key feature of high-risk genotypes. This is the first time this type of analysis has been performed. It is believed that the bipartite DNA binding by the E2 dimer takes

place with E2 monomers binding to both half-site (ACCG or CGGT) of the palindromic consensus sequence [104]; therefore, any variation in this sequence will alter the E2-DNA complex formation as well as binding affinity.

### **C. Bioinformatics analysis of E2 binding sites**

These findings prompted us to perform a comprehensive analysis of the E2 DNA binding site sequences from all known HPVs. The International Agency for Research on Cancer categorizes HPVs as high-risk/carcinogenic, probable high-risk/carcinogenic and low-risk/non-carcinogenic. These sequence data were analyzed and presented as a bioinformatics analysis of the E2 binding sites in the origin of HPVs in all three groups in Figure 16. Each E2 binding site within the origin is a 12 bp palindrome with a four bp intervening sequence separating the conserved ACCG and CGGT sequences at the 5' and 3' ends respectively. Sequence alignment of the four E2 DNA binding sites indicates that the two half-site motifs (ACCG and CGGT) in the binding site palindrome remained highly conserved amongst all four binding sites of the low-risk group (Figure 16a). All high-risk types, that were examined, possess at least one single nucleotide variation in the 3' half-site (CGGT) (Figure 16b). These SNVs were most commonly observed in BS3 of the high-risk group, as shown in Figure 16b. However, SNVs in E2 half-sites were also occasionally observed in carcinogenic and probable carcinogenic subtypes at the same position in BS2 as well. Similar SNVs in BS2 were never observed in any of the known low-risk genotypes (Figure 16a). Hence, a common feature amongst the high-risk genotypes was demonstrated, which was that they harbor a SNV in an E2 binding site (BS2 or

BS3) with concomitant reduced DNA binding affinity, suggesting a pathway of development of HPV carcinogenicity. The SNVs in E2 binding sites in the origin appeared to correlate well with the level of HPV-associated cancer risk. This is the first time this type of analysis has been performed.

#### **D. Effect of E2 multimer formation on the origin with one or more SNVs**

This study has also identified the recurrence of the same SNVs at one or more binding sites in several “probable” high-risk HPVs. High-risk HPV-51 and probable high-risk HPV-82 both have nucleotide variations in BS2 (**gGGT**) and BS3 (**gtGT**) concurrently. To explore the functional relevance, the effect of multiple SNVs on E2 binding was investigated by incorporating them into a single oligonucleotide probe for use in EMSA. The spatial arrangement and multiple E2 binding sites in the viral origin are highly conserved among HPVs and are likely significant in the activation of DNA replication initiation. Multi-protein E2-DNA complex formation were assessed with a double-stranded oligonucleotide probe containing three binding sites, BS1, BS2, and BS3, *in vitro*. Three separate E2-DNA complexes (C1, C2, and C3) were observed with the DNA containing three consensus E2 binding sites, representative of the low-risk HPV-11 origin (Figure 15a). C1 was assumed to represent the initial binding of E2 dimer to the highest affinity site of the three available sites. Higher E2 concentration allowed for binding to the second highest affinity site, with a possible cooperative interaction (Figure 15a). This resulted in C2 formation, signifying two E2 dimers binding side-by-side. Further increase in E2 concentration facilitated the formation of C3, which was likely the binding of three

dimers on the DNA (Figure 15a). However, alteration of wild-type BS3, with the nucleotide variation found in high-risk HPV-16 BS3, resulted in a decrease of C3 formation (Figure 15b). Simultaneous SNVs in both BS2 and BS3, as in HPV-51, led to attenuation in C2 and C3 formation, most likely as a consequence of the BS2 and BS3 nucleotide variations (Figure 15c). This is the first time this type of analysis has been performed. This data appeared to suggest that HPV origins with more than one SNV site may have less E2-DNA complex formation, and I speculate may be less persistent due to decreased DNA replication efficiency, consistent with recent epidemiological data [85].

#### **E. Low-risk E2 protein cannot bind E2 binding site containing high-risk SNV**

Binding sites from all low-risk genotypes display complete half-site consensus sequences in all four binding sites as shown in Fig. 16a. No variation was observed in the consensus sequence conforming to ACCG(N)<sub>4</sub>CGGT. Analysis of 11 established high-risk genotypes revealed variation in the consensus half-site of at least one of the E2 binding sites (Fig. 16b). The SNV in the conserved half-site was examined to determine the effect on DNA binding by low-risk E2 protein. Interestingly, no DNA binding by low-risk E2 protein was observed with oligonucleotides containing the high-risk SNV in the E2 binding site (Figure 18b-d). However, the low-risk E2 protein was observed to bind other low-risk E2 binding sites (Figure 17). This suggests that a substitution of cytosine in the CGGT sequence, observed in high-risk BS3, led to dramatic attenuation of the binding affinity of low-risk E2. It is tempting to speculate that high-risk E2 protein has evolutionarily

adapted to be able to bind the variant BS3, albeit at a lower level than the other E2 binding site. Low-risk viruses do not contain this variation in BS3 (Figure 16a); therefore, there is no need for the low-risk virus to contain the necessary adjustment in its E2 protein. However, an adaptation may have been necessary to propagate the high-risk virus. In other words, in order to replicate, over time the virus must have modified the E2 protein to adjust for this variation in the HPV origin to some extent.

#### **F. Structural analysis of E2 on HPV origin by Atomic Force Microscopy (AFM)**

Although previous studies have visualized individual E2-DNA structures, [33, 86, 87], there have not been any comparative analyses of the low-risk and high-risk E2-DNA structures. The results with AFM presented in this thesis demonstrated clear differences between structures of both low-risk and high-risk E2-DNA complexes. Two linear DNA substrates containing either HPV-11 or HPV-16 long control regions were used as substrates for their respective E2 proteins (Figure 19). Using AFM, low-risk E2-DNA complexes formed structures with E2 bound mostly at the end of the DNA molecule (Figure 20b, Figure 20c). The observed position of the low-risk E2 was consistent with the known locations of the BS1, BS2, and BS3 cluster (Figure 19a). E2 binding to these three sites may lead to the cooperative binding of E2 and then formation of a larger complex. This complex was likely three E2 dimers on the DNA. Some low-risk E2 was observed as large masses bound to the DNA (Figure 20c). It is possible that this complex contained multiple E2 dimers. When E2 is bound to BS1, BS2, and BS3, it is possible that the DNA undergoes great torsional strain and may facilitate the melting of the DNA. The putative E1 binding

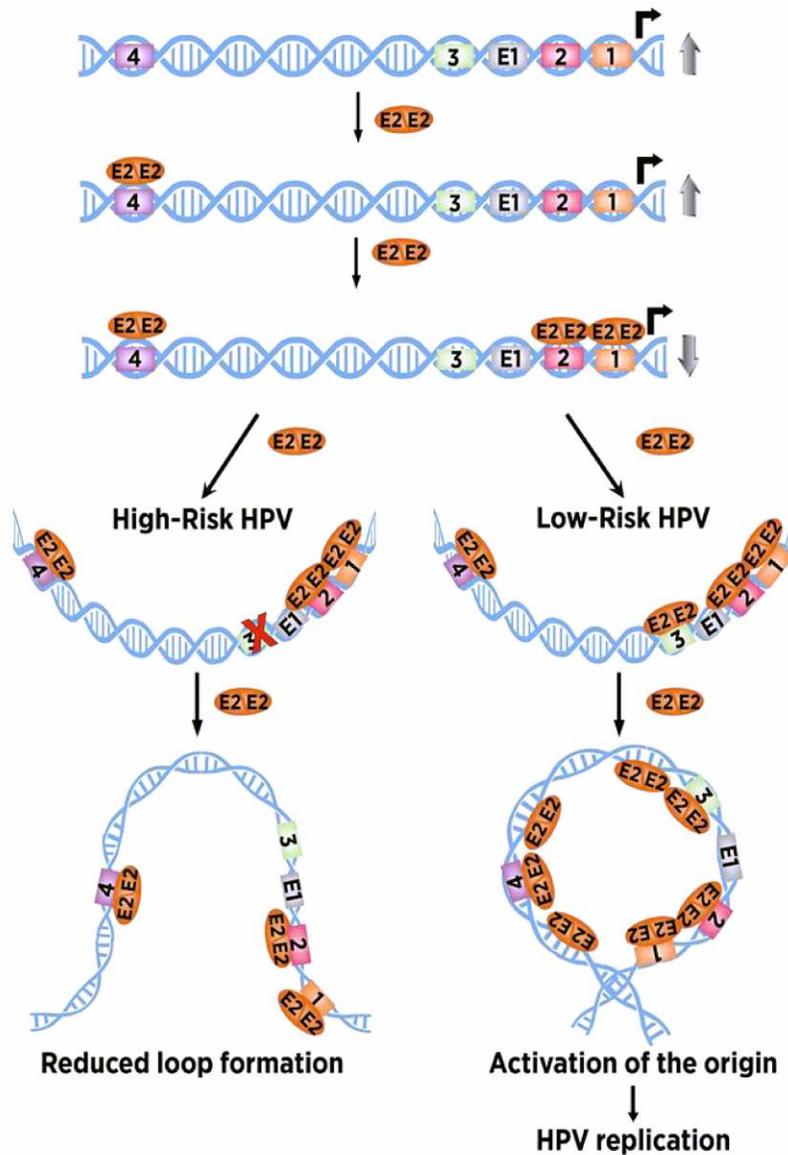
site is located between BS2 and BS3, and this remodeling of the origin by E2 could facilitate recruitment of the E1 protein and melting of the DNA during viral replication. This process is reminiscent of the *E. coli* replication initiator protein, DnaA, binding to its origin and forming a higher-order complex, during the initiation of DNA replication [30-33]. Incidentally, the lambda bacteriophage O protein and eukaryotic origin recognition complex also recognize and remodel their respective origins in a similar fashion [34-36].

Unlike the low-risk particles, high-risk E2 seemed to form a smaller complex on the DNA (Figure 20e). Also, the location of high-risk E2 on the DNA was different than the position of the low-risk E2 on the DNA (Figure 20e). More specifically, high-risk E2-DNA complexes were frequently observed with E2 bound DNA in the center of the DNA substrate (Figure 20e). This position was consistent with the known location of high affinity BS4 on this DNA substrate (Figure 19a). This high-risk E2-DNA complex was likely a single E2 dimer bound to BS4. Unlike the low-risk E2-DNA complex, the high-risk E2 was binding in a linear fashion to the DNA, which likely does not favor initiation of replication. In this scenario, E2 may not facilitate the recruitment of E1 as well as DNA melting.

#### **G. A putative model for HPV E2 DNA binding and activation of replication at the low-risk and high-risk origins**

Based on the evidence presented here, a scheme for the interplay of E2 and its binding sites in HPV DNA replication and modulation of carcinogenicity was proposed (Figure 29). The highest affinity site, BS4, will be predominantly bound at low concentration of high-risk E2 protein because of its low  $K_D$  (Figure 29). Notably,

this is the only binding site which was observed to have no SNV from the bioinformatics study (Figure 16). Increasing concentrations of E2 should promote E2 binding to BS1 and BS2, which are also involved in transcriptional repression of *E6* and *E7* oncogenes. E2 acts as the repressor protein, inhibiting transcription of oncogenes *E6* and *E7*. This is in agreement with previous studies demonstrating the repression of *E6* and *E7* transcription by increased E2 expression [98]. The presence of E2 at BS1 and BS2 hinders the binding of cellular transcription factors Sp1 and TFIID [42-45]. This is reminiscent of the lac operon bound by the lac repressor, inhibiting the transcription of lac operon genes [105]. A recent electron microscopy study found that the inclusion of all four E2 binding sites are required for DNA loop formation, a prerequisite for helicase loading and the initiation of DNA replication - a model similar to that demonstrated for *E. coli* DNA replication by Bramhill and Kornberg [30, 31, 33]. Based on the data presented here and previous studies, I propose that the decreased binding of BS3 or BS2, due to SNV in high-risk genotypes, would lead to attenuated initiation of DNA replication as a consequence of decreased E2-DNA complex formation. HPV types with higher affinity E2 sites, mainly BS3, may be linked to a lower frequency in carcinogenesis. The low-risk HPVs have high binding affinity for all four binding sites including BS3 and are also rarely found in cancerous lesions. Earlier reports showed that low-risk HPVs replicated better than high-risk types [20]. High-risk HPV-16 specifically displayed less initial viral amplification than other high-risk types in replication experiments [47].



**Fig. 29. A putative model for HPV E2 DNA binding and activation of replication at the low-risk and high-risk origins.** The structure of the viral DNA replication origin with E2 binding sites is shown (not drawn to scale) in this illustration. E2 binding sites 1-4 are represented by individual boxes on the DNA specified as “1” (BS1), “2” (BS2), “3” (BS3) and “4” (BS4) in their relative positions (not drawn to scale) to the E6/E7 promoter shown as a bent arrow pointing to the right. The E1 binding site (E1BS) is also included in the native origin between BS2 and BS3. At

low E2 concentration, the highest affinity binding site (BS4) would be bound by the dimeric E2 protein. Active transcription is represented by the gray arrow pointing upward. Increasing E2 concentration would allow for the intermediate affinity sites (BS1 and BS2) to then be bound by E2 dimers. Repressed transcription is indicated by gray arrow pointing downward. At the low-risk origin, higher E2 concentration would allow for the efficient binding of BS3 and subsequent bending of the DNA. In the low-risk scenario, further E2 interaction with the origin facilitates remodeling into a looped DNA structure. This is a critical step for the activation of the origin and subsequent HPV replication. The high-risk origin contains a SNV in BS3, which results in ineffective binding to BS3 in comparison to the binding of low-risk BS3. In the high-risk origin, the inefficient binding to the final E2 binding site (with the SNV) results in less bending of the DNA and reduced loop formation. In comparison to low-risk HPV, there may be diminished initiation of DNA replication in high-risk HPVs.

High-risk HPV integration into the host genome is tightly linked to the progression of cervical cancer [52]. Integration often results in the disruption of the E2 gene [55]. I speculate that the decreased replication of high-risk HPVs could be linked to lower levels of E2, resulting in higher expression of E6 and E7 oncoproteins, and in turn HPV-linked carcinogenesis. Carcinogenesis requires both HPV integration and the loss of inhibitory E2-expressing episomes, as suggested by a current report [51]. In addition, a recent study has shown that efficient viral DNA replication inactivates transcription of viral oncogenes [106]. I believe that efficient replication of episomal HPV DNA results in high gene dosage and enhanced expression of E2 and higher repression of E6 and E7 expression. It is tempting to speculate that HPV episomal replication may act as a preventative mechanism for HPV-linked cancer. HPVs that replicate very efficiently should have higher levels of episomal viral DNA and E2 protein, which suppresses HPV-mediated oncogenesis. On the other hand, HPVs which replicate extremely poorly may cause oncogenesis, but produce fewer infective virions and a lower rate of HPV infection. Thus, HPVs which have an intermediate efficiency in replication may cause carcinogenesis and produce enough virions to continue active cycles of HPV infection. The latter are likely the HPVs frequently observed in patients diagnosed with cancer. HPV-16 may confer the optimal balance of E2 binding necessary for both persistence and replication required for carcinogenesis.

### **Section III: E1-E2 interaction at the origin**

The development of new anti-virals against HPV is limited due to the difficulty of growing the virus in culture and producing the HPV proteins *in vitro*. E1 protein is required for replication of HPV; therefore, understanding its mechanism of action is essential for improved anti-viral therapies. Most of our current knowledge of the mechanism of activation by E1 is based on studies of full-length bovine papillomavirus or truncated HPV E1 polypeptides [68]. Multiple sequence alignment of human and bovine papillomavirus E1 proteins demonstrates limited homology between the two proteins. Therefore, the native full-length low-risk HPV-11 E1 protein was cloned, expressed, purified and evaluated for its interaction with E2 and the HPV origin.

#### **A. Purification of HPV E1 DNA helicase**

A comprehensive characterization of E1 and E2 activity at the origin is lacking due to the difficulty in isolating sufficient quantities of purified E1 protein. Thus, development of an optimized expression and purification scheme for HPV E1 was needed to perform mechanistic, structural, and functional studies for the full-length protein. In my work, the full-length recombinant HPV-11 E1 protein was purified to reveal its biochemical and functional properties. E1 had a tendency to precipitate with increase in concentration, which was overcome in the presence of high salt concentration. This made it difficult to isolate this protein in sufficient quantities. A purification scheme was developed that allows purification of E1 at

high salt conditions. Interestingly, E2 also required high salt conditions to maintain solubility in aqueous solution.

### **B. DNA unwinding activity of E1 helicase**

To further confirm E1 was active, a standard DNA helicase assay was used. DNA unwinding was determined by the release of  $^{32}\text{P}$ -labeled 60-mer from partially duplex M13mp19 ssDNA. E1 was able to unwind the 60 bp duplex and the unwinding increased with increasing amounts of protein (Figure 23). Helicase activity was dependent on E1 protein concentration and sensitive to high salt concentrations (Figure 23). Together, these results indicated that the purified E1 was active, by its known enzymatic function, helicase activity.

### **C. Effect of E2 on E1 Helicase activity**

Since E1 and E2 are required for viral DNA replication, this prompted me to examine whether E2 could have an effect on E1 helicase activity. However, even the highest concentrations of E2 protein evaluated had no effect on E1 helicase activity (Figure 24b). As a control, E2 alone was also analyzed for helicase activity (Figure 24a, Lanes 8-11). No helicase activity was observed for E2 in the absence of E1 (Figure 24a, Lanes 8-11). The results presented in Figure 24 clearly demonstrated that E2 had no effect on E1 DNA helicase activity. Mechanistically, this suggests that E2 may be required prior to E1 helicase activity, possibly in an earlier step in DNA replication initiation.

## **D. Characterization of E1-E2-DNA complex formation**

### **1. E1 DNA helicase bound to E2-DNA complex but not free double-stranded DNA**

Sequence-specific DNA binding is an important function of replication proteins. E2 and E1 are both required for efficient viral DNA replication. Previous foot-printing assays indicated E1 binding to a region of the origin very close to two E2 binding sites and was termed as E1 binding site or E1BS [64]. The binding of E1 to the origin was examined using EMSA of DNA fragment containing the native E1BS and two E2 binding sites as they appear in the HPV origin, (Figure 25). The titration of E1 in the binding assay with this oligonucleotide did not result in a slower migrating band of E1-DNA complex (Figure 25a, Lanes 2-5). However, E2 was observed to form a complex with the same DNA, as seen in my previous experiments (Figure 25b, Lane 7). As expected, E2 alone produced two complexes (Figure 25b, Lane 7), as mentioned earlier with experiments involving cooperative E2-DNA binding (Figure 7). Complex 1 (C1) is likely an E2 dimer binding to one E2 binding site. Complex 2 (C2) is two E2 dimers, each binding to an E2 binding sites. The increasing E1 concentrations in the reaction, containing E2 and DNA, led to the formation of a much larger and higher order complex (C3) (Figure 25b, Lanes 8-11). At low concentrations of E1 this complex was not observed (Figure 25b, Lanes 8-9). Higher E1 protein concentrations resulted in the super-shifted C3 band, representative of the E1-E2-DNA complex (C3) (Figure 25b, Lanes 10-11). This association of E1

with the DNA was only observed in the presence of E2. These results demonstrate that, in the presence of E2, E1 forms a complex with the DNA, which otherwise does not form. This indicates a critical activity of E2 for DNA replication. E2 is likely responsible for recruiting E1 to the HPV origin. Also, the fact that the low-risk E1 protein can interact with high-risk E2 protein argues that E1 functions are qualitatively conserved.

## **2. E1-E2-DNA complex formation did not require the E1 binding site**

Since the results suggested that E2 is required for E1 to bind to the E2-DNA complex, the requirement of the E1 binding site in the formation of the E1-E2-DNA complex was explored. As shown earlier, the E1-E2-DNA complex can form with oligonucleotide “BS1+2+E1BS”, containing the E1 binding site and two E2 binding sites (Figure 26a). A similar experiment with oligonucleotide “BS1+2” (Table 4), possessing only the two E2 binding sites, was performed (Figure 26b). The titration of E1 with E2 and this DNA, with no E1 binding site, yielded similar results to the DNA with the E1 binding site. In other words, the E1-E2-DNA complex was not dependent on the presence of the E1 binding site (Figure 26).

## **3. E1-E2-DNA complex formation required at least two E2 binding sites**

These results suggested that E1 binding to the origin was dependent upon E2 binding to the sites in the origin. E1-E2-DNA complex formation was evaluated with DNA containing a single E2 binding site (Figure 27). Although E2 alone formed a single complex with the DNA, titration with E1 did not affect E2 binding (Figure 27,

Lanes 3-8). In addition, no higher order complex was formed with increasing concentrations of E1 (Figure 27, Lanes 3-8). This suggests that the E1-E2-DNA complex requires at least two E2 binding sites or binding of two E2 dimers. Together, these data suggest that E1 requires two E2 dimers in close proximity on the origin to provide a platform for E1 binding. This DNA binding pattern is indicative of cooperative DNA binding of E1 and E2 proteins at the origin. This result provides a mechanistic basis of the finding that that a minimum of two E2 binding sites are required for initiation of HPV replication [90, 91].

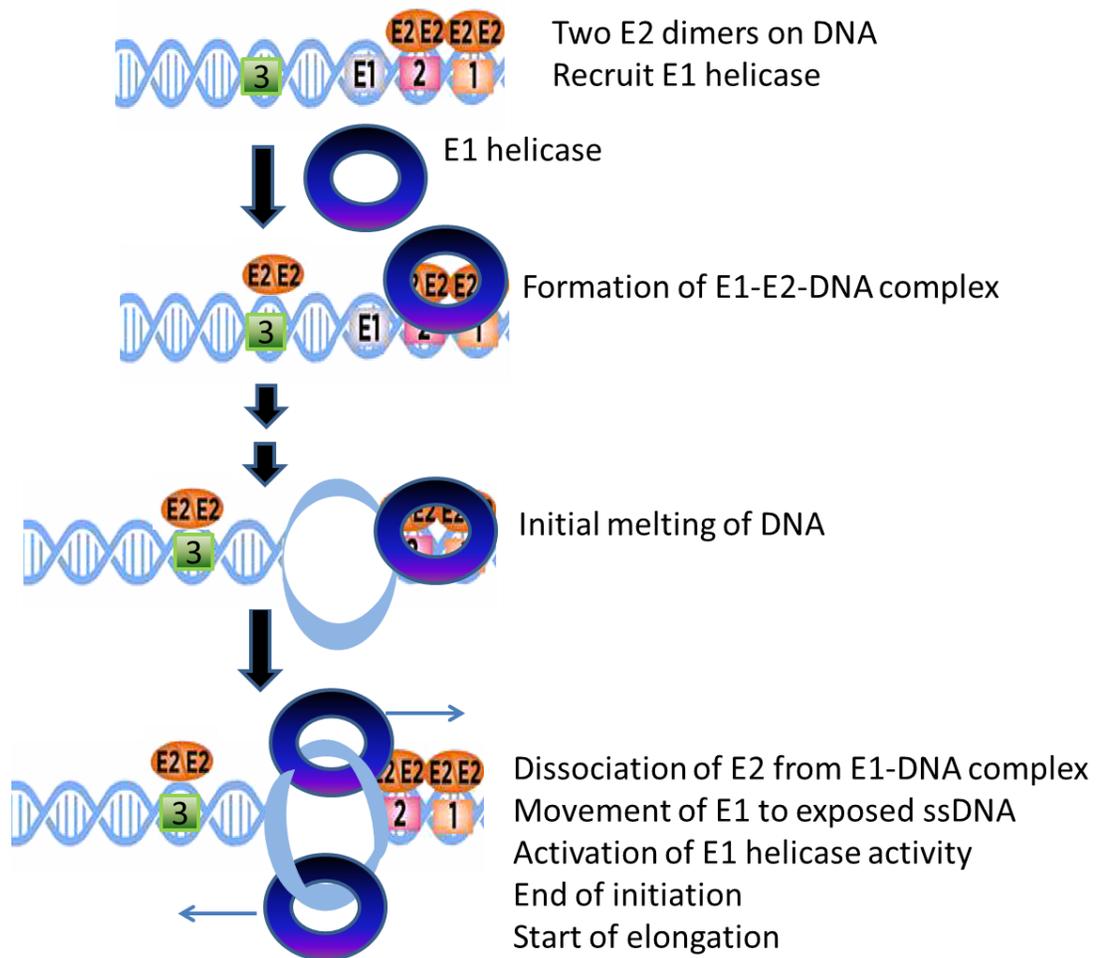
#### **E. Role of ssDNA on E1-E2-DNA complex**

Since E1 can also bind ssDNA, it was addressed whether E1 could bind to ssDNA and E2 simultaneously. Competition experiments were performed to determine if ssDNA inhibited E1-E2-DNA complex formation (Figure 28). Oligo(dT)<sub>30</sub> ssDNA had very little effect on E1-E2-DNA complex formation, even with the highest concentration tested (Figure 28, Lanes 3-5). Poly(dI-dC) had no effect on the formation of the E1-E2-DNA complex (Figure 28, Lanes 6-8). The helicase substrate, used previously in the helicase assay mentioned above, had a modest effect on E1-E2-DNA formation only at the highest concentration tested (100 ng) (Figure 28, Lanes 9-11). Overall, however, very little inhibition was observed if non-radio-labeled ssDNA was to the reaction with E1, E2 and the radio-labeled probe (Figure 28). In other words, once E1 is bound to E2 and DNA, it likely changes conformation that is unfavorable for binding with ssDNA.

## **F. Putative model of E1-E2 interaction at the origin for initiation of HPV replication**

Based on the evidence presented here, a scheme for the interplay of E1 and E2 at the HPV origin was proposed (Figure 30). The initial step would be the binding of E2 dimers at the E2 binding sites located in the origin. These suggest that the binding of at least two E2 binding sites is required for the recruitment of the E1 helicase (Figure 26, Figure 27). E1 is likely localized to the DNA by E2 bound to the DNA, through E1-E2 protein interactions. This is in agreement with a study which found E1-E2 protein interactions using co-expressed crude extracts [107]. Interestingly, the formation of the E1-E2-DNA complex did not require the presence of the E1 binding site (Figure 26). This may suggest that initially E1 does not directly bind the DNA, but instead is bound to the E2-DNA complex. Since the E1 binding site is positioned close to the cluster of E2 BS1, BS2, and BS3, the binding of E2 at DNA in this position may be essential to recruit E1 in close proximity to the E1 binding site, specifically. Eventually, the torsional stress of the proteins binding in this area may initiate localized melting of the DNA. This melting will likely expose ssDNA, which may be bound by E1 and initiate helicase activity. E1 recruits cellular DNA replication proteins to the replication fork [67, 108]. E1 interacts with the host polymerase alpha primase and the transactivation domain of E2, in a mutually exclusive manner [21, 66]. Since E2 did not affect E1 helicase activity, E2 may no longer be required at this stage. Dissociation of E2 from the E1-DNA complex may signal the end of viral replication initiation and the start of replication elongation. E1

has been shown to work bi-directionally as a double hexamer [109]. Together, the data presented here supports the essential E1-E2 interaction at the origin in HPV DNA replication initiation.



**Figure 30. Putative model of E1-E2 interaction at the origin for initiation of HPV replication.** Outlined is a potential scheme for the interplay of E1 and E2 at the HPV origin. The initial step would be the binding of E2 dimers at the E2 binding sites located in the origin. E2 will recruit E1, leading to the formation of the E1-E2-DNA complex. These steps will lead to localized melting of the DNA. This melting will likely expose ssDNA, and initiate the movement or placement of E1 in this position. E1 helicase activity will be initiated. E2 may no longer be required at this stage. Dissociation of E2 from the E1-DNA complex may signal the end of viral replication initiation and the start of replication elongation.

## **SUMMARY AND CONCLUSIONS**

This study reported the development of a methodology for purification of recombinant full-length HPV E2 protein. The studies with DNA binding properties of E2 protein led to important insights into the regulatory mechanism of HPV DNA replication and transcription. E2 binding sites have specific sequences that are required for DNA recognition and high affinity binding. DNA sequence features, such as unique regions within and surrounding individual E2 DNA binding sites, resulted in small but distinctive DNA binding affinities. Full-length low-risk E2 protein is a dimer in solution, but also has a natural tendency to oligomerize under various conditions like DNA binding. E2 displayed cooperative DNA binding with spatially-conserved E2 binding sites. Association of E2 with the HPV origin resulted in multimer formation, a prerequisite for the initiation of loop formation. The E2-DNA interaction is thermodynamically driven by hydrophobic interactions. Together, the results of these studies reveal unique features involved in the mechanism of E2-DNA interaction, essential in both the repression of transcription and initiation of replication in low-risk HPVs.

This study provides combined bioinformatics and molecular evidence distinguishing high-risk and many probable high-risk types from low-risk HPV subtypes. The systematic examination of the E2 binding sites of low-risk and high-risk HPV types revealed differences among the E2 binding sites between the two groups. Undoubtedly, there is a common theme amongst the high-risk HPVs, which is a nucleotide variation(s) in the consensus sequence. Nucleotide sequence variation

could potentially be utilized as a useful tool for assessing cancer risk. The evidence presented here should help further studies on a new paradigm of HPV-linked carcinogenesis and a useful predictor of cancer risk.

In this report, the enzymatic activities of HPV-11 E1 protein were evaluated. The fundamental biochemical properties of the E1-E2 interaction with the HPV origin were outlined. The binding of E2 dimers at the E2 binding sites located in the origin is necessary for E1 interaction. More specifically, the formation of the E1-E2-DNA complex required the binding of at least two E2 binding sites for the recruitment of the E1 helicase, but did not require the presence of the E1 binding site. E1 is localized to the DNA by E2 bound to the DNA, using E1-E2 protein interactions. Together, the data presented here supports the essential E1-E2 interaction at the origin in HPV DNA replication initiation. Further studies may elaborate upon the work presented here for the prevention of HPV-induced cancers.

## REFERENCES

- [1] E.R. Myers, D.C. McCrory, K. Nanda, L. Bastian, D.B. Matchar, Mathematical model for the natural history of human papillomavirus infection and cervical carcinogenesis, *American journal of epidemiology*, 151 (2000) 1158-1171.
- [2] N. Munoz, F.X. Bosch, S. de Sanjose, R. Herrero, X. Castellsague, K.V. Shah, P.J. Snijders, C.J. Meijer, G. International Agency for Research on Cancer Multicenter Cervical Cancer Study, Epidemiologic classification of human papillomavirus types associated with cervical cancer, *The New England journal of medicine*, 348 (2003) 518-527.
- [3] C.J. Lacey, C.M. Lowndes, K.V. Shah, Chapter 4: Burden and management of non-cancerous HPV-related conditions: HPV-6/11 disease, *Vaccine*, 24 Suppl 3 (2006) S3/35-41.
- [4] G.I. Sanchez, R. Jaramillo, G. Cuello, K. Quintero, A. Baena, A. O'Byrne, A.J. Reyes, C. Santamaria, H. Cuello, A. Arrunategui, A. Cortez, G. Osorio, J.C. Reina, W.G. Quint, N. Munoz, Human papillomavirus genotype detection in recurrent respiratory papillomatosis (RRP) in Colombia, *Head & neck*, 35 (2013) 229-234.
- [5] W.C. Reeves, S.S. Ruparelia, K.I. Swanson, C.S. Derkay, A. Marcus, E.R. Unger, National registry for juvenile-onset recurrent respiratory papillomatosis, *Archives of otolaryngology--head & neck surgery*, 129 (2003) 976-982.
- [6] M.F. Janicek, H.E. Averette, Cervical cancer: prevention, diagnosis, and therapeutics, *CA: a cancer journal for clinicians*, 51 (2001) 92-114; quiz 115-118.
- [7] A. Goodman, Role of routine human papillomavirus subtyping in cervical screening, *Current opinion in obstetrics & gynecology*, 12 (2000) 11-14.
- [8] M. Schiffman, G. Clifford, F.M. Buonaguro, Classification of weakly carcinogenic human papillomavirus types: addressing the limits of epidemiology at the borderline, *Infectious agents and cancer*, 4 (2009) 8.
- [9] I.W.G.o.t.E.o.C.R.t. Humans, Biological agents. Volume 100 B. A review of human carcinogens, IARC monographs on the evaluation of carcinogenic risks to humans, 100 (2012) 1-441.

- [10] Centers for Disease Control and Prevention, Genital HPV Infection-Fact Sheet, in, vol. 2017, 2017.
- [11] A.K. Chaturvedi, E.A. Engels, R.M. Pfeiffer, B.Y. Hernandez, W. Xiao, E. Kim, B. Jiang, M.T. Goodman, M. Sibug-Saber, W. Cozen, L. Liu, C.F. Lynch, N. Wentzensen, R.C. Jordan, S. Altekruze, W.F. Anderson, P.S. Rosenberg, M.L. Gillison, Human papillomavirus and rising oropharyngeal cancer incidence in the United States, *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 29 (2011) 4294-4301.
- [12] H.U. Bernard, The clinical importance of the nomenclature, evolution and taxonomy of human papillomaviruses, *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*, 32 Suppl 1 (2005) S1-6.
- [13] H.U. Bernard, R.D. Burk, Z. Chen, K. van Doorslaer, H. zur Hausen, E.M. de Villiers, Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments, *Virology*, 401 (2010) 70-79.
- [14] H. zur Hausen, E.M. de Villiers, Human papillomaviruses, *Annual review of microbiology*, 48 (1994) 427-447.
- [15] P.M. Day, D.R. Lowy, J.T. Schiller, Papillomaviruses infect cells via a clathrin-dependent pathway, *Virology*, 307 (2003) 1-11.
- [16] E.K. Yim, J.S. Park, The role of HPV E6 and E7 oncoproteins in HPV-associated cervical carcinogenesis, *Cancer research and treatment : official journal of Korean Cancer Association*, 37 (2005) 319-324.
- [17] C.M. Sanders, A. Stenlund, Recruitment and loading of the E1 initiator protein: an ATP-dependent process catalysed by a transcription factor, *The EMBO journal*, 17 (1998) 7044-7055.
- [18] L. Yang, R. Li, I.J. Mohr, R. Clark, M.R. Botchan, Activation of BPV-1 replication in vitro by the transcription factor E2, *Nature*, 353 (1991) 628-632.

- [19] M. Ustav, A. Stenlund, Transient replication of BPV-1 requires two viral polypeptides encoded by the E1 and E2 open reading frames, *The EMBO journal*, 10 (1991) 449-457.
- [20] F. Sverdrup, S.A. Khan, Replication of human papillomavirus (HPV) DNAs supported by the HPV type 18 E1 and E2 proteins, *Journal of virology*, 68 (1994) 505-509.
- [21] P.J. Masterson, M.A. Stanley, A.P. Lewis, M.A. Romanos, A C-terminal helicase domain of the human papillomavirus E1 protein binds E2 and the DNA polymerase alpha-primase p68 subunit, *Journal of virology*, 72 (1998) 7407-7419.
- [22] Y. Han, Y.M. Loo, K.T. Militello, T. Melendy, Interactions of the papovavirus DNA replication initiator proteins, bovine papillomavirus type 1 E1 and simian virus 40 large T antigen, with human replication protein A, *Journal of virology*, 73 (1999) 4899-4907.
- [23] A. Stenlund, E1 initiator DNA binding specificity is unmasked by selective inhibition of non-specific DNA binding, *The EMBO journal*, 22 (2003) 954-963.
- [24] I.J. Mohr, R. Clark, S. Sun, E.J. Androphy, P. MacPherson, M.R. Botchan, Targeting the E1 replication protein to the papillomavirus origin of replication by complex formation with the E2 transactivator, *Science*, 250 (1990) 1694-1699.
- [25] E.J. Androphy, D.R. Lowy, J.T. Schiller, Bovine papillomavirus E2 transactivating gene product binds to specific sites in papillomavirus DNA, *Nature*, 325 (1987) 70-73.
- [26] A.A. McBride, The papillomavirus E2 proteins, *Virology*, 445 (2013) 57-79.
- [27] M. Remm, R. Brain, J.R. Jenkins, The E2 binding sites determine the efficiency of replication for the origin of human papillomavirus type 18, *Nucleic acids research*, 20 (1992) 6015-6021.
- [28] S.Y. Chan, H.U. Bernard, C.K. Ong, S.P. Chan, B. Hofmann, H. Delius, Phylogenetic analysis of 48 papillomavirus types and 28 subtypes and variants: a showcase for the molecular evolution of DNA viruses, *Journal of virology*, 66 (1992) 5714-5725.

- [29] P. Hawley-Nelson, E.J. Androphy, D.R. Lowy, J.T. Schiller, The specific DNA recognition sequence of the bovine papillomavirus E2 protein is an E2-dependent enhancer, *The EMBO journal*, 7 (1988) 525-531.
- [30] D. Bramhill, A. Kornberg, Duplex opening by dnaA protein at novel sequences in initiation of replication at the origin of the *E. coli* chromosome, *Cell*, 52 (1988) 743-755.
- [31] D. Bramhill, A. Kornberg, A model for initiation at origins of DNA replication, *Cell*, 54 (1988) 915-918.
- [32] R.S. Fuller, J.M. Kaguni, A. Kornberg, Enzymatic replication of the origin of the *Escherichia coli* chromosome, *Proceedings of the National Academy of Sciences of the United States of America*, 78 (1981) 7370-7374.
- [33] J. Sim, S. Ozgur, B.Y. Lin, J.H. Yu, T.R. Broker, L.T. Chow, J. Griffith, Remodeling of the human papillomavirus type 11 replication origin into discrete nucleoprotein particles and looped structures by the E2 protein, *Journal of molecular biology*, 375 (2008) 1165-1177.
- [34] M. Dodson, R. McMacken, H. Echols, Specialized nucleoprotein structures at the origin of replication of bacteriophage lambda. Protein association and disassociation reactions responsible for localized initiation of replication, *The Journal of biological chemistry*, 264 (1989) 10719-10725.
- [35] C. Alfano, R. McMacken, Ordered assembly of nucleoprotein structures at the bacteriophage lambda replication origin during the initiation of DNA replication, *The Journal of biological chemistry*, 264 (1989) 10699-10708.
- [36] M. Gaczynska, P.A. Osmulski, Y. Jiang, J.K. Lee, V. Bermudez, J. Hurwitz, Atomic force microscopic analysis of the binding of the *Schizosaccharomyces pombe* origin recognition complex and the spOrc4 protein with origin DNA, *Proceedings of the National Academy of Sciences of the United States of America*, 101 (2004) 17952-17957.

- [37] R. Li, J. Knight, G. Bream, A. Stenlund, M. Botchan, Specific recognition nucleotides and their DNA context determine the affinity of E2 protein for 17 binding sites in the BPV-1 genome, *Genes & development*, 3 (1989) 510-526.
- [38] C.L. Bedrosian, 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*, 174 (1990) 557-575.
- [39] M.T. Chin, R. Hirochika, H. Hirochika, T.R. Broker, L.T. Chow, Regulation of human papillomavirus type 11 enhancer and E6 promoter by activating and repressing proteins from the E2 open reading frame: functional and biochemical studies, *Journal of virology*, 62 (1988) 2994-3002.
- [40] T.P. Cripe, T.H. Haugen, J.P. Turk, F. Tabatabai, P.G. Schmid, 3rd, M. Durst, L. Gissmann, A. Roman, L.P. Turek, Transcriptional regulation of the human papillomavirus-16 E6-E7 promoter by a keratinocyte-dependent enhancer, and by viral E2 trans-activator and repressor gene products: implications for cervical carcinogenesis, *The EMBO journal*, 6 (1987) 3745-3753.
- [41] W.C. Phelps, P.M. Howley, Transcriptional trans-activation by the human papillomavirus type 16 E2 gene product, *Journal of virology*, 61 (1987) 1630-1638.
- [42] B.A. Bernard, C. Bailly, M.C. Lenoir, M. Darmon, F. Thierry, M. Yaniv, The human papillomavirus type 18 (HPV18) E2 gene product is a repressor of the HPV18 regulatory region in human keratinocytes, *Journal of virology*, 63 (1989) 4317-4324.
- [43] G. Dong, T.R. Broker, L.T. Chow, Human papillomavirus type 11 E2 proteins repress the homologous E6 promoter by interfering with the binding of host transcription factors to adjacent elements, *Journal of virology*, 68 (1994) 1115-1127.
- [44] S.Y. Hou, S.Y. Wu, T. Zhou, M.C. Thomas, C.M. Chiang, Alleviation of human papillomavirus E2-mediated transcriptional repression via formation of a TATA binding protein (or TFIID)-TFIIB-RNA polymerase II-TFIIF preinitiation complex, *Molecular and cellular biology*, 20 (2000) 113-125.
- [45] S.H. Tan, B. Gloss, H.U. Bernard, During negative regulation of the human papillomavirus-16 E6 promoter, the viral E2 protein can displace Sp1 from a proximal promoter element, *Nucleic acids research*, 20 (1992) 251-256.

- [46] G. Steger, S. Corbach, Dose-dependent regulation of the early promoter of human papillomavirus type 18 by the viral E2 protein, *Journal of virology*, 71 (1997) 50-58.
- [47] M.J. Lacey, J.R. Anson, A.J. Klingelutz, J.H. Lee, A.D. Bossler, T.H. Haugen, L.P. Turek, Human papillomavirus (HPV) type 18 induces extended growth in primary human cervical, tonsillar, or foreskin keratinocytes more effectively than other high-risk mucosal HPVs, *Journal of virology*, 83 (2009) 11784-11794.
- [48] L. Pirami, V. Giache, A. Becciolini, Analysis of HPV16, 18, 31, and 35 DNA in pre-invasive and invasive lesions of the uterine cervix, *Journal of clinical pathology*, 50 (1997) 600-604.
- [49] A.P. Cullen, R. Reid, M. Champion, A.T. Lorincz, Analysis of the physical state of different human papillomavirus DNAs in intraepithelial and invasive cervical neoplasm, *Journal of virology*, 65 (1991) 606-612.
- [50] J.D. Khoury, N.M. Tannir, M.D. Williams, Y. Chen, H. Yao, J. Zhang, E.J. Thompson, T. Network, F. Meric-Bernstam, L.J. Medeiros, J.N. Weinstein, X. Su, Landscape of DNA virus associations across human malignant cancers: analysis of 3,775 cases using RNA-Seq, *Journal of virology*, 87 (2013) 8916-8926.
- [51] M. Pett, N. Coleman, Integration of high-risk human papillomavirus: a key event in cervical carcinogenesis?, *The Journal of pathology*, 212 (2007) 356-367.
- [52] N. Wentzensen, S. Vinokurova, M. von Knebel Doeberitz, Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dysplasia and invasive cancer of the female lower genital tract, *Cancer research*, 64 (2004) 3878-3884.
- [53] G. Hudelist, M. Manavi, K.I. Pischinger, T. Watkins-Riedel, C.F. Singer, E. Kubista, K.F. Czerwenka, Physical state and expression of HPV DNA in benign and dysplastic cervical tissue: different levels of viral integration are correlated with lesion grade, *Gynecologic oncology*, 92 (2004) 873-880.
- [54] H. Arias-Pulido, C.L. Peyton, N.E. Joste, H. Vargas, C.M. Wheeler, Human papillomavirus type 16 integration in cervical carcinoma in situ and in invasive cervical cancer, *Journal of clinical microbiology*, 44 (2006) 1755-1762.

- [55] G. Badaracco, A. Venuti, A. Sedati, M.L. Marcante, HPV16 and HPV18 in genital tumors: Significantly different levels of viral integration and correlation to tumor invasiveness, *Journal of medical virology*, 67 (2002) 574-582.
- [56] H. Romanczuk, P.M. Howley, Disruption of either the E1 or the E2 regulatory gene of human papillomavirus type 16 increases viral immortalization capacity, *Proceedings of the National Academy of Sciences of the United States of America*, 89 (1992) 3159-3163.
- [57] H. zur Hausen, Intracellular surveillance of persisting viral infections. Human genital cancer results from deficient cellular control of papillomavirus gene expression, *Lancet*, 2 (1986) 489-491.
- [58] F. Mantovani, L. Banks, The human papillomavirus E6 protein and its contribution to malignant progression, *Oncogene*, 20 (2001) 7874-7887.
- [59] M.E. McLaughlin-Drubin, K. Munger, The human papillomavirus E7 oncoprotein, *Virology*, 384 (2009) 335-344.
- [60] G. Chen, A. Stenlund, Characterization of the DNA-binding domain of the bovine papillomavirus replication initiator E1, *Journal of virology*, 72 (1998) 2567-2576.
- [61] M.S. Longworth, L.A. Laimins, Pathogenesis of human papillomaviruses in differentiating epithelia, *Microbiology and molecular biology reviews : MMBR*, 68 (2004) 362-372.
- [62] A.E. Gorbalenya, E.V. Koonin, Y.I. Wolf, A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses, *FEBS letters*, 262 (1990) 145-148.
- [63] S. Titolo, A. Pelletier, A.M. Pulichino, K. Brault, E. Wardrop, P.W. White, M.G. Cordingley, J. Archambault, Identification of domains of the human papillomavirus type 11 E1 helicase involved in oligomerization and binding to the viral origin, *Journal of virology*, 74 (2000) 7349-7361.

- [64] Y.N. Sun, J.Z. Lu, D.J. McCance, Mapping of HPV-11 E1 binding site and determination of other important cis elements for replication of the origin, *Virology*, 216 (1996) 219-222.
- [65] P.W. White, A. Pelletier, K. Brault, S. Titolo, E. Welchner, L. Thauvette, M. Fazekas, M.G. Cordingley, J. Archambault, Characterization of recombinant HPV6 and 11 E1 helicases: effect of ATP on the interaction of E1 with E2 and mapping of a minimal helicase domain, *The Journal of biological chemistry*, 276 (2001) 22426-22438.
- [66] K.L. Conger, J.S. Liu, S.R. Kuo, L.T. Chow, T.S. Wang, Human papillomavirus DNA replication. Interactions between the viral E1 protein and two subunits of human dna polymerase alpha/primase, *The Journal of biological chemistry*, 274 (1999) 2696-2705.
- [67] V.G. Wilson, M. West, K. Woytek, D. Rangasamy, Papillomavirus E1 proteins: form, function, and features, *Virus genes*, 24 (2002) 275-290.
- [68] S. Titolo, K. Brault, J. Majewski, P.W. White, J. Archambault, Characterization of the minimal DNA binding domain of the human papillomavirus e1 helicase: fluorescence anisotropy studies and characterization of a dimerization-defective mutant protein, *Journal of virology*, 77 (2003) 5178-5191.
- [69] O. Danos, I. Giri, F. Thierry, M. Yaniv, Papillomavirus genomes: sequences and consequences, *The Journal of investigative dermatology*, 83 (1984) 7s-11s.
- [70] Y. Sun, H. Han, D.J. McCance, Active domains of human papillomavirus type 11 E1 protein for origin replication, *The Journal of general virology*, 79 ( Pt 7) (1998) 1651-1658.
- [71] E.P. Dixon, G.L. Pahl, W.J. Rocque, J.A. Barnes, D.C. Lobe, M.H. Hanlon, K.A. Alexander, S.F. Chao, K. Lindley, W.C. Phelps, The E1 helicase of human papillomavirus type 11 binds to the origin of replication with low sequence specificity, *Virology*, 270 (2000) 345-357.
- [72] N. Moscufo, F. Sverdrup, D.E. Breiding, E.J. Androphy, Two distinct regions of the BPV1 E1 replication protein interact with the activation domain of E2, *Virus research*, 65 (1999) 141-154.

- [73] P. Monini, S.R. Grossman, B. Pepinsky, E.J. Androphy, L.A. Laimins, Cooperative binding of the E2 protein of bovine papillomavirus to adjacent E2-responsive sequences, *Journal of virology*, 65 (1991) 2124-2130.
- [74] A.A. Amin, S. Titolo, A. Pelletier, D. Fink, M.G. Cordingley, J. Archambault, Identification of domains of the HPV11 E1 protein required for DNA replication in vitro, *Virology*, 272 (2000) 137-150.
- [75] J. Sambrook, Fritsch, E.F., Maniatis, T., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, NY, (1989).
- [76] E.E. Biswas, P.H. Chen, S.B. Biswas, Overexpression and rapid purification of biologically active yeast proliferating cell nuclear antigen, *Protein expression and purification*, 6 (1995) 763-770.
- [77] E. Harlow, Lane, D., *Antibodies: A Laboratory Manual.* , Cold Spring Harbor Laboratory Press, 1988.
- [78] E.E. Biswas, C.M. Ewing, S.B. Biswas, Characterization of the DNA-dependent ATPase and a DNA unwinding activity associated with the yeast DNA polymerase alpha complex, *Biochemistry*, 32 (1993) 3020-3026.
- [79] C. Brown, K. Campos-Leon, M. Strickland, C. Williams, V. Fairweather, R.L. Brady, M.P. Crump, K. Gaston, Protein flexibility directs DNA recognition by the papillomavirus E2 proteins, *Nucleic acids research*, 39 (2011) 2969-2980.
- [80] C.S. Hines, C. Meghoo, S. Shetty, M. Biburger, M. Brenowitz, R.S. Hegde, DNA structure and flexibility in the sequence-specific binding of papillomavirus E2 proteins, *Journal of molecular biology*, 276 (1998) 809-818.
- [81] M. Ustav, Jr., F.R. Castaneda, T. Reinson, A. Mannik, M. Ustav, Human Papillomavirus Type 18 cis-Elements Crucial for Segregation and Latency, *PloS one*, 10 (2015) e0135770.
- [82] S.H. Tan, L.E. Leong, P.A. Walker, H.U. Bernard, The human papillomavirus type 16 E2 transcription factor binds with low cooperativity to two flanking sites and represses the E6 promoter through displacement of Sp1 and TFIID, *Journal of virology*, 68 (1994) 6411-6420.

- [83] J.B. Finley, M. Luo, X-ray crystal structures of half the human papilloma virus E2 binding site: d(GACCGCGGTC), *Nucleic acids research*, 26 (1998) 5719-5727.
- [84] V. Bouvard, R. Baan, K. Straif, Y. Grosse, B. Secretan, F. El Ghissassi, L. Benbrahim-Tallaa, N. Guha, C. Freeman, L. Galichet, V. Cogliano, W.H.O.I.A.f.R.o.C.M.W. Group, A review of human carcinogens--Part B: biological agents, *The lancet oncology*, 10 (2009) 321-322.
- [85] J.D. Combes, P. Guan, S. Franceschi, G.M. Clifford, Judging the carcinogenicity of rare human papillomavirus types, *International journal of cancer. Journal international du cancer*, 136 (2015) 740-742.
- [86] J.D. Knight, R. Li, M. Botchan, The activation domain of the bovine papillomavirus E2 protein mediates association of DNA-bound dimers to form DNA loops, *Proceedings of the National Academy of Sciences of the United States of America*, 88 (1991) 3204-3208.
- [87] E.E. Hernandez-Ramon, J.E. Burns, W. Zhang, H.F. Walker, S. Allen, A.A. Antson, N.J. Maitland, Dimerization of the human papillomavirus type 16 E2 N terminus results in DNA looping within the upstream regulatory region, *Journal of virology*, 82 (2008) 4853-4861.
- [88] F. Muller, M. Sapp, Domains of the E1 protein of human papillomavirus type 33 involved in binding to the E2 protein, *Virology*, 219 (1996) 247-256.
- [89] F.J. Hughes, M.A. Romanos, E1 protein of human papillomavirus is a DNA helicase/ATPase, *Nucleic acids research*, 21 (1993) 5817-5823.
- [90] J.Z. Lu, Y.N. Sun, R.C. Rose, W. Bonnez, D.J. McCance, 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, *Journal of virology*, 67 (1993) 7131-7139.
- [91] F. Sverdrup, S.A. Khan, Two E2 binding sites alone are sufficient to function as the minimal origin of replication of human papillomavirus type 18 DNA, *Journal of virology*, 69 (1995) 1319-1323.

- [92] K.A. Alexander, W.C. Phelps, A fluorescence anisotropy study of DNA binding by HPV-11 E2C protein: a hierarchy of E2-binding sites, *Biochemistry*, 35 (1996) 9864-9872.
- [93] R.S. Fuller, A. Kornberg, Purified dnaA protein in initiation of replication at the Escherichia coli chromosomal origin of replication, *Proceedings of the National Academy of Sciences of the United States of America*, 80 (1983) 5817-5821.
- [94] A.A. Antson, J.E. Burns, O.V. Moroz, D.J. Scott, C.M. Sanders, I.B. Bronstein, G.G. Dodson, K.S. Wilson, N.J. Maitland, Structure of the intact transactivation domain of the human papillomavirus E2 protein, *Nature*, 403 (2000) 805-809.
- [95] Y. Wang, R. Coulombe, D.R. Cameron, L. Thauvette, M.J. Massariol, L.M. Amon, D. Fink, S. Titolo, E. Welchner, C. Yoakim, J. Archambault, P.W. White, Crystal structure of the E2 transactivation domain of human papillomavirus type 11 bound to a protein interaction inhibitor, *The Journal of biological chemistry*, 279 (2004) 6976-6985.
- [96] G. Dell, K.W. Wilkinson, R. Tranter, J. Parish, R. Leo Brady, K. Gaston, Comparison of the structure and DNA-binding properties of the E2 proteins from an oncogenic and a non-oncogenic human papillomavirus, *Journal of molecular biology*, 334 (2003) 979-991.
- [97] K.M. Carr, J.M. Kaguni, Stoichiometry of DnaA and DnaB protein in initiation at the Escherichia coli chromosomal origin, *The Journal of biological chemistry*, 276 (2001) 44919-44925.
- [98] F. Stubenrauch, I.M. Leigh, H. Pfister, E2 represses the late gene promoter of human papillomavirus type 8 at high concentrations by interfering with cellular factors, *Journal of virology*, 70 (1996) 119-126.
- [99] A. Thain, K. Webster, D. Emery, A.R. Clarke, K. Gaston, DNA binding and bending by the human papillomavirus type 16 E2 protein. Recognition of an extended binding site, *The Journal of biological chemistry*, 272 (1997) 8236-8242.
- [100] S.M. Rotoli, E. Biswas-Fiss, S.B. Biswas, Quantitative analysis of the mechanism of DNA binding by Bacillus DnaA protein, *Biochimie*, 94 (2012) 2764-2775.

- [101] J.M. Zimmerman, L.J. Maher, 3rd, Solution measurement of DNA curvature in papillomavirus E2 binding sites, *Nucleic acids research*, 31 (2003) 5134-5139.
- [102] C. Weigel, A. Schmidt, H. Seitz, D. Tungler, M. Welzeck, W. Messer, The N-terminus promotes oligomerization of the Escherichia coli initiator protein DnaA, *Molecular microbiology*, 34 (1999) 53-66.
- [103] B.A. Spalholz, J.C. Byrne, P.M. Howley, Evidence for cooperativity between E2 binding sites in E2 trans-regulation of bovine papillomavirus type 1, *Journal of virology*, 62 (1988) 3143-3150.
- [104] R.S. Hegde, S.R. Grossman, L.A. Laimins, P.B. Sigler, Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target, *Nature*, 359 (1992) 505-512.
- [105] F. Jacob, J. Monod, Genetic regulatory mechanisms in the synthesis of proteins, *Journal of molecular biology*, 3 (1961) 318-356.
- [106] X. Wang, H. Liu, H.K. Wang, C. Meyers, L. Chow, Z.M. Zheng, HPV18 DNA replication inactivates the early promoter P55 activity and prevents viral E6 expression, *Virologica Sinica*, 31 (2016) 437-440.
- [107] G.L. Bream, C.A. Ohmstede, W.C. Phelps, Characterization of human papillomavirus type 11 E1 and E2 proteins expressed in insect cells, *Journal of virology*, 67 (1993) 2655-2663.
- [108] E.T. Fouts, X. Yu, E.H. Egelman, M.R. Botchan, Biochemical and electron microscopic image analysis of the hexameric E1 helicase, *The Journal of biological chemistry*, 274 (1999) 4447-4458.
- [109] J.S. Liu, S.R. Kuo, A.M. Makhov, D.M. Cyr, J.D. Griffith, T.R. Broker, L.T. Chow, Human Hsp70 and Hsp40 chaperone proteins facilitate human papillomavirus-11 E1 protein binding to the origin and stimulate cell-free DNA replication, *The Journal of biological chemistry*, 273 (1998) 30704-30712.

## ABBREVIATIONS LIST

AFM: Atomic Force Microscopy

bp: base pair

BPV: Bovine Papillomavirus

BS1: E2 binding site 1

BS2: E2 binding site 2

BS3: E2 binding site 3

BS4: E2 binding site 4

C1: Complex 1

C2: Complex 2

C3: Complex 3

DBD: DNA-binding domain

DLS: Dynamic Light Scattering

dsDNA: double-stranded DNA

*E. coli* : *Escherichia coli*

EDTA: Ethylenediamine tetra-acetic acid

EMSA: Electrophoretic Mobility Shift Assay

E1BS: E1 binding site

LCR: long control region

LC-MS/MS: liquid chromatography-tandem mass spectrometry

HPV: Human Papillomavirus

IPTG: Isopropyl- $\beta$ -D-thiogalactopyranoside

JORRP: Juvenile-onset recurrent respiratory papillomatosis

$K_D$ : dissociation constant

kDa: kilodalton

mM: millimolar

nM: nanomolar

ORC: origin recognition complex

ORF: open reading frame

PaVE: Papillomavirus Genome Database

PCR: polymerase chain reaction

RRP: recurrent respiratory papillomatosis

SDS-PAGE: sodium dodecyl sulfate polyacrylamide gel electrophoresis

SNV: single nucleotide variation

ssDNA: single-stranded DNA

TBE: Tris-Borate-EDTA

## **ATTRIBUTES OF THE THESIS**

All figures and experiments contained in this thesis were performed by Gulden Yilmaz.

## LIST OF FIGURES

	PAGE
<b>Figure 1. Structure of low-risk E2 protein, HPV-11 genome and its predicted crystal structure</b>	<b>35</b>
<b>Figure 2. E2 protein purity, oligomeric structure, mass spectrometric data, and DNA preference</b>	<b>37</b>
<b>Figure 3. Differential binding of E2 protein to its four binding sites</b>	<b>40</b>
<b>Figure 4. EMSA challenge assay verifies E2 hierarchy of binding</b>	<b>42</b>
<b>Figure 5. Binding of E2 to DNA binding site with altered spacer lengths and sequence</b>	<b>45</b>
<b>Figure 6. Effect of flanking sequences on DNA binding by E2 protein</b>	<b>48</b>
<b>Figure 7. Analysis of cooperative DNA binding of E2 protein</b>	<b>52</b>
<b>Figure 8. Analysis of ionic strength on E2-DNA complex formation</b>	<b>54</b>
<b>Figure 9. Thermodynamic parameters of E2-DNA complex formation</b>	<b>56</b>
<b>Figure 10. SDS-PAGE analysis and mass spectrometric data of high-risk E2 protein used in this study</b>	<b>59</b>
<b>Figure 11. EMSA analysis of high-risk E2 DNA binding activity and specificity</b>	<b>61</b>

<b>Figure 12. High-risk E2 binds differentially to its four binding sites in the origin</b>	<b>63</b>
<b>Figure 13. Single nucleotide variation (SNV) in E2 binding site results in attenuated binding affinity</b>	<b>65</b>
<b>Figure. 14. SNV leads to decreased E2 DNA binding</b>	<b>67</b>
<b>Figure 15. Multi-protein E2-DNA complex formation in the HPV origin containing three binding sites</b>	<b>71</b>
<b>Figure 16. Bioinformatics analysis of E2 binding sites and detection of SNVs in the consensus sequence</b>	<b>76</b>
<b>Figure 17. Low-risk HPV-11 E2 can bind to other low-risk BS3</b>	<b>78</b>
<b>Figure 18. Single nucleotide variation (SNV) in E2 binding site results in attenuated binding affinity of low-risk E2</b>	<b>79</b>
<b>Figure 19. DNA substrates for E2-DNA structural analysis using AFM</b>	<b>83</b>
<b>Figure 20. Structural analysis of low-risk and high-risk E2-DNA complex using AFM</b>	<b>84</b>
<b>Figure 21. E1 protein purity and mass spectrometric data</b>	<b>88</b>
<b>Figure 22. No DnaB helicase was present in E1 protein preparation</b>	<b>89</b>
<b>Figure 23. DNA helicase activity of E1 protein</b>	<b>91</b>
<b>Figure 24. E2 had no effect on DNA helicase activity of E1</b>	<b>93</b>

<b>Figure 25. E1 binds DNA in the presence of E2</b>	<b>96</b>
<b>Figure 26. E1-E2-DNA complex formation does not require E1 binding site</b>	<b>99</b>
<b>Figure 27. E1-E2-DNA complex formation requires at least two E2 binding sites</b>	<b>100</b>
<b>Figure 28. Effect of ssDNA on E1-E2-DNA complex formation</b>	<b>102</b>
<b>Fig. 29. A putative model for HPV E2 DNA binding and activation of replication at the low-risk and high-risk origins</b>	<b>118</b>
<b>Figure 30. Putative model of E1-E2 interaction at the origin for initiation of HPV replication</b>	<b>128</b>

## LIST OF TABLES

	<b>PAGE</b>
<b>Table 1. Oligonucleotide sequences used in this study</b>	<b>32</b>
<b>Table 2. Sequences for oligonucleotides used as low-risk E2 binding sites</b>	<b>39</b>
<b>Table 3. Sequences for flanking sequence oligonucleotides used as E2 binding sites</b>	<b>47</b>
<b>Table 4. Sequences of oligonucleotides utilized in cooperative DNA binding assays</b>	<b>51</b>
<b>Table 5. Thermodynamic analysis of binding affinities for E2 protein</b>	<b>57</b>
<b>Table 6. Sequences for oligonucleotides used as high-risk E2 binding sites</b>	<b>68</b>
<b>Table 7. Sequences for oligonucleotides used in multi-protein E2-DNA complex formation</b>	<b>72</b>