4-12-2017

Biophysical Studies of Hairpin Polyamides with Broad-Spectrum Activity Against High-Risk Human Papillomaviruses

Carlos H. Castaneda
University of Missouri-St. Louis, chcvdc@umsl.edu

Biophysical Studies of Hairpin Polyamides with Broad-Spectrum Activity Against High-Risk Human Papillomaviruses

Carlos Hernando Castañeda
Bachelor of Science in Chemistry, Southeast Missouri State University, 2005
Master of Science in Biochemistry and Biotechnology, University of Missouri-St. Louis, 2009
Master of Science in Chemistry, University of Missouri-St. Louis, 2015

A Dissertation Submitted to The Graduate School at the University of Missouri-St. Louis
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Chemistry and Biochemistry

May 2017

Advisory Committee

James K. Bashkin, Ph.D.
Chairperson

Wesley R. Harris, Ph.D.

Chung F. Wong, Ph.D.

Bethany K. Zolman, Ph.D.

For my children, Isabela and Carlos

Acknowledgements

First and foremost, I am deeply indebted to my family for their unconditional love and support. To my beautiful daughter, Isabela, whose artistic disposition and inquisitiveness show me the beauty and simplicity with which all scientific exploration and inquiry begins. To my handsome son, Carlos, whose curiosity and stamina never cease to amaze me. The way he questions mundane ideas in such complex ways encourages me to look beyond the obvious and think more objectively. The infectious joy and unfailing optimism of my children warms my heart and has strengthened my resolve during my most trying times. I would like to thank my wife, Rebecca, for all of her love, support and encouragement. More than any single person, she is the foundation of my pursuits and is my courage to advance when I falter; she is my rock in stormy waters. She has also supported our family during my time as a graduate student. I would also like to thank her for proofreading this thesis. To my mother, Maria, for her constant reminder of how much pride she holds for my accomplishments and life path. I am extremely appreciative of her hard work and the struggles she endured during my formative years to ensure that I lead a life of integrity, respectability, intellect and success. To my sister, Blanca, who has pursued her own career with inspirational fervor. Her profound example of persistence and professional dedication are a guiding light for my own pursuits. My entire family has been exceptionally patient throughout the time that I have invested in pursuing this degree. They encouraged me to pursue a direction that is unusual for someone who has spent so many years out of school, working in industry. Without their trust, this degree would still be a dream rather than the reality that has been realized during these past five years.

I would like to express my immeasurable gratitude to my advisor, Dr. James K. Bashkin, for his support, encouragement and guidance during my time in his research group. His leadership has been the backbone for my work and the level of interest he has for this subject is one that I admire and respect greatly. I am most grateful for his insights and assistance in the development of my hypotheses. It has been an honor to work under a man so dedicated to his field and his work.

In addition, I would also like to acknowledge and thank the faculty of the Department of Chemistry and Biology for the valuable knowledge they have provided through their lectures and discussions. The time I was allowed to spend as a teaching assistant and helping to train incoming students renewed my passion for learning. I am especially grateful for those experiences.

I would like to extend my appreciation to my committee members for the time they have invested in me. I am grateful to Dr. Wesley Harris for his encouragement to pursue a doctorate in chemistry and for allowing me to conduct research as a Master's student in his laboratory. I would like to express my gratitude to Dr. Chung Wong for his assistance with bioinformatics applications and his insightful observations during seminars. Special thanks go to Dr. Bethany Zolman for her valuable input and useful insights.

Furthermore, I would like to thank all the members of Dr. Bashkin's lab, past and present, including Dr. Gaofei He, Dr. Dave Harris, Dr. Kevin Koeller, Karl Aston, Silke Evdokimov, Faten Tamimi, Priyanka Bapat, José Scuderi, Dr. Edith Csiki-Fejer, Magnus Creed, Brandon Wood, Rachel Schafer and Deborah Kirchner for their assistance and helpful discussions. In particular, I would like to thank Dr. He for his guidance and initial training upon joining Dr. Bashkin's group. I would like to also extend my gratitude to

José Scuderi for her biophysical studies which led to a joint publication in MedChemComm. Also, I would like to thank Brandon Wood for his help with performing in-cell hydroxyl radical footprinting and COSMIC-Seq (Crosslinking of Small Molecules for Isolation of Chromatin) experiments. I was fortunate to work alongside this great group of scientists. Much of this work came to be through strong communication and group efficiency. We worked as a team and became family as we pursued different directions in search of a common goal. I greatly appreciate them for their assistance, especially during the late nights, long weekends and missed holidays that led to the completion of my work.

I would particularly like to extend my gratitude to our collaborators at NanoVir, LLC, Terri Edwards and Dr. Chris Fisher for conducting the antiviral activity assays and mechanism of action studies. I would like to thank Dr. William Miller, Carl Herbold and Gregory Gunn from the University of Missouri Research Reactor (MURR) for performing gamma-radiation on plasmid and cell samples. I would like to thank Dr. R. Michael Roberts, Dr. Toshihiko Ezashi, Dr. Aihua Dai and Dennis F. Reith for allowing me to use their BSL-2 facilities at the University of Missouri. I would like to extend my gratitude to Nathan Bivens, Ellen Kesler, Mingyi Zhou, and Karen Bromert from the DNA Core Facility at the University of Missouri for performing DNA Capillary Electrophoresis Fragment Analysis and for sequencing constructed High-throughput (Illumina) Sequencing DNA libraries. I greatly appreciate Dr. Christopher Bottoms and Dr. Scott Givan from the Informatics Research Core Facility at the University of Missouri for demultiplexing the raw sequencing data and for writing the CountNicks.pm script to count DNA nicks for *in vitro* and in-cell hydroxyl radical footprinting experiments. I would also like to express my appreciation to Dr. Cynthia Dupureur and all of her group members for their helpful discussions and guidance on biophysical analysis. I greatly appreciate Dr. Aseem Ansari and his group members, Dr. Graham Erwin, Dr. Asfa Ali, Matthew Grieshop, Dr. Asuka Eguchi, Dr. Devesh Bhimsaria and Mackenzie Spurgat for their hospitality during my visit to the University of Wisconsin-Madison and for providing me with the opportunity to study our compounds by means of COSMIC-Seq and CSI (Cognate Site Identification).

I greatly appreciate my friends, in particular Andrew Lutes, for his assistance with various computational efforts and his careful proofreading of this document. His keen eye and scientific knowledge helped me to look at things from previously unexplored perspectives. Many times, he believed in me when I had trouble believing in myself. His encouragement has grounded me more times than he will ever understand and for that I am eternally in his debt.

For everyone, mentioned and not, I offer you my innumerable thanks for playing a part in the successes involved in my pursuit. I truly believe that we are all a product of our environment and I have no doubt that the people that have come into my life have, in one way or another, assisted in helping me achieve this goal.

Finally, I would like to also extend my gratitude to the National Institute of Health, the National Science Foundation, the UMSL ORA Research Grant Program and the University of Missouri Interdisciplinary Intercampus (IDIC) Research Funding Program for their financial support. Thank you to UMSL for taking the chance on me. You provided me with the resources, the guidance, and the support that helped me hold my head high in confidence throughout this path.

Abstract

Human papillomavirus (HPV) is a small double-stranded DNA (dsDNA) virus that infects mucosal and cutaneous epithelial tissues. Persistent infection with high-risk (HR) HPV is the main etiological agent in the development of cervical cancer worldwide. Although prophylactic vaccines against HPV are available, these preventative measures are type-specific and are ineffective against existing infections. Thus, there is a pressing need for antiviral drugs with a broad-spectrum activity against HPV to eradicate existing infections, no matter the subtype.

Our group and collaborators have synthesized an extensive library of novel *N*-methylpyrrole/*N*-methylimidazole (Py/Im) hairpin polyamides (PAs) with broad-spectrum activities against three prevalent HR-HPV types (HPV16, HPV18 and HPV31) without apparent cytotoxicity. Py/Im hairpin polyamides are cell-permeable, synthetic DNA ligands and higher homologues of the natural antibiotics netropsin and distamycin A. Because Py/Im polyamides can be rationally designed to bind the minor groove of double-stranded DNA (dsDNA) in a sequence-dependent manner, these small molecules are attractive candidates as modulators of gene expression, as molecular probes for diagnostics and as antiviral agents for the selective elimination of dsDNA viruses. Despite the promise of applying these agents to treat HPV infections, much regarding their mechanism of action remains unexplained. Therefore, the overall goal of this dissertation is to investigate the DNA-binding properties of potent anti-HPV PAs under cell-free conditions and in HPV-harboring keratinocytes.

Because PA-binding events impart structural perturbations to DNA, these DNA-targeting agents can disrupt viral protein-DNA interactions in the viral genome and thus inhibit their functions. Consequently, displacement of viral proteins, as well as cellular transcription factors (TFs) and replication machinery, from their canonical binding sites on HPV would presumably lead to episomal instability followed by the elimination of the viral genome. **Chapter 2** describes biophysical studies of anti-HPV hairpin polyamides (**PA1** and **PA25**) using deoxyribonuclease I (DNase I) footprinting and affinity cleavage (AC) to interrogate the sequence specificity, binding sites and dissociation constants within the HPV18 Long Control Region (LCR). Analysis of the DNA-binding properties of **PA1** and **PA25** demonstrated that these anti-HPV polyamides bind avidly to the minor groove of A/T-rich sequences in the LCR of HPV18 with dissociation constants in the nanomolar range, and tolerate PA-DNA mismatches without a significant decrease in binding affinity. **Chapter 3** extends these analyses to a new structural class of hairpin polyamides with tetramethylguanidinium (**PA30**) and guanidinium (**PA31**) substitutions on the N-terminus of **PA1**. In contrast to **PA1** and **PA30**/**PA31**, **PA25** has the ability to tolerate a higher number of PA-DNA mismatches and exhibits a more extended coverage of the LCR. Because the HPV LCR harbors the early promoter and regulatory sequences vital to HPV's lifecycle, the improved antiviral activity of **PA25** against HPV18 may arise from more abundant PA occupancy within vital protein binding sites in the LCR. **Chapter 4** presents a detailed procedure to determine the genome-wide binding events of DNA-binding molecules in cell-free conditions using hydroxyl radical footprinting coupled to massively parallel DNA sequencing (·OH-Seq). As a proof of concept, a genome-wide binding map of **PA1** across the supercoiled HPV16 episome is presented and compared against previously reported binding sites within the HPV16 LCR using DNase I footprinting with an automated capillary electrophoresis. The binding sites

obtained with ·OH-Seq are nearly identical to those observed in the capillary electrophoresis DNase I method within the HPV16 LCR.

Because the viral genome is negatively supercoiled, compacted into a chromatin-like structure and associated with numerous viral and host proteins *in vivo*, some of the potential DNA-binding sequences may be inaccessible by polyamides. To this end, ·OH-Seq and COSMIC-Seq (**Chapter 5**) have been employed to assess the binding occupancies of anti-HPV hairpin polyamides across the viral genomes in keratinocytes harboring HPV16 episomes.

A deeper understanding of the molecular underpinnings of the ability of antiviral hairpin PAs to eliminate the HPV viral load may facilitate the development of a new generation of broad-spectrum antiviral treatments against this deadly virus.

Table of Contents

List of Figure and Tables

Chapter 1 **Page**

Chapter 2 **Page**

Chapter 3 **Page**

Chapter 4 **Page**

Chapter 5 **Page**

Abbreviations

ACN	Acetonitrile
ATCC	American Type Culture Collection
ATP	Adenosine Triphosphate
AU	Activity Units
β	Beta-Alanine
bp	Base Pairs
BamHI	*Bacillus amyloliquefaciens* H Restriction Enzyme
BSA	Bovine Serum Albumin
CD	Circular Dichroism
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
ChIP	Chromatin Immunoprecipitation
CIN	Cervical Intraepithelial Neoplasia
COSMIC	Crosslinking of Small Molecules for Isolation of Chromatin
CpG	Cytosine-phosphate-Guanine
CSI	Cognate Site Identification
Cq	Quantification Cycle
DDR	DNA Damage Response
DIEA	*N*,*N*-Diisopropylethylamine
DMEM	Dulbecco's Modified Eagle's Medium
DMF	Dimethyl Formamide
DMSO	Dimethyl Sulfoxide
Dp	3-(dimethylamino)propylamine
dsDNA	Double-Stranded Deoxyribonucleic Acid
dsRNA	Double-Stranded Ribonucleic Acid
E Genes	Early Genes
EBV	Epstein-Barr Virus
EBNA1	Epstein-Barr Virus Nuclear Antigen 1
EcoRI	*Escherichia coli* RY13 Restriction Enzyme
EDTA	Ethylenediaminetetraacetic Acid
EGF	Epidermal Growth Factor
Ets	E26 Transformation-Specific
FAM	6-carboxyfluorescein
FBS	Fetal Bovine Serum
γ	Gamma-aminobutyric acid
H1-hESC	H1 Human Embryonic Stem Cells
HATU	1-[Bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium 3-oxid hexafluorophosphate
HBES	HEPES Buffered Earle's Salts
HEX	Hexachloro-fluorescein
HLA	Human Leukocyte Antigen
HPLC	High-Performance Liquid Chromatography
HPV	Human Papillomavirus
HR-HPV	High-Risk Human Papillomavirus
HRMS-ESI	High-Resolution Mass Spectrometry-Electrospray Ionization
IC_{50}	Polyamide Concentration Leading to 50% Decrease in Viral DNA
IC_{90}	Polyamide Concentration Leading to 90% Decrease in Viral DNA
Im	*N*-methylimidazole

L Genes	Late Genes
LB	Lysogenic Broth
LCR	Long Control Region
LLC	Limited Liability Company
MACS	Model-Based Analysis of ChIP-Seq
MeOH	Methanol
MU	University of Missouri
MURR	Missouri University Research Reactor
NCBI	National Center for Biotechnology Information
NGS	Next-Generation Sequencing
NMP	*N*-Methyl-2-Pyrrolidinone
NMR	Nuclear Magnetic Resonance
·OH	Hydroxyl Radical
ORF	Open Reading Frame
oriR	Origin of Replication
oriP	Origin of Plasmid Replication
PA	Polyamide
Pap	Papanicolaou
PBS	Phosphate Buffer Saline
PB	Psoralen-Biotin Moiety
PCR	Polymerase Chain Reaction
Poly(dI-dC)	Poly(deoxyinosinic-deoxycytidylic) Acid Polymer
PMSF	Phenylmethanesulfonylfluoride
Py	*N*-methylpyrrole
Q-PCR	Quantitative-PCR
Rb	Retinoblastoma
S-Phase	Synthesis Phase
SDS	Sodium Dodecyl Sulfate
Seq	Next-Generation Sequencing
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
SPRI	Solid Phase Reversible Immobilization
ssRNA	Single-Stranded Ribonucleic Acid
T_3	3,3',5-triiodo-L-thyronine
Ta	3, 3'-diamino-*N*-methyldipropylamine
TAE	Tris-acetate, EDTA
TFA	Trifluoroacetic Acid
TKMC	Tris-HCl, KCl, $MgCl_2$, $CaCl_2$
TMB	3,3′,5,5′-tetramethylbenzidine
TF	Transcription Factor
UMSL	University of Missouri-St. Louis
UV-Vis	Ultraviolet-Visible
V	Volts
VLP	Virus-Like Particle

Safety

All work was carried out under BSL-2 conditions under the approval of the Biosafety Committees and from the Environmental Health and Safety (EHS) Department from UMSL and MU. Furthermore, rules and standard operating procedures (SOPs) set in placed by the Chemistry and Biochemistry Department were followed as outlined in the UMSL Lab Safety Plan on the EHS website. Refresher training was performed yearly to remain compliant with the departmental SOPs.

Chapter 1

Introduction to Human Papillomavirus and N-Methylpyrrole/N-Methylimidazole Hairpin Polyamides

1.1 HUMAN PAPILLOMAVIRUS (HPV)

Human papillomavirus (HPV) is a small dsDNA virus with a predilection (tropism) toward basal keratinocytes of cutaneous and mucosal tissues.[1] HPV is classified into five genera (*α-, β-, γ-, μ-, ν-papillomavirus*) within the *Papillomaviridae* family, and sub-divided into types based on the genomic sequence of the major capsid protein.[2,3] HPV infections remain a major health concern with most clinically important types belonging to the *α-, β-,* and *γ-papillomavirus* genera.[3] HPV infections can lead to proliferative epithelial lesions, ranging from benign skin and genital warts to cancer development of the penis, anus, vagina, oropharynx, mouth and cervix.[4-6] Accordingly, HPV types are classified as low- and high-risk (HR) based on their propensity to develop into malignant tumors.[2,3] To date, over 180 types have been isolated of which approximately 15 oncogenic types have been identified.[7-9] The development of cervical cancer, the second most prevalent cancer affecting women worldwide, is closely associated with persistent infections of HR-HPV types.[10-12] Specifically, HR-HPV DNA has been detected in 99.7 % of cervical cancers, as well as in ~90 % of anal cancers and ~60 % of oropharyngeal (mouth and throat) cancers.[5] In particular, HPV types 16 and 18 account for ~70 % of cervical cancer cases worldwide.[13] Infection with HR-HPV is required, but not solely responsible, for the development of cervical cancer (**Figure 1.1**).[6,10]

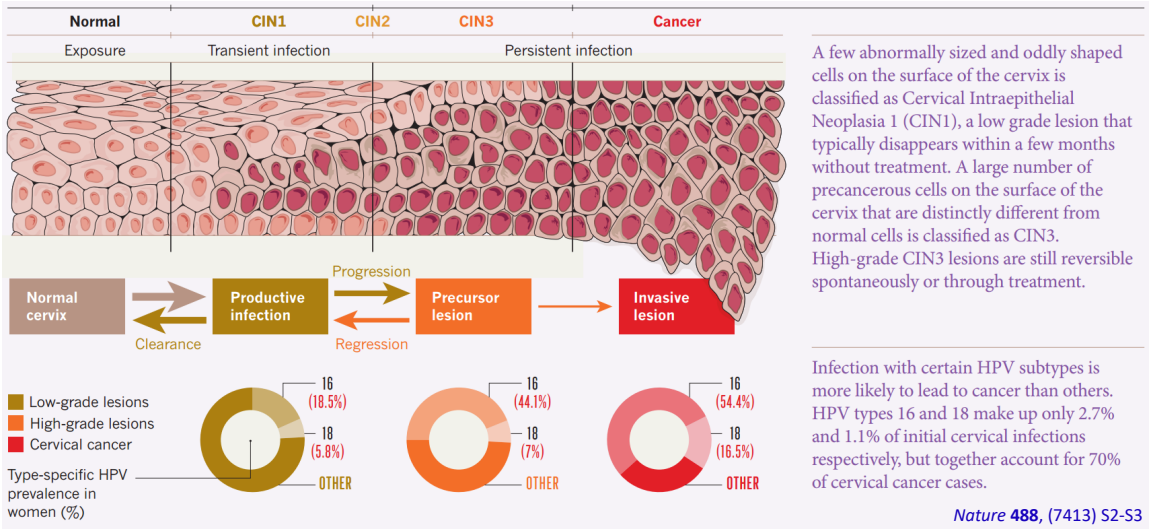


Figure 1.1. Stages of Cervical Cancer Development and HPV Types Associated with Each Stage. The development of cervical cancer, the second most prevalent cancer affecting women worldwide, is closely associated with persistent infections of HPV types 16 and 18. Figure adapted from Crow 2012 (*permission # 4087191013835*).

Most HPV infections are asymptomatic and are cleared by the immune system.[14] In developed countries, the prevention of cervical cancer has benefited from screening programs and prophylactic vaccines,[15] yet more than 35,000 HPV-related cancer cases were reported in 2009 in the United States alone.[5] In third world countries, where these preventive measures are less available, HPV infections remain a significant health burden; the approximately 500,000 new cases diagnosed annually contribute to 290,000 deaths within the same time frame.[12]

1.2 HPV LIFECYCLE

At the cellular level, infection is initiated through microlesions of cutaneous or mucosal epithelia, where the viral capsid can bind to its host receptor and gain entry to the basal cells.[16] Entry of the virion is believed to occur through clathrin-dependent endocytosis by interactions of the viral L1 capsid protein with heparan sulfate and a second unidentified receptor (co-receptor) on the surface of basal epithelial cells.[16,17]

Structurally, HPV is a non-enveloped virus whose closed-circular genome consists of roughly 8,000 base pairs of dsDNA that encode for eight well-characterized proteins. Based on their function within the lifecycle, the HPV genome can be divided into three regions: a late region (L), a non-coding region (LCR; long control region) and an early region (E) (**Figure 1.2**).[18] The encoded non-structural proteins from the early region primarily serve regulatory functions including DNA replication, transcription, cell signaling and immune evasion. The HPV episome is protected by an icosahedral capsid composed of the late-region major capsid protein L1 and minor protein L2.[19] L1 is necessary for the assembly of the virion and its genomic sequence serves as the basis for HPV type classification; L2 alone cannot assemble into a virion.[19] Nonetheless, L2 is essential for viral infection as cleavage at its C-terminus allows for both endosomal escape and nuclear targeting of the viral DNA.[17]

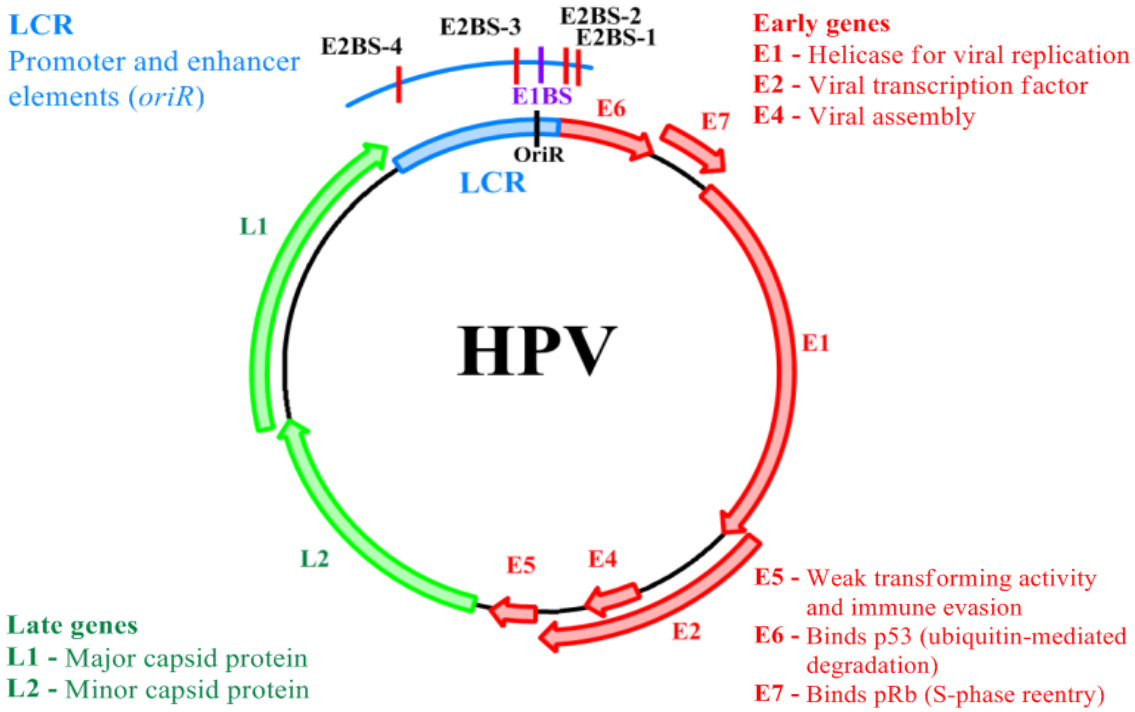


Figure 1.2. Characteristic Organization of the Mucosal High-Risk HPV Genome. Coding regions and functions for each gene are given with their relative positions in the genome. The origin of replication (*oriR*), E2 (E2BS) and E1 (E1BS) binding sites within the LCR are illustrated above the map.

In its infectious cycle, HPV benefits from the normal differentiation of keratinocytes (**Figure 1.3**).[20] Once the viral genome reaches the nucleus, it is maintained as a low-copy circular extrachromosomal DNA (episome) of ~50-100 copies/cell in an E1 and E2

dependent manner.[20] HPV DNA is compacted into a chromatin-like structure that is not as condensed as human chromatin, though the actual structure is a topic of active investigation. Expression of the early proteins (E6, E7, E1, E2, E4, E5) is minimal until the HPV-carrying keratinocytes begin differentiation, upon which expression is upregulated.[18] This is advantageous to the virus because it allows HPV to remain undetected by the host immune system in the primitive compartment of the epithelium, as this compartment is in close proximity to stromal dendritic cells.[20] As basal keratinocytes migrate to upper layers of the epithelium, where they terminally differentiate, they carry the viral payload which is amplified to a high-copy number of at least a thousand copies per cell.[18,20] Because the viral infection is strictly intraepithelial, it produces no event that can be detected by the cell-mediated immune system such as inflammation[21] or blood-circulating antigens (viremia);[22] thus, it has the ability to migrate and proliferate amongst the immediate epithelial tissue.[20,22] In addition to these passive immunoevasion strategies, HPV has also evolved several active strategies to thwart the host immune system. For example, HPV infections inhibit the presentation of human leukocyte antigen (HLA) class I viral peptide complexes and dampen several pathways of the innate immune system.[23-28]

The small genome of HPV only encodes for one enzyme, E1.[29] The nuclear-localized E1 protein is the essential ATPase helicase of papillomaviruses that binds to HPV LCR after it is recruited by the DNA-binding E2 protein. The latter binds the major groove of the palindromic sequence ACCgN_4cGGT (upper case designates required nucleotides, lower case represents preferred nucleotides, and N denotes no nucleotide preference) near the viral origin of replication (*oriR*).[30,31] There, E1 recruits the host replication machinery to initiate viral DNA replication.[1] E2 is also a transcription factor that tightly regulates, both positively and negatively, the expression of E6 and E7 oncoproteins.[32] In addition, E2 is also important for proper viral episome segregation during cell division of infected cells.[30,33] Because HPV has a tropism for basal keratinocytes, which eventually differentiate and exit the cell cycle, the oncogenic proteins E6 and E7 permit the reentry to S-phase which is necessary for proliferation and the synthesis of infectious particles.[30] To achieve this result, E6 orchestrates the ubiquitin-proteasomal degradation of p53, which evades the induction of apoptosis as a result of unscheduled S-phase reentry. The latter is promoted by the sequestration of the retinoblastoma (pRb) family of tumor suppressors upon E7 binding.[16] Thus, integration of the E6 and E7 ORF into the host chromosome along with the loss of their transcription regulator (E2) leads to the overexpression of these oncoproteins, culminating in cancerous growth.[12,16,34]

Virion packaging occurs in the upper layers of the epithelium, as the capsid-associated L1 gene expression relies on the usage of rare mammalian codons.[35,36] Moreover, the late promoter region becomes more accessible to transcriptional factors as HPV31-infected cells differentiate.[37] The release of progeny virions coincides with desquamation of keratinocytes; hence, virion release is not associated with a lytic phase.[38] These features are advantageous for immunoevasion as virion release occurs far from immunosurveillance compartments and circumvents the requirement of a viremic phase.[36] In this manner, the HPV lifecycle terminates with the release of new infectious particles that can infect naïve individuals (**Figure 1.3**).[1]

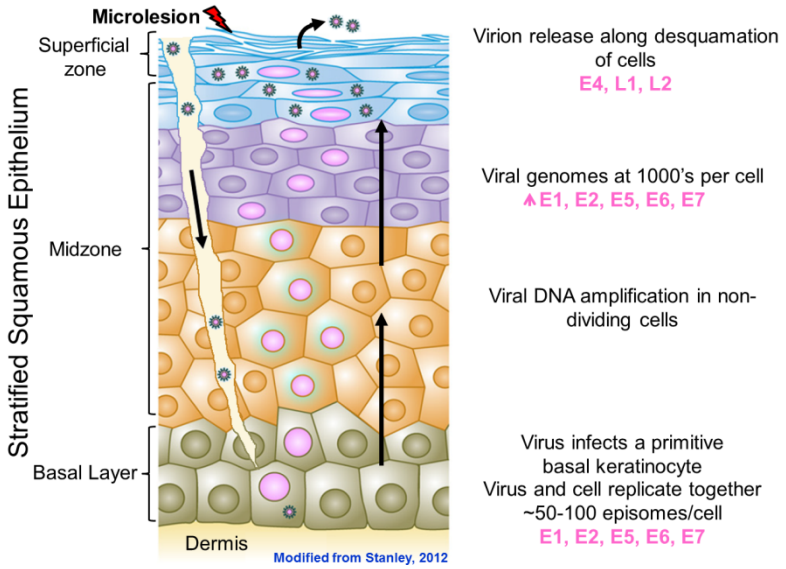


Figure 1.3. HPV Lifecycle. HPV infection is initiated through microlesions of cutaneous or mucosal epithelia, where the viral capsid binds to its host receptor. In the nucleus, it is maintained as a low-copy circular extrachromosomal DNA of ~50-100 copies/cell in an E1 and E2 dependent manner, although replication mechanisms have also been shown to not require these proteins. Expression of the early proteins (E6, E7, E1, E2, E4, E5) is minimal until the HPV-carrying keratinocytes begin differentiation, upon which expression is upregulated. Virion packaging occurs in the upper layers of the epithelium upon E4, L1 and L2 expression, and the released infectious particles can then infect naïve individuals. Figure adapted from Stanley 2012 (*permission to account #3001138983*).

1.3 PREVENTATIVE MEASURES AND TREATMENTS AGAINST HPV INFECTIONS

The late region capsid proteins, in particular L1, are the focus in the development of HPV vaccines.[19,39] Three prophylactic vaccines have been successfully implemented to immunize at-risk individuals against HR-HPV16 and HR-HPV18 (Cervarix® developed by GlaxoSmith-Kline), these high-risk types plus low-risk HPV6 and HPV11 (Gardasil® developed by Merck)[19,39,40] or all of these plus the HR-HPV types 31, 33, 45, 52 and 58 (Gardasil 9® developed by Merck). With the addition of these five HR-HPV types, Gardasil 9® has the potential to prevent over 90 % of HPV-induced cancers.[41] However, these preventative measures are type-specific and ineffective against existing infections.[42] Moreover, the prevention of cervical cancer has also benefited from Papanicolau (Pap) screening, which is responsible for an approximately 50 % reduction of cases in the developed world.[15,43] Currently, the available therapies against HPV infections rely on the surgical excision and cryogenic destruction of infected tissues. Additionally, non-surgical treatments that are exclusively approved against cutaneous infections are available. These include topical formulations with cytotoxic chemicals like trichloroacetic acid, anti-proliferative natural products such as podophyllotoxin, or immunomodulatory agents like imiquimod. However, these treatments are not HPV-specific and are often associated with numerous negative side-effects including pain, bleeding, tissue scarring, and severe irritation of the treated area. Also, these therapeutic approaches are cytodestructive and relatively inefficient with high frequencies of viral recurrence.[44]

There is a pressing need for antiviral drugs with a broad-spectrum activity against the various HPV types that lack cytotoxicity. Most antiviral agents primarily target viral enzymes; however, HPV only encodes one enzyme (E1)[29] and hijacks the host replication machinery to establish a productive infection. As a result, the development of novel HPV-specific antiviral therapies has been hindered due to the limited number of conventional targets. Therefore, our group and collaborators have pursued a novel strategy to eliminate viral loads of episome-harboring keratinocytes by targeting the HPV genome with hairpin polyamides.

1.4 *N*-METHYLPYRROLE/*N*-METHYLIMIDAZOLE HAIRPIN POLYAMIDES (Py/Im PAs)

The rational design of small molecules to selectively target predetermined DNA sequences is a major goal in the development of novel therapeutics. The ability to selectively regulate gene expression and DNA replication with limited toxicity has potential applications in the treatment of cancer and virus-associated diseases. *N*-methylpyrrole/*N*-methylimidazole (Py/Im) hairpin polyamides (PAs) are synthetic heteroaromatic compounds inspired by the chemical structures of the natural antibiotics netropsin and distamycin A.[45-50] These crescent-shaped molecules, collectively known as polyamides, bind with high affinity (nano- to picomolar dissociation constants) to the minor groove of dsDNA in a sequence-selective fashion.[51-53]

The pioneering work from the Dickerson, Wemmer and Dervan groups, among others, provided the impetus and paved the way in the research area of sequence-selective DNA recognition by polyamide molecules. Initial efforts were focused on explaining the molecular origin of the specificity toward A/T-rich sequences exhibited by distamycin A and netropsin. These natural DNA ligands are composed of repeating *N*-methylpyrrole heterocycles linked by amide bonds and display a broad-spectrum of antibacterial, antiviral and antitumor activities. However, unlike their synthetic counterparts, they exhibit dramatic cytotoxicity, rendering them unfit for therapeutic applications.[54-56] Nonetheless, their ability to preferentially bind A/T-rich sequences in a non-intercalative manner has received considerable attention.[46]

In the mid-1970s, Wells *et al.* proposed that the sequence preference of netropsin for A/T-rich DNA sequences was governed by steric interactions generated by the 2-amino group of guanine and stabilized by specific hydrogen bond interactions.[57] About a decade later, these molecular interactions were confirmed by separate structural studies of netropsin-DNA complexes conducted by Dickerson and Patel; Dickerson *et al.* solved the X-ray crystal structure of a 1:1 complex of netropsin and DNA, whereas Patel *et al.* investigated this complex using NMR.[46,58] These structural studies suggested that the crescent shape of netropsin provided the optimum fit with the inherent curvature of the minor groove of B-DNA, affording favorable van der Waals and hydrogen bond interactions. The narrowness of the minor groove in A/T-rich sequences and the optimal spacing of netropsin's *N*-methylpyrrole building blocks were found to provide for specific bifurcated H-bonds to form between the amide NHs with the N3 of adenine and O2 of thymine in the floor of the minor groove of B-DNA (**Figure 1.4**). These interactions were suggested to be further stabilized by the electrostatic interactions afforded by the positive termini of netropsin and the negatively charged phosphates in the DNA backbone. However, the exclusion of G·C base pairs in the canonical recognition

sequence of netropsin and distamycin A was proposed to be governed mainly by the steric hindrance that arises by the exocyclic amino group of guanine in the minor groove of G·C base pairs, impeding hydrogen bonding between the ligand and DNA.

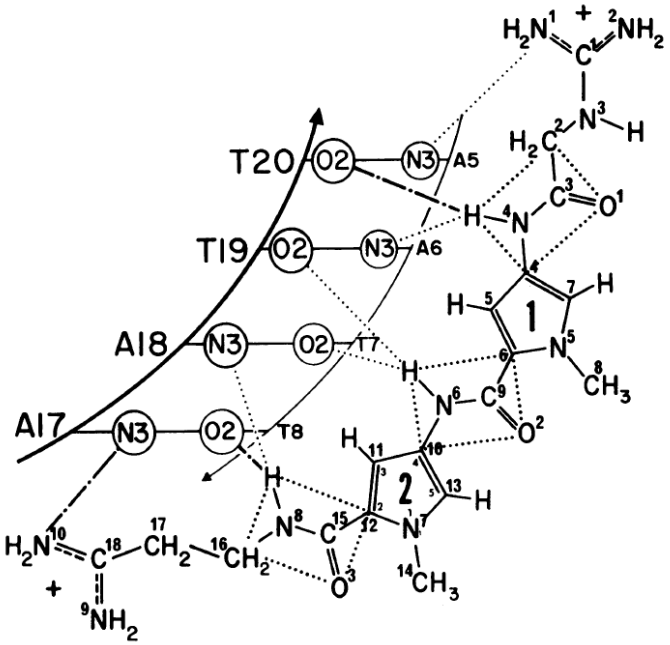


Figure 1.4. Chemical Scheme of a Netropsin Recognizing an A/T-Rich Binding Site. The minor groove atoms that interact with netropsin are shown in the DNA ladder. Standard distances of H-bonds are shown as dot-dash lines, while distances longer than 3.2 Å are indicated as dotted lines. Figure adapted from Dickerson *et al*. 1985.

Additionally, Dickerson *et al.* proposed that replacing *N*-methylpyrrole with *N*-methylimidazole would furnish a H-bond acceptor and remove the steric hindrance generated by the exocyclic amino group of guanine, allowing for the expansion of the permitted targeting sequences to include G·C pairs (**Figure 1.5**).[46]

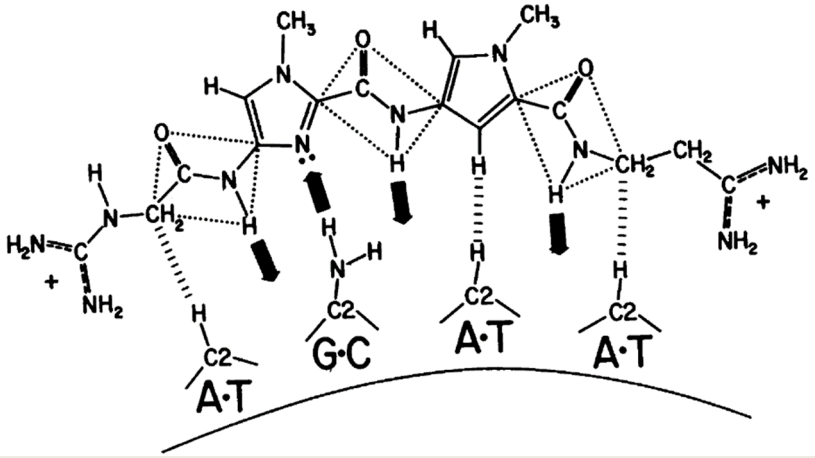


Figure 1.5. Chemical Scheme of a Lexitropsin Ligand Recognizing a G·C Base Pair as Proposed by Dickerson *et al.* The replacement of *N*-methylpyrrole with *N*-methylimidazole provides an H-bond acceptor and relieves the steric hindrance with guanine's exocyclic amino group, affording the recognition of G·C base pairs. The minor groove of DNA is represented in the bottom of the figure. H-bonds are shown as filled in arrows, while van der Waals contacts are indicated as short parallel lines. Figure adapted from Dickerson *et al.* 1985.

In the breakthrough findings by Wemmer *et al.*, it was demonstrated that in addition to the 1:1 ratio of ligand to DNA complex, distamycin A, a close relative of netropsin, could also bind DNA with a 2:1 stoichiometry.[59] Using NMR spectroscopy, the authors showed that the two ligands bound side-by-side in an anti-parallel fashion in the minor groove of an AAATT duplex, allowing each distamycin A molecule to interact independently with each anti-parallel strand of DNA (**Figure 1.6**).[59]

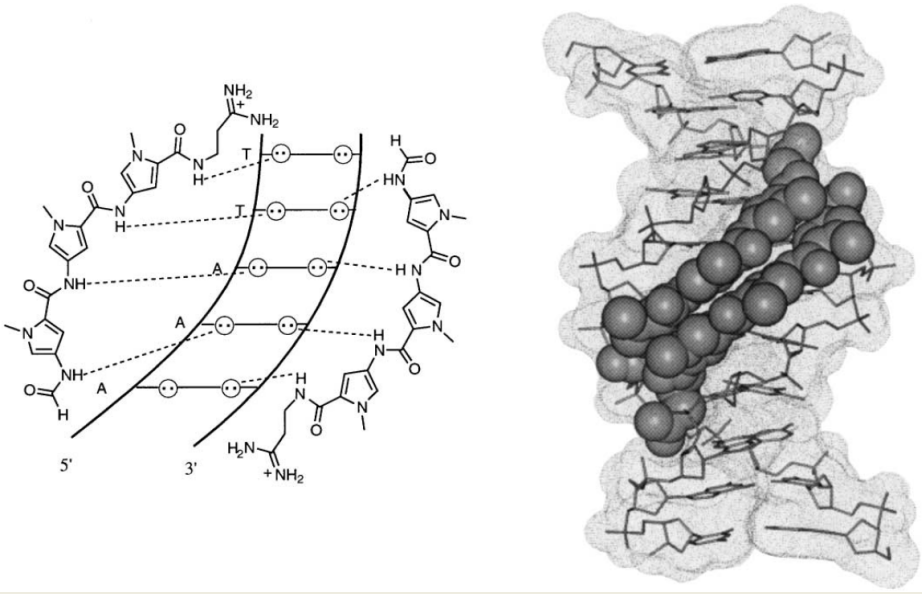


Figure 1.6. Chemical Scheme of the 2:1 Distamycin-DNA Complex. Each distamycin A molecule recognizes the DNA strands independently in an antiparallel manner. Figure adapted from Dervan *et al*. 2001 (*permission # 4087220270140*).

Consistent with these observations, Dervan *et al.* confirmed that the introduction of a *N*-methylimidazole subunit into a synthetic PA (ImPyPy) bound the 5'-WGWCW-3' (where W represents A or T) sequence as a dimer, suggesting that the antiparallel pairing of Im/Py could discriminate G·C from C·G base pairs.[60-62] Additionally, Dervan demonstrated that linking the C-terminus of a PA with the N-terminus of a second PA by γ-aminobutyric acid (GABA) preserved the appropriate ring pairing, and thus improved binding affinity and sequence specificity.[45,63]

The novel hairpin motif introduced by Dervan's group established the paradigm of antiparallel side-by-side ring pairing to recognize the sequence-dependent H-bonds presented by the Watson-Crick bases in the minor groove of B-DNA (**Figure 1.7**).[45] Specifically, a binary code has been established where a Py/Py pair H-bonds to the degenerate A·T and T·A base pairs, as each Py heterocycle recognizes nucleotides (A, T, or C) that present H-bond acceptors to the minor groove.[60-62,64] An Im/Py or Py/Im pair preferentially binds G·C or C·G base pairs, respectively. This selectivity arises from the steric hindrance between pyrrole substituents and the exocyclic amino group of guanine found in the minor groove, and is relieved by imidazole substituents which provide both room and an H-bond acceptor to this NH_2 group.[60-62,64] Notably, substitution of Py heterocycles with the flexible β-alanine (β) substituent relieves the inherent overcurvature observed in polyamides with four or more contiguous heterocycles and resets the H-bond

register between the ligand and DNA.[65,66] The GABA hairpin turn recognizes the degenerate A·T and T·A base pairs preferentially due to steric factors.[67] The C-terminal positively-charged tail recognizes A·T and T·A base pairs and mimics the cationic moieties found in distamycin A and netropsin.[46,67] Furthermore, hairpin polyamides have been shown to bind B-form DNA exclusively, with little or no affinity for dsRNA, A-DNA or Z-DNA.[46,57,68]

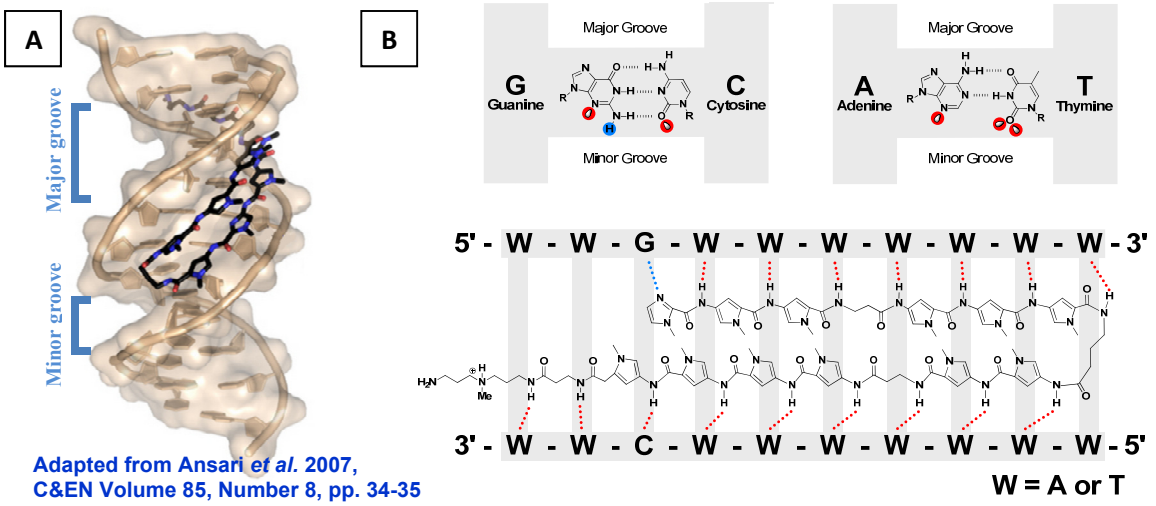


Figure 1.7. Sequence-Dependent Recognition in the Minor Groove of dsDNA by Hairpin Py/Im Polyamides. Pyrrole/imidazole hairpin polyamides can be synthesized to recognize specific DNA sequences by following the reported PA-DNA pairing rules, which delineate the Watson-Crick base pairs recognized by each heterocycle pair in a hairpin polyamide. (**A**) Model of a pyrrole/imidazole hairpin polyamide bound in the minor groove of dsDNA (Figure adapted from Ansari *et al*. 2007). (**B**) Chemical structure of the Watson-Crick base pairs with H-bond acceptors (circled in blue) and donors (circled in red). Binding model of anti-HPV **PA1** in the minor groove of dsDNA to its canonical binding sequence. Dashed lines represent H-bonds between DNA and polyamide.

Because of their modular nature, hairpin PAs afford the rational design of different contiguous combinations of their building blocks, linked by amide bonds, to target predetermined sequences for the control of gene expression and DNA replication. As a result, these small, cell-permeable molecules are attractive novel candidates as modulators of gene expression,[69-74] as molecular probes for diagnostics in living cells[75-78] and as antiviral agents for the selective elimination of dsDNA viruses.[32,79-81] Gene transcription is orchestrated by the intricate interplay between chromatin accessibility[37] and the recruitment of RNA polymerases by DNA-TF (transcription factor) complexes in canonical regulatory elements found in gene promoters.[47] Hairpin polyamides exhibit binding affinities similar to transcription factors.[67] Thus, polyamides can be designed to downregulate specific gene transcripts by direct or allosteric competition with TFs for DNA sites, depending on whether the DNA-binding protein recognizes the major or minor groove of DNA, respectively.[67] PAs have been shown to inhibit numerous DNA-binding proteins, including Ets-1,[70] TFIIIA,[82] HTLV-1 Tax,[83] EBNA1[79] and HPV18 E2.[32] Because gene expression and DNA replication are regulated via specific DNA-binding proteins,[84] hairpin polyamides are adept at specifically downregulating the propagation of viral genomes and related gene products, while avoiding toxicity to both infected and adjacent healthy cells.[80,81]

Our group and collaborators have synthesized an extensive library of hairpin PAs with broad-spectrum activities against three prevalent HR-HPV types (HPV16, HPV18 and HPV31).[80,85] Compared to the reported short hairpin polyamides, anti-PAs exhibit greater binding promiscuity.[81,86,87] Specifically, these large anti-HPV PAs tolerate PA-DNA mismatches without a significant decrease in binding affinity and require a binding site of $\geq$10 base pairs.[86,87] Recent findings by our collaborators at NanoVir revealed that effective episomal elimination by anti-HPV PAs is accompanied by a significant change in expression of members from the DNA Damage Response (DDR) pathways.[81] Most notably, the authors demonstrated that modulation of DDR pathways was dependent on the presence of HPV in the episomal form, as SiHa cells harboring integrated copies of HPV16 and HPV-negative cells did not elicit such response. Likewise, the inactive eight-ring hairpin polyamide **PA11** had no effect on the DDR transcripts studied.[81] Thus, the mode of action by which antiviral PAs eliminate HPV episomes is more complex than a simple allosteric inhibition of DNA-binding proteins, and it may comprise the concerted interplay of multiple cellular pathways.

1.5 BIBLIOGRAPHY

(1) Doorbar, J. *Clin Sci (Lond)* **2006**, *110*, 525.
(2) de Villiers, E. M.; Fauquet, C.; Broker, T. R.; Bernard, H. U.; zur Hausen, H. *Virology* **2004**, *324*, 17.
(3) Bernard, H.-U.; Burk, R. D.; Chen, Z.; van Doorslaer, K.; Hausen, H. z.; de Villiers, E.-M. *Virology* **2010**, *401*, 70.
(4) Giuliano, A. R.; Tortolero-Luna, G.; Ferrer, E.; Burchell, A. N.; de Sanjose, S.; Kjaer, S. K.; Muñoz, N.; Schiffman, M.; Bosch, F. X. *Vaccine* **2008**, *26*, K17.
(5) Ryser, M. D.; McGoff, K.; Herzog, D. P.; Sivakoff, D. J.; Myers, E. R. *Epidemics* **2015**, *11*, 32.
(6) zur Hausen, H. *Cancer Research* **1989**, *49*, 4677.
(7) Bottalico, D.; Chen, Z.; Kocjan, B. J.; Seme, K.; Poljak, M.; Burk, R. D. *J Gen Virol* **2012**, *93*, 1774.
(8) Muñoz, N.; Bosch, F. X.; Castellsagué, X.; Díaz, M.; de Sanjose, S.; Hammouda, D.; Shah, K. V.; Meijer, C. J. L. M. *International Journal of Cancer* **2004**, *111*, 278.
(9) Archambault, J.; Melendy, T. *Antivir Ther* **2013**, *18*, 271.
(10) Walboomers, J. M. M.; Jacobs, M. V.; Manos, M. M.; Bosch, F. X.; Kummer, J. A.; Shah, K. V.; Snijders, P. J. F.; Peto, J.; Meijer, C. J. L. M.; Muñoz, N. *J Pathol* **1999**, *189*, 12.
(11) Organization, W. H. Comprehensive Cervical Cancer Prevention and Control Programme Guidance for Countries. [Online Early Access]. Published Online: 2012.
(12) Faridi, R.; Zahra, A.; Khan, K.; Idrees, M. *Virol J* **2011**, *8*, 269.
(13) Clifford, G.; Franceschi, S.; Diaz, M.; Munoz, N.; Villa, L. L. *Vaccine* **2006**, *24 Suppl 3*, S3/26.
(14) Baseman, J. G.; Koutsky, L. A. *J Clin Virol* **2005**, *32 Suppl 1*, S16.
(15) Adams, H. P.; Carnright, E. L. *Clinician Reviews* **2013**, *23*, 42.
(16) Doorbar, J. *J Clin Virol* **2005**, *32 Suppl 1*, S7.

(17) Richards, R. M.; Lowy, D. R.; Schiller, J. T.; Day, P. M. *Proc Natl Acad Sci U S A* **2006**, *103*, 1522.
(18) Stanley, M. A. P., M.M.; Coleman N. *Biochemical Society Transaction* **2007**, *35*.
(19) Wang, J. W.; Roden, R. B. *Virology* **2013**, *445*, 175.
(20) Stanley, M. A. *Clin Microbiol Rev* **2012**, *25*, 215.
(21) Karim, R.; Meyers, C.; Backendorf, C.; Ludigs, K.; Offringa, R.; van Ommen, G.-J. B.; Melief, C. J. M.; van der Burg, S. H.; Boer, J. M. *PLoS ONE* **2011**, *6*, e17848.
(22) Stanley, M. *Vaccine* **2006**, *24 Suppl 1*, S16.
(23) Hasan, U. A.; Bates, E.; Takeshita, F.; Biliato, A.; Accardi, R.; Bouvard, V.; Mansour, M.; Vincent, I.; Gissmann, L.; Iftner, T.; Sideri, M.; Stubenrauch, F.; Tommasino, M. *The Journal of Immunology* **2007**, *178*, 3186.
(24) Karim, R.; Meyers, C.; Backendorf, C.; Ludigs, K.; Offringa, R.; van Ommen, G. J.; Melief, C. J.; van der Burg, S. H.; Boer, J. M. *PLoS One* **2011**, *6*, e17848.
(25) Barnard, P.; Payne, E.; McMillan, N. A. *Virology* **2000**, *277*, 411.
(26) Ronco, L. V.; Karpova, A. Y.; Vidal, M.; Howley, P. M. *Genes & Development* **1998**, *12*, 2061.
(27) Gruener, M.; Bravo, I. G.; Momburg, F.; Alonso, A.; Tomakidi, P. *Virol J* **2007**, *4*, 116.
(28) Ashrafi, G. H.; Haghshenas, M. R.; Marchetti, B.; O'Brien, P. M.; Campo, M. S. *Int J Cancer* **2005**, *113*, 276.
(29) White, P. W.; Faucher, A. M.; Massariol, M. J.; Welchner, E.; Rancourt, J.; Cartier, M.; Archambault, J. *Antimicrob Agents Chemother* **2005**, *49*, 4834.
(30) Hegde, R. S. *Annu Rev Biophys Biomol Struct* **2002**, *31*, 343.
(31) Lehoux, M.; Fradet-Turcotte, A.; Lussier-Price, M.; Omichinski, J. G.; Archambault, J. *J Virol* **2012**, *86*, 3486.
(32) Schaal, T. D.; Mallet, W. G.; McMinn, D. L.; Nguyen, N. V.; Sopko, M. M.; John, S.; Parekh, B. S. *Nucleic Acids Res* **2003**, *31*, 1282.
(33) Van Tine, B. A.; Dao, L. D.; Wu, S.-Y.; Sonbuchner, T. M.; Lin, B. Y.; Zou, N.; Chiang, C.-M.; Broker, T. R.; Chow, L. T. *Proc Natl Acad Sci U S A* **2004**, *101*, 4030.
(34) Hausen, H. z. *J Natl Cancer Inst* **2000**, *92*, 690.
(35) Disbrow, G. L.; Sunitha, I.; Baker, C. C.; Hanover, J.; Schlegel, R. *Virology* **2003**, *311*, 105.
(36) Tindle, R. W. *Nat Rev Cancer* **2002**, *2*, 6.
(37) Peña, L. d. M.; Laimins, L. A. *J Virol* **2001**, *75*, 10005.
(38) Wang, H. K.; Duffy, A. A.; Broker, T. R.; Chow, L. T. *Genes & Development* **2009**, *23*, 181.
(39) Han, K. T.; Sin, J. I. *Clin Exp Vaccine Res* **2013**, *2*, 106.
(40) Stanley, M. *Gynecologic Oncology* **2008**, *109*, S15.
(41) Kirby, T. *The Lancet Oncology* **2015**, *16*, e56.
(42) Hildesheim, A.; Herrero, R.; Wacholder, S.; et al. *JAMA* **2007**, *298*, 743.
(43) Graham, S. V. *Future microbiology* **2010**, *5*, 1493.
(44) Stanley, M. A. *Journal of General Virology* **2012**, *93*, 681.
(45) Mrksich, M.; Parks, M. E.; Dervan, P. B. *Journal of the American Chemical Society* **1994**, *116*, 7983.

(46) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc Natl Acad Sci U S A* **1985**, *82*, 1376.
(47) Dervan, P. B.; Edelson, B. S. *Current Opinion in Structural Biology* **2003**, *13*, 284.
(48) Chen, X. R., Boopathy; Rao, Sambhorao T.; Sundaralingam, Muttaiya *Nat Struct Mol Biol* **1994**, *1*, 7.
(49) Marky, L. A.; Breslauer, K. J. *Proc Natl Acad Sci U S A* **1987**, *84*, 4359.
(50) Finlay, A. C.; Hochstein, F. A.; Sobin, B. A.; Murphy, F. X. *Journal of the American Chemical Society* **1951**, *73*, 341.
(51) Herman, D. M.; Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1998**, *120*, 1382.
(52) Pilch, D. S.; Poklar, N.; Gelfand, C. A.; Law, S. M.; Breslauer, K. J.; Baird, E. E.; Dervan, P. B. *Proceedings of the National Academy of Sciences* **1996**, *93*, 8306.
(53) White, S.; Baird, E. E.; Dervan, P. B. *Biochemistry* **1996**, *35*, 12532.
(54) Schultz, P. G.; Dervan, P. B. *Journal of Biomolecular Structure and Dynamics* **1984**, *1*, 1133.
(55) Zimmer, C.; Puschendorf, B.; Grunicke, H.; Chandra, P.; Venner, H. *European Journal of Biochemistry* **1971**, *21*, 269.
(56) Zimmer, C.; Reinert, K. E.; Luck, G.; Wähnert, U.; Löber, G.; Thrum, H. *J Mol Biol* **1971**, *58*, 329.
(57) Wartell, R. M.; Larson, J. E.; Wells, R. D. *Journal of Biological Chemistry* **1974**, *249*, 6719.
(58) Patel, D. J. *Proc Natl Acad Sci U S A* **1982**, *79*, 6424.
(59) Pelton, J. G.; Wemmer, D. E. *Proceedings of the National Academy of Sciences* **1989**, *86*, 5723.
(60) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Journal of the American Chemical Society* **1992**, *114*, 8783.
(61) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proceedings of the National Academy of Sciences* **1992**, *89*, 7586.
(62) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385.
(63) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559.
(64) White, S.; Baird, E. E.; Dervan, P. B. *Chemistry & Biology* **1997**, *4*, 569.
(65) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *Journal of the American Chemical Society* **1996**, *118*, 6160.
(66) Wang, C. C. C. E., U.; Dervan, P. B. *Bioorg Med Chem* **2001**, *9*, 5.
(67) Dervan, P. B. D., R.M.; Marques, M.A. *Current Medicinal Chemistry - Anti-Cancer Agents* **2005**, *5*.
(68) Chenoweth, D. M.; Meier, J. L.; Dervan, P. B. *Angew Chem Int Ed Engl* **2013**, *52*, 415.
(69) Arora, P. S.; Ansari, A. Z.; Best, T. P.; Ptashne, M.; Dervan, P. B. *Journal of the American Chemical Society* **2002**, *124*, 13067.
(70) Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan, P. B. *Proc Natl Acad Sci U S A* **1998**, *95*, 12890.
(71) Mapp, A. K.; Ansari, A. Z.; Ptashne, M.; Dervan, P. B. *Proceedings of the National Academy of Sciences* **2000**, *97*, 3930.

(72) Nickols, N. G.; Szablowski, J. O.; Hargrove, A. E.; Li, B. C.; Raskatov, J. A.; Dervan, P. B. *Molecular Cancer Therapeutics* **2013**, *12*, 675.
(73) Raskatov, J. A.; Meier, J. L.; Puckett, J. W.; Yang, F.; Ramakrishnan, P.; Dervan, P. B. *Proceedings of the National Academy of Sciences* **2012**, *109*, 1023.
(74) Wang, X.; Nagase, H.; Watanabe, T.; Nobusue, H.; Suzuki, T.; Asami, Y.; Shinojima, Y.; Kawashima, H.; Takagi, K.; Mishra, R.; Igarashi, J.; Kimura, M.; Takayama, T.; Fukuda, N.; Sugiyama, H. *Cancer Sci* **2010**, *101*, 759.
(75) Nozeret, K.; Loll, F.; Escudé, C.; Boutorine, A. S. *ChemBioChem* **2015**, *16*, 549.
(76) Chenoweth, D. M.; Viger, A.; Dervan, P. B. *Journal of the American Chemical Society* **2007**, *129*, 2216.
(77) Maeshima, K.; Janssen, S.; Laemmli, U. K. *The EMBO Journal* **2001**, *20*, 3218.
(78) Janssen, S.; Durussel, T.; Laemmli, U. K. *Molecular Cell* **2000**, *6*, 999.
(79) Yasuda, A.; Noguchi, K.; Minoshima, M.; Kashiwazaki, G.; Kanda, T.; Katayama, K.; Mitsuhashi, J.; Bando, T.; Sugiyama, H.; Sugimoto, Y. *Cancer Sci* **2011**, *102*, 2221.
(80) Edwards, T. G.; Koeller, K. J.; Slomczynska, U.; Fok, K.; Helmus, M.; Bashkin, J. K.; Fisher, C. *Antiviral Res* **2011**, *91*, 177.
(81) Edwards, T. G.; Vidmar, T. J.; Koeller, K.; Bashkin, J. K.; Fisher, C. *PLoS One* **2013**, *8*, e75406.
(82) Gottesfeld, J. M.; Neely, L.; Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1997**, *387*, 202.
(83) Lenzmeier, B. A.; Baird, E. E.; Dervan, P. B.; Nyborg, J. K. *J Mol Biol* **1999**, *291*, 731.
(84) Demeret, C.; Le Moal, M.; Yaniv, M.; Thierry, F. *Nucleic Acids Res* **1995**, *23*, 4777.
(85) Bashkin, J. K.; Koeller, K. J.; Edwards, T. G.; Fisher, C.; Google Patents: 2007.
(86) He, G.; Vasilieva, E.; Harris Jr, G. D.; Koeller, K. J.; Bashkin, J. K.; Dupureur, C. M. *Biochimie* **2014**, *102*, 83.
(87) Koeller, K. J.; Harris, G. D.; Aston, K.; He, G.; Castaneda, C. H.; Thornton, M. A.; Edwards, T. G.; Wang, S.; Nanjunda, R.; Wilson, W. D.; Fisher, C.; Bashkin, J. K. *Medicinal Chemistry* **2014**, *4*, 338.

Chapter 2

Determination of the DNA-Binding Properties of Antiviral Hairpin Polyamides PA1 and PA25 on Natural DNA Sequences Corresponding to the Long Control Region (LCR) of HPV18 Genome

2.1 ABSTRACT

Human papillomavirus (HPV) is the main etiological agent in the development of cervical cancer worldwide, with HPV18 accounting for approximately 20 % of the reported cases. *N*-methylpyrrole/*N*-imidazole hairpin polyamides **PA1** and **PA25** are among a large library of antiviral compounds capable of significantly decreasing HPV18 episomes in cell culture (**PA1** IC_{50} = 0.7 μM; **PA25** IC_{50} = 0.06 μM) without cytotoxicity. To establish whether the basis of the antiviral activity differences is a result of dissimilar DNA-binding properties, the sequence specificity, binding sites and equilibrium dissociation constants were interrogated by DNase I footprinting and affinity cleavage experiments using DNA sequences corresponding to the HPV18 regulatory regions under cell-free conditions. Although these DNA-ligands were designed to target specific A/T-rich sequences (**PA1** = W_2GW_7 and **PA25** = $W_2GW_5GW_4$, where W can be A or T), we find an extensive coverage of the studied DNA sequences with a population of binding sites with multiple polyamide-DNA mismatches. Specifically, both large hairpin polyamides tolerate as many as 2 (**PA1**) or 3 (**PA25**) DNA-polyamide mismatches without a significant change in binding affinity. We also report that these minor groove binders bind avidly to A/T-rich sequences corresponding to the HPV18 genome with equilibrium dissociation constants ranging from 1.1 nM to 2.6 nM for **PA1** and 0.4 nM to 1.2 nM for **PA25**.

2.2 INTRODUCTION

2.2.1 Human Papillomavirus (HPV)

Human papillomavirus (HPV) is a small double-stranded DNA virus with a tropism for mucosal and cutaneous epithelial tissues.[1] HPV is the most prevalent sexually transmitted infection in the United States,[2] and remains a significant health concern as persistent infections with an oncogenic HPV type can lead to the development of cervical cancer, in addition to oral and other head and neck cancers.[3-5] To date, over 180 types have been isolated of which approximately 15 oncogenic types have been identified.[6-8] Of these types, HPV16 and HPV18 are the most prevalent and together account for approximately 70 % of the reported cervical cancers worldwide.[9]

At the molecular level, the dsDNA genome of HPV18 consists of 7,857 base pairs that encode for eight well-characterized proteins. Based on their function within the lifecycle, the HPV18 genome can be divided into three regions: a late region (L; 4244-7136 bp), a non-coding region (LCR, long control region; 7137-104 bp) and an early region (E; 105-4157 bp) (**Figure 2.1**).[10] The encoded non-structural proteins from the early region primarily serve regulatory functions including DNA replication, transcription, immune evasion and cell signaling. The proteins encoded from the late region are necessary for the assembly of infectious virions that can infect naïve individuals.[11] On the other hand, the LCR genomic region does not encode for any proteins, but it harbors the early promoter and regulatory sequences vital to HPV's lifecycle.[10] Thus, the DNA sequences studied here correspond to the LCR region of the HPV18 genome.

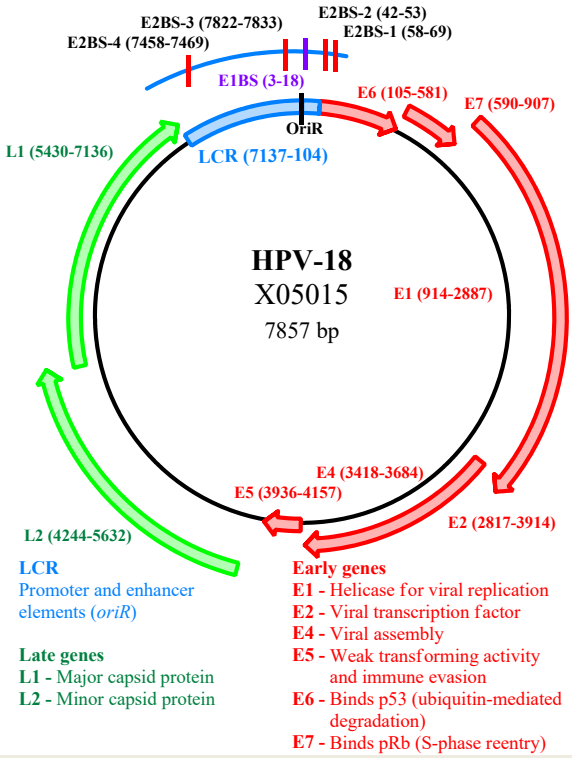


Figure 2.1. Structure of the HPV18 Genome. The circular genome of HPV18 (GenBank Accession No. X05015) encodes for early (E) and late (L) genes and consists of 7857 bp of dsDNA. Coding regions for each gene are annotated with their respective positions in the genome. The E2 and E1 binding sites within the long control region (LCR) are illustrated above the map. The functions for the HPV18 encoded genes are also given below the map.

Prevention of HPV-related cervical cancer has benefited from Pap smears and related screening programs.[12,13] Although prophylactic vaccines introduced about ten years ago have also made an impact, vaccines are ineffective against existing infections. Furthermore, levels of vaccination have not yet achieved projected values.[11,14,15]

Most HPV infections are asymptomatic and are cleared by the immune system.[16] However, more than 35,000 HPV-related cancer cases were reported in 2009 in the United States alone.[5] In third world countries, where these preventive measures are less available, HPV infections remain a significant health burden; the approximately 500,000 new cases diagnosed annually contribute to 290,000 deaths within the same time frame.[17] Therefore, a need for antivirals against HPV is dramatic and such drugs would be lifesaving. Toward this goal, we have designed and synthesized an extensive library of large *N*-methylpyrrole/*N*-methylimidazole hairpin polyamides with broad-spectrum activities against three prevalent HR-HPV types (HPV16, HPV18 and HPV31).[18,19]

2.2.2 *N*-Methylpyrrole/*N*-Methylimidazole Hairpin Polyamides (Py/Im PAs)

N-methylpyrrole/*N*-methylimidazole (Py/Im) hairpin polyamides (PAs) are synthetic heteroaromatic compounds inspired by the chemical structures of the natural antibiotics distamycin A and netropsin, which bind to the minor groove of A/T-rich DNA sequences.[20-23] These crescent-shaped molecules, collectively known as polyamides, bind with high affinity (nano- to picomolar dissociation constants) to the minor groove of dsDNA in a sequence-selective fashion.[24-26] According to the literature, Py/Im PAs can

be rationally designed to target predetermined DNA sequences based on the established PA-DNA recognition rules.[27-31] These recognition rules delineate the Watson-Crick base pairs recognized by each antiparallel side-by-side ring pairing in a hairpin polyamide. A Py/Im pair H-bonds to a G·C (but not a C·G), whereas a Py/Py H-bonds to the degenerate A·T and T·A base pairs. While replacing a Py heterocycle with the flexible β-alanine substituent relieves the inherent overcurvature observed in polyamides with four or more contiguous heterocycles and resets the H-bond register between the ligand and DNA.[27,28] The gamma-aminobutyric turn recognizes the degenerate A·T and T·A base pairs preferentially due to steric factors.[29] The C-terminal positively-charged tail recognizes A·T and T·A base pairs and mimics the cationic moieties found in distamycin A and netropsin.[29,30] This reported specificity is achieved by the optimal positioning of the H-bond acceptors and donors, van der Waals interactions with the walls of the minor groove and appropriate curvature of the polyamide with respect to the minor groove of B-DNA.[31]

Because of their modular nature, hairpin PAs afford the rational design to target predetermined sequences for the control of gene expression and DNA replication. As a result, these small, cell-permeable molecules are attractive novel candidates as modulators of gene expression,[32-37] as molecular probes for diagnostics in living cells[38-41] and as antiviral agents for the selective elimination of dsDNA viruses.[18,42-44] Because polyamide-binding events impart significant structural perturbations to DNA,[45] these synthetic DNA-ligands can allosterically repress gene transcription and DNA replication by competing with the endogenous DNA-binding proteins and transcription factors.[29] Of interest, Yasuda *et al.* designed Py-Im hairpin polyamides targeting the Epstein-Barr Nuclear Antigen 1 (EBNA1) binding sites in the origin of plasmid replication (*oriP*) within the viral genome and demonstrated an efficient inhibition of EBNA1-binding and reduced recruitment of host replication machinery to the *oriP* in chromatin immunoprecipitation (ChIP) assays.[42] Schaal *et al*. showed that tandem hairpin polyamides were able to displace the HPV18 E2, a major groove binding protein, from its canonical E2BS4 (E2 Binding Site 4) DNA binding site. Consistent with the allosteric model, the authors suggested that these tandem hairpin polyamides prevented the bending of DNA structure required for E2 binding.[43]

Displacement of viral proteins as well as cellular transcription factors and the replication machinery from their viral, canonical binding sites by anti-HPV PAs would presumably cause episomal instability followed by the elimination of the HPV genome. In fact, the multifunctional E2 protein forms a complex with E1 to initiate viral DNA replication within the HPV LCR,[1] while E2 alone tightly regulates transcription of viral proteins[43] and ensures proper episomal segregation during cell division of infected cells.[46,47] Because the HPV LCR harbors the early promoter and regulatory sequences vital to HPV's lifecycle, our group designed **PA1** to target specific viral DNA sequences near the origin of replication in the LCR of HPV16. Despite the heterogeneity of the DNA sequences among these types, **PA1** and **PA25** exhibit broad-spectrum antiviral activities against these prevalent high-risk types (**Table 2.1**). Our novel anti-HPV polyamides potentiate a substantial and rapid decrease in episomal levels in human keratinocytes and tissue cultures without measurable cytotoxicity *via* MTT assays.[18,19,48] Remarkably, the loss of episomal DNA is not associated with viral integration into the host genome.[18] Specifically, **PA1** and **PA25** exhibit IC_{50} values of 0.7 μM and 0.06 μM against HPV18, respectively.

Table 2.1. IC_{50} and IC_{90} Values of **PA1** and **PA25** Against High-Risk HPV16, HPV18 and HPV31 Types.

	HPV16			HPV18			HPV31		
	IC_{50} (μM)	**IC_{90} (μM)**	**N**	**IC_{50} (μM)**	**IC_{90} (μM)**	**n**	**IC_{50} (μM)**	**IC_{90} (μM)**	**n**
PA1	0.1(1)	1.1(8)	4	0.7(4)	>10	3	0.1(1)	1.0(5)	4
PA25	0.036(1)	0.351	3	0.056(5)	1.462	6	0.030(1)	0.510	3

The numbers in parentheses are standard deviations; n is the number of independent measurements.

Py/Im **PA1** has the sequence dIm-Py-Py-β-Py-Py-Py-γ-Py-Py-β-Py-Py-Py-Py-β-Ta (dIm = desamino-imidazole, Py = pyrrole, β = β-alanine, γ = γ-aminobutyric acid and Ta = $CH_3N(CH_2CH_2CH_2NH_2)_2$).[49] According to the prediction rules, **PA1** recognizes the DNA sequence 5'-W_2GW_7-3', where W (Weak) can be A or T. On the other hand, Py/Im **PA25** has the sequence dIm-Py-Py-β-Py-Py-Im-β-Py-Py-γ-Py-Py-β-Py-Py-Py-β-Py-Py-Py-β-Ta and a cognate DNA binding motif of 5'-$W_2GW_5GW_4$-3'.[44] The chemical structures and cognate binding sites for **PA1** and **PA25** are shown in **Figure 2.2**.

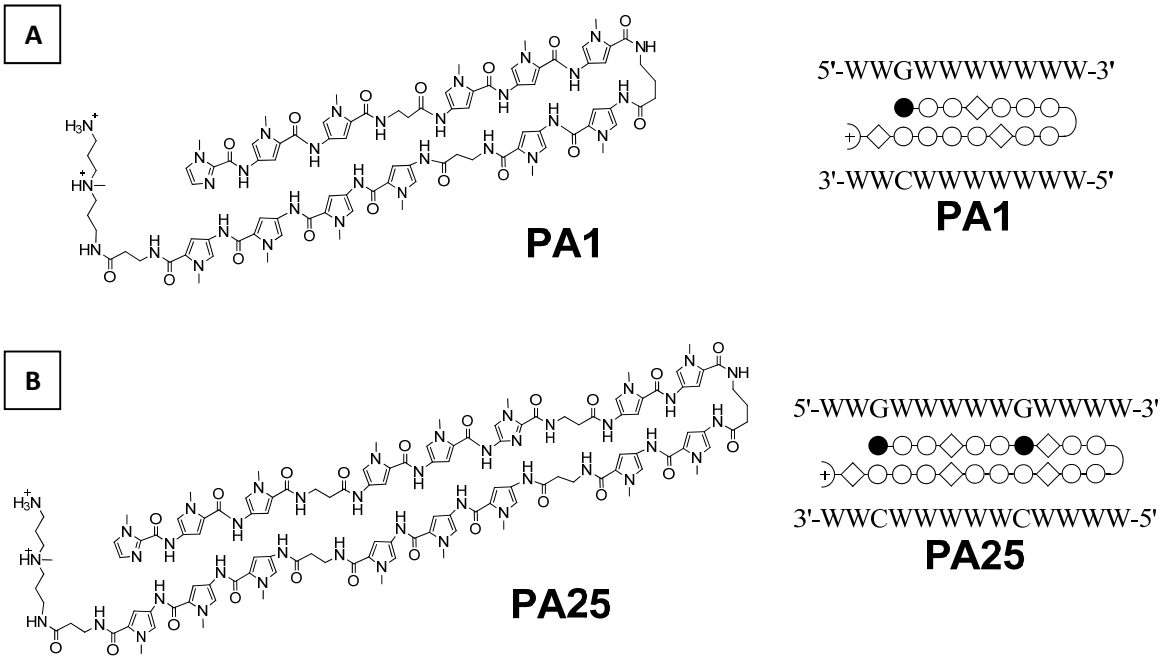


Figure 2.2. Chemical Structures for PA1 and PA25. Chemical structure of anti-HPV **(A) PA1** and **(B) PA25** are provided. According to the PA-DNA recognition rules, **PA1** targets the DNA sequence 5'-W_2GW_7-3', where W (Weak) can be A or T. On the other hand, the cognate DNA binding motif for **PA25** is 5'-$W_2GW_5GW_4$-3'. The polyamide building blocks are represented by open circles for *N*-methylpyrrole, filled circles for *N*-methylimidazole, diamonds for β-alanine, + for the Ta tail and ⊃ for γ-aminobutyric acid.

To establish whether the basis of these antiviral activity differences between **PA1** and **PA25** is a result of dissimilar DNA-binding properties, the sequence specificity, binding orientation and dissociation constants were interrogated under cell-free conditions for the HPV18 LCR genomic region (7479-157 bp) by quantitative DNase I (deoxyribonuclease I) footprinting and affinity cleavage. Binding events to the specific regulatory sequences (*i.e.*, E1-BS, E2BS, *oriR*; see **Figure 2.1**) in the LCR will be assessed to investigate potential allosteric competition with DNA-binding proteins. Because the anti-HPV polyamides are much larger than those reported in the literature, it is of considerable

interest to determine whether the pairing rules established with small eight-ring PAs also apply to our large hairpin PAs.

2.2.3 Deoxyribonuclease I (DNase I) Footprinting

DNase I footprinting pioneered by Galas and Schmitz[50] provides a powerful yet simple technique for deciphering the specific contacts in DNA-ligand complexes. DNase I footprinting measures the protection of DNA by a ligand from the enzymatic cleavage by DNase I endonuclease.[50,51] This method has been successfully employed to determine the sequence selectivity of numerous DNA-ligands including intercalators,[52-55] transcription factors,[56-58] and minor groove DNA binders.[52,53,55,59,60]

The principle of DNase I footprinting is summarized schematically in **Figure 2.3**. A dsDNA fragment harboring the sequence of interest is end-labelled, either radioactively with ^{32}P or fluorescently for visualization purposes. The choice of the end-labelling depends on the method of detection of the generated fragments. In this chapter, we have relied on separation and visualization by automated capillary electrophoresis using two different fluorescent dyes covalently attached to the 5'-ends of the strands in the DNA duplex. The dual fluorescently-labeled DNA fragment is then incubated in the presence and absence of the DNA-binding ligand (*e.g.*, polyamide). For quantitative DNase I footprinting, multiple solutions of DNA are incubated with increasing concentrations of the DNA-binding ligand. After equilibration of the ligand-DNA complex, the ensemble of DNA molecules are subjected to mild digestion by DNase I endonuclease under conditions of single-hit kinetics, that is, one cut per DNA molecule on average.[51] DNase I binds to the minor groove and cleaves the phosphodiester backbone of DNA, leading to the formation of cleavage products with a 3'-hydroxyl group and a 5'-phosphate.[61] The fragmented DNA is then separated by automated capillary electrophoresis and the cleavage patterns are compared between the naked and ligand-bound DNA samples. Binding of the ligand to its recognition site, or cognate sequence, leads to increased protection of DNA at that particular location, providing a 'footprint' (**Figure 2.3A**).[51,61] In the case of quantitative DNase I footprinting, monitoring the decrease of cleavage products at the particular fragment length where the polyamide binds as a function of increasing ligand concentrations allows for the determination of the equilibrium dissociation constants (**Figure 2.3B**).[62]

Although this technique can reveal both the binding sites and the binding constants of DNA-binding ligands, DNase I footprinting has some disadvantages. Specifically, the cleavage agent exhibits sequence-dependent structural motifs of the DNA duplex, leading to low signal-to-noise ratio at some regions of the DNA fragment, precluding the determination of binding constants at these sites. In addition, DNase I is relatively large, with a DNA-binding surface of approximately 10 bp, which causes overestimation of ligand binding site sizes and poor resolution of multiple ligands bound in close proximity of each other.[51,61,63]

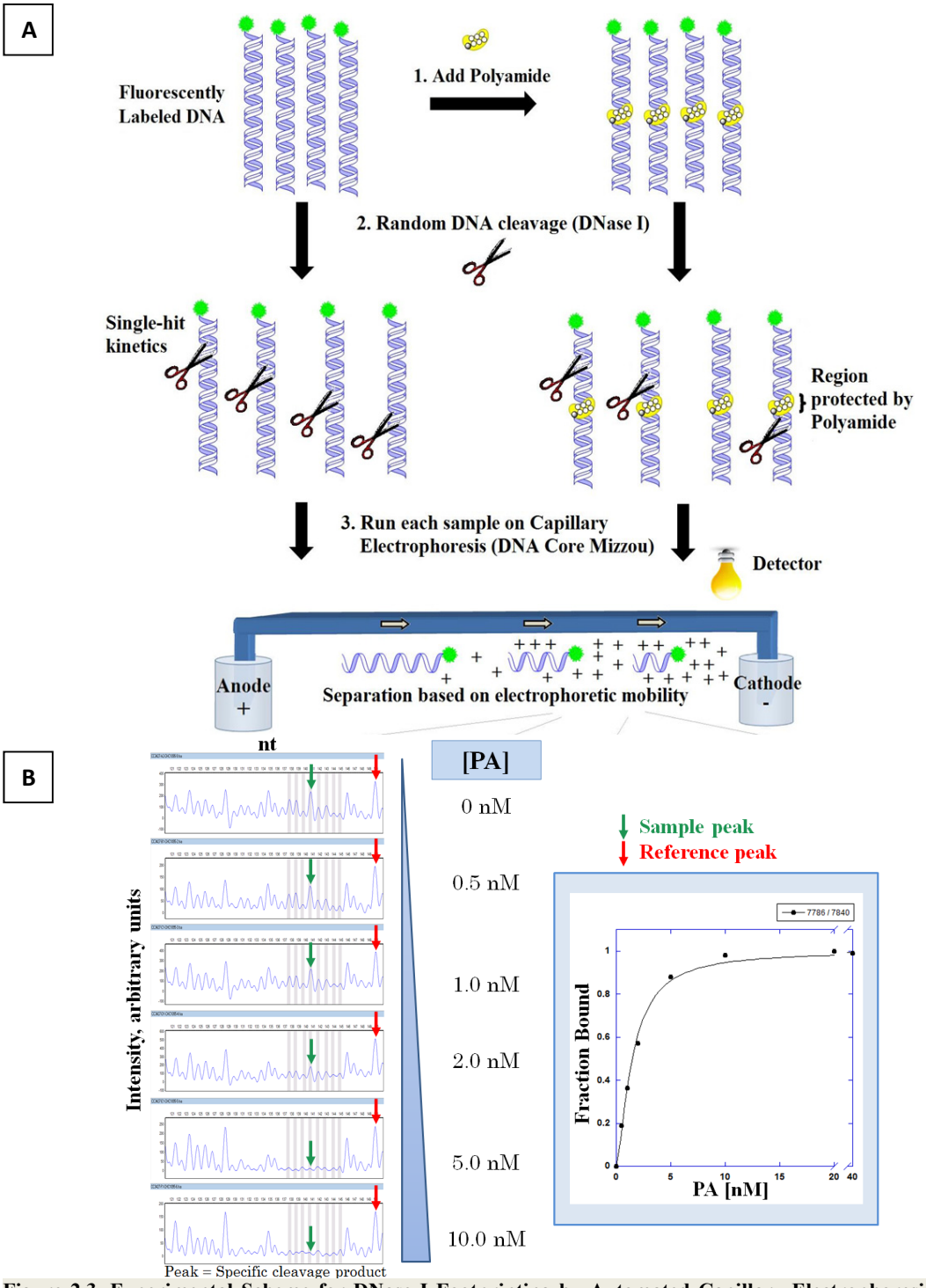


Figure 2.3. Experimental Scheme for DNase I Footprinting by Automated Capillary Electrophoresis. (A) For reasons of clarity, only one 5' end of the DNA fragments is shown as fluorescently-labeled in this illustration. Briefly, the dual fluorescently-labeled DNA fragment is incubated in the absence and presence of different PA amounts for at least 4 hours at 37 ºC. At the end of the incubation period, treatment with the appropriate concentration of DNase I is performed to yield single-hit kinetics or one cut per DNA molecule on average. **(B)** Quantitative analysis of polyamide footprints. Because the bound polyamides protect the DNA from endonuclease cleavage, automated capillary electrophoresis analysis of the control and polyamide-treated samples permits for the determination of polyamide binding sites and equilibrium dissociation constants. Green arrow – integration peak within polyamide footprint; red arrow – integration peak not bound by polyamide, thereby, it is used to normalize the data. nt – nucleotide.

2.2.4 Affinity Cleavage

The affinity cleavage method relies on the sequence-specific cleavage of DNA by a DNA-binding molecule (*e.g.*, polyamide) that has been conjugated with EDTA-Fe^{2+}. In the case of hairpin polyamides, the EDTA moiety is covalently attached onto its C-terminus (*i.e.*, Ta tail) *via* an amide bond. In the presence of ferrous iron, oxygen and the reducing agent dithiothreitol (DTT), the catalytic formation of diffusible hydroxyl radicals occurs at the C-terminal EDTA and leads to oxidative cleavage of DNA near this position (**Figure 2.4A**). The cleavage pattern yields multiple neighboring cleavage sites with a Gaussian-like distribution due to the ability of the hydroxyl radicals to diffuse along the DNA backbone. Furthermore, the pattern is also asymmetrical compared to the opposite strand, exhibiting a 3' shift from the location of the EDTA-Fe^{2+} moiety. The 3' shift is a result of a 2 bp offset of the deoxyribose to its complement on the 3' side of the minor groove of B-form DNA **(Figure 2.4B)**.[64] Affinity cleavage experiments were performed to decipher the polyamide binding sites and relative orientations in the fragments corresponding to the HPV18 LCR.

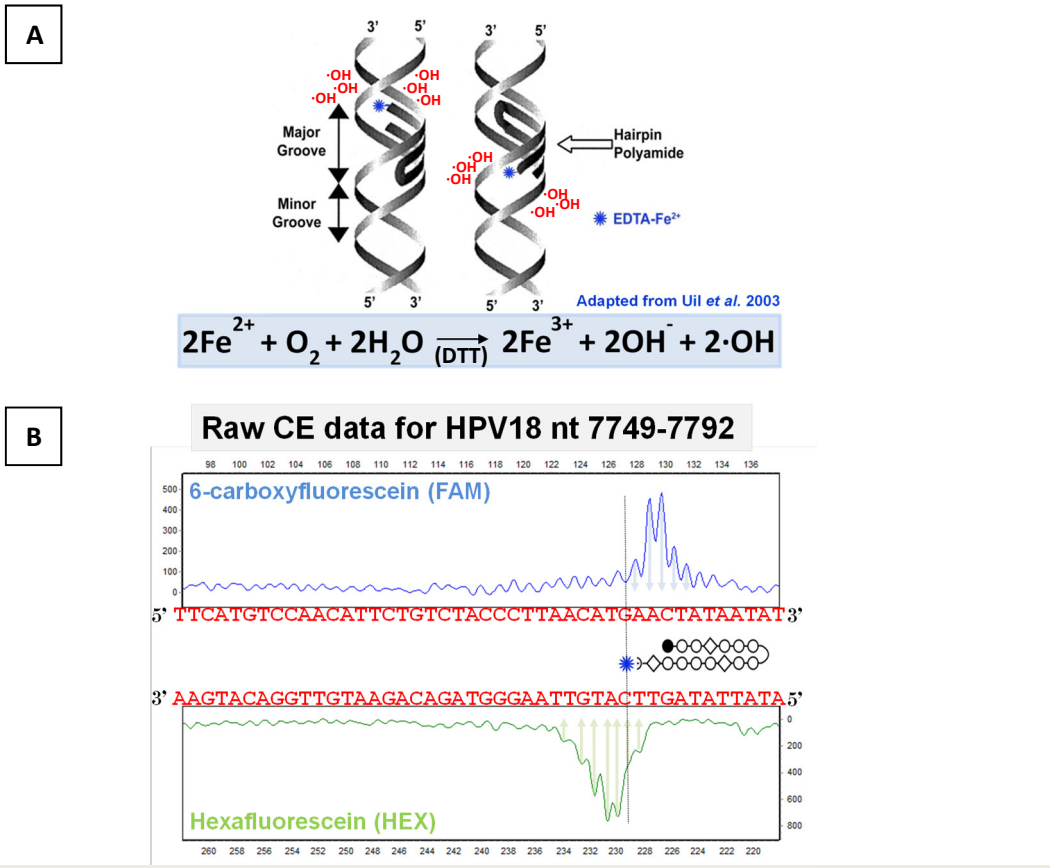


Figure 2.4. Cartoon Representation of PA-EDTA-Fe^{2+} Analog in the Minor Groove of DNA and Representative Electropherograms. (A) Affinity cleavage is performed using PA-EDTA conjugates. Sequence-specific cleavage of DNA by a hairpin polyamide covalently attached with EDTA onto its C-terminus (*i.e.* Ta tail). In the presence of ferrous iron, oxygen and the reducing agent dithiothreitol (DTT), the catalytic formation of diffusible hydroxyl radicals occurs at the C-terminal EDTA and leads to oxidative cleavage of the DNA near this position (Figure adapted from Uil *et al.* 2003; *permission # 4087221444733*). **(B)** Representative FAM/HEX channel electropherograms. Cleavage pattern yields multiple neighboring cleavage sites with a Gaussian-like distribution due to the ability of the hydroxyl radical to diffuse along the backbone of the DNA. Asymmetrical pattern is generated compared to the opposite strand with a 3' shift from the location of the EDTA-Fe^{2+} moiety.

2.3 MATERIALS AND METHODS

2.3.1 Buffers and Reagents

Starting reagents were used without further purification, unless specifically noted. Autoclaved MilliQ H_2O (18.2 MΩ·cm at 25 °C, Millipore Integral 10) was used in the preparation of all reagents. HPV18/pBR322 plasmid was purchased from ATCC, Catalog # 45152. Lysogenic Broth (LB) Agar was obtained from Sigma, Catalog # L7533-6X500ML. Glucose was purchased from Aldrich, Catalog # 158968-1KG. Ampicillin was purchased from Fisher Scientific, Catalog #BP1760-25. Calcium chloride was obtained from Alfa Aesar, Catalog # 10680. *Bam*HI and *Eco*RI restriction enzymes (Catalog # R0136S, 20,000 U/mL / Catalog # R0101S, 20,000 U/mL), 1X NEBuffer 3.1 (Catalog # R7203S, 10X concentration), Gel Loading Dye, Blue (6X, Catalog # B7021S), 1 kb DNA ladder (Catalog # N3232S, 500 µg/mL) and 100 bp DNA ladder (Catalog # N3231S, 500 µg/mL) were purchased from New England Biolabs. Ethylenediaminetetraacetic acid (EDTA) was purchased from Sigma, Catalog # E5134-100G. 1X TAE buffer consisted of 40 mM Tris-acetate, 1 M EDTA, pH 8.0. Ethidium bromide was obtained from Spectrum (Catalog # E1031, 1 % solution). Oligonucleotides for DNase I footprinting, affinity cleavage and circular dichroism experiments were purchased from Integrated DNA Technologies. Oligonucleotides for X-ray crystallography screening were obtained from Midland Certified Reagent Company. Deoxynucleotide (dNTP) Solution Mix (Catalog # N0447S, 10 mM), 10X Standard *Taq* Reaction Buffer (Catalog # B9014S) and *Taq* DNA Polymerase (Catalog # M0273L, 5,000 U/mL) were purchased from New England Biolabs. Sodium acetate (Catalog # S7545-250G) was obtained from Sigma. Isopropanol (Catalog # BP2618-1) was purchased from Fisher Scientific. Thermo Sequence Dye Primer Manual Cycle Sequencing Kit USB (Catalog # 79260 1 KT) was obtained from Affymetrix. DNase I endonuclease (Catalog # M610A, 1,000 units) was purchased from Promega. QIAGEN QIAprep Spin Miniprep Kit (Catalog # 27104) and QIAquick PCR Purification Kit (Catalog # 28106) was purchased from QIAGEN. Dimethyl sulfoxide was purchased from Sigma Life Science, Catalog # D2650. TKMC buffer consisted of 10 mM Tris, 10 mM KCl, 5 mM $MgCl_2$ and 5 mM $CaCl_2$. Ammonium iron(II) sulfate hexahydrate was obtained from Aldrich, Catalog # 203505-25G. 0.22 µm MILLEX-GP filters were purchased from Merck Millipore Ltd. 100 % ethanol was obtained from Decon Laboratories, Inc., Catalog # 2716. Agarose was obtained from Sigma, Catalog # A9539-100G. Natrix HT crystallization matrix (Catalog # HR2-131) was purchased from Hampton Research.

2.3.2 Polyamide Synthesis

Hairpin polyamide synthesis was performed by Dr. K. J. Koeller and Dr. G. D. Harris, Jr. **PA1**, **PA25** and their EDTA conjugates were synthesized by Boc solid-phase methods[65] as previously reported in the literature.[18,66]

2.3.3 HPV18/pBR322 Plasmid DNA Amplification and Purification

2.3.3.1 Transformation of Escherichia coli JM109 Cells with HPV18/pBR322 Plasmid

HPV18 DNA amplification was accomplished by transforming *E. coli* JM109 cells with a plasmid construct consisting of the full HPV18 genome (GenBank Accession No. X05015) cloned into the *Eco*RI restriction site of the pBR322 vector (ATCC, Catalog # 45152, total length = 12,219 bp). Because the *Eco*RI restriction site (which cleaves the 5' adenine on each DNA strand of the sequence 5'-GAATTC-3') occurs within the open reading frame of E1, cloning of the HPV18 genome into the pBR322 vector leads to the disruption of the E1 coding region.[67,68]

In order to avoid contamination, aseptic techniques were implemented when handling media and bacterial colonies. In particular, surfaces and disposable gloves were treated with 70 % ethanol, as well as the equipment (*i.e.*, inoculation loop) was sterilized with 70 % ethanol followed by heating with an open flame. Sterile Lysogenic Broth (LB) agar (Sigma, Catalog # L7533-6X500ML) was melted in a microwave and poured into three plates (approximately half full). The LB agar was allowed to solidify and the plates were labeled 1 JM109, 2 JM109 and Control. 1 JM109 and 2 JM109 were each divided into two sections. In one of the halves from 1 JM109, an *E. coli* JM109 stab was plated followed by serial dilution of the bacterial colonies into the three remaining halves (**Figure 2.5**). The plates were placed upside-down in a Thermo Scientific MaxQ 4000 incubator and incubated overnight at 37 °C. There were no cells in the Control plate after the overnight incubation period.

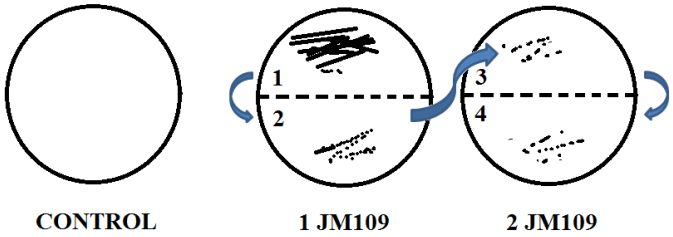


Figure 2.5. Streaking LB Agar Plates. Illustration of the streaking procedure. The *E. coli* JM109 stab was streaked directly onto Section 1 of Plate 1 JM109 followed by sterilization of the inoculation loop. Some colonies were then transferred onto Section 2 of Plate 1 JM109 by gently touching the sterile inoculation loop to a streaked region of Section 1 followed by streaking onto Section 2. This procedure was then repeated for Section 3 (using colonies from Section 2) and Section 4 (using colonies from Section 3) of Plate 2 JM109. Individual colonies were obtained from Section 2 of Plate 1 JM109. CONTROL = no *E. coli* plated.

The following day, a single colony from 1 JM109 plate Section 2 was inoculated into a single sterile test tube containing 3 mL of LB solution and labeled CHC 01-14-13 JM109. Similarly, a control was prepared and labeled CHC 01-14-13 CTL. Both test tubes were placed in a Brunswick G Series 25 incubator shaker at 37 °C and 250 RPM overnight. The next morning, 1 mL of LB solution was transferred into a sterile flask containing 50 mL sterile LB media. The *E. coli* cells were allowed to grow in a Brunswick G Series 25 incubator at 37 °C and 250 RPM until an optical density of 0.4 $\pm$ 0.05 at 590 nm was achieved. Once the optical density was reached, the LB medium harboring the *E. coli* cells was poured into a sterile 50 mL conical centrifuge tube and placed on ice for 40 min. The cells were pelleted by centrifugation for 10 min at 2,800 RPM, 4 °C in a Thermo Scientific Sorvall ST16R centrifuge. The supernatant was

decanted and the tube was placed upside down for ~1 min to remove any remaining media. The cell pellet was gently resuspended in 10 mL of sterile 0.1 M $CaCl_2$ (Alfa Aesar, Catalog # 10680) and placed in an ice bath. After three minutes of cold shock, the suspended *E. coli* cells were spun for 10 min at 2,800 RPM, 4 °C in a Thermo Scientific Sorvall ST16R centrifuge. The supernatant was decanted and 2 mL of sterile 0.1 M $CaCl_2$ were added to the cell pellet. Four separate 200 µL aliquots were transferred into individual autoclaved 1.5 mL microcentrifuge tubes. These were then placed on ice and stored at 4 °C.

After about three hours at 4 °C, competent cells and HPV18/pBR322 (ATCC Catalog No. 45152; stock concentration = 1.10 µg/µL) were thawed on ice. A 40X dilution was prepared by diluting 2 µL of the DNA plasmid stock in 78 µL of autoclaved MilliQ H_2O. A 1 µL aliquot of this solution was then added to one of the 1.5 mL microcentrifuge tubes containing 200 µL of competent cells, while a separate tube without DNA plasmid was used as a control. These tubes were labeled CHC 01-15-13 (+) and CHC 01-15-13 (-) respectively and incubated on ice for 30 min. In the meantime, ampicillin/LB agar plates were prepared. Briefly, 30 µL of 150 mg/mL ampicillin (Fisher Scientific, Catalog # BP1760-25) solution was separately added to three LB agar plates and evenly spread aseptically. Upon completion of the 30 min incubation on ice, the tubes containing the competent *E. coli* cells were transferred into a 42 °C water bath for 45 s, followed by 90 s in an ice bath. The cells were inoculated in growth media consisting of 1 % glucose (Aldrich 158968-1KG) in LB media. In order for the cured cells to develop ampicillin resistance, the control (competent cells without HPV18/pBR322) and the tubes containing competent cells with DNA plasmid were incubated for 1 hour and 15 min at 37 °C and 250 RPM in a Brunswick G Series 25 incubator. A 100 µL aliquot from each of the control and competent cells subjected to HPV18/pBR322 were plated out separately onto individual ampicillin/LB agar plates. On a third plate, a 200 µL aliquot of the competent cells with HPV18/pBR322 was also plated out. These three plates were incubated overnight at 37 °C in a Thermo Scientific MaxQ 4000 incubator.

The next morning, the ampicillin/LB agar plates were inspected for bacterial growth. The control plate did not grow any colonies and minimal growth was observed on the plate with 100 µL aliquot of JM109. Single colonies covered the plate with 200 µL aliquot of JM109. A single colony from the 200 µL plate was inoculated into a sterile flask with 50 mL of LB solution supplemented with 50 µL of 150 mg/mL ampicillin solution. This was repeated once to yield two separate cell suspensions. The flasks were labeled CHC 01-16-13 HPV18-1 and CHC 01-16-13 HPV18-2. Both samples were incubated overnight at 37 °C and 250 RPM in a Brunswick G Series incubator.

2.3.3.2 HPV18/pBR322 Plasmid Purification Using QIAprep Spin Miniprep Kit

HPV18/pBR322 plasmid DNA[67,68] purification from transformed *E. coli* JM109 cells was achieved with a QIAGEN QIAprep Spin Miniprep Kit (Catalog # 27104) as recommended by the manufacturer. Briefly, after overnight incubation, a 15 mL aliquot from each flask (CHC 01-16-13 HPV18-1 and CHC 01-16-13 HPV18-2) was separately centrifuged for 15 min at 4,000*x g*. The supernatant was discarded and each cell pellet was resuspended in 2 mL of Buffer P1. Each sample was separately aliquoted into five 1.5 mL microcentrifuge tubes (400 µL per tube). Cell lysis was performed by adding 400 µL of Buffer P2 to each tube and was mixed by inverting the samples gently 4-6 times.

To each sample, a 560 µL aliquot of Buffer N3 was added and the resulting slurry was mixed immediately by inverting the tube 4-6 times. The samples were centrifuged for 10 min at 17,900*x g*.

The resulting supernatants were applied to QIAprep spin columns (3 columns per sample; 6 columns total) and centrifuged for 60 s at 17,900*x g*. The flow-through was discarded, followed by washing of the columns with 750 µL of Buffer PE and were centrifuged for 60 s at 17,900*x g*. The flow-through was discarded and the columns were centrifuged for an additional minute at 17,900*x g* to remove the residual Buffer PE. Each column was then placed into a clean 1.5 mL microcentrifuge tube and the DNA was eluted by adding 50 µL of Buffer EB onto the center of the column. The columns were allowed to stand for 1 min, followed by centrifugation at 17,900*x g* for 60 s. The eluted plasmid DNA in Buffer EB were pooled together into two different samples (HPV18-1 CHC1007-1 and HPV18-2 CHC1007-2) and stored at -20 °C. Plasmid DNA concentrations were determined by measuring the absorbance at 260 nm using a Shimadzu UV1800 UV spectrophotometer.

2.3.3.3 Entry of HPV18/pBR322 Clones into Plasmid Library and Stabs

In order to ensure that the plasmid library and stabs in Dr. Dupureur's lab are maintained, 20 µL of HPV18-1 CHC1007-1 DNA were entered into the plasmid library and 1 mL of LB agar stabbed with *E. coli* JM109 cells.

2.3.3.4 Restriction Digests of HPV18/pBR322 Clones

The purified HPV18/pBR322 DNA from transformed *E. coli* JM109 cells was subjected to restriction digests using *Bam*HI (NEB Catalog # R0136S, 20,000 U/mL) and *Eco*RI (NEB Catalog # R0101S, 20,000 U/mL) as recommended by the manufacturer. Briefly, a 30 µL aliquot of purified HPV18/pBR322 DNA was diluted with 48 µL of autoclaved MilliQ H_2O to make a working standard (WS). The *Bam*HI restriction digest was performed by incubating 26 µL aliquot of the HPV18/pBR322 WS with 3 µL 10X NE Buffer 3, 1 µL *Bam*HI and 0.3 µL 100X bovine serum albumin (BSA; NEB) at 37 °C for 1 hour. The *Eco*RI restriction digest was performed by incubating 26 µL aliquot of the HPV18/pBR322 WS with 3 µL 10X NE Buffer 3, 1 µL EcoRI and 0.3 µL 100X BSA at 37 °C for 1 hour. The restriction digest fragments were electrophoresed on a 1 % agarose gel. To prepare the agarose gel, 1 gram of agarose powder (Sigma-Aldrich, Catalog # A9539-100G) was added to 100 mL of autoclaved MilliQ H_2O and microwaved for 1 min. The solution was mixed well, allowed to cool to ~70 °C and poured onto a Thermo Scientific Owl Easycast B1-BP casting plate with 10-well comb. The reaction products produced by restriction enzyme treatments were separately mixed with Gel Loading Dye, Blue (6X, NEB, Catalog # B7021S) and loaded onto the agarose gel. After sample loading, gel electrophoresis was performed in 1X TAE buffer (40 mM Tris-acetate, 1 M EDTA, pH 8.0) at 110 V, room temperature for 1 h. After electrophoresis, the gel was stained with 0.5 µg/mL ethidium bromide (Spectrum E1031, 1 % solution) for 20 min and washed with autoclaved MilliQ H_2O for 5 min in a staining box. The fragmented DNA and 1 kb DNA ladder (NEB, Catalog # N3232S, 500 µg/mL) were imaged using a FOTO/Convertible UV-light box equipped with an ethidium bromide filter (FOTODYNE Incorporated). Molecular weights of linearized DNA

fragments were estimated using NEB 1 kb DNA ladder (NEB N3232S, 500 µg/mL). The agarose gel was imaged and analyzed with FOTO/Analyst PC Image version 5.00.

2.3.4 Polymerase Chain Reaction (PCR): Amplification of HPV18 LCR Fragments Corresponding to Positions 7479-7783 (305 bp) and 7647-157 (368 bp)

2.3.4.1 PCR Reagent Amounts and Parameters

Two DNA fragments corresponding to 7479-7783 (305 bp) and 7647-157 (368 bp) from the HPV18 LCR were PCR-amplified for subsequent quantitative DNase I footprinting and affinity cleavage experiments with anti-HPV hairpin polyamides **PA1** and **PA25**. The LCR region is important for both viral transcription and replication,[10] and as a result it was chosen to study polyamide binding. The 368 bp fragment harbors HPV18's origin of replication, the E2 binding sites (E2BS-1 7822-7833; E2BS-2 42-53; E2BS-3 58-69) and the E1 binding site (E1BS 3-18) (**Figure 2.6**). The viral E1 protein is a helicase that unwinds DNA using its ATPase activity. While the viral E2 protein can repress or activate transcription depending on the viral infection stage, its interaction with E1 leads to the initiation of DNA replication by the host cell machinery.[69]

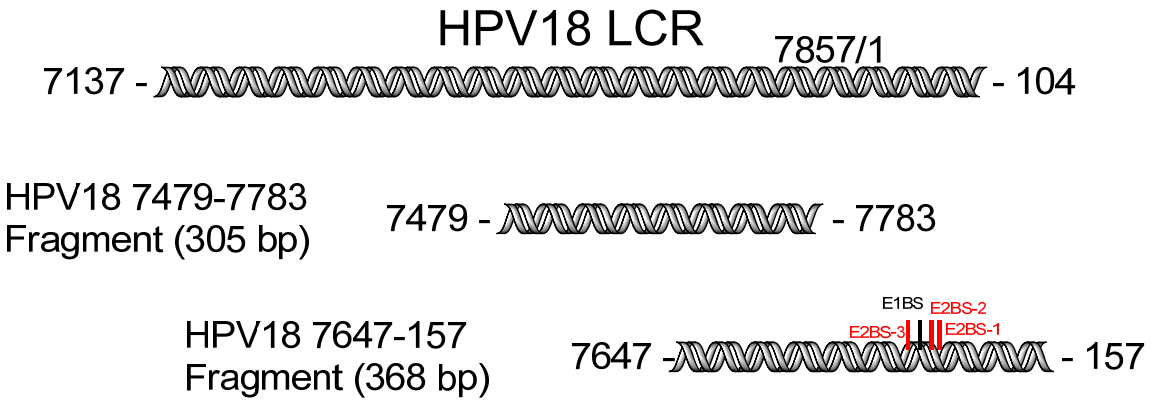


Figure 2.6. Map of HPV18 DNA Fragments Used in DNA-Binding Studies. Top duplex represents entire HPV18 LCR (Genbank Accession number X05015). Two fragments corresponding to 7479-7783 (305 bp) and 7647-157 (368 bp) from HPV18 LCR were used for quantitative DNase I footprinting and affinity cleavage experiments. The 368 bp fragment harbors HPV18's origin of replication, three out of four E2 binding sites (red) and the E1 binding site (black).

Two different fluorescent dyes were attached to the 5'-end of each primer to visualize the fragmentation patterns obtained from quantitative DNase I footprinting and affinity cleavage experiments using capillary electrophoresis (CE). The top strand was labeled with 6-carboxyfluorescein (6-FAM) and the bottom strand with hexachloro-fluorescein (HEX). These oligonucleotides were purchased from Integrated DNA Technologies. The primer sequences with their respective fluorescein dyes are provided below:

<u>Primers used to amplify the LCR region covering HPV18 7479-7783:</u>

Forward Primer (SLE1004F FAM): 5′-/56-FAM/CT TAT GTC TGT GGT TTT CTG
20 base pairs, 40 % GC content, T_m 55.9 °C

Reverse Primer (SLE1004R HEX): 5′-/5HEX/TT CAT GTT AAG GGT AGA CAG
20 base pairs, 40 % GC content, T_m 56.3 °C

Primers used to amplify the LCR/E6 region covering HPV18 7647-157:

Forward Primer (CHC1017FD): 5′-/56-FAM/TG CAT AAC TAT ATC CAC TCC
20 base pairs, 40 % GC content, T_m 56 °C

Reverse Primer (CHC1017RD): 5′-/5HEX/CA CAG ATC AGG TAG CTT GTA
20 base pairs, 45 % GC content, T_m 58.3 °C

(T_m values were determined using OligoAnalyzer 3.1 http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/ with Na^+ concentration of 50 mM and Mg^{++} concentration of 1.5 mM)

PCR reactions were performed on a Mastercycler Nexus Thermal Gradient (Eppendorf) using Carlos/mid frag.cyc for 7479-7783 or Carlos/HPV18 CHC1020.cyc for 7647-157. The cocktail and parameters used for the PCR reactions are given in **Table 2.2**.

Table 2.2. PCR Reagent Amounts and Run Parameters. The amounts given are for a 50 µL reaction of PCR.

Program	Reagent	Volume (µL)	PCR Step	Time (min)	
Carlos/mid frag.cyc (7479-7783 fragment)	MilliQ H_2O	41.45	Polymerase Activation at 95 °C	3	
	10X Standard *Taq* Buffer	5.0	Denaturation at 95 °C	0.5	30 X
	10 mM dNTP Mix	1.0	Annealing at 48 °C	0.5	
	10 µM Forward Primer	1.0	Extension at 68 °C	0.5	
	10 µM Reverse Primer	1.0	Extension at 68 °C	10	
	Template DNA (200X HPV18-1)	0.3	Hold at 6 °C	∞	
	Taq DNA Polymerase	0.25			
Carlos/HPV18 CHC1020.cyc (7647-157 fragment)	MilliQ H_2O	41.45	Polymerase Activation at 95 °C	3	
	10X Standard *Taq* Buffer	5.0	Denaturation at 95 °C	0.5	30 X
	10 mM dNTP Mix	1.0	Annealing at 55 °C	1	
	10 µM Forward Primer	1.0	Extension at 68 °C	0.5	
	10 µM Reverse Primer	1.0	Extension at 68 °C	10	
	Template DNA (200X HPV18-1)	0.3	Hold at 6 °C	∞	
	Taq DNA Polymerase	0.25			

Deoxynucleotide (dNTP) Solution Mix (Catalog # N0447S, 10 mM), 10X Standard *Taq* Reaction Buffer (Catalog # B9014S) and *Taq* DNA Polymerase (Catalog # M0273L, 5,000 U/mL) were purchased from New England Biolabs.

Upon completion of PCR, aliquots from the reactions were verified by gel electrophoresis using a 1 % agarose gel in 1X TAE buffer at 100 V, room temperature for about 1 hour or until the gel loading dye had migrated approximately ¾ of the gel length. The gel was stained with 0.5 µg/mL ethidium bromide for 20 min and washed with MilliQ H_2O for 5 min in a staining box. The PCR-amplified DNA and 100 bp DNA ladder (NEB, Catalog # N3231S, 500 µg/mL) were visualized with a FOTO/Convertible UV-light box equipped with an ethidium bromide filter (FOTODYNE Incorporated). The agarose gel was imaged and analyzed with FOTO/Analyst PC Image version 5.00.

2.3.4.2 Purification of PCR-Amplified LCR Fragments

Purification of the PCR-amplified fragments was initiated by precipitating the DNA. Briefly, 1/10 volume of 3 M sodium acetate, pH 5.2 was added to the PCR reaction and mixed well by inverting the vial ~ 6 times. To this solution, 3X volume of cold 100 % ethanol was added and mixed again by inversion. The DNA sample was then placed in a -20 °C freezer overnight. The next morning, the microcentrifuge vial was centrifuged at 17,900*x g* and 4 °C for 30 min. The supernatant was carefully removed using a pipettor, being careful not to disturb the DNA pellet. The pellet was washed with 400 µL of cold (4 °C) 70 % ethanol without mixing. This solution was centrifuged at 17,900*x g* and 4 °C for 15 min. The supernatant was again removed and the pellet was allowed to air dry until no solvent was observed (~20 min). The DNA pellet was dissolved in 30 µL of Buffer EB (10 mM Tris·Cl, pH 8.5; QIAGEN) and subjected to 1 % agarose gel electrophoresis as previously discussed. The PCR-amplified DNA fragments were then excised from the agarose gel using a clean razor blade and each lane was purified using a QIAGEN Gel Extraction Kit (Catalog # 28706). Briefly, 3X volume of Buffer QG were added to 1X volume of the gel (volume determined by weight of gel, where 100 µL is equivalent to approximately 100 mg of gel). The suspension was then heated to 50 °C for 10 min, vortexing every 2 min to ensure complete dissolution of agarose gel. After ensuring that the color of the solution remained yellow upon dissolution, 1X volume of isopropanol (Fisher BioReagents, Catalog # BP2618-1) was added and mixed by vortexing. This solution was carefully transferred into a QIAquick spin column. The column was then centrifuged for 60 s at 17,900*x g* and the flow-through was discarded. A 500 µL aliquot of Buffer QG was then added to the column, centrifuged for 60 s at 17,900*x g* and the flow-through was discarded. The column was then washed with 750 µL of Buffer PE, allowed to stand for 2 min and centrifuged for 60 s at 17,900*x g*. The flow-through was discarded and the wash step was repeated once. The column was centrifuged for an extra 60 s at 17,900*x g* to remove any residual Buffer PE. The QIAquick column was then placed in a clean 1.5 mL microcentrifuge tube. The DNA was eluted from the column by adding 50 µL of Buffer EB to the center of the column membrane and the column was allowed to incubate for 5 min at room temperature. The column was then centrifuged for 60 s at 17,900*x g*. The purified DNA fragments were aliquoted and stored at -20 °C.

2.3.4.3 Sanger DNA Sequencing

To ensure that the PCR-amplified fragments corresponded to the HPV 18 LCR regions from 7479 to 7783 bp and from 7647 to 157 bp, Sanger DNA sequencing was performed. In this sequencing method, four different reactions are prepared with the four 2'-deoxynucleotide triphosphates plus one of the four 2', 3'-dideoxynucleotide triphosphates in each reaction at the proper ratio. Since DNA polymerization requires a hydroxyl group on the 3' carbon of the sugar ring, the DNA polymerase will irreversibly terminate upon insertion of the particular 2', 3'-dideoxynucleotide yielding a population of fragments terminating with each nucleotide that can then be separated and sequenced using electrophoresis.[70] With the advancements in capillary electrophoresis and dye chemistry, the sequencing of a DNA fragment can be deduced from a single reaction with 2', 3'-dideoxynucleotide triphosphate terminators with different fluorescent dyes and detected via automated capillary electrophoresis.[71]

The two HPV18 fragments (305 bp, 7479-7783 and 368 bp, 7647-157) without 5' dyes were submitted to the DNA Core Facility at University of Missouri for Sanger DNA sequencing to ensure that the correct fragments were PCR-amplified. Submission guidelines can be found at http://biotech.missouri.edu/dnacore/sangersequencing.html. The procedure is delineated below:

1. Samples are submitted in a 1.5 mL microcentrifuge tube.
2. The total volume of the template and primer mix in the 1.5 mL microcentrifuge tube must be 16 µL. MilliQ H_2O is used as the diluent.
3. For PCR products, the total amount of DNA is 50-275 ng.
4. An aliquot of 15-26 pmol of a single primer (without 5' dye) is added to each reaction.

The sequencing data were processed using FinchTV version 1.4.0 (Geospiza Inc.).

2.3.5 Biophysical Analysis of DNA-PA Interactions in the HPV18 LCR

Quantitative DNase I footprinting experiments with **PA1** and **PA25**, as well as affinity cleavage experiments with the EDTA conjugates of **PA1** and **PA25**, were performed on the 305 bp and 368 bp fragments corresponding to 7479-7783 bp and 7647-157 bp of the HPV18 LCR, respectively. Capillary electrophoresis analysis was conducted by MU DNA Core using an ABI 3730XI DNA Analyzer with an internal size standard consisting of Genescan 600 LIZ.

2.3.5.1 CE Size Indexing for the 305 bp (7479-7783) and 368 bp (7647-157) Using Sanger Sequencing Chemistry

In order to accurately map the **PA1** and **PA25** binding sites on the HPV18 fragments (7479-7783 and 7647-157), the correct nucleotide position on the HPV18 genome must be accurately correlated to the relative CE mobility of the fragment terminating with the nucleotide in question. Both the sequencing chemistry and the nature of 5' dye influence the CE mobility of the DNA, and as a result the appropriate indexing method must be chosen. Particularly for quantitative DNase I footprinting, the cleavage products have a neutral hydroxyl group on the 3' end and display a similar CE mobility to the neutral fragments generated by Sanger chemistry (**Figure 2.7A** and **C**).[72]

Indexing was performed for both the top and bottom strands of the duplex since the former bears a 5' FAM dye and the latter has a 5' HEX dye. Sanger sequencing was conducted using a Thermo Sequence Dye Primer Manual Cycle Sequencing Kit USB (Affymetrix, Catalog # 79260 1 KT) as per the manufacturer's instructions. The sequencing reactions were performed at two different DNA template concentrations: 20 ng/µL and 2 ng/µL. Briefly, 4 sets of four 0.2 mL PCR tubes were labeled G, A, T or C and placed on ice. Master reaction mix stocks for both DNA concentrations and both the fluorescently labeled forward and reverse primers were prepared (**Table 2.3**). USB sequencing was performed at both 20 ng/µL and 2 ng/µL in order to optimize the output signal.

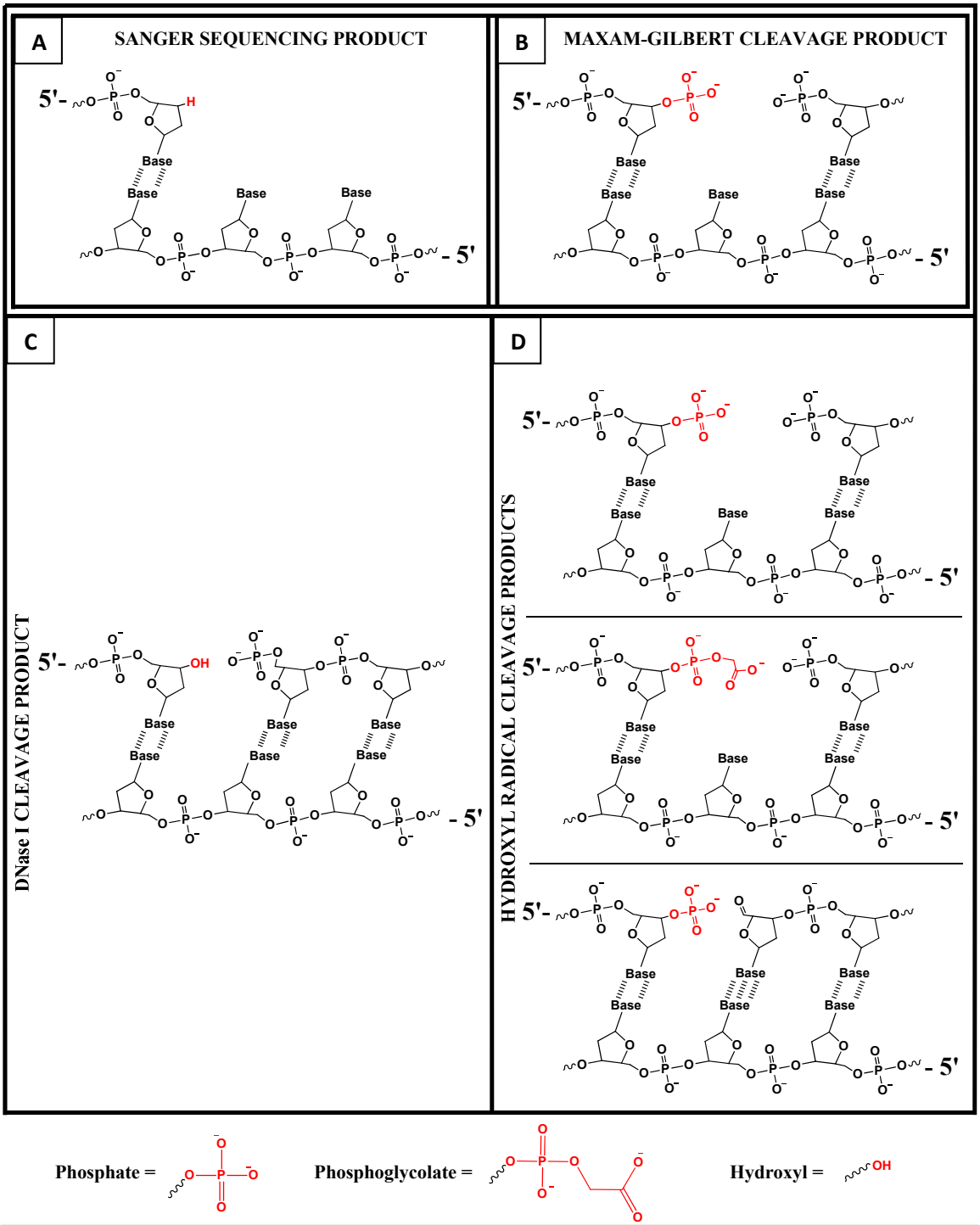


Figure 2.7. Chemical Structures of the Cleavage Products Generated by Various Sequencing and Footprinting Chemistries. DNA fragments are 5' end-labelled with fluorescent dyes and cleavage is only shown for the top strand of the DNA duplex. **(A & C)** DNase I cleavage products generate neutral hydroxyl group on the 3' end and display a similar CE mobility to the neutral fragments generated by Sanger chemistry. **(D)** In radical chemistry, abstraction of a deoxyribose hydrogen atom by a hydroxyl radical produces single-stranded nicks and gaps along the DNA duplex. Depending on the site of hydroxyl radical attack, the resulting reaction products consist of a single-nucleoside gap with a 3' phosphate (*Top*) or a single-nucleoside gap with a 3' phosphoglycolate (*Middle*), or a single-stranded nick with a 3' phosphate (*Bottom*). **(B & D)** Because Maxam-Gilbert and hydroxyl radical chemistries lead to the generation of negatively charged 3' phosphates, their cleavage products display similar CE mobility. P = phosphate, PG = phosphoglycolate, OH = hydroxyl.

Table 2.3. USB Sequencing Reagents and Amounts for Master Reaction Mix Stocks.

Reagent	Volumes for 20 ng/µL DNA Template Master RXN Mix (µL)		Volumes for 2 ng/µL DNA Template Master RXN Mix (µL)	
	5'-FAM Primer	5'-HEX Primer	5'-FAM Primer	5'-HEX Primer
HPV18 7479-7783 (305 bp) no dye OR HPV18 7647-157 (368 bp) no dye	1.0	1.0	1.0	1.0
Reaction Buffer (conc.)	2.2	2.2	2.2	2.2
2 µM CHC1017 FD	1.0	NA	1.0	NA
2 µM CHC1017 RD	NA	1.0	NA	1.0
MilliQ H_2O	12.8	12.8	12.8	12.8
Thermo Sequenase (20 U/µL)	1.0	1.0	1.0	1.0
TOTAL	**18.0**	**18.0**	**18.0**	**18.0**

A total of four master reaction mix stocks were prepared. Each DNA template concentration was prepared twice containing either the 5'-FAM primer (top strand) or the 5'-HEX primer (bottom strand).

Next, 4 µL aliquots from each of the master reaction mix stocks were transferred into the 0.2 mL PCR tubes labeled G, A, T or C, previously cooled on ice. To each of these tubes, 1 µL of the respective termination mix was added (*i.e.*, ddGTP termination mix was added to the tubes labeled G). Immediately, the reactions were placed in a Mastercycler Nexus Thermal Gradient and the PCR program Carlos/DNA Index CE.cyc was started. **Table 2.4** provides the PCR parameters. The samples were submitted to MU DNA Core for fragmentation analysis and the results were analyzed using GeneMarker Version 2.4.0 (SoftGenetics LLC.).

Table 2.4. USB Sequencing PCR Thermal Profile.

PCR Step	Time (min)	
Polymerase Activation at 95 °C	3	
Denaturation at 95 °C	0.5	**30 X**
Annealing at 55 °C	0.5	
Extension at 72 °C	1	
Extension at 68 °C	10	
Hold at 6 °C	∞	

2.3.5.2 CE Size Indexing for the 305 bp (7479-7783) and 368 bp (7647-157) Using Maxam-Gilbert Sequencing Chemistry

As previously discussed, it is imperative to accurately index the CE fragment size to the actual nucleotide on the sequence of interest in order to correctly map the binding site of DNA-binding ligands. To accurately determine the cleavage sites in affinity cleavage experiments, the DNA fragment must be indexed by Maxam-Gilbert sequencing chemistry. In this technique, the DNA is cleaved by specific base modifications, followed by eviction of the modified bases and strand cleavage at these positions.[73] The cleavage products from the Maxam-Gilbert sequencing reactions have a 3'-phosphate group. Similarly, the cleavage products for affinity cleavage experiments and hydroxyl radical footprinting have a negatively charged phosphate group or a negatively charged phosphoglycolate group on their 3' end, and display a similar CE mobility to the fragments generated by Maxam-Gilbert chemistry (**Figure 2.7B** and **D**).[72]

Indexing was performed for both the top and bottom strands of the duplex. Indexing of the HVP18 fragments (7647-157 and 7479-7783) were determined using the Maxam-

Gilbert protocol found in Molecular Cloning 2nd Edition.[74] Briefly, two separate reactions were prepared for each HPV 18 fragment and subjected to either piperidine formate (pH 2.0) or hydrazine. The former will react with purines and the latter will react with pyrimidines. After incubation with these chemicals, phosphate bond cleavage at the modified base is accomplished by the addition of hot 1 M piperidine solution. These two reactions were also performed at different DNA concentrations and incubation times (see **Section 2.71**). The samples were submitted to MU DNA Core for fragmentation analysis and the results were processed with GeneMarker Version 2.4.0.

2.3.5.3 Quantitative DNase I Footprinting

Quantitative DNase I footprinting experiments were performed to determine the **PA1** and **PA25** binding sites in the HPV18 LCR, as well as the respective equilibrium dissociation constants associated with the observed binding events.

PA1 (KG1068) and **PA25** (KTK6006) polyamides were first lyophilized into single-use aliquots by accurately weighing 1 mg of powder, diluting in 1 mL of 50 % H_2O (0.1 % TFA) : 50 % ACN (0.1 % TFA) and transferring a 100 µL aliquot of this solution into 9 different 1.5 mL Eppendorf microcentrifuge tubes. The 10 vials were lyophilized overnight and stored at -20 °C. The concentration was then determined prior to use. Briefly, a lyophilized aliquot was dissolved in 100 µL of 100 % DMSO and the absorbance was measured at the λ_{max} (~305 nm) using a Thermo Scientific Evolution 260 Bio UV-visible spectrophotometer, and molar extinction coefficients used were 88,235 $M^{-1}cm^{-1}$ and 147,400 $M^{-1}cm^{-1}$ for **PA1** and **PA25**, respectively.

2.3.5.3.1 Control DNase I Fragmentation Reaction

To determine the appropriate DNase I concentration for subsequent footprinting experiments, a DNase I endonuclease (Promega RQ1 RNase-free DNase I, Catalog # M610A) titration from 0-0.08 Activity Units (AU) was carried out. Quantitative DNase I footprinting reactions must be conducted under conditions of single-hit kinetics, that is, one cut per DNA molecule on average.[51] The control fragmentation reactions were performed at 200 pM of DNA in TKMC buffer (10 mM Tris, 10 mM KCl, 5 mM $MgCl_2$ and 5 mM $CaCl_2$), 2 % v/v DMSO and 10 mM 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) for a total reaction volume of 250 µL. Eight aliquots from this solution were transferred into separate 1.5 mL microcentrifuge tubes and the respective DNase I in QIAGEN Buffer EB (10 mM Tris-HCl, pH 8.5) amount was added into each. The reactions were incubated for 5 min at 37 °C and followed by quenching the reaction by the addition of 10 µL 200 mM ethylenediaminetetraacetic acid (EDTA) and denaturation of the endonuclease at 100 °C for 2 min. The fragments were purified with a QIAquick PCR Purification Kit following the manufacturer's instructions with a slight modification. Because salt anions compete with DNA during electrokinetic injection into the capillary,[75] the filter-bound DNA samples were always washed twice with Buffer PE, allowing the buffer to stand for one minute before discarding. Furthermore, DNA elution was always performed by incubating the Buffer EB in the column membrane for 5 min. The samples were submitted to MU DNA Core for fragment analysis and the data were processed with GeneMarker Version 2.4.0.

2.3.5.3.2 PA1 Quantitative DNase I Footprinting

Quantitative DNase I footprinting experiments were performed on the 368 bp (HPV18 7647-157) fragment to determine the **PA1** binding sites and equilibrium binding constants in this region of the viral genome. To determine the equilibrium dissociation constant (K_d) values for **PA1**, quantitative DNase I footprinting reactions were performed at 200 pM DNA at 250 µL per reaction. Briefly, HPV18 DNA fragment stock was prepared in TKMC buffer and 10 mM CHAPS to yield a final DNA concentration of 200 pM. Eight aliquots from this solution were transferred into separate 1.5 mL Eppendorf microcentrifuge tubes. A titration was then performed by increasing the concentration from 0 to 40 nM **PA1** per tube, keeping the total DMSO concentration per reaction at 2 % v/v. Polyamide binding to the DNA fragment was carried out at 37 °C for 4 h to overnight.

After the polyamide-DNA incubation period, DNase I digestion was performed at 37 °C for 5 min with 0.01 AU of DNase I (as determined by a control DNase I fragmentation reaction). Each reaction was quenched by adding 10 µL of 200 mM EDTA, followed by denaturation of the enzyme at 100 °C for 2 min. The fragments were purified with a QIAquick PCR Purification Kit (Catalog # 28106) and submitted to MU DNA Core for fragment analysis. The data were processed with GeneMarker Version 2.4.0. Quantitative DNase I footprinting was repeated at least three times. Equilibrium dissociation constants were determined separately per binding site and reported as the average of the separate trials.

2.3.5.3.3 PA25 Quantitative DNase I Footprinting

Quantitative DNase I footprinting experiments were performed on both the 305 bp (HPV18 7479-7783) and 368 bp (HPV18 7647-157) DNA fragments to determine the **PA25** binding sites and equilibrium dissociation constants in the LCR region of the viral genome. To determine the equilibrium dissociation constants for **PA25**, quantitative DNase I footprinting reactions were performed at 200 pM of DNA (250 µL/reaction), as well as at 50 pM (1000 µL/reaction) due to the observed low K_d values at 200 pM DNA. Experiments at 10 pM DNA were also carried out; however, DNase I cutting displayed significant uneven cleavage throughout the DNA fragment (data not shown). **PA25** quantitative DNase I footprinting experiment were performed by preparing HPV18 DNA fragment stocks in TKMC buffer and 10 mM CHAPS to yield final DNA concentrations of 200 pM and 50 pM. Twelve aliquots from each solution were transferred into separate 1.5 mL microcentrifuge tubes. Titrations were then performed by increasing the concentration from 0 to 40 nM **PA25** per tube for samples containing 200 pM DNA, or 0 to 10 nM **PA25** for 50 pM DNA, keeping the total DMSO concentration per reaction at 2 % v/v. Polyamide binding to the DNA fragment was carried out at 37 °C for 2.5 h to overnight.

After the polyamide-DNA incubation period, DNase I digestion was performed at 37 °C for 5 min with 0.06 AU of DNase I for the 305 bp fragment at 200 pM DNA and 0.01 AU of DNase I for the 368 bp fragment (as determined by a control DNase I fragmentation reaction). For experiments conducted at 50 pM, 0.2 AU DNase I for the 305 bp fragment and 0.1 AU of DNase I for the 368 bp fragment were used. Each

reaction was quenched by adding 50 µL of 200 mM EDTA, followed by denaturation of the enzyme at 100 °C for 2 min. Because the experiments at the lower 50 pM DNA concentration were carried out at 1000 µL per reaction, each sample was lyophilized overnight to reduce the sample volume prior to column purification. The fragments were purified with a QIAquick PCR Purification Kit and submitted to MU DNA Core for fragment analysis. The data were processed with GeneMarker Version 2.4.0. The analysis was repeated at least three times. Equilibrium dissociation constants were determined separately per binding site and reported as the average of the separate trials.

2.3.5.3.4 Determination of Equilibrium Dissociation Constants (K_d Values)

Increasing the ligand concentration and monitoring the decrease of cleavage product (as the integrated area under the curve) at particular nucleotide positions along the 305 bp and 368 bp HPV18 DNA fragments reveals the polyamide binding sites. Once these binding sites were established, a reference peak on the electropherogram was chosen in order to normalize the cleavage data. This reference peak must not fall within a polyamide binding site or a DNase I hypersensitive site. For the 305 bp DNA fragment corresponding to 7479-7783 bp of the HPV18 LCR, the reference peak at nucleotide 7629 and nucleotide 7625 were chosen as the reference peaks for the 5'-FAM- and 5'-HEX-labelled strands, respectively. For the 368 bp DNA fragment corresponding to 7647-157 bp of the HPV18 LCR, the reference peak at nucleotide 7840 and nucleotide 7837 were chosen as the reference peaks for the 5'-FAM- and 5'-HEX-labelled strands, respectively.

In order to calculate the K_d values of the observed PA-binding events, the integrated area of at least one peak within a polyamide binding site—preferably one with a high signal-to-noise ratio found in the middle of the protected region—and the area of the respective reference peak were tabulated at each polyamide concentration. The ratio of the sample peak area (polyamide binding site peak) to the reference peak area was calculated. Since each sample was submitted as technical triplicates, the average of these ratios was used to calculate the normalized bound fraction of polyamide to DNA. The data were fit to both Langmuir and Hill equations (**Equations 2.1** and **2.2**) using Kaleidagraph 4.1.1 (Synergy Software).

Langmuir equation

$$\Theta = \frac{K_a[PA]}{1 + K_a[PA]} \tag{2.1}$$

Hill equation

$$\Theta = \frac{K_a[PA]^n}{1 + (K_a[PA])^n} \tag{2.2}$$

(θ = fraction bound, K_a = equilibrium association constant, n = Hill coefficient, [PA] = polyamide concentration)

2.3.5.4 Affinity Cleavage Experiments

In order to determine the correct binding orientation of the polyamides within the protected regions on the HPV18 DNA fragments, affinity cleavage experiments were conducted using the EDTA conjugates of **PA1** and **PA25**. This method relies on the sequence-specific cleavage of DNA by a hairpin polyamide covalently attached with EDTA onto its C-terminus (*i.e.*, Ta tail) **(Figure 2.8)**. In the presence of ferrous iron, oxygen and the reducing agent dithiothreitol (DTT), the catalytic formation of diffusible hydroxyl radicals occurs at the C-terminal EDTA and leads to oxidative cleavage of the DNA near this position. This cleavage pattern yields multiple neighboring cleavage sites with a Gaussian-like distribution due to the ability of the hydroxyl radicals to diffuse along the backbone of the DNA. Furthermore, the pattern is also asymmetrical compared to the opposite strand, exhibiting a 3' shift from the location of the EDTA-Fe^{2+} when hydroxyl radicals are generated in the minor groove. The 3' shift is a result of a 2 bp offset of the deoxyribose to its complement on the 3' side of the minor groove of B-form DNA **(Figure 2.9)**.[64]

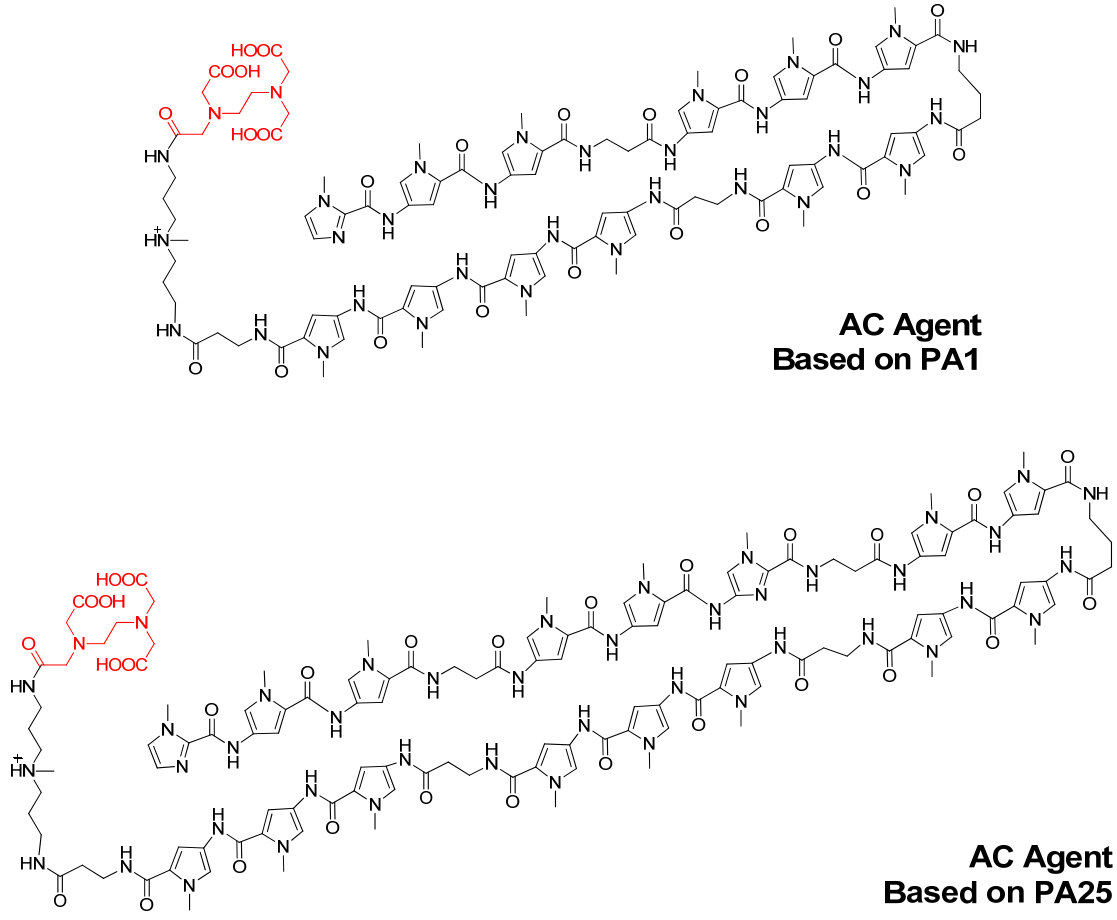


Figure 2.8. Chemical Structures of the PA1- and PA25-EDTA Affinity Cleavage (AC) Agents. Sequence-specific DNA-binding polyamides **PA1** and **PA25** covalently modified at the C-terminus with EDTA moiety (red).

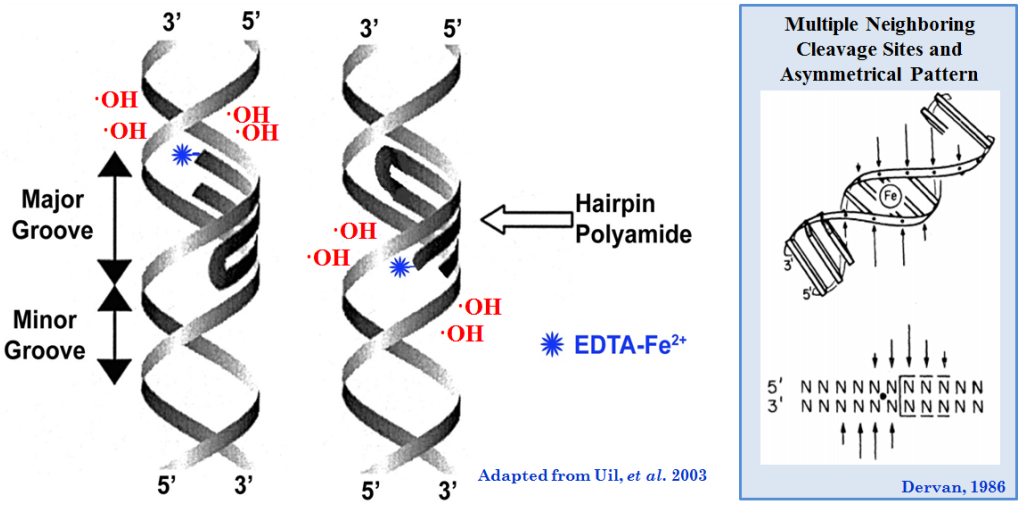


Figure 2.9. Model for the Asymmetric Pattern Observed in Affinity Cleavage. Ferrous iron (Fe^{2+}) is chelated by the covalently attached EDTA group on the C-terminal Ta tail of the polyamide. The cleavage pattern of the deoxyribose complement of the nucleotide closest to the ferrous iron on the top strand (5' – 3' from left to right) is shifted by 2 bp in the 3' direction. Figures adapted from Uil, *et al.* 2003 (*permission # 4087221444733*) and Dervan, 1986 (*permission # 4087230173714*).

Affinity cleavage experiments were performed using **PA1-EDTA** and **PA25-EDTA** to determine the PA binding orientations on the observed footprinting regions for DNase I footprinting experiments. The affinity cleavage reactions were performed using 1 nM of DNA (40 µL/reaction) in Tris-HCl buffer (10 mM Tris-HCl, pH 7.5) and 10 mM CHAPS. The cleavage agent, PA-EDTA conjugate, was complexed with 0.8 equivalent of Fe(II) in the form of ammounium iron(II) sulfate hexahydrate (Aldrich, Catalog # 203505-25G). A titration was then performed by increasing the concentration from 0 to 200 nM PA-EDTA per 1.5 mL microcentrifuge tube, keeping the total DMSO concentration per reaction at 2 % v/v. Polyamide binding to the DNA fragment was carried out at 37 °C for 4 h to overnight. After the polyamide-DNA incubation, the radical reaction was initiated by the addition of 5 µL 100 mM DTT to each reaction. DNA cleavage was performed at room temperature for 30 min. Each reaction was quenched by performing DNA purification with a QIAquick PCR Purification Kit. The samples were submitted to MU DNA Core for fragmentation analysis and the results were processed with GeneMarker Version 2.4.0.

2.3.6 X-ray Crystallography Screening Matrix of Oligonucleotide with PA25

DNA oligomers were purchased from Midland Certified Reagent Company (top strand: 5'-GGGATATTTAGTTTGCGGCG-3' and bottom strand: 5'-CCGCAAACTAAAATATCCCCG-3') and annealed by heating equimolar concentrations of these oligos at 95 °C for 5 min, followed by slow cooling to room temperature. DNA duplex used for X-ray crystallization screening is given in **Figure 2.10**.

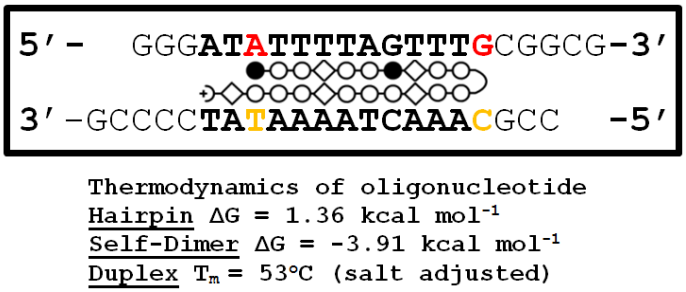


Figure 2.10. DNA Fragment Used for X-ray Crystallography. PA25 is shown in between complementary DNA strands of an oligomer used in X-ray crystallography screening experiments. The match binding site for **PA25** is 5'-$W_2G_7W_4$-3'. The DNA mismatched nucleotides for **PA25** are colored in red fonts; orange font illustrates nucleotides that by themselves are not mismatched. The polyamide building blocks are represented by open circles for *N*-methylpyrroles, filled circle for *N*-methylimidazole, diamonds for β-alanines, + for the Ta tail and a loop for γ-aminobutyric acid. Thermodynamic values obtained for hairpin, self-dimer and DNA melting temperature from http://mfold.rna.albany.edu/?q=mfold/DNA-Folding-Form, http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/ and http://www.basic.northwestern.edu/biotools/OligoCalc.html, respectively (DNA = 0.5 mM, NaCl = 50 mM, Mg^{++} = 1 mM).

The Natrix HT crystallization matrix (Hampton Research, Catalog # HR2-131) and a matrix generated from reagent conditions published in polyamide literature were used to screen crystallization conditions. The Natrix HT kit comes in a user-friendly 96-well format that provides a systematic variation in pH, buffer composition, salt and precipitant. On the other hand, the generated matrix was based on the crystallization of an 8-ring cyclic polyamide to its cognate DNA binding site.[45] The crystallization conditions reported were 21 % of 2-methyl-2,4-pentanediol (MPD), 35 mM calcium acetate and 10 mM Tris pH 7.5.[45] The generated matrix optimizes the MPD concentration from 50 to 10 %, the calcium acetate concentration from 100 to 30 mM and two pH values (7.5 and 8.0) (see **Table SI2.1**).

To perform the screening matrixes, four solution stocks were prepared. These stocks contained either a 1:1.3 DNA:**PA25**(formate) (KJK6076 ~98 % purity), 1:2 DNA:**PA25**(formate), 1:1.3 DNA:**PA25**(TFA) (KTK6006 ~95 % purity) or 1:2 DNA:**PA25**(TFA). The formate salt of **PA25** was included in these experiments as it exhibits a higher solubility than its TFA analog. **PA25** was added slowly in 1 µL intervals every 10 min at 40 °C to minimize polyamide aggregation. Once the correct amount of **PA25** was added into the DNA sample in a microcentrifuge tube, the solution was incubated overnight at 37 °C. The final DNA and **PA25** concentrations were 0.39 mM and 0.51 mM for 1:1.3 DNA:**PA25** solutions; and 0.39 mM and 0.78 mM for 1:2 DNA:**PA25** solutions, respectively. The next day, 40 µL of each crystallization reagent from either the Natrix HT matrix or the generated matrix were transferred into the appropriate reservoir of the crystallization plate. Using a new pipet tip, 1 µL of the crystallization reagent from the reservoir and 1 µL of either 1:1.3 or 1:2 DNA:PA solutions were added into each sitting drop well. The plates were sealed and carefully placed in a Thermo Scientific HERAtherm at 19 °C. After seven weeks, the plates were examined for crystal growth under a stereo microscope.

2.3.7 Circular Dichroism (CD) Spectroscopy

Circular dichroism is a spectroscopic technique that measures the absorption difference between left-handed and right-handed circularly polarized light by chiral

molecules.[76] Because this technique allows for the determination of interactions and conformations of biological macromolecules with their respective ligands, CD was used to measure the interactions of **PA1** and **PA2** to different DNA duplexes.

DNA oligomers were purchased from IDT (CHC2050-1 top strand: 5'-CCCAACCTATTTCGGT-3' and bottom strand: 5'-ACCGAAATAGGTTGGG-3'; CHC2050-2 top strand: 5'-GCGAAGTTAATAGGCG-3' and bottom strand: 5'-CGCCTATTAACTTCGC-3'; CHC2050-3 top strand: 5'-GCGAACTTAATAGGCG-3' and bottom strand: 5'-CGCCTATTAAGTTCGC-3'; CHC2050-GH6084C top strand: 5'-GCGAAGTTAATATGCG-3' and bottom strand: 5'-CGCATATTAACTTCGC; CHC2050-GH6084D top strand: 5'-GCGAACTTAATATGCG-3' and bottom strand: 5'-CGCATATTAAGTTCGC-3') and separately dissolved in MilliQ H_2O to make a solution of approximately 500 μM using the manufacturer's determined nanomoles. The actual concentrations were determined by UV-Vis using the provided molar extinction coefficients. The complementary oligos were then annealed by heating equimolar concentrations (100 μM) at 95 °C for 30 s and allowed to slowly cool to room temperature.

CHC2050-1 duplex corresponds to a **PA1** binding site (Site #14 in **Figure 2.22**) found in the HPV18 368 bp fragment. A titration with **PA1** was performed into a solution of the CHC2050-1 duplex. Site #14 can be explained by **PA1** binding to a forward triple-base-pair mismatch site or a reverse double-base-pair mismatch site **(Figure 2.11)**.

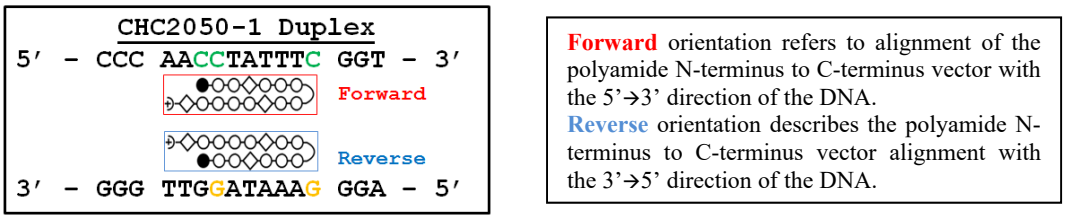


Figure 2.11. DNA Sequence for CHC2050-1 Duplex Used in CD Experiments to Evaluate PA1 Binding to Site #14. CHC2050-1 duplex presents a forward triple-base-pair-mismatch binding site and a reverse double-pair-mismatch binding site to **PA1**. Forward and reverse binding orientations are boxed in red and blue, respectively. Mismatches for the forward binding orientation are shown on the top strand in green, while mismatches for the reverse binding orientation are highlighted on the bottom strand in orange. **PA1** match binding site is 5'-W_2G_7-3', where the letter W (weak) stands for A or T. The polyamide building blocks are represented by open circles for *N*-methylpyrroles, filled circle for *N*-methylimidazole, diamonds for β-alanines, + for the Ta tail and a loop for γ-aminobutyric acid.

DNA sequences used for these experiments were flanked by GC nucleotides to decrease the possibility of multiple polyamide binding sites. Two titrations with **PA1** were conducted to determine whether binding of **PA1** to a forward single-base-pair mismatch site exhibited qualitatively different spectra than binding to a reverse single-base-pair mismatch site on a 16-mer DNA duplex (**Figure 2.12**). The sequences were designed with a single-base-pair mismatch at the nucleotides that interact with the γ-turn of **PA1**. The rationale was that **PA1** would prefer the least number of PA-DNA mismatches and would have a slight preference for the forward orientation. Consequently, the forward binding orientation of **PA1** on the CHC2050-2 duplex would lead to a single-base-pair mismatch at the γ-turn, where the reverse binding orientation (180° rotation of **PA1** along its horizontal axis in **Figure 2.12**) results in a double-base-pair-mismatch. On the other hand, binding of **PA1** in the reverse orientation to the CHC2050-3 duplex might be more favorable because this conformation results in a

single-base-pair mismatch at the γ-turn, in contrast to a double-base-pair mismatch in the forward binding orientation.

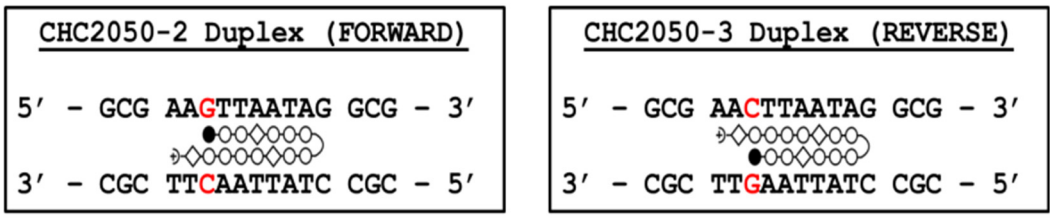


Figure 2.12. DNA Sequences for CHC2050-2 and -3 Duplexes Used in CD Experiments to Evaluate PA1 Binding Orientation. CHC2050-2 duplex presents a forward single-base-pair-mismatch binding site to **PA1**, whereas the CHC2050-3 duplex contains a reverse **PA1** single-base-pair-mismatch binding site. Mismatches are located on the nucleotides that make contact with the γ-turn from **PA1**. Red font highlights the nucleotides that are swapped in the sequences. **PA1** match binding site is 5'-W_2G_7-3', where the letter W (weak) stands for A or T. The polyamide building blocks are represented by open circles for *N*-methylpyrroles, filled circle for *N*-methylimidazole, diamonds for β-alanines, + for the Ta tail and a loop for γ-aminobutyric acid.

The same experiment was performed using **PA2** and two different match binding sequences. **PA2** shares the same sequence as **PA1**, except that it has a chiral (R)-NH_2 in the γ-turn. As a result, it recognizes the same DNA sequence (5'-W2G7-3') as **PA1**; however, the chiral (R)-substitution on the γ-turn is expected to increase the binding preference for the forward orientation.[77] The purpose of this experiment was to determine whether binding of **PA2** to a forward match site exhibited qualitatively different spectra than binding to an oligomer with both a forward single-base-pair mismatch site and a reverse match site (**Figure 2.13**).

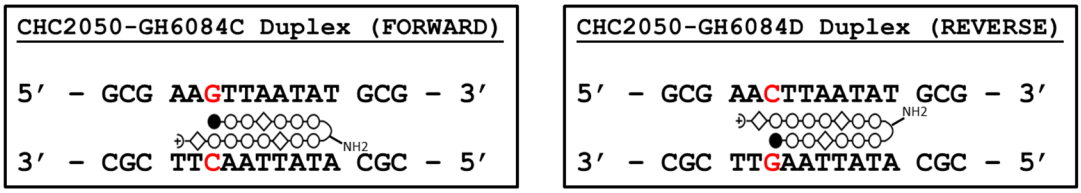


Figure 2.13. DNA Sequences for CHC2050-GH6084C and -D Duplexes Used in CD Experiments to Evaluate PA2 Binding Orientation. CHC2050-GH6084C Duplex presents a forward match binding site to **PA2**, whereas the CHC2050-GH6084D duplex contains a reverse match binding site and a forward single-base-pair-mismatch binding site. Red font highlights the nucleotides that are swapped in the sequences. **PA2** match binding site is 5'-W_2G_7-3', where the letter W (weak) stands for A or T. The polyamide building blocks are represented by open circles for *N*-methylpyrroles, filled circle for *N*-methylimidazole, diamonds for β-alanines, + for the Ta tail and a loop for γ-aminobutyric acid.

CD measurements were performed in a 10 mm Quartz cuvette using a Jasco J-1500 CD Spectrometer. Spectra were recorded from 450-200 nm. The baseline consisted of 192 μL of Buffer EB. To this solution, 8 μL of 100 μM DNA duplex was added to the cuvette and mixed using a pipettor. The CD spectrum for this sample was acquired. A 100 μM PA in DMSO was prepared. Titration of the polyamide was then performed by adding 1.6 μL aliquots of 100 μM PA into the solution of 4.0 μM DNA. After each aliquot addition, the solution was mixed using a pipettor and allowed to equilibrate for 5 min before data collection. The resulting ratios ranged from 0.2 to 3.0 PA to DNA. All experiments were conducted at 20 °C and two scans were collected per sample using the CD parameter file Para1 in Spectra Manager Version 2.10.01.

2.4 RESULTS

We relied on DNase I footprinting and affinity cleavage experiments to map the binding sites of **PA1** and **PA25** on DNA fragments corresponding to the HPV18 LCR genomic region under cell-free conditions. We report that anti-HPV hairpin polyamides **PA1** and **PA25** bind avidly to the minor groove of A/T-rich sequences in the HPV18 LCR with dissociation constants ranging from 0.4 to 2.6 nM.

2.4.1 Restriction Digests of HPV18/pBR322 Clones

In order to confirm that the correct plasmid DNA was transformed and amplified in JM109 *E. coli* cells, restriction digests using *Bam*HI and *Eco*RI restriction enzymes were performed on the purified DNA. The *Bam*HI restriction digest of HPV18/pBR322 is expected to generate three DNA fragments with sizes of 1.1 kb, 4.8 kb and 6.6 kb, respectively. On the other hand, the restriction digest of HPV18/pBR322 with *Eco*RI should produce two DNA fragments with sizes of 4.5 kb and 8.0 kb, respectively. The restriction enzyme profile of the fragmented plasmid DNA separated by agarose gel electrophoresis is presented in **Figure 2.14**. The observed restriction products are in agreement with the expected DNA bands for HPV18/pBR322 digestion with *Bam*HI and *Eco*RI restriction enzymes.

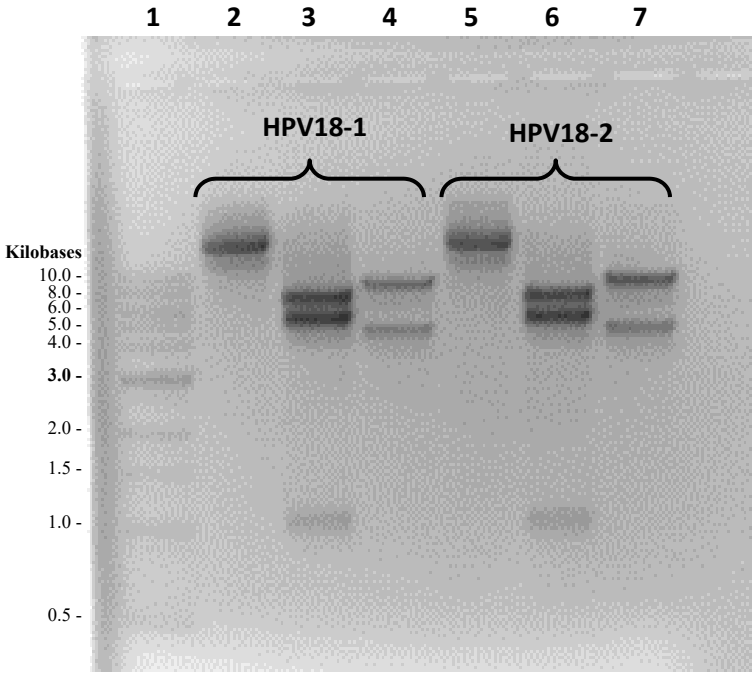


Figure 2.14. Agarose Gel Electrophoresis of Restriction Digest for HPV18/pBR322. *Bam*HI and *Eco*RI restriction digests performed on two separate HPV18/pBR322 clones purified from *E. coli* JM109 cells. Lane 1 corresponds to a 1 kb DNA molecular weight ladder with the expected size in kilobases given on the left of the picture. Lanes 2-4 and Lanes 5-7 illustrate the results for HPV18-1 and HPV18-2, respectively. Lanes 2 and 5 correspond to the uncut DNA (12.219 kb). Lanes 3 and 6 correspond to restriction digest products of HPV18-1 and HPV18-2 samples treated with *Bam*HI restriction enzyme, respectively. Lanes 4 and 7 correspond to restriction digest products of HPV18-1 and HPV18-2 samples treated with *Eco*RI restriction enzyme, respectively.

2.4.2 Sanger DNA Sequencing of the PCR-amplified DNA Fragments Corresponding to 7479-7783 bp and 7647-157 bp Regions within the HPV18 LCR

The approximate sizes of the PCR-amplified DNA fragments were first evaluated by agarose gel electrophoresis. For the DNA fragment corresponding to HPV18 7479-7783 bp, the expected size is 305 bp; while the DNA fragment corresponding to HPV18 7647-157 bp, the expected size is 368 bp. The observed apparent fragment sizes of the PCR-amplified products separated by agarose gel electrophoresis are in agreement with the expected 305 and 368 bp fragment sizes **(Figure 2.15)**.

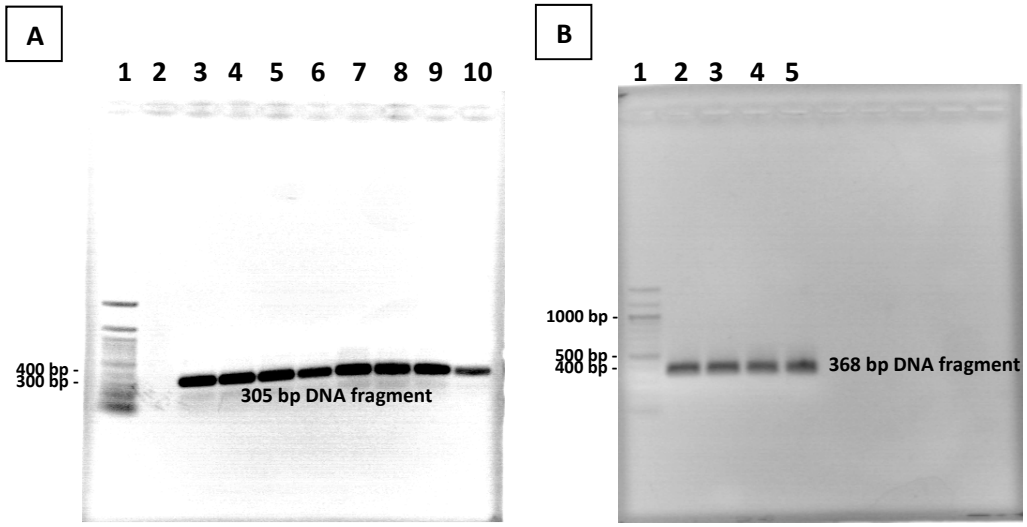


Figure 2.15. Agarose Gel Electrophoresis of PCR Amplification of HPV18 LCR 7479-7783 (305 bp) and 7647-157 (368 bp) Fragments. (A) HPV18 LCR 7479-7783; Lane 1 was loaded with a 100 bp DNA ladder size standard. Lanes 3-10 contain the 305 bp fragment with 5' fluorescent dyes. **(B)** HPV18 LCR 7647-157; Lane 1 was loaded with the 100 bp DNA ladder size standard. Lanes 2 and 3 correspond to the PCR products of a reaction using primers without 5' fluorescent dyes for Sanger sequencing. Lanes 4 and 5 contain the 368 bp fragment with 5' fluorescent dyes.

To ensure that the PCR-amplified fragments corresponded to the HPV 18 LCR regions from 7479 to 7783 bp and from 7647 to 157 bp, Sanger DNA sequencing was performed. The Sanger sequencing data matched the DNA sequences for both regions of the HPV18 LCR ((GenBank Accession No. X05015) **(Figure 2.16)**.

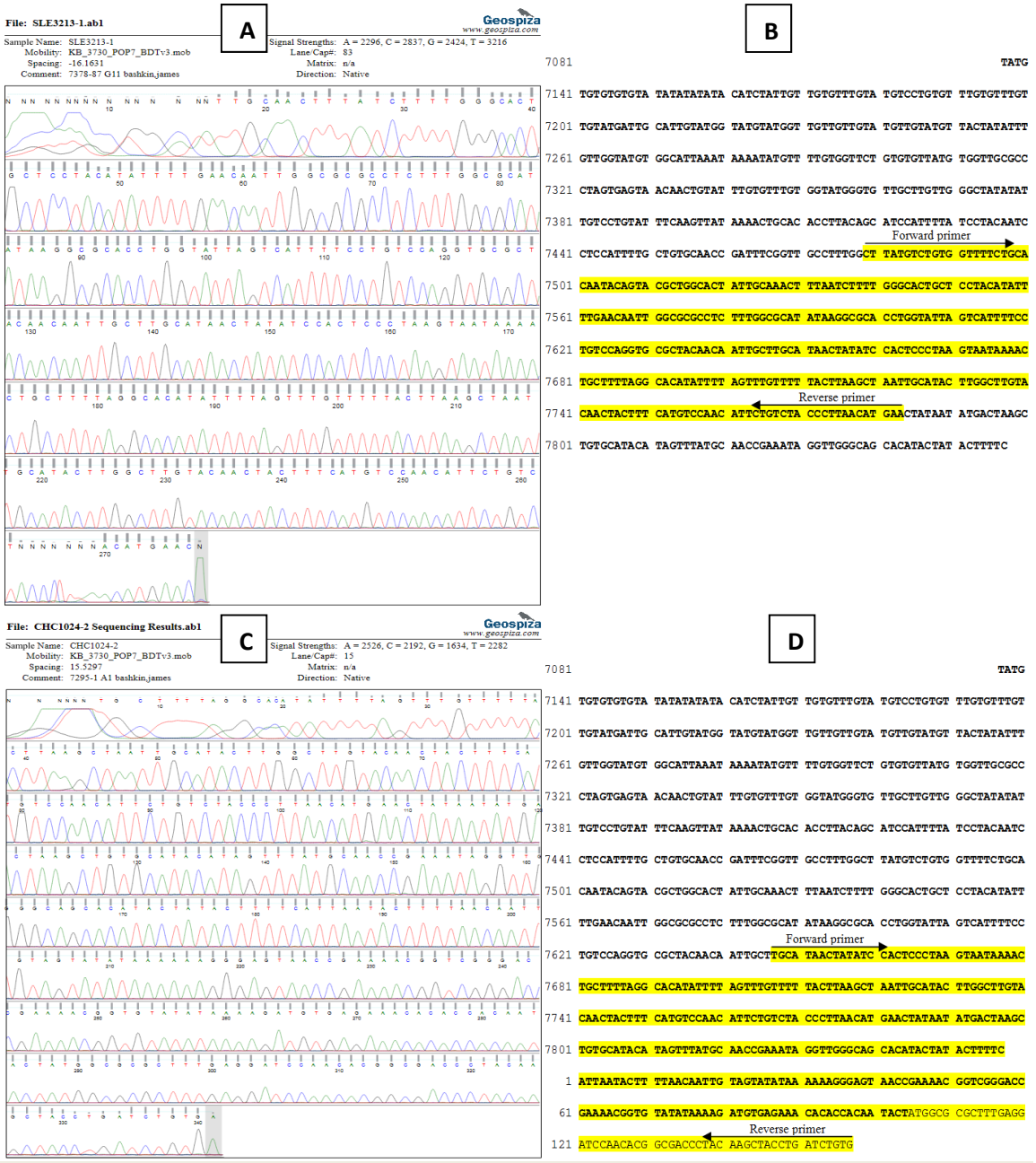
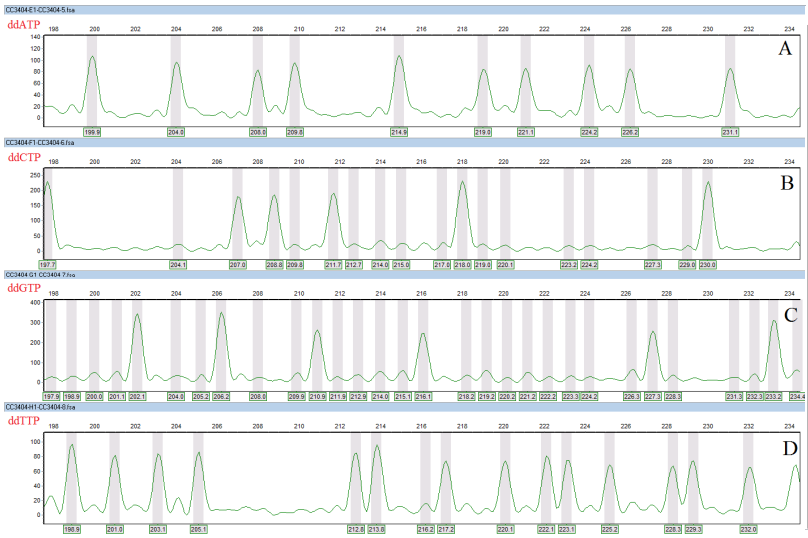


Figure 2.16. Sanger Sequencing Electropherograms and Regions Amplified from HPV18 LCR Sequence. Sanger sequencing results for (**A**) 305 bp fragment 7479-7783 and (**C**) 368 bp fragment 7647-157 from HPV18. (**B**) HPV18 LCR sequence (7479-7783) with the expected sequence for the 305 bp fragment (highlighted). (**D**) HPV18 LCR sequence (7137-104) is annotated with bold letters and the expected sequence for the 368 bp fragment has been highlighted. Sanger sequencing results matched the DNA sequences of HPV18 (GenBank Accession No. X05015). Data were processed using FinchTV v1.4.0.

2.4.3 DNase I Footprinting – CE Size Indexing for the 305 bp (7479-7783) and 368 bp (7647-157) Using Sanger Sequencing Chemistry

In order to accurately map the **PA1** and **PA25** binding sites on the HPV18 fragments (7479-7783 and 7647-157), the correct nucleotide position on the HPV18 genome must be accurately correlated to the relative CE mobility of the fragment terminating with the

nucleotide in question. Both the sequencing chemistry and the nature of 5' dye influence the CE mobility of the DNA, and as a result the appropriate indexing method must be chosen. Particularly for quantitative DNase I footprinting, the cleavage products have a neutral hydroxyl group on the 3' end and display a similar CE mobility to the neutral fragments generated by Sanger chemistry. **Figure 2.17** provides the indexing of a short region and respective electropherogram plot for the USB Sanger sequencing results. In this technique, fragments terminating at each nucleotide along the DNA molecule are generated upon addition of a particular 2', 3'-dideoxynucleotide by DNA polymerase. The generated fragments were in agreement with the viral genomic sequence of HPV18 (GenBank Accession No. X05015). Nucleotide position on the HPV18 genome was then matched to the relative CE mobility of the fragment terminating with the nucleotide in question.



CE Fragment Mobility vs. Real Position

CE position	Sequence	Fragment position	HPV18 position	CE - Fragment position
197.8	C	202	7813	-4.2
198.9	T	203	7812	-4.1
199.9	A	204	7811	-4.1
201.0	T	205	7810	-4.0
202.0	G	206	7809	-4.0
203.1	T	207	7808	-3.9
204.0	A	208	7807	-4.0
205.1	T	209	7806	-3.9
206.2	G	210	7805	-3.8
207.1	C	211	7804	-3.9
208.0	A	212	7803	-4.0
208.8	C	213	7802	-4.2
209.8	A	214	7801	-4.2
210.8	G	215	7800	-4.2
211.7	C	216	7799	-4.3
212.8	T	217	7798	-4.2
213.8	T	218	7797	-4.2
214.9	A	219	7796	-4.1
216.0	G	220	7795	-4.0
217.2	T	221	7794	-3.8
218.0	C	222	7793	-4.0
219.0	A	223	7792	-4.0
220.1	T	224	7791	-3.9
221.1	A	225	7790	-3.9
222.1	T	226	7789	-3.9
223.1	T	227	7788	-3.9
224.1	A	228	7787	-3.9
225.2	T	229	7786	-3.8
226.1	A	230	7785	-3.9
227.3	G	231	7784	-3.7
228.0	T	232	7783	-4.0
229.3	T	233	7782	-3.7
230.1	C	234	7781	-3.9
231.0	A	235	7780	-4.0
232.1	T	236	7779	-3.9
233.2	G	237	7778	-3.8

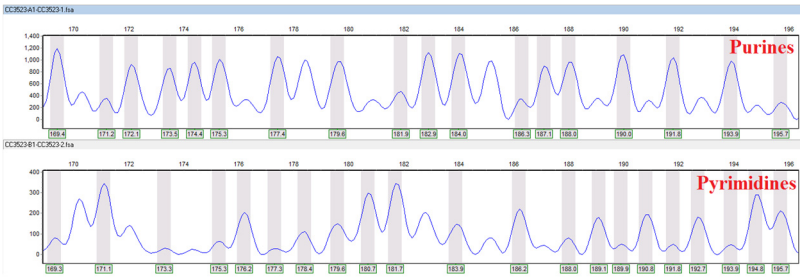
Figure 2.17. Representative USB Sequencing Electropherogram Results Indexed to a Short HPV18 Region (7813-7778). Sample electropherogram (HEX channel) from USB sequencing of a region within the 368 bp fragment (position in HPV18 from 7813 to 7778 bp). USB sequencing PCR reactions (20 ng/µL template DNA) with (**A**) ddATP termination mix, (**B**) ddCTP termination mix, (**C**) ddGTP termination mix and (**D**) ddTTP termination mix. The table on the right shows the apparent CE position *versus* the real position of the nucleotide in the fragment of the plotted region. Data analyzed using GeneMarker Version 2.4.0 and Microsoft Excel 2010.

Fragment indexing data used in DNase I footprinting experiments to correlate the CE size calling to sequencing results for both the top (5'-FAM) and bottom (5'-HEX) strands of the 305 bp (7479-7783) and 368 bp (7647-157) HPV18 fragments are provided in **Figures SI2.2A, -B, -C** and **–D**. The USB sequencing reactions performed with 20 ng/µL of template DNA yielded better output signal than those with 2 ng/µL of template DNA.

2.4.4 Affinity Cleavage – CE Size Indexing for the 305 bp (7479-7783) and 368 bp (7647-157) Using Maxam-Gilbert Sequencing Chemistry

As previously discussed, it is imperative to accurately index the CE fragment size to the actual nucleotide on the sequence of interest in order to correctly map the binding site of DNA-ligands. To accurately determine the cleavage site in hydroxyl radical footprinting and affinity cleavage experiments, the DNA fragment must be indexed by

Maxam-Gilbert sequencing chemistry. In this technique, the DNA is cleaved by specific base modifications, followed by eviction of the modified bases and strand cleavage at these positions.[73] The cleavage products from the Maxam-Gilbert sequencing reactions have a 3'-phosphate group. Similarly, the cleavage products for affinity cleavage experiments and hydroxyl radical footprinting have either a negatively charged phosphate group or a phosphoglycolate group on their 3' end and display a similar CE mobility to the fragments generated by Maxam-Gilbert chemistry.[72] **Figure 2.18** presents a sample electropherogram region for the results from Maxam-Gilbert sequencing. The electropherogram clearly demonstrates the base modification preference of piperidine formate (pH 2.0) for purines and the preference of hydrazine for pyrimidines. In addition, the generated fragmentation peaks are in agreement with the viral genomic sequence of HPV18 (GenBank Accession No. X05015). Nucleotide position on the HPV18 genome was matches to the relative CE mobility of the fragment terminating with the nucleotide in question.



CE Fragment Mobility vs. Real Position

CE position	Sequence	Fragment position	HPV18 position	CE - Fragment position
169.3	A	176	7822	-6.7
170.3	C	177	7823	-6.7
171.2	C	178	7824	-6.8
172.0	G	179	7825	-7.0
173.4	A	180	7826	-6.6
174.3	A	181	7827	-6.7
175.2	A	182	7828	-6.8
176.3	T	183	7829	-6.7
177.3	A	184	7830	-6.7
178.3	G	185	7831	-6.7
179.5	G	186	7832	-6.5
180.8	T	187	7833	-6.2
181.8	T	188	7834	-6.2
182.8	G	189	7835	-6.2
183.9	G	190	7836	-6.1
185.1	G	191	7837	-5.9
186.3	C	192	7838	-5.7
187.0	A	193	7839	-6.0
187.9	G	194	7840	-6.1
189.2	C	195	7841	-5.8
189.9	A	196	7842	-6.1
190.9	C	197	7843	-6.1
191.7	A	198	7844	-6.3
192.8	T	199	7845	-6.2
193.8	A	200	7846	-6.2
194.9	C	201	7847	-6.1

Figure 2.18. Representative Maxam-Gilbert Sequencing Electropherogram Results Indexed to a Short HPV18 Region (7822-7847). Sample electropherogram (FAM channel) from Maxam-Gilbert sequencing of a region within the 368 bp fragment (position in HPV18 from 7822 to 7847 bp). Upper electropherogram illustrates the cleavage pattern from DNA subjected to piperidine in formic acid (purine base modification); whereas the lower electropherogram pattern was obtained from hydrazine cleavage (pyrimidine base modification). The table on the right tabulates the apparent CE position vs the real position of the nucleotide in the fragment of the plotted region. Data analyzed using GeneMarker Version 2.4.0 and Microsoft Excel 2010.

Fragment indexing data used in affinity cleavage experiments to correlate the CE size calling to sequencing results for both the top (5'-FAM) and bottom (5'-HEX) strands of the 305 bp (7479-7783) and 368 bp (7647-157) HPV18 fragments are given in **Figures SI2.3A, -B, -C** and **–D**). The Maxam-Gilbert sequencing reactions performed for 15 and 30 min yielded peaks with high signal-to-noise ratios.

The difference between Sanger and Maxam-Gilbert fragment mobility were plotted (**Figure 2.19**). The results for the DNA fragments are in agreement with the results published by Gaofei *et al*.[72] Particularly, the difference is significant (up to 8 nucleotides) for fragments of about 100 nucleotides and smaller. On the other hand, for fragments of 200 nucleotides and larger, this difference reduces to about 2 to 0 nucleotides. Thus, the aforementioned observations provide evidence to the importance of indexing the DNA sequence prior to the determination of binding sites by either DNase I footprinting, affinity cleavage experiments or hydroxyl radical footprinting.

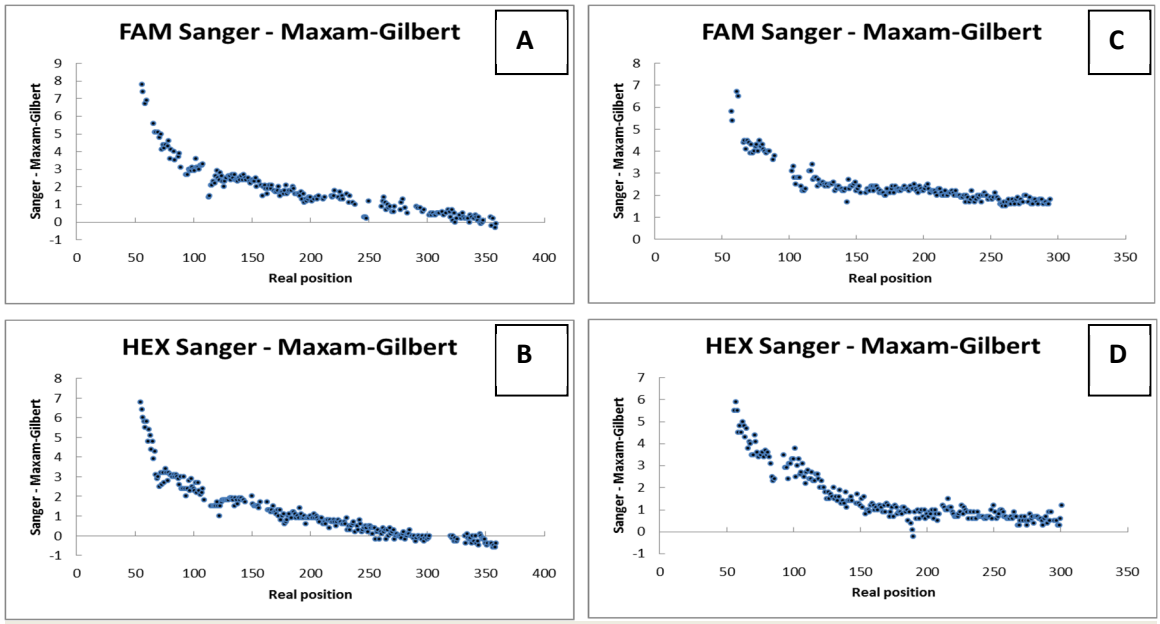


Figure 2.19. Difference between Sanger and Maxam-Gilbert Sequencing Fragment Mobility for HPV18 Sequences. Mobility difference between Sanger and Maxam-Gilbert (Sanger – Maxam-Gilbert) for **(A)** the top strand (5'-FAM labeled) and **(B)** the bottom strand (5'-HEX labeled) of HPV18 7647-157, **(C)** the top strand (5'-FAM labeled) and **(D)** the bottom strand (5'-HEX labeled) of HPV18 7479-7783.

2.4.5 Quantitative DNase I Footprinting: Control DNase I Fragmentation

Quantitative DNase I footprinting reactions must be conducted under conditions of single-hit kinetics.[51] If the cleavage reaction is not performed under these conditions, the cleavage pattern will be biased toward small DNA fragments making it difficult to determine ligand footprints.

Figure 2.20 shows sample electropherograms for the control fragmentation results of the 305 bp and 368 bp fragments. Analysis of the intact fluorescently labelled DNA fragments in the absence of DNase I did not display any fragmented products, suggesting that the DNA was not degraded prior to the cleavage reaction (**Figure 2.20** top panels of **A** and **B**). On the other hand, high DNase I concentrations of 0.03 and 0.05 AU led to overdigestion of the 368 bp DNA fragment. The optimal cleavage under single-kit kinectic conditions were achieved at DNase I concentrations of 0.06 AU for the 305 bp fragment and 0.01 AU for the 368 bp fragment. The difference in the AU between the two experiments is most likely due to the age of the DNase I enzyme, as the control fragmentation experiment for the 305 bp DNA fragment was conducted months after the experiment for the 368 bp DNA fragment. At these DNase I concentrations, the cleavage peaks displayed relative fluorescence intensities many folds (~100X) lower than the relative fluorescence intensities for the uncut peaks, suggesting appropriate experimental conditions for single-hit kinetics. Consequently, these DNase I concentrations were used in subsequent quantitative DNase I footprinting experiments at 200 pM DNA.

Figure 2.20. Representative Electropherograms of Control DNase I Fragmentation. Electropherograms for **(A)** 305 bp (HPV18 7479-7783) and **(B)** 368 bp (HPV18 7647-157). Respective DNase I Activity Units (AU) are given in the upper left corner of each electropherogram. No cleavage pattern with 0 AU was observed for both fragments, which serves as a control to check for fragment contaminants or DNA degradation. At 0.03 and 0.05 AU DNase I, the 368 bp fragment is over-digested. Optimal cleavage observed at DNase I concentration of 0.06 AU and 0.01 AU for 305 bp and 368 bp fragments, respectively. Full length fragments are indicated with vertical orange arrows.

2.4.6 DNase I Footprinting and Affinity Cleavage

We relied on quantitative DNase I footprinting to determine the equilibrium dissociation constants of PA-DNA interactions and affinity cleavage experiments to assign the PA-binding orientations.

2.4.6.1 PA1 Binding Sites on HPV18 7647-157 bp Determined by DNase I Footprinting and Affinity Cleavage Experiments

The putative **PA1** binding sites on HPV18 7647-157 were determined by following the published PA-DNA recognition rules.[20,25,26] **Figure SI2.4** maps the expected **PA1** binding sites on the 368 bp fragment covering match binding sites (5'-W_2GW_7-3') to double-base-pair mismatch binding sites in both reverse and forward orientations. Forward orientation refers to alignment of the polyamide N-terminus to C-terminus vector with the 5'→3' direction of the DNA, whereas a reverse orientation describes the polyamide N→C vector alignment with the 3'→5' direction of the DNA **(Figure 2.21A)**.[77]

The 368 bp fragment used in this analysis harbors three putative forward match binding sites and four reverse match binding sites for **PA1**. Additionally, there are 37 predicted single-base-pair mismatch binding sites, while many more putative double-base-pair mismatch binding sites. Many of these predicted binding sites cluster around

the E1 binding region between the E2 binding sites E2BS-1 (7822-7833) and E2BS-2 (42-53) (see **Figure SI2.4**).

Representative results from quantitative DNase I footprinting and affinity cleavage experiments corresponding to an HPV18 region bound by **PA1** and **PA1-EDTA** are presented in **Figure 2.21**. The process of determining **PA1** binding sites are shown in this representative region. To correctly assign the PA-binding orientations within these fragments, affinity cleavage experiments were conducted as described previously in the literature. **Panel A** shows both the forward and reverse binding orientations for **PA1**. **Panel C** provides the raw electropherograms for quantitative DNase I footprinting and affinity cleavage experiments. **Panel B** maps the predicted and observed binding sites for **PA1** in the HPV18 region corresponding to 7770-7805 bp. Interestingly, **PA1** binding to the single-base-pair mismatch sequence 5'-ATGAACTATA-3' (mismatch underlined), as predicted following the reported DNA recognition rules, was not observed. On the other hand, **PA1** binding to another single-base-pair mismatch site, 5'-TAGTCATATT-3', was observed (furthest to the right). Furthermore, a second polyamide binding event was observed to the sequence 5'-AACTATAATA-3', but the polyamide binding orientation cannot be determined unequivocally. As a result, both of the possible binding modes are illustrated (labelled as Reverse match site or Single-base-pair mismatch). **PA1** binding in the forward orientation leads to a single-base-pair mismatch at *N*-methylimidazole, while the reverse orientation presents the match site for **PA1**.

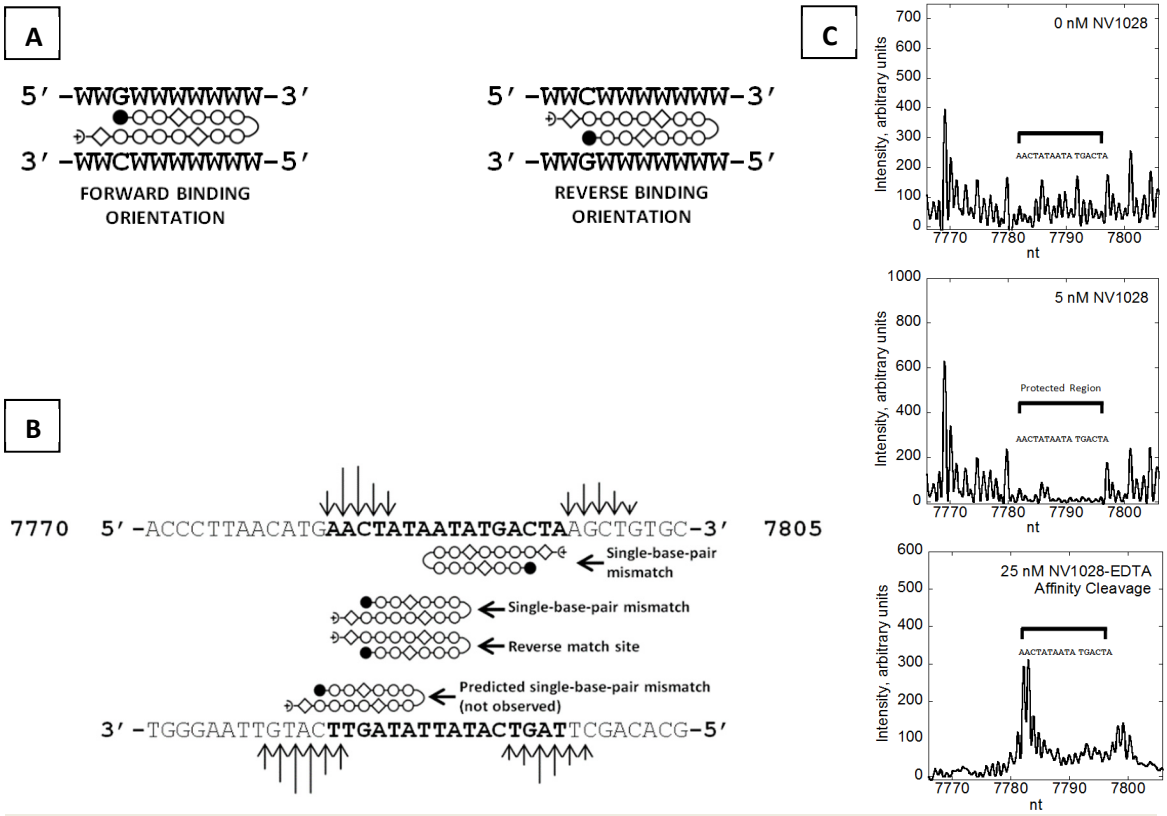


Figure 2.21. *Caption on next page.*

Figure 2.21. Schematic Representation of the Predicted and Observed Binding Sites in HPV18 LCR DNA from 7770-7805 bp for PA1. (A) Binding orientations for **PA1** are depicted in between complementary DNA strands. The DNA match site for **PA1** is 5'-W_2G_7-3', where the letter W (weak) stands for A or T. The polyamide building blocks are represented by open circles for *N*-methylpyrroles, filled circle for *N*-methylimidazole, diamonds for β-alanines, + for the Ta tail and a loop for γ-aminobutyric acid. **(B)** Map of the predicted and observed binding sites along with their corresponding binding orientations with respect to the Ta tail for **PA1** as determined by quantitative DNase I footprinting and affinity cleavage experiments. Bold sequences depict the observed binding sites as determined from the plots in Panel C (compare 0 nM and 5 nM polyamide). The respective cleavage patterns are shown as arrows above or below the DNA sequences. Affinity cleavage patterns were determined from the lower plot from Panel C for the top strand (affinity cleavage electropherogram for the bottom strand is not shown). Interestingly, **PA1** binding to the single-base-pair mismatch sequence 5'-ATGAACTATA-3' (mismatch underlined), as predicted following the reported DNA recognition rules, was not observed. On the other hand, polyamide binding was observed for 5'-TAGTCATATT-3' (furthest to the right). Furthermore, a second polyamide binding event was observed to the sequence 5'-AACTATAATA-3', but the polyamide binding orientation cannot be determined unequivocally. As a result, both of the possible binding modes are illustrated (labelled as Reverse match site or Single-base-pair mismatch). **PA1** binding in the forward orientation leads to a single-base-pair mismatch at *N*-methylimidazole, while the reverse orientation presents the match site for **PA1**. **(C)** Raw electropherograms of the HPV18 LCR 7770-7805 sequence using DNase I footprinting and affinity cleavage. The top panel shows the control DNase I fragmentation (0 nM polyamide) and the middle panel illustrates the protected region (solid line above pattern) upon addition of 5 nM of **PA1**. Affinity cleavage was performed with EDTA-modified **PA1** at its Ta tail.

The **PA1** binding events determined from quantitative DNase I footprinting and affinity cleavage experiments in the HPV18 LCR corresponding to 7647-157 bp are mapped in **Figure 2.22**. The sequence map summarizes the **PA1** binding sites with their respective equilibrium dissociation values calculated using the Hill equation followed by the standard deviations in parentheses. **Table 2.5** summarizes only the higher affinity sites (**PA1** DNase I footprinting at 5 nM PA / affinity cleavage patterns at 50 nM PA conducted at 200 pM DNA and at 1 nM DNA, respectively).

The binding events are assigned assuming that **PA1** has a preference for the lowest number of PA-DNA mismatches and prefers the forward orientation binding mode. Because the binding orientation (forward *vs.* reverse) cannot be determined unequivocally by our experiments, both the forward orientations and their respective reverse, lower mismatch partners are shown in **Figure 2.22**. This map also includes yellow highlighted sequences for observed **PA1** footprinting regions at 5 nM and grey highlighted sequences for footprinting regions that appeared at 10 nM **PA1**. The 3' shift of the footprinting patterns observed in our experiments is consistent with DNA cleavage by DNase I originating on the minor groove of the B-form DNA.[63]

Using the indexing obtained from Maxam-Gilbert sequencing experiments, the actual nucleotide positions of the hydroxyl radical attack sites were determined. The obtained affinity cleavage patterns revealed an asymmetrical 3' shift between opposite DNA strands, which agrees with **PA1**-binding to the minor groove of B-form DNA. A total of 24 distinct affinity cleavage sites were observed by **PA1-EDTA** affinity cleavage experiments and these were labeled 1 through 24 based on their position within the fragment. Out of these sites, 16 were observed at 25 and 50 nM **PA1** in affinity cleavage experiments. The affinity cleavage patterns were obtained for both strands of the binding sites except for Site #1 because it is positioned near the beginning of the fragment in the FAM channel. Similarly, DNase I footprints were not shown for the first and last 50 nucleotides of the map. Binding events were determined by the presence of DNase I footprints with affinity cleavage flanking at least one end of these protected regions. While all of the predicted **PA1** match binding sites were in agreement with the affinity cleavage and DNase I footprints, various predicted single-base-pair mismatch **PA1**

binding sites did not produce DNase I footprints and/or did not agree with the observed affinity cleavage patterns **(Figure SI2.5** & **Table SI2.2)**.

Each observed site will be briefly discussed in this section. Site #1 is best described by **PA1** binding in the forward orientation to the match site at genomic position 7669-7678, while Site #2 corresponds to a single-base-pair mismatch site (7673-7682). Although the map shows these two **PA1** hairpin polyamides overlapping in their recognition sequences, it is unlikely that two hairpin polyamides would fit concurrently side-by-side in the minor groove of the DNA duplex. Instead, this apparent overlapping is due to the experimental conditions, which require an ensemble of DNA and polyamide molecules, leading to an averaged-out determination of DNA-PA interactions. In this manner, a population of DNA molecules would be cleaved by **PA1**-EDTA at Site #1 while others would be cleaved at Site #2. This rationale applies to other sites on this map and **PA25** maps.

Site #3 can be explained by three different forward double-base-pair mismatch binding events to the genomic sequence spanning 7694 to 7705. However, the affinity cleavage patterns suggest that only one polyamide may bind at this site with the polyamide in the middle (7695-7704) displaying the best fit with the top and bottom affinity cleavage patterns. In the case of Site #4, the affinity cleavage pattern and DNase I footprinting results agree with **PA1** binding to either a reverse match site or a forward single-base-pair mismatch on 7695-7704, as well as **PA1** binding to a forward single-base-pair mismatch site (7704-7713). It is not clear at this point which of these binding sites and conformations (forward *vs.* reverse) are occupied by the polyamide. The dissociation constant shared by **PA1** in Site #3 and Site #4 was determined as 1.5(7) nM with a Hill coefficient (n) of 1.76 (see **Figure SI2.6** for sample binding isotherms).

Site #5 is best described by **PA1** binding in a forward orientation to a single-base-pair mismatch site with a K_d of 1.2(4) (n = 1.76) from the FAM channel and 1.4(1) nM (n = 1.51) from the HEX channel. Similar to Site #4, Site #9 has two possible **PA1** binding conformations: a reverse match site and a forward single-base-pair mismatch site on the same DNA sequence (7782-7791), which cannot be differentiated by our experiments. The determined K_d values at Site #9 are 1.9(4) nM (n = 1.63) and 1.2(2) nM (n = 1.45) for FAM and HEX channels, respectively.

Site #10 is in agreement with two predicted single-base-pair mismatch sites (reverse 7786-7795 and forward 7788-7797) and one forward double-base-pair mismatch site (7786-7795). However, the latter fits the highest intensity affinity cleavage pattern better than the former. Furthermore, the diffuse nature of the low intensity affinity cleavage pattern suggests that **PA1** can populate multiple, adjacent double-base-pair mismatch sites. The K_d vlues for Site #10 were determined as 1.5(2) nM (n = 2.14) and 1.2(2) nM (n = 1.45) for FAM and HEX channels, respectively.

Unlike all of the other high affinity cleavage sites, Site #12 and #13 are best explained by **PA1** binding in the forward orientation to two double-base-pair mismatch sites (7811-7820 and 7813-7822). However, the generated patterns were low in intensity and the K_d values were slightly higher with values of 2.9(3) nM (n = 3.20) and 2.5(3) nM (n = 1.72) for FAM and HEX channels, respectively. These two sites bind in close proximity (Site #12) or overlap by one nucleotide (Site #13) with E2BS-1 (7822-7833).

Sites #15, #16, #17, #18 and #19 bind to the E1-binding region of HPV18.[78] Site #15 can be explained by **PA1** binding to a forward single-base-pair mismatch site on 7845-

7854, a reverse match site (7855-7) or a forward single-base-pair mismatch site (7855-7). The forward single-base-pair mismatch site at 7845-7854 is bound by **PA1** because these nucleotides displayed DNase I footprinting at 5 nM PA with a calculated K_d values of 3.1(5) nM (n = 2.25), 2.6(3) nM (n = 2.70), 2.3(5) nM (n = 2.52) and 2.1(5) nM (2.25) from both CE channels. On the other hand, the binding orientation for **PA1** on the HPV18 sequence from 7855 to 7 cannot be determined based on the data; therefore, both the forward single-base-pair mismatch and the reverse match binding modes are represented. Furthermore, it is not completely clear if **PA1** even binds this sequence because Site #15 can be generated by **PA1** binding to 7845-7854. Sites #16 and #17 are best described by **PA1** binding to a forward single-base-pair mismatch site and a forward match binding site, respectively. The dissociation constant values shared by **PA1** in Sites #15, #16 and #17 are 1.2(1) nM (n = 1.71), 1.1(1) nM (n = 1.58) and 1.8(6) nM (n=1.60). Site #18 can be explained by **PA1** binding in the forward orientation to the match site at genomic position 21-30, while Site #19 corresponds to two adjacent single-base-pair mismatch sites (24-33 and 25-34). The latter two binding events are within a DNA turn from the E2 Binding Site 2 (E2BS-2 42-53). The shared K_d values for Sites #18 and #19 were determined as 1.4(2) nM (n = 1.93) and 1.9(1) nM (n = 2.41) for FAM and HEX channels, respectively.

Sites #22 and #23 are in close proximity to E2 binding site 3 (E2BS-3 7822-7833). Site #22 is best described by **PA1** binding in the forward orientation to a single-base-pair mismatch at genomic position 68-77 with a K_d value of 1.6(4) nM (n = 1.85), while Site #23 corresponds to **PA1** recognizing either a forward single-base-pair mismatch site and/or a reverse match binding site (73-82) with K_d values of 1.6(4) nM (n = 1.85) and 1.5(4) nM (n = 1.49).

On the other hand, the affinity cleavage patterns for Sites #6-8, #11, #14, #20, #21 and #24 were only observed at high **PA1** concentrations ($\geq$100 nM) and their calculated K_d values were the highest for this fragment. Furthermore, these sites also exhibited large Hill coefficient values with Site #11, #14 and #24 displaying the high binding cooperativity (n = 2.81, n = 3.76, n = 4.34, respectively).

In summary, all of the predicted **PA1** match binding sites are in agreement with DNase I footprints and affinity cleavage pattern(s) flanking these protected regions. In contrast, various predicted single-base-pair mismatch **PA1** binding sites did not produce DNase I footprints and/or did not agree with the observed affinity cleavage patterns, indicating that these sites were not bound by the hairpin polyamide (see **Figure SI2.5** & **Table SI2.2**). Perhaps these observations are due to the fact that the published PA-DNA recognition rules, while they provide an overall prediction of polyamide binding, do not account for subtle DNA structural differences between seemingly equivalent DNA sequences. Furthermore, these rules fail to include the flanking sequences around the predicted binding site and other binding sites nearby that may induce a conformational change in the local DNA structure. Thus, recognition of the minor groove by **PA1** is not simply based on the established PA-DNA recognition rules.

The high affinity DNase I footprinting (5 nM PA) sites and affinity cleavage patterns (25 and 50 nM PA) can be explained with binding sites corresponding to match, single- and double-base-pair mismatches. On the other hand, low affinity sites—DNase I footprints at 10 nM PA and affinity cleavage patterns at 100 nM PA—required polyamide binding to sites with double-, triple- and even quadruple-base-pair

mismatches. The equilibrium dissociation constants determined by the Hill-calculated binding isotherms ranged from 1.1 to 1.9 nM for match binding sites and from 1.2 to 3.1 nM for single-base-pair mismatch binding sites. These results indicated that a single-base-pair mismatch in the binding sequence was tolerated and conferred little to no effect on the equilibrium dissociation constant.

Binding of **PA1** to the HPV18 origin of replication may be of importance to the antiviral activity of **PA1** against HPV18. The 368 bp fragment harbors the viral origin of replication, the E2 binding sites (E2BS-1 7822-7833; E2BS-2 42-53; E2BS-3 58-69) and the E1 binding site (E1BS 3-18). Because **PA1** binds to the E1 binding region and in close proximity to the E2 binding sites, this DNA minor groove binder may preclude binding of the essential viral proteins E1 and E2 in infected cells to the origin of replication by steric hindrance or by modifying the local viral DNA structure.

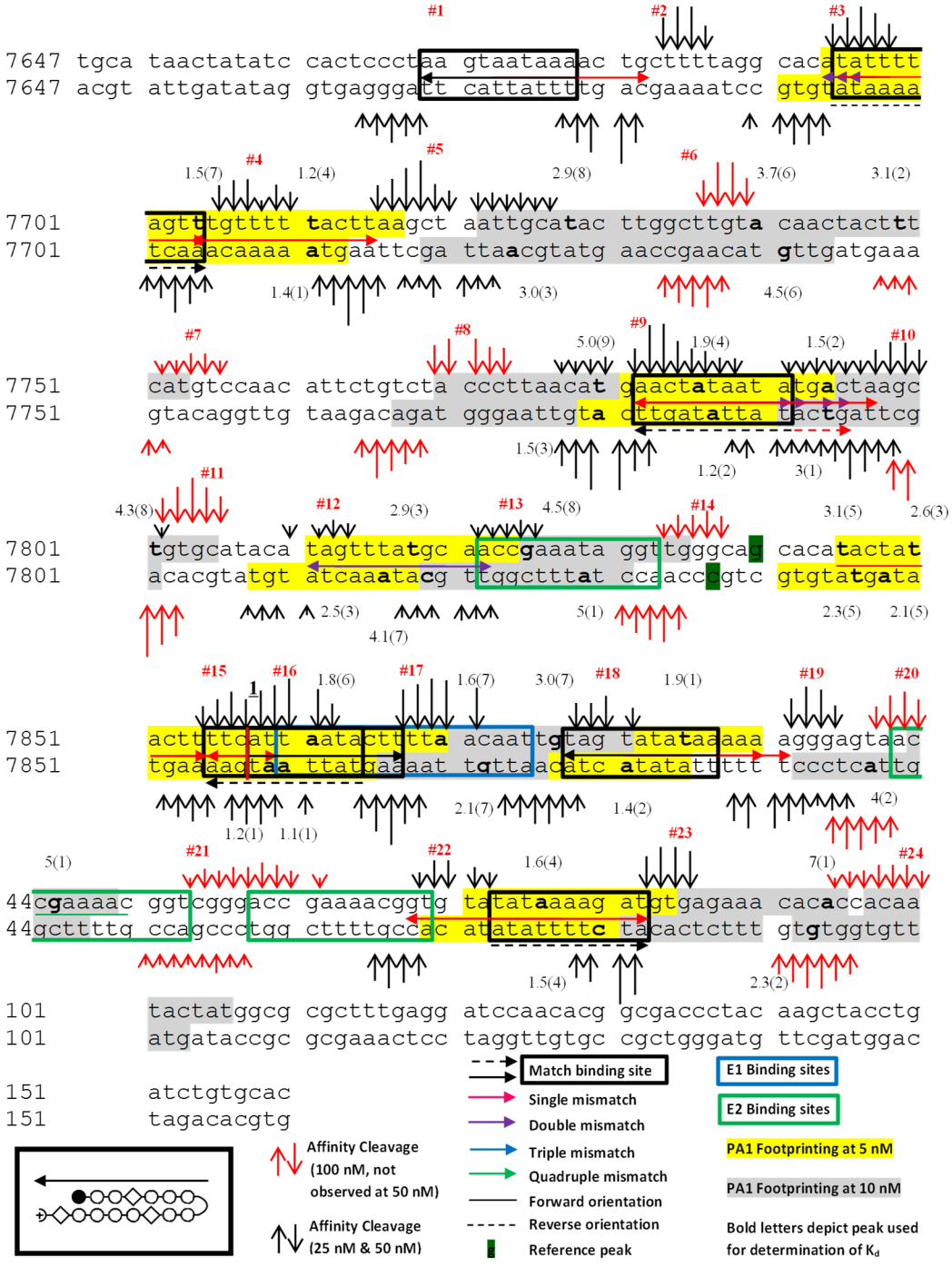


Figure 2.22. PA1 Binding Sites and Hill-Calculated Equilibrium Dissociation Constants on the HPV18 Region Corresponding to 7647-157 bp. The DNA duplex for the 368 bp HPV18 DNA fragment is shown along with **PA1** represented as arrows at its respective binding sites. **PA1** is depicted with either a solid arrow (forward orientation) or a dashed arrow (reverse orientation). HPV E2 binding sites are depicted in green boxes. HPV E1 binds in the region (HPV18 3-19 bp) flanked by the first and second E2 binding sites. The origin of replication is marked with 1. Numbers on the left correspond to the nucleotide position in HPV18 genome. Equilibrium dissociation constants calculated using the Hill equation from the FAM and HEX channels of CE at the particular nucleotide (bold) are given. The sample standard deviations are in parentheses. Affinity cleavage patterns at 25-50 nM (black arrows) and 100 nM (red arrows) **PA1-EDTA** are illustrated. At high polyamide concentrations of 100 nM **PA1-EDTA** and above, new binding sites in the 368 bp fragment are populated by the polyamide. Yellow highlighted sequences represent footprinting regions observed at 5 nM **PA1**, whereas gray highlighted sequences represent footprinting regions observed at 10 nM **PA1**.

Table 2.5. PA1 Sites on 368 bp Fragment Corresponding to HPV18 7647-157 bp.

AC site #	Sequence	Position	Site type	Binding orientation	Integration nt	*Kd* (nM) DNase I	Hill Coeff.
1	TCCCT AA**G**TAATAAA ACTGC	7669-7678	Perfect	Forward	*ND*	*ND*	*ND*
2	AAAAG CA**G**TTTTATT ACTTA	7682-7673	Single	Forward	*ND*	*ND*	*ND*
3[a]	GGCAC AT**A**TTTTAGT TTGTT	7694-7703	Double	Forward	7704	1.5 ± 0.7	1.76
3[a]	GCACA TA**T**TTTAGTT TGTTT	7695-7704	Double	Forward			
3[a]	CACAT AT**T**TTAGTTT GTTTT	7696-7705	Double	Forward			
4	AAACA AA**C**TAAAATA TGTGC	7704-7695	Single Perfect	Forward Reverse			
4	TTAGT TT**G**TTTTTAC TTAAG	7704-7713	Single	Forward	7711	1.4 ± 0.1	1.51
5	AGCTT AA**G**TAAAAAC AAACT	7715-7706	Single	Forward			
9	ACATG AA**C**TATAATA TGACT	7782-7791	Single Perfect	Forward Reverse	7787	1.2 ± 0.2	1.45
10	GCTTA GT**C**ATATTAT AGTTC	7795-7786	Single	Reverse	7794	1.5 ± 0.2	2.14
10	CAGCT TA**G**TCATATT ATAGT	7797-7788	Single	Forward			
12	ATACA TA**G**TTTATGC AACCG	7811-7820	Double	Forward	7818	2.9 ± 0.3	3.20
13	TCGGT TG**C**ATAAACT ATGTA	7822-7813	Double	Forward			
15	ATGAA AA**G**TATAGTA TGTGC	7854-7845	Single	Forward	7850	2.6 ± 0.3	2.70
15	TACTT TT**C**ATTAATA CTTTT	7855-7	Single Perfect	Forward Reverse	2	1.2 ± 0.1	1.71
16	TATTA AT**G**AAAAGTA TAGTA	2-7850	Single	Forward			
17	GTTAA AA**G**TATTAAT GAAAA	10-1	Perfect	Forward			
18	AATTG TA**G**TATATAA AAAAG	21-30	Perfect	Forward	28	1.9 ± 0.1	2.41
19	TCCCT TT**T**TTATATA CTACA	33-24	Single	Forward			
19	CTCCC TT**T**TTTATAT ACTAC	34-25	Single	Forward			
22	AAACG GT**G**TATATAA AAGAT	68-77	Single	Forward	76	1.6 ± 0.4	1.85
23	CTCAC AT**C**TTTTATA TACAC	82-73	Single	Forward			

The affinity cleavage (AC) site # corresponds to the AC patterns in the map provided in the previous page. The **PA1** binding sequences are given in a 5' to 3' direction and include 5 nucleotide (nt) flanking regions. The bold nt highlights the position where the polyamide imidazole/pyrrole pair binds. Position corresponds to the HPV18 genomic nucleotides. The forward binding orientation refers to alignment of the polyamide N-terminus to C-terminus vector with the 5'→3' direction of the DNA, whereas a reverse orientation describes the polyamide N→C vector alignment with the 3'→5' direction of the DNA.

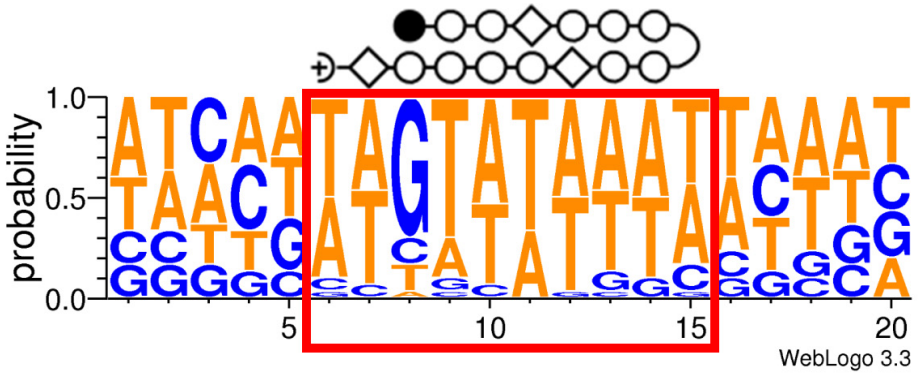
ND, Not determined. Sites 1 and 2 are found in the beginning of the fragment. All of the predicted **PA1** match (perfect) sequences where supported by both affinity cleavage patterns and DNase I footprinting results. [a] A single binding event cannot be assigned due to significant overlap of the predicted **PA1** recognition sequences.

2.4.6.2 PA1 WebLogo 3.3 Motif Analysis

In order to create a graphical representation of the consensus DNA binding sequence for **PA1**, the observed binding sites were inputted into WebLogo 3.3 (**http://weblogo.threeplusone.com/create.cgi**). WebLogo 3.3 is a sequence logo generator that displays the nucleotide conservation at a particular position for multiple user-generated aligned sequences.[79] Sequences with DNase I footprinting and flanked with the appropriate affinity cleavage pattern (observed in 25 nM and 50 nM **PA1**-EDTA) were used to generate the WebLogo plot shown in **Figure 2.23** (sample size of n = 30). The sequence input consisted of DNA sequences from the strand recognized by the N-terminal, top strand of **PA1** with polyamide orientations (forward or reverse) leading to the lowest DNA-PA mismatches.

WebLogo analysis (**Figure 2.23**) confirms that **PA1** binds to A/T-rich sequences with a strong preference of the imidazole unit of **PA1** recognizing the guanine base in the duplex DNA. Furthermore, base-pair mismatches are allowed in all of the positions in the binding region except the position occupied by a β-alanine on the top strand of **PA1** in

this data set; however, data from other HPV18 LCR fragments suggest that this position can also accommodate mismatches. Because the DNA sequences presented to the **PA1** are limited, these observations can be verified using high-throughput method coupled with massively parallel sequencing (see **Chapter 5**). Thus, random oligonucleotides with all the possible DNA sequence permutations in a redundant manner can be interrogated.



TCCCT AAGTAATAAA ACTGC	TTTGT TTGATTTTAT ACACG	TGTAC TTGATATTAT ACTGA
ATGAA AAGTAATTAT GAAAA	GTTAA AAGTATTAAT GAAAA	AATTG TAGTATATAA AAAAG
GAGTG TAGAAAATAT ATGTG	AAAAG CAGTTTTATT ACTTA	AAACA AACTAAAATA TGTGC
TTAGT TTGTTTTTAC TTAAG	AGCTT AAGTAAAAAC AAACT	ACATG AACTATAATA TGACT
CGAAT CAGTATAATA TCAAG	CAGCT TAGTCATATT ATAGT	ATGAA AAGTATAGTA TGTGC
TATTA ATGAAAAGTA TAGTA	TACTT TTCATTAATA CTTTT	TCCCT TTTTTATATA CTACA
CTCCC TTTTTTATAT ACTAC	AAACG GTGTATATAA AAGAT	CTCAC ATCTTTTATA TACAC
GGCAC ATATTTTAGT TTGTT	GCACA TATTTTAGTT TGTTT	CACAT ATTTTAGTTT GTTTT
GTTAA TCGAATTCAT TTTTG	ATACA TAGTTTATGC AACCG	AGCCA ACGTATTTGA TACAT
TTCGG TTGCATAAAC TATGT	ATACG TTGGCTTTAT CCAAC	ACGGG TTGGATAAAG CCAAC

Red = polyA tracts

Figure 2.23. WebLogo Plot for PA1 Binding Sites Observed on the HPV18 Region Corresponding to 7647-157 bp. The relative probability of a given nucleotide to be found in a particular position is represented by the height of the letter (A = Adenine, T = Thymine, G = Guanine and C = Cytosine). Red box encloses the **PA1** binding site. The input sequences including the flanking DNA sequences (5 bp) are given below the plot. No obvious trend is observed for the presence or relative position of polyA tracts in or immediately adjacent to the **PA1** binding site.

2.4.6.3 PA25 Binding Sites Determined by DNase I Footprinting and Affinity Cleavage Experiments

The putative **PA25** binding sites on HPV18 7479-7783 bp and HPV18 7647-157 bp were determined by following the published PA-DNA recognition rules.[20,25,26] **Figures SI2.7** and **SI2.8** map the expected **PA25** binding sites on the 305 bp and 368 bp fragments covering match binding sites (5'-$W_2GW_5GW_4$-3') to triple-base-pair mismatch binding sites in both reverse and forward orientations. For the two HPV18 LCR fragments studied, there are zero **PA25** predicted match binding sites.

Figures 2.24-2.27 (tabulated data in **Tables 2.6-2.9**) summarize the binding sites, affinity cleavage patterns and K_d values for **PA25** determined on the 305 bp (7479-7783 bp) and 368 bp (7647-157 bp) fragments.

The equilibrium dissociation constants were determined using both the FAM and HEX channels from the CE results. The data were fitted to both Langmuir and Hill equations using Kaleidagraph 4.1.1. Some equilibrium dissociation constants could not be calculated due to been positioned at the beginning of the DNA fragments or because of low signal-to-noise ratios. Thus, the two fragments were designed to overlap in order to obtain the binding sites at the beginning of the 368 bp fragment. In line with **PA1**, the

assumption is that **PA25** has a preference for the lowest number of mismatches, thus **PA25** may bind in both reverse and forward orientations. However, the binding orientation (forward *vs*. reverse) cannot be determined unequivocally by our experiments, thereby, both the forward orientations and their respective reverse, lower mismatch partners are shown in **Figures 2.24-2.27**. These figures include cases where forward orientations can also bind in the reverse orientation with one lower mismatch. These maps also include yellow highlighted sequences for observed **PA25** footprinting regions at 2 nM **PA25** and grey highlighted sequences for footprinting regions that appeared at 5 nM **PA25** for titrations performed with 200 pM DNA (**Figures 2.24 and 2.26**). For titrations performed with 50 pM DNA, the yellow highlighted sequences correspond to footprints observes at 0.5 nM **PA25** and the grey highlighted sequences show footprints that appeared at 1 nM **PA25** (**Figures 2.25 and 2.27**). Analogous to the previously discussed **PA1** analysis, the 3' shift of the footprinting patterns is consistent with the DNase I-mediated cleavage of DNA occurring in the minor groove. DNase I footprints were not shown for the first and last 50 nucleotides of the map. Binding events were determined by the presence of DNase I footprints with affinity cleavage flanking at least one end of these protected regions.

2.4.6.4 PA25 Binding Sites Mapped on HPV18 7479-7783 bp

A total of 20 distinct affinity cleavage sites were observed by **PA25-EDTA** affinity cleavage experiments and these were labeled 1 through 20 based on their position in the fragment. Out of these sites, 16 were observed in affinity cleavage experiments with 25 and 50 nM **PA25**. The affinity cleavage patterns were obtained for both strands of the binding sites except for Site #20 because it is positioned near the end of the fragment in the FAM channel. There are zero predicted match binding sites for **PA25** on this fragment. And similar to **PA1** results, a predicted single-base-pair mismatch **PA25** binding site (7721-7709) did not generate an affinity cleavage pattern (**Figure SI2.9** and **Table SI2.3**).

Refer to **Figures 2.24 and 2.25**, as each observed site will be briefly discussed in this section. Site #1 is best described by **PA25** binding to two forward triple-base-pair mismatch sites (7531-7519 and 7528-7540) and a reverse double-base-pair mismatch (7528-7540), while Site #2 corresponds to a forward double-base-pair mismatch site (7538-7526). It is not clear at this point which of these sites and conformations are bound by **PA25**. Although the affinity cleavage pattern for Sites #1 and #2 do not fit perfectly with the described **PA25** binding, moving **PA25** to 7518-7530 and 7528-7540 leads to a quintuple-base-pair mismatch site and a quadruple-base-pair mismatch site, respectively. Instead, the disagreement between the affinity cleavage and the described **PA25** binding may be due to a difference between the fragment indexing and the size calling in the experiment as this site is positioned at the beginning/end of the fragment. The shared K_d values for Sites #1 and #2 were determined as 0.4(1) nM (n = 1.91) and 0.5(2) nM (n = 2.21) at 200 pM DNA, and 0.3(2) nM (n = 1.99) and 0.11(1) nM (n = 1.48) at 50 pM DNA. Representative binding isotherms are provided in **Figure SI2.10**.

Site #3 can be explained by three forward triple-base-pair mismatch sites (7553-7565, 7555-7567 and 7558-7570) and two reverse double-base-pair mismatch sites corresponding to 7553-7565 and 7558-7570. Of the sites mentioned, the forward triple-

base-pair mismatch site at 7555-7567 fits best the observed affinity cleavage pattern. In the case of Site #5, the affinity cleavage pattern and DNase I footprinting results agree with **PA25** binding to a forward double-base-pair mismatch site at 7568-7556 and a reverse triple-base-pair mismatch site at 7571-7559. The former fits the affinity cleavage pattern and DNase I footprint better than the site at 7571-7559. The **PA25** dissociation constants for Sites #3 and #5 were determined to be 0.7(3) nM (n = 2.13) and 0.6(1) nM (n = 1.88) at 200 pM DNA, and 0.43(9) nM (n = 3.09) and 0.24(9) nM (n = 1.35) at 50 pM DNA.

Site #8 is best described by **PA25** binding in a reverse orientation to a triple-base-pair mismatch site (7605-7617), but it may also bind in a forward orientation to a double-base-pair mismatch site nearby on the sequence corresponding to 7603-7615. Similarly, Site #9 is better described by **PA25** binding to a reverse triple-base-pair mismatch site on 7617-7605 than **PA25** binding to a forward double-base-pair mismatch site on 7621-7609. The determined K_d values at Sites #8 and #9 are 0.9(1) nM (n = 2.40) at 200 pM DNA and 0.35(5) nM (n = 1.91) at 50 pM DNA.

Site #10 is in agreement with a reverse triple-base-pair mismatch site (7637-7649), while Site #11 can be explained as **PA25** binding to a forward triple-base-pair mismatch site (7651-7639). Although a reverse double-base-pair mismatch site is found at 7652-7664, the DNase I footprint does not suggest **PA25** binding to this region. The K_d values shared by Sites #10 and #11 were determined to be 1.0(4) nM (n = 2.87) and 0.9(3) nM (n = 2.24) at 200 pM DNA, and 0.24(1) nM (n = 1.51) and 0.3(1) nM (n = 1.69) at 50 pM DNA.

Site #12 is best described by **PA25** binding in a forward orientation to a single-base-pair mismatch site on 7662-7650 with dissociation constants of 0.9(3) nM (n = 2.23) at 200 pM DNA and 0.32(4) nM (n = 1.78) at 50 pM DNA.

Because the end of the 305 bp fragment (7479-7783) overlaps with the beginning of the 368 bp fragment (7647-157), the remaining Sites will be discussed in the context of the 368 bp fragment (see below). The affinity cleavage patterns and DNase I footprints in the overlapping region of the two fragments are in agreement, except for Site #17 which is observed for the 305 bp fragment, but not the 368 bp fragment. Because Sites #13-16 were too close to the beginning of the 368 bp (7647-157) fragment, the K_d values were determined using the 305 bp fragment (7479-7783). Site #13 can be explained by **PA25** binding in the forward orientation to the double-base-pair mismatch site at genomic position 7669-7681, while Site #15 can be explained by two different triple-base-pair mismatch binding events in the forward conformations to positions 7685-7673 and 7688-7676. The **PA25** K_d value for the region shared by Site #13 and #15 was determined as 0.9(5) nM (n = 1.30) and 0.8(5) nM (n = 3.78) at 200 pM DNA, while at 50 pM DNA these values decreased to 0.5(3) nM (n = 1.89) and 0.2(1) nM (n = 1.39).

For Site #16, the affinity cleavage pattern and DNase I footprinting results agree with **PA25** binding to a forward double-base-pair mismatch site on 7694-7706. Site #17 was only observed in the 305 bp fragment and can be explained by **PA25** binding to two different forward triple-base-pair mismatch sites and two reverse double-base-pair mismatch sites. The K_d values shared by Sites #16 and #17 were determined as 0.4(3) nM (n = 1.84) at 200 pM DNA and 0.11(3) nM (n = 1.28) at 50 pM DNA. Site #18 is best described by **PA25** binding in a reverse orientation to a single-base-pair mismatch site, while Site #19 can be explained by a forward double-base-pair mismatch site on 7715-

7703. Site #20 fits the predicted forward triple-base-pair mismatch site on 7727-7715. The determined K_d values at the region shared by these sites were 0.43(8) nM (n = 1.72) at 200 pM DNA and 0.14(9) nM (n = 1.40) at 50 pM DNA.

On the other hand, affinity cleavage patterns for Sites #4, #6-7 and #14 were only observed at high **PA25** concentrations ($\geq$100 nM). Furthermore, Sites #6 and #7 exhibited large Hill coefficients values (n = 3.54) and high K_d values (2.7 $\pm$ 0.6 nM).

A predicted single-base-pair mismatch **PA25** binding site did not produce an affinity cleavage pattern, indicating that this site was not bound by the hairpin polyamide. Thus, recognition of the minor groove by **PA25**, similar to **PA1**, is not simply based on the established recognition rules.

In summary, the equilibrium dissociation constants determined by the Hill equation at a DNA concentration of 200 pM ranged from 0.4 nM to 1.2 nM for binding events that generated affinity cleavage patterns at 25 and 50 nM of **PA25-EDTA**. When the DNA concentration was lowered to 50 pM, these equilibrium dissociation constants decreased to 0.1-0.5 nM.

The high affinity DNase I footprinting sites (2 nM **PA25**) and affinity cleavage patterns (25 and 50 nM **PA25-EDTA**) can be explained with binding sites corresponding to single-, double-, and triple-base-pair mismatches. On the other hand, low affinity sites, DNase I footprints at 5 nM **PA25** and affinity cleavage patterns at 100 nM **PA25-EDTA**, required polyamide binding to sites with double-, triple-, quadruple- and quintuple-base-pair mismatches. In contrast to **PA1**, **PA25** has the ability to tolerate a higher number of base-pair mismatches, probably due to its expected larger DNA binding site. Analogous to the results obtained for **PA1**, a predicted **PA25** single-base-pair mismatch site did not agree with the observed affinity cleavage patterns or produced low affinity cleavage patterns close to the noise level, indicating that these sites were not bound or were weakly bound by the hairpin polyamide.

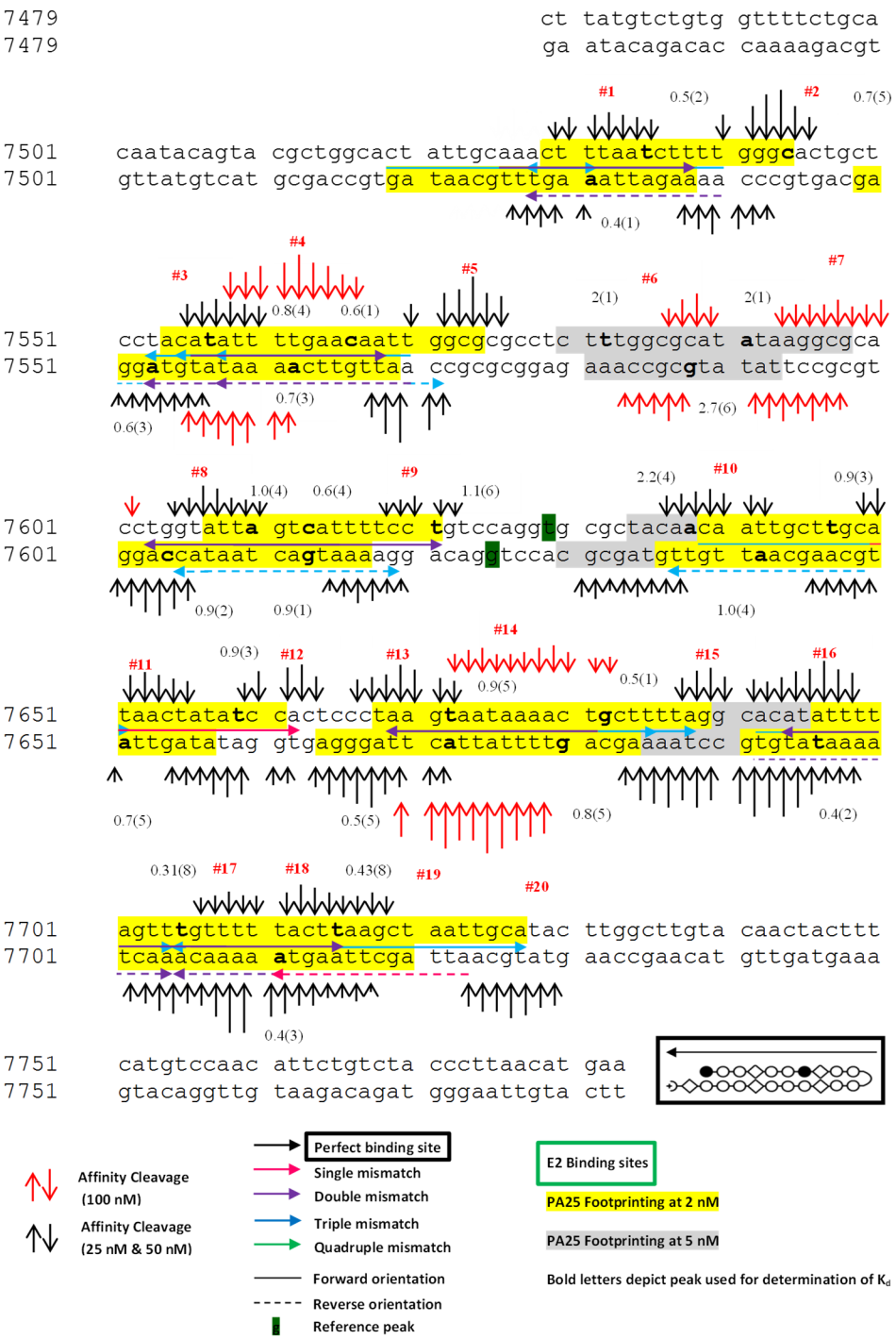


Figure 2.24. PA25 Binding Sites and Hill-calculated Equilibrium Dissociation Constants (200 pM DNA) on the HPV18 Region Corresponding to 7479-7783 bp. The DNA duplex for the 305 bp HPV18 DNA fragment is shown along with **PA25** represented as arrows at its respective binding sites. **PA25** is depicted with either a solid arrow (forward orientation) or a dashed arrow (reverse orientation). This fragment does not harbor any HPV E2 binding sites. Numbers on the left correspond to the nucleotide position in HPV18 genome. Equilibrium dissociation constants calculated using the Hill equation from the FAM and HEX channels of CE at the particular nucleotide (bold) are given. The sample standard deviations are in parentheses. Affinity cleavage patterns at 25-50 nM (black arrows) and 100 nM (red arrows) **PA25-EDTA** are illustrated. At high polyamide concentrations of 100 nM **PA25-EDTA** and above, new binding sites in the 305 bp fragment are populated by the polyamide. Yellow highlighted sequences represent footprinting regions observed at 2 nM **PA25**, whereas gray highlighted sequences represent footprinting regions observed at 5 nM **PA25**.

Table 2.6. PA25 Sites on 305 bp Fragment (200 pM DNA) Corresponding to HPV18 7479-7783 bp.

AC site #	Sequence	Position	Site type	Binding orientation	Integration nt	*Kd* (nM) DNase I	Hill Coeff.
1	GATTA AA**G**TTTGC**A**ATAG TGCCA	7531-7519	Triple	Forward	7531	0.4 ± 0.1	1.91
	TGCAA AC**T**TTAAT**C**TTTT GGGCA	7528-7540	Triple	Forward			
			Double	Reverse			
2	CCCAA AA**G**ATTAA**A**GTTT GCAAT	7538-7526	Double	Forward			
3[a]	GCTCC TA**C**ATATT**T**TGAA CAATT	7553-7565	Triple	Forward	7562	0.7 ± 0.3	2.13
			Double	Reverse			
3[a]	TCCTA CA**T**ATTTT**G**AACA ATTGG	7555-7567	Triple	Forward			
3[a]	TACAT AT**T**TTGAA**C**AATT GGCGC	7558-7570	Triple	Forward			
			Double	Reverse			
5	GCCAA TT**G**TTCAA**A**ATAT GTAGG	7568-7556	Double	Forward			
5	CGCGC CA**A**TTGTT**C**AAAA TATGT	7571-7559	Triple	Reverse			
8	GCACC TG**G**TATTA**G**TCAT TTTCC	7603-7615	Double	Forward	7613	0.9 ± 0.1	2.40
8	ACCTG GT**A**TTAGT**C**ATTT TCCTG	7605-7617	Triple	Reverse			
9	GACAG GA**A**AATGA**C**TAAT ACCAG	7617-7605	Triple	Reverse			
9	TGGAC AG**G**AAAAT**G**ACTA ATACC	7621-7609	Double	Forward			
10	GCTAC AA**C**AATTG**C**TTGC ATAAC	7637-7649	Triple	Reverse	7647	0.9 ± 0.3	2.24
11	TAGTT AT**G**CAAGC**A**ATTG TTGTA	7651-7639	Triple	Forward			
12	GGGAG TG**G**ATATA**G**TTAT GCAAG	7662-7650	Single	Forward	7659	0.9 ± 0.3	2.23
13	TCCCT AA**G**TAATA**A**AACT GCTTT	7669-7681	Double	Forward	7672	0.9 ± 0.5	1.30
15	CCTAA AA**G**CAGTT**T**TATT ACTTA	7685-7673	Triple	Forward	7682	0.5 ± 0.1	2.43
15	GTGCC TA**A**AAGCA**G**TTTT ATTAC	7688-7676	Triple	Forward			
16	GGCAC AT**A**TTTTA**G**TTTG TTTTT	7694-7706	Double	Forward	7696	0.4 ± 0.3	1.84
17	AAACA AA**C**TAAAA**T**ATGT GCCTA	7704-7692	Triple	Forward			
			Double	Reverse			
17	TAGTT TG**T**TTTTA**C**TTAA GCTAA	7705-7717	Triple	Forward	7715	0.43 ± 0.08	1.72
			Double	Reverse			
18	GTTTT TA**C**TTAAG**C**TAAT TGCAT	7711-7723	Single	Reverse			
19	AGCTT AA**G**TAAAA**A**CAAA CTAAA	7715-7703	Double	Forward			
20	AAGTA TG**C**AATTA**G**CTTA AGTAA	7727-7715	Triple	Forward			

The affinity cleavage (AC) site # corresponds to the AC patterns in the map provided in **Figure 2.24**. The **PA25** binding sequences are given in a 5' to 3' direction and include 5 nucleotide (nt) flanking regions. The bold nt highlights the position where the polyamide imidazole/pyrrole pair binds. Position corresponds to the HPV18 genomic nucleotides. The forward binding orientation refers to alignment of the polyamide N-terminus to C-terminus vector with the 5'→3' direction of the DNA, whereas a reverse orientation describes the polyamide N→C vector alignment with the 3'→5' direction of the DNA.

[a] A single binding event cannot be assigned due to significant overlap of the predicted **PA25** recognition sequences.

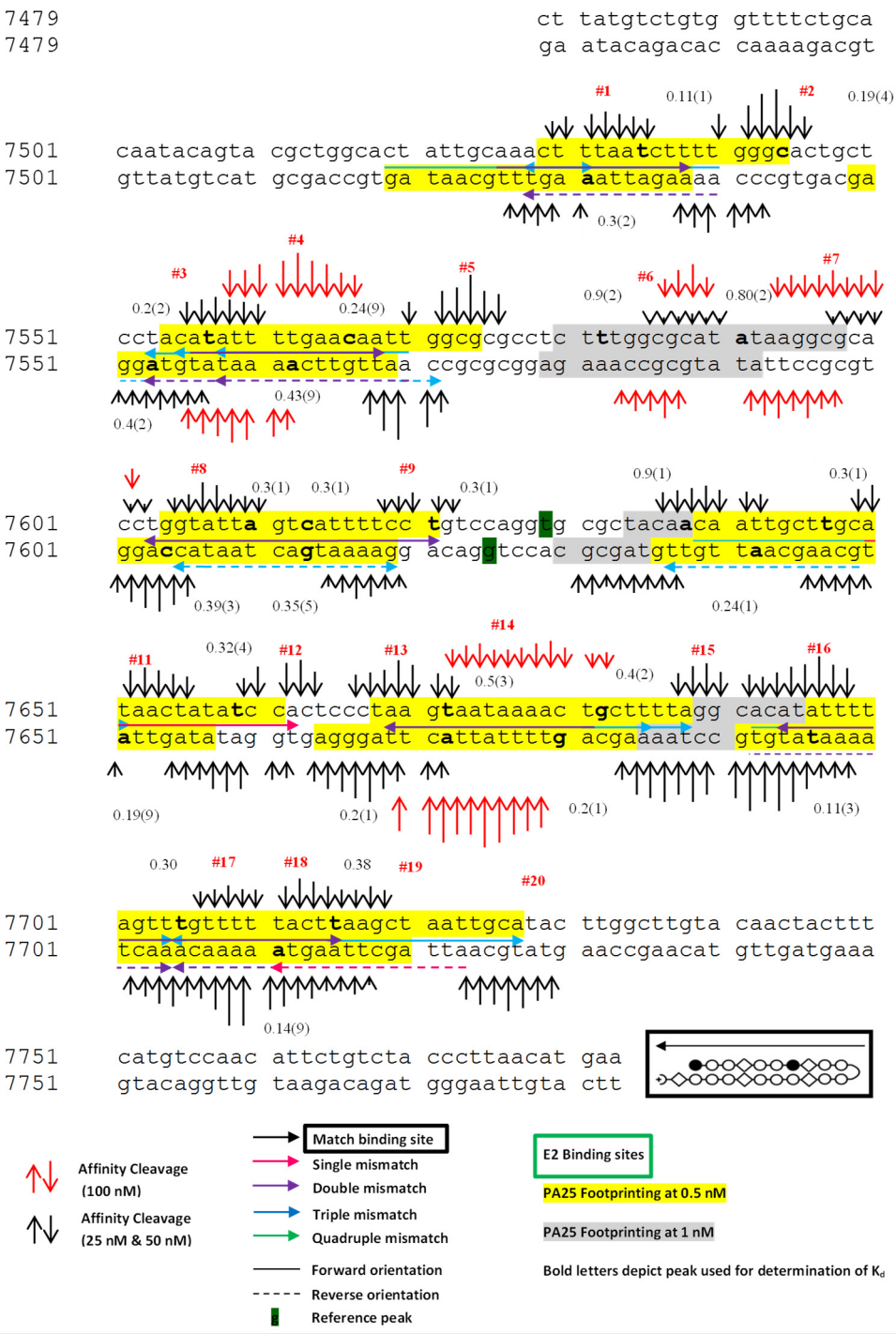


Figure 2.25. PA25 Binding Sites and Hill-calculated Equilibrium Dissociation Constants (50 pM DNA) on the HPV18 Region Corresponding to 7479-7783 bp. The DNA duplex for the 305 bp HPV18 DNA fragment is shown along with **PA25** represented as arrows at its respective binding sites. **PA25** is depicted with either a solid arrow (forward orientation) or a dashed arrow (reverse orientation). This fragment does not harbor any HPV E2 binding sites. Numbers on the left correspond to the nucleotide position in HPV18 genome. Equilibrium dissociation constants calculated using the Hill equation from the FAM and HEX channels of CE at the particular nucleotide (bold) are given. The sample standard deviations are in parentheses. Affinity cleavage patterns at 25-50 nM (black arrows) and 100 nM (red arrows) **PA25-EDTA** are illustrated. At high polyamide concentrations of 100 nM **PA25-EDTA** and above, new binding sites in the 305 bp fragment are populated by the polyamide. Yellow highlighted sequences represent footprinting regions observed at 0.5 nM **PA25**, whereas gray highlighted sequences represent footprinting regions observed at 1 nM **PA25**.

Table 2.7. PA25 Sites on 305 bp Fragment (50 pM DNA) Corresponding to HPV18 7479-7783 bp.

AC site #	Sequence	Position	Site type	Binding orientation	Integration nt	*Kd* (nM) DNase I	Hill Coeff.
1	GATTA AA**G**TTTGC**A**ATAG TGCCA	7531-7519	Triple	Forward	7531	0.3 ± 0.2	1.99
	TGCAA AC**T**TTAAT**C**TTTT GGGCA	7528-7540	Triple	Forward			
			Double	Reverse			
2	CCCAA AA**G**ATTAA**A**GTTT GCAAT	7538-7526	Double	Forward			
3[a]	GCTCC TA**C**ATATT**T**TGAA CAATT	7553-7565	Triple	Forward	7562	0.43 ± 0.09	3.09
			Double	Reverse			
3[a]	TCCTA CA**T**ATTTT**G**AACA ATTGG	7555-7567	Triple	Forward			
3[a]	TACAT AT**T**TTGAA**C**AATT GGCGC	7558-7570	Triple	Forward			
			Double	Reverse			
5	GCCAA TT**G**TTCAA**A**ATAT GTAGG	7568-7556	Double	Forward			
5	CGCGC CA**A**TTGTT**C**AAAA TATGT	7571-7559	Triple	Reverse			
8	GCACC TG**G**TATTA**G**TCAT TTTCC	7603-7615	Double	Forward	7613	0.3 ± 0.1	1.65
8	ACCTG GT**A**TTAGT**C**ATTT TCCTG	7605-7617	Triple	Reverse			
9	GACAG GA**A**AATGA**C**TAAT ACCAG	7617-7605	Triple	Reverse			
9	TGGAC AG**G**AAAAT**G**ACTA ATACC	7621-7609	Double	Forward			
10	GCTAC AA**C**AATTG**C**TTGC ATAAC	7637-7649	Triple	Reverse	7647	0.3 ± 0.1	1.69
11	TAGTT AT**G**CAAGC**A**ATTG TTGTA	7651-7639	Triple	Forward			
12	GGGAG TG**G**ATATA**G**TTAT GCAAG	7662-7650	Single	Forward	7659	0.32 ± 0.04	1.78
13	TCCCT AA**G**TAATA**A**AACT GCTTT	7669-7681	Double	Forward	7672	0.5 ± 0.3	1.89
15	CCTAA AA**G**CAGTT**T**TATT ACTTA	7685-7673	Triple	Forward	7680	0.2 ± 0.1	1.39
15	GTGCC TA**A**AAGCA**G**TTTT ATTAC	7688-7676	Triple	Forward			
16	GGCAC AT**A**TTTTA**G**TTTG TTTTT	7694-7706	Double	Forward	7696	0.11 ± 0.03	1.28
17	AAACA AA**C**TAAAA**T**ATGT GCCTA	7704-7692	Triple	Forward			
			Double	Reverse			
17	TAGTT TG**T**TTTTA**C**TTAA GCTAA	7705-7717	Triple	Forward	7711	0.14 ± 0.09	1.40
			Double	Reverse			
18	GTTTT TA**C**TTAAG**C**TAAT TGCAT	7711-7723	Single	Reverse			
19	AGCTT AA**G**TAAAA**A**CAAA CTAAA	7715-7703	Double	Forward			
20	AAGTA TG**C**AATTA**G**CTTA AGTAA	7727-7715	Triple	Forward			

The affinity cleavage (AC) site # corresponds to the AC patterns in the map provided in **Figure 2.25**. The **PA25** binding sequences are given in a 5' to 3' direction and include 5 nucleotide (nt) flanking regions. The bold nt highlights the position where the polyamide imidazole/pyrrole pair binds. Position corresponds to the HPV18 genomic nucleotides. The forward binding orientation refers to alignment of the polyamide N-terminus to C-terminus vector with the 5'→3' direction of the DNA, whereas a reverse orientation describes the polyamide N→C vector alignment with the 3'→5' direction of the DNA.

[a] A single binding event cannot be assigned due to significant overlap of the predicted **PA25** recognition sequences.

2.4.6.5 PA25 Binding Sites Mapped on HPV18 7647-157 bp

A total of 25 distinct affinity cleavage sites were observed in **PA25-EDTA** affinity cleavage experiments, and these were labeled 1 through 25 based on their position within this fragment. Out of these sites, 16 were observed in the affinity cleavage experiments with 25 and 50 nM **PA25**. The affinity cleavage patterns were obtained for both strands of the binding sites except for Sites #1 and #25 because these are positioned near the beginning/end of the fragment. Unlike **PA1**, there were zero match binding sites for **PA25** on this fragment. But similarly to the results in the 305 fragment, the predicted **PA25** single-base-pair mismatch site at genomic position 7721-7709 did not produce an affinity cleavage pattern (**Figure SI2.11** and **Table SI2.4**).

Refer to **Figures 2.26 and 2.27**, as each observed site will be briefly discussed in this section. Site #1 is best described by **PA25** binding in the forward orientation to the double-base-pair mismatch site at genomic position 7669-7681, while Site #3 can be explained by two different triple-base-pair mismatch binding events in the forward conformations to positions 7685-7673 and 7688-7676. It is not clear at this point which of these sites and conformations are bound by the polyamide; however the former fits the observed affinity cleavage pattern best.

In the case of Site #4, the affinity cleavage pattern and DNase I footprinting results agree with **PA25** binding to a forward double-base-pair mismatch site at 7694-7704. The **PA25** dissociation constants for Site #4 were determined to be 0.4(2) nM with a Hill coefficient (n) of 2.57 at 200 pM DNA and 0.13(2) nM (n = 1.70) at 50 pM DNA. Representative binding isotherms are provided in **Figure SI2.12**.

Site #5 is best described by **PA25** binding in a reverse orientation to a single-base-pair mismatch site (or forward triple-base-pair mismatch) with a K_d value of 1.1(6) nM (n = 2.17) at 200 pM DNA and 0.5(1) nM (n = 2.69) at 50 pM DNA. Whereas Site #6 agrees with **PA25** binding to a forward double-base-pair mismatch site on 7703-7715. The determined K_d values at Site #6 are 0.47(5) nM (n= 2.02) at 200 pM DNA and 0.21(8) nM (n = 1.45) at 50 pM DNA. While Site #7 is in agreement with a forward triple-base-pair mismatch site (7727-7715) with K_d values of 1.1(6) nM (n = 2.17) and 1.2(4) nM (n = 2.32) at 200 pM DNA, and 0.5(1) nM (n = 2.69) at 50 pM DNA. These values overlap with **PA25** binding to Site #5.

Unlike all of the other high affinity cleavage sites, Site #12 generated a diffused low-intensity affinity cleavage pattern that is best explained by **PA25** binding in various conformations to the HPV18 genome. These conformations include reverse and forward single-base-pair mismatch sites, reverse and forward double-base-pair mismatch sites, and a forward triple-base-pair mismatch site. The K_d values determined at 200 pM DNA for Site #12 are: 0.8(3) nM (n = 1.66), 0.7(3) nM (n = 2.69), 0.7(3) nM (n = 2.62), 0.7(3) nM (n = 2.91) and 0.6(3) nM (n = 2.62). On the other hand, the equivalent K_d values determined at 50 pM DNA for Site #12 are: 0.24(9) nM (n = 1.58), 0.27(6) nM (n = 2.09), 0.20(4) nM (n = 1.81), 0.23(5) nM (n = 2.08) and 0.18(4) nM (n = 1.84). Most of these K_d values are shared with Site #13 which is best explained by a forward double-base-pair mismatch and generates an affinity cleavage pattern with higher intensity than the pattern observed for Site #12.

Site #14 is located just upstream of the E2BS-1 (7822-7833). This site is best explained by **PA25** binding in a forward conformation to a triple-base-pair mismatch site at 7817-7805. The K_d values determined at 200 pM DNA for Site #14 are: 0.9(2) nM (n = 2.10), 0.8(4) nM (n = 2.96) and 0.9(4) nM (n = 2.78). While the equivalent K_d values determined at 50 pM DNA are: 0.33(6) nM (n = 2.83), 0.30(5) nM (n = 2.02) and 0.28(9) nM (n = 2.17).

Sites #16, #17, #18, #19 and #20 bind to the E1-binding region of HPV18.[78] Site #16 can be explained by **PA25** binding either in the forward conformation to two adjacent triple-base-pair mismatch sites (7849-4 and 7850-5), or in the reverse conformations to the same sequences to generate double-base-pair mismatch sites. Site #17 is best described by **PA25** binding to a forward double-base-pair mismatch site (7853-8), but two forward triple-base-pair mismatch sites and their equivalent reverse double-base-pair mismatch sites cannot be ruled out (7849-4 and 7850-5). The K_d values shared by Sites

#16 and #17 at 200 pM DNA are 0.7(4) nM (n = 2.44), 0.6(3) nM (n = 2.45) and 0.7(3) nM (n = 2.39). While these K_d values are 0.20(4) nM (n = 2.09), 0.18(4) nM (n = 2.38) and 0.25(3) nM (n = 2.20) at 50 pM DNA. Site #18 can be explained by a forward double-base-pair mismatch site (7855-10) or two forward triple-base-pair mismatch sites (1-13 and 12-24). Site #19 can be described by **PA25** binding to a forward single-base-pair mismatch site (21-33), two forward triple-base-pair mismatch sites (10-22 and 11-23) and a reverse double-base-pair mismatch site (10-22). Site #20 has four possibilities: a reverse double-base-pair mismatch (19-31) and three different forward triple-base-pair mismatch sites (19-31, 21-33 and 22-34). The K_d values shared by Sites #18 and #19 at 200 pM DNA are 0.6(2) nM (n = 2.15) and 0.6(2) nM (n = 2.24); while at 50 pM these K_d values are 0.21(5) nM (n = 2.34) and 0.19(7) nM (n = 2.05). The K_d values shared by Sites #19 and #20 at 200 pM DNA are 0.7(3) nM (n = 2.43), 0.5(2) nM (n = 2.19) and 0.6(2) nM (n = 2.40), while at 50 pM these K_d values are 0.24(6) nM (n = 1.47), 0.13(3) nM (n = 1.84) and 0.17(4) nM (n = 1.54). Site #20 is within a DNA helical turn from E2 Binding Site 2 (E2BS-2 42-53).

Another two binding events, Sites #23 and #24, were observed in close proximity to E2 Binding Site 3 (E2BS-3 7822-7833). Site #23 is best described by **PA25** binding in the forward orientation to a double-base-pair mismatch at genomic position 72-84, while Site #23 corresponds to **PA25** recognizing either a forward triple-base-pair mismatch site and/or a reverse double-base-pair mismatch site (70-82). The K_d values shared by Sites #23 and #24 at 200 pM DNA are 0.7(3) nM (n = 1.86) and 0.5(3) nM (n = 2.26), while at 50 pM these K_d values are 0.22(7) nM (n = 1.84) and 0.13(4) nM (n = 1.49).

On the other hand, the affinity cleavage patterns for Sites #2, #8-11, #15, #21, #22 and #25 were only observed at high **PA25** concentrations ($\geq$100 nM).

Some single-base-pair mismatch sites bound by **PA25** generated affinity cleavage with low-intensities, relative to adjacent sites with higher base-pair mismatches (Sites #12 and #19). In agreement with the results observed for the 305 bp fragment, the predicted **PA25** single-base-pair mismatch at genomic position 7721-7709 did not generate an affinity cleavage pattern. Thus, recognition of the minor groove by **PA25**, similar to **PA1**, is not simply based on the established recognition rules.

Binding of **PA25** to the HPV18 origin of replication may be of importance to the robust antiviral action of **PA25** against HPV18 in monolayer keratinocytes. The 368 bp fragment harbors the viral origin of replication, the E2 binding sites (E2BS-1 7822-7833; E2BS-2 42-53; E2BS-3 58-69) and the E1 binding site (E1BS 3-18). Because **PA25** and **PA1** bind avidly to the E1 binding region and in close proximity to the E2 binding sites, these DNA minor groove binder may preclude binding of the essential viral proteins E1 and E2 to the origin of replication by steric hindrance or by modifying the local DNA structure. Indeed, the equilibrium dissociation constant for HPV18 E2dbd DNA binding domain is 4.5 nM.[43] Because our hairpin polyamides exhibit affinities to this region of the viral genome comparable to the vital HPV18 E2 protein, it is plausible that they can function has inhibitors of viral replication, gene expression and/or proper episomal segregation during cell division of infected cells. The higher antiviral activity of **PA25** against HPV18 as compared to **PA1** may be due to its lower dissociation constant, as well as a more extensive coverage of this genomic region important for viral replication, particularly in the E1 binding region.

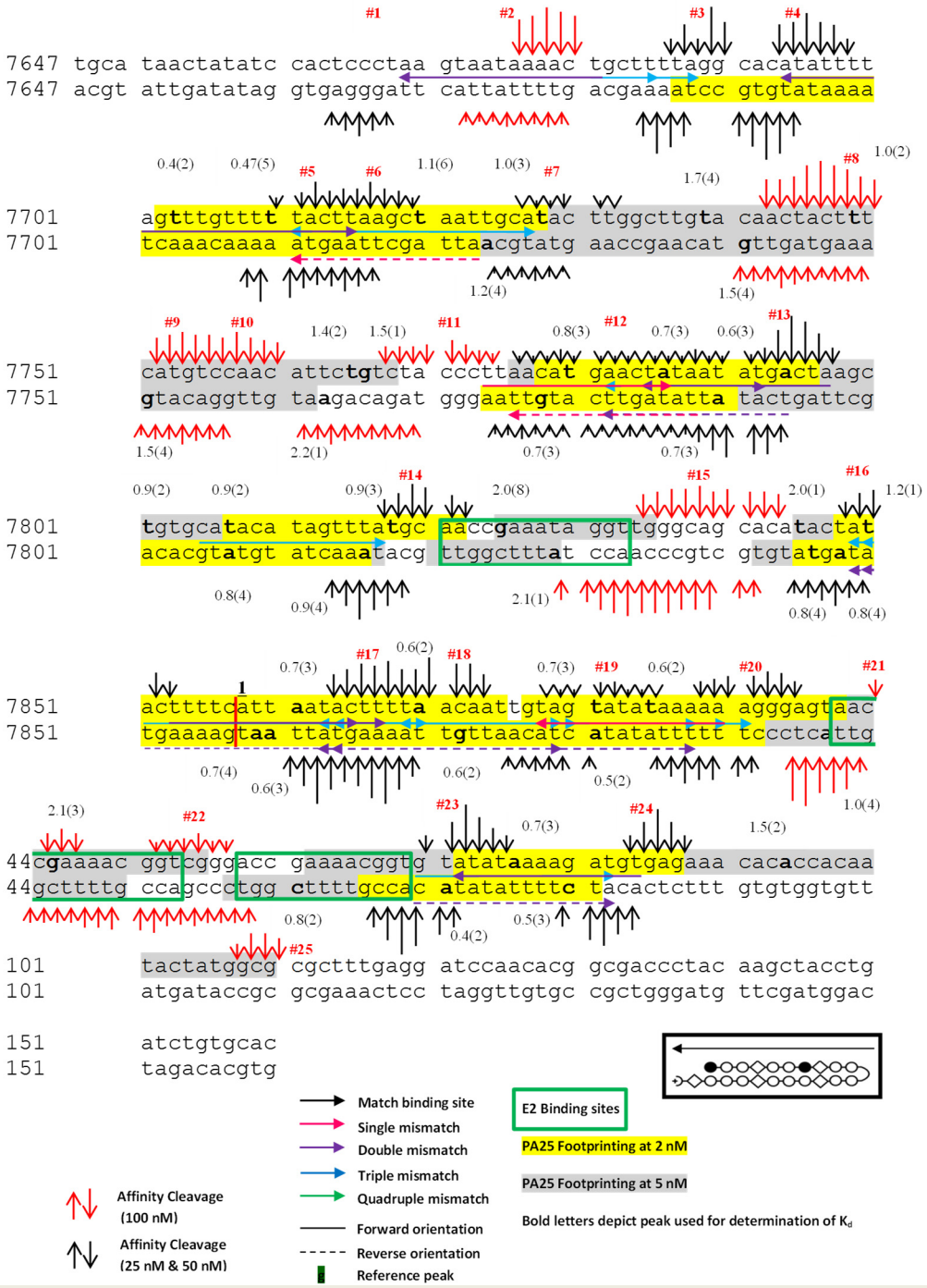


Figure 2.26. PA25 Binding Sites and Hill-calculated Equilibrium Dissociation Constants (200 pM DNA) on the HPV18 Region Corresponding to 7647-157 bp. The DNA duplex for the 368 bp HPV18 DNA fragment is shown along with **PA25** represented as arrows at its respective binding sites. **PA25** is depicted with either a solid arrow (forward orientation) or a dashed arrow (reverse orientation). HPV E2 binding sites are depicted in green boxes. HPV E1 binds in the region (HPV18 3-19 bp) flanked by the first and second E2 binding sites. The origin of replication is marked with 1. Numbers on the left correspond to the nucleotide position in HPV18 genome. Equilibrium dissociation constants calculated using the Hill equation from the FAM and HEX channels of CE at the particular nucleotide (bold) are given. The sample standard deviations are in parentheses. Affinity cleavage patterns at 25-50 nM (black arrows) and 100 nM (red arrows) **PA25-EDTA** are illustrated. At high polyamide concentrations of 100 nM **PA25-EDTA** and above, new binding sites in the 368 bp fragment are populated by the polyamide. Yellow highlighted sequences represent footprinting regions observed at 2 nM **PA25**; whereas, gray highlighted sequences represent footprinting regions observed at 5 nM **PA25**.

Table 2.8. PA25 Sites on 368 bp Fragment (200 pM DNA) Corresponding to HPV18 7647-157 bp.

AC site #	Sequence	Position	Site type	Binding orientation	Integration nt	*Kd* (nM) DNase I	Hill Coeff.
1	TCCCT AA**G**TAATA**A**AACT GCTTT	7669-7681	Double	Forward	*ND*	*ND*	*ND*
3	CCTAA AA**G**CAGTT**T**TATT ACTTA	7685-7673	Triple	Forward			
3	GTGCC TA**A**AAGCA**G**TTTT ATTAC	7688-7676	Triple	Forward			
4	GGCAC AT**A**TTTTA**G**TTTG TTTTT	7694-7706	Double	Forward	7703	0.4 ± 0.2	2.57
6	AGCTT AA**G**TAAAA**A**CAAA CTAAA	7715-7703	Double	Forward			
5	GTTTT TA**C**TTAAG**C**TAAT TGCAT	7711-7723	Triple	Forward	7720	1.1 ± 0.6	2.17
			Single	Reverse			
7	AAGTA TG**C**AATTA**G**CTTA AGTAA	7727-7715	Triple	Forward			
12 [a]	TATTA TA**G**TTCAT**G**TTAA GGGTA	7786-7774	Single	Forward	7790	0.7 ± 0.3	2.91
12 [a]	CCCTT AA**C**ATGAA**C**TATA ATATG	7776-7788	Single	Reverse			
12 [a]	ACATG AA**C**TATAA**T**ATGA CTAAG	7782-7794	Triple	Forward			
			Double	Reverse			
12 [a]	TGAAC TA**T**AATAT**G**ACTA AGCTG	7785-7797	Double	Forward			
13	TAGTC AT**A**TTATA**G**TTCA TGTTA	7792-7780	Double	Forward			
14	TTGCA TA**A**ACTAT**G**TATG CACAG	7817-7805	Triple	Forward	7807	0.9 ± 0.2	2.10
16	ATACT AT**A**CTTTT**C**ATTA ATACT	7849-4	Triple	Forward	4	0.7 ± 0.3	2.39
			Double	Reverse			
16	TACTA TA**C**TTTTC**A**TTAA TACTT	7850-5	Triple	Forward			
			Double	Reverse			
17	TAAAA GT**A**TTAAT**G**AAAA GTATA	8-7853	Double	Forward			
18	GTTAA AA**G**TATTA**A**TGAA AAGTA	10-7855	Double	Forward			
18	ATTGT TA**A**AAGTA**T**TAAT GAAAA	13-1	Triple	Forward			
17	ATTAA TA**C**TTTTA**A**CAAT TGTAG	6-18	Triple	Forward	15	0.6 ± 0.2	2.24
			Double	Reverse			
17	TTAAT AC**T**TTTAA**C**AATT GTAGT	7-19	Triple	Forward			
			Double	Reverse			
18	ACTTT TA**A**CAATT**G**TAGT ATATA	12-24	Triple	Forward			
19	TATAC TA**C**AATTG**T**TAAA AGTAT	22-10	Triple	Forward			
			Double	Reverse			
19	ATATA CT**A**CAATT**G**TTAA AAGTA	23-11	Triple	Forward			
19	AATTG TA**G**TATAT**A**AAAA AGGGA	21-33	Single	Forward	28	0.6 ± 0.2	2.40
20 [a]	CCTTT TT**T**ATATA**C**TACA ATTGT	31-19	Triple	Forward			
			Double	Reverse			
20 [a]	TCCCT TT**T**TTATA**T**ACTA CAATT	33-21	Triple	Forward			
20 [a]	CTCCC TT**T**TTTAT**A**TACT ACAAT	34-22	Triple	Forward			
23	GGTGT AT**A**TAAAA**G**ATGT GAGAA	72-84	Double	Forward	80	0.5 ± 0.3	2.26
24	CTCAC AT**C**TTTTA**T**ATAC ACCGT	82-70	Triple	Forward			
			Double	Reverse			

The affinity cleavage (AC) site # corresponds to the AC patterns in the map provided in **Figure 2.26**. The **PA25** binding sequences are given in a 5' to 3' direction and include 5 nucleotide (nt) flanking regions. The bold nt highlights the position where the polyamide imidazole/pyrrole pair binds. Position corresponds to the HPV18 genomic nucleotides. The forward binding orientation refers to alignment of the polyamide N-terminus to C-terminus vector with the 5'→3' direction of the DNA, whereas a reverse orientation describes the polyamide N→C vector alignment with the 3'→5' direction of the DNA.

ND, Not determined. Sites 1 and 3 are found in the beginning of the fragment.

[a] A single binding event cannot be assigned due to significant overlap of the predicted **PA25** recognition sequences.

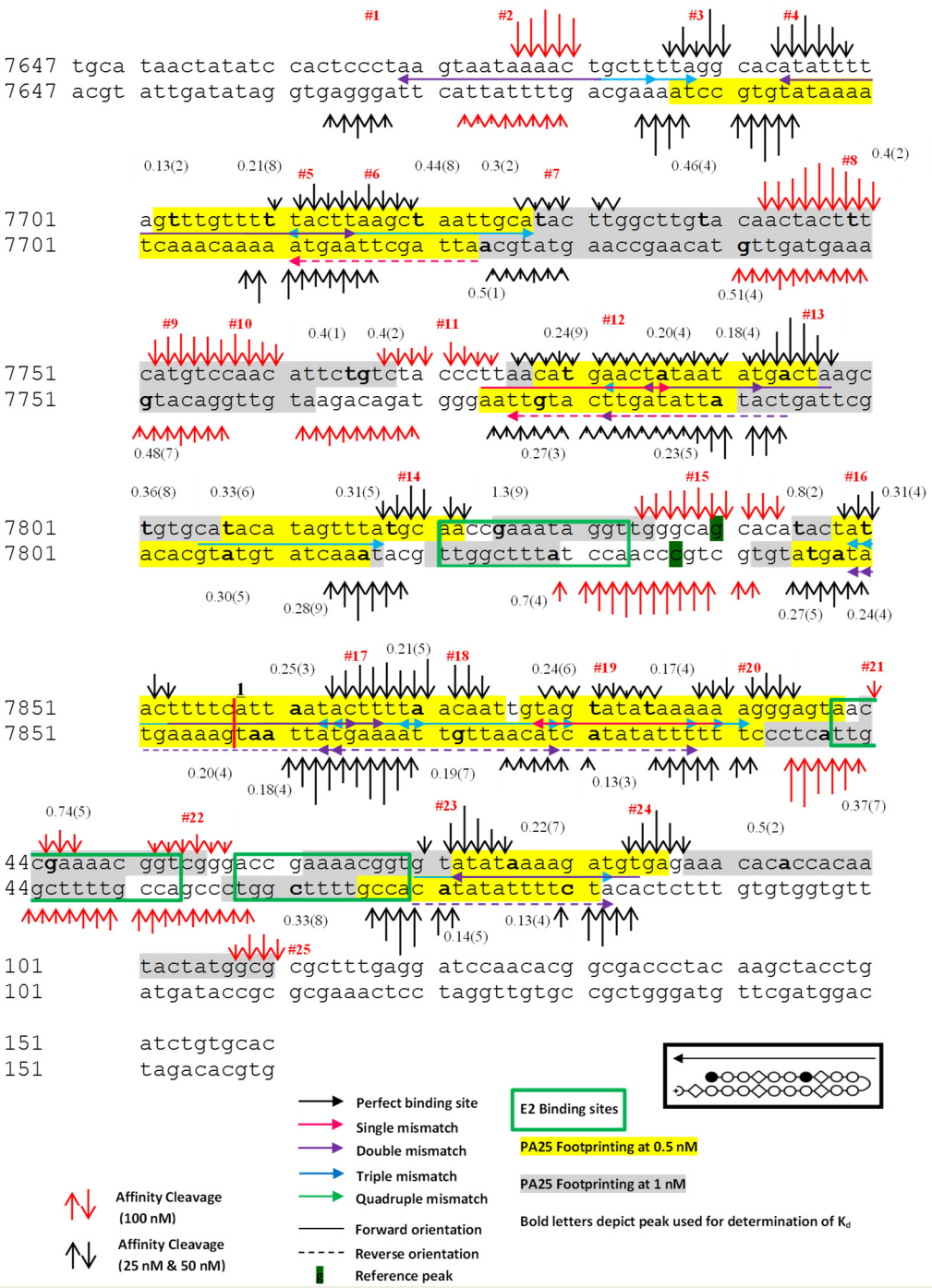


Figure 2.27. PA25 Binding Sites and Hill-calculated Equilibrium Dissociation Constants (50 pM DNA) on the HPV18 Region Corresponding to 7647-157 bp. The DNA duplex for the 368 bp HPV18 DNA fragment is shown along with **PA25** represented as arrows at its respective binding sites. **PA25** is depicted with either a solid arrow (forward orientation) or a dashed arrow (reverse orientation). HPV E2 binding sites are depicted in green boxes. HPV E1 binds in the region (HPV18 3-19 bp) flanked by the first and second E2 binding sites. The origin of replication is marked with 1. Numbers on the left correspond to the nucleotide position in HPV18 genome. Equilibrium dissociation constants calculated using the Hill equation from the FAM and HEX channels of CE at the particular nucleotide (bold) are given. The sample standard deviations are in parentheses. Affinity cleavage patterns at 25-50 nM (black arrows) and 100 nM (red arrows) **PA25-EDTA** are illustrated. At high polyamide concentrations of 100 nM **PA25-EDTA** and above, new binding sites in the 368 bp fragment are populated by the polyamide. Yellow highlighted sequences represent footprinting regions observed at 0.5 nM **PA25**; whereas, gray highlighted sequences represent footprinting regions observed at 1 nM **PA25**.

Table 2.9. PA25 Sites on 368 bp Fragment (50 pM DNA) Corresponding to HPV18 7647-157 bp.

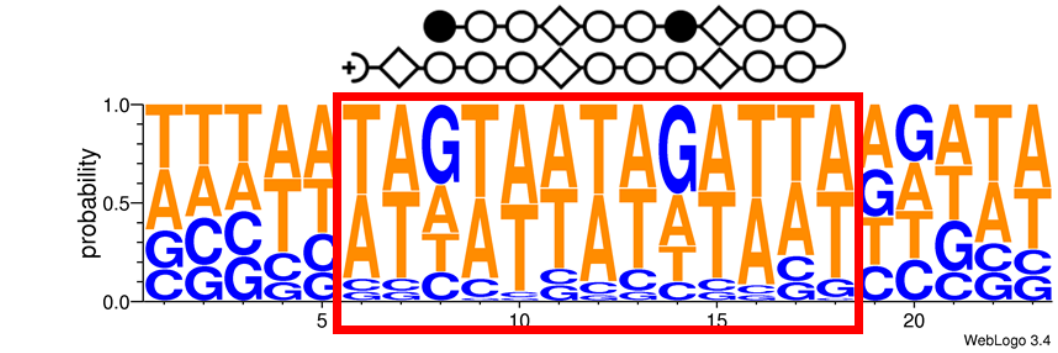
AC site #	Sequence	Position	Site type	Binding orientation	Integration nt	*Kd* (nM) DNase I	Hill Coeff.
1	TCCCT AA**G**TAATA**A**AACT GCTTT	7669-7681	Double	Forward	*ND*	*ND*	*ND*
3	CCTAA AA**G**CAGTT**T**TATT ACTTA	7685-7673	Triple	Forward			
3	GTGCC TA**A**AAGCA**G**TTTT ATTAC	7688-7676	Triple	Forward			
4	GGCAC AT**A**TTTTA**G**TTTG TTTTT	7694-7706	Double	Forward	7703	0.13 ± 0.02	1.70
6	AGCTT AA**G**TAAAA**A**CAAA CTAAA	7715-7703	Double	Forward			
5	GTTTT TA**C**TTAAG**C**TAAT TGCAT	7711-7723	Triple	Forward	7724	0.5 ± 0.1	2.69
			Single	Reverse			
7	AAGTA TG**C**AATTA**G**CTTA AGTAA	7727-7715	Triple	Forward			
12 [a]	TATTA TA**G**TTCAT**G**TTAA GGGTA	7786-7774	Single	Forward	7790	0.23 ± 0.05	2.08
12 [a]	CCCTT AA**C**ATGAA**C**TATA ATATG	7776-7788	Single	Reverse			
12 [a]	ACATG AA**C**TATAA**T**ATGA CTAAG	7782-7794	Triple	Forward			
			Double	Reverse			
12 [a]	TGAAC TA**T**AATAT**G**ACTA AGCTG	7785-7797	Double	Forward			
13	TAGTC AT**A**TTATA**G**TTCA TGTTA	7792-7780	Double	Forward			
14	TTGCA TA**A**ACTAT**G**TATG CACAG	7817-7805	Triple	Forward	7807	0.33 ± 0.06	2.83
16	ATACT AT**A**CTTTT**C**ATTA ATACT	7849-4	Triple	Forward	4	0.25 ± 0.03	2.20
			Double	Reverse			
16	TACTA TA**C**TTTTC**A**TTAA TACTT	7850-5	Triple	Forward			
			Double	Reverse			
17	TAAAA GT**A**TTAAT**G**AAAA GTATA	8-7853	Double	Forward			
18	GTTAA AA**G**TATTA**A**TGAA AAGTA	10-7855	Double	Forward			
18	ATTGT TA**A**AAGTA**T**TAAT GAAAA	13-1	Triple	Forward			
17	ATTAA TA**C**TTTTA**A**CAAT TGTAG	6-18	Triple	Forward	15	0.19 ± 0.07	2.05
			Double	Reverse			
17	TTAAT AC**T**TTTAA**C**AATT GTAGT	7-19	Triple	Forward			
			Double	Reverse			
18	ACTTT TA**A**CAATT**G**TAGT ATATA	12-24	Triple	Forward			
19	TATAC TA**C**AATTG**T**TAAA AGTAT	22-10	Triple	Forward			
			Double	Reverse			
19	ATATA CT**A**CAATT**G**TTAA AAGTA	23-11	Triple	Forward			
19	AATTG TA**G**TATAT**A**AAAA AGGGA	21-33	Single	Forward	28	0.17 ± 0.04	1.54
20 [a]	CCTTT TT**T**ATATA**C**TACA ATTGT	31-19	Triple	Forward			
			Double	Reverse			
20 [a]	TCCCT TT**T**TTATA**T**ACTA CAATT	33-21	Triple	Forward			
20 [a]	CTCCC TT**T**TTTAT**A**TACT ACAAT	34-22	Triple	Forward			
23	GGTGT AT**A**TAAAA**G**ATGT GAGAA	72-84	Double	Forward	80	0.13 ± 0.04	1.49
24	CTCAC AT**C**TTTTA**T**ATAC ACCGT	82-70	Triple	Forward			
			Double	Reverse			

The affinity cleavage (AC) site # corresponds to the AC patterns in the map provided in **Figure 2.27**. The **PA25** binding sequences are given in a 5' to 3' direction and include 5 nucleotide (nt) flanking regions. The bold nt highlights the position where the polyamide imidazole/pyrrole pair binds. Position corresponds to the HPV18 genomic nucleotides. The forward binding orientation refers to alignment of the polyamide N-terminus to C-terminus vector with the 5'→3' direction of the DNA, whereas a reverse orientation describes the polyamide N→C vector alignment with the 3'→5' direction of the DNA.

ND, Not determined. Sites 1 and 3 are found in the beginning of the fragment.
[a] A single binding event cannot be assigned due to significant overlap of the predicted **PA25** recognition sequences.

2.4.6.6 PA25 WebLogo 3.3 Motif Analysis

WebLogo 3.3 was also used to create a graphical representation of the consensus DNA binding sequence for the observed binding sequences of **PA25** in the studied HPV18-derived fragments. The WebLogo analysis (**Figure 2.28**) shows that **PA25**, like **PA1**, binds to A/T-rich sequences (sample size of n = 61) with a slight preference of the imidazole units of the hairpin polyamide recognizing the guanine base in duplex DNA. Furthermore, base-pair mismatches are allowed in all the positions in the polyamide binding region.

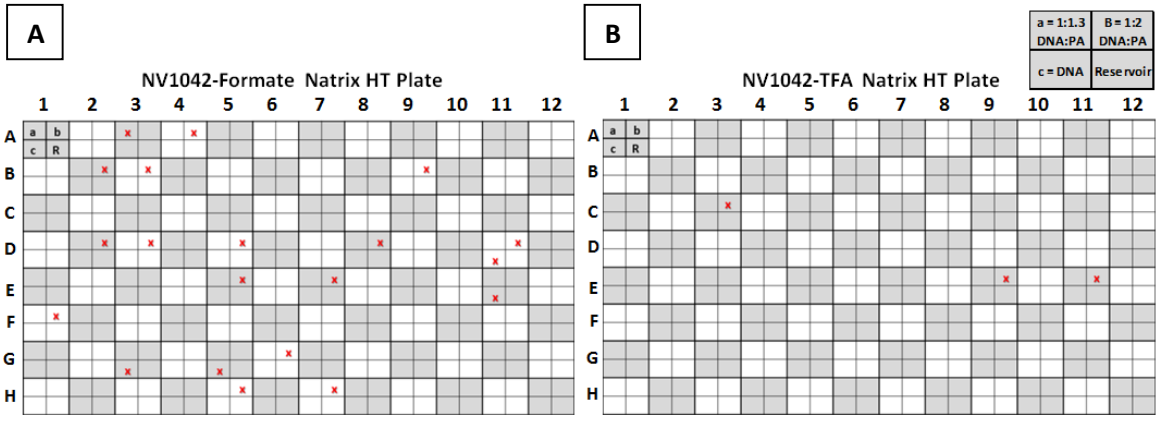


CAAAA ATGAATTCGATTA ACGTA	TATTA TAGTTCATGTTAA GGGTA	GGGAA TTGTACTTGATAT TATAC
AATTG TAGTATATAAAAA AGGGA	TCCCT AAGTAATAAAACT GCTTT	GGCAC ATATTTTAGTTTG TTTTT
AGCTT AAGTAAAAACAAA CTAAA	TGTAC TTGATATTATACT GATTC	TGAAC TATAATATGACTA AGCTG
TAGTC ATATTATAGTTCA TGTTA	ACCCG TCGTGTATGATAT GAAAA	TATGA TATGAAAAGTAAT TATGA
ATGAT ATGAAAAGTAATT ATGAA	TAATT ATGAAAATTGTTA ACATC	AATTA TGAAAATTGTTAA CATCA
TAAAA GTATTAATGAAAA GTATA	GTTAA AAGTATTAATGAA AAGTA	TTAAC ATCATATATTTTT TCCCT
ATATG ATGTTAACAATTT TCATA	GGAAA AAATATATGATGT TAACT	GGTGT ATATAAAAGATGT GAGAA
GAGTG TAGAAAATATATG TGGCA	CCTAA AAGCAGTTTTATT ACTTA	GTGCC TAAAAGCAGTTTT ATTAC
GTTTT TACTTAAGCTAAT TGCAT	AAGTA TGCAATTAGCTTA AGTAA	ACATG AACTATAATATGA CTAAG
ATAAT ATGACTAAGCTGT GCATA	TTGCA TAAACTATGTATG CACAG	ATACT ATACTTTTCATTA ATACT
TACTA TACTTTTCATTAA TACTT	ATTAA TACTTTTAACAAT TGTAG	TTAAT ACTTTTAACAATT GTAGT
ACTTT TAACAATTGTAGT ATATA	ATTGT TAAAAGTATTAAT GAAAA	TATAC TACAATTGTTAAA AGTAT
ATATA CTACAATTGTTAA AAGTA	CCTTT TTTATATACTACA ATTGT	TCCCT TTTTTATATACTA CAATT
CTCCC TTTTTTATATACT ACAAT	AAAGG GAGTAACCGAAAA CGGTC	CTCAC ATCTTTTATATAC ACCGT
CCCAA AAGATTAAAGTTT GCAAT	ACGTT TGAAATTAGAAAA CCCGT	CGAGG ATGTATAAAACTT GTTAA
GCCAA TTGTTCAAAATAT GTAGG	ATGTA TAAAACTTGTTAA CCGCG	ACGTA TTGATATAGGTGA GGGAT
TTTGT TTGATTTTATACA CGGAT	ATCAA ACAAAAATGAATT CGATT	GATTA AAGTTTGCAATAG TGCCA
TGCAA ACTTTAATCTTTT GGGCA	GCTCC TACATATTTTGAA CAATT	TCCTA CATATTTTGAACA ATTGG
TACAT ATTTTGAACAATT GGCGC	GCGCG GTTAACAAGTTTT ATACA	TGGAC CATAATCAGTAAA AGGAC
CTGTC CTTTTACTGATTA TGGTC	CGATG TTGTTAACGAACG TATTG	TAGTT ATGCAAGCAATTG TTGTA
CCTAA AAGCAGTTTTATT ACTTA		Red = polyA tracts

Figure 2.28. WebLogo Plot for PA25 Binding Sites Observed on the HPV18 Regions Corresponding to 7479-7783 bp and 7647-157 bp. The relative probability of a given nucleotide to be found in a particular position is represented by the height of the letter (A = Adenine, T = Thymine, G = Guanine and C = Cytosine). Red box encloses the **PA25** binding site. Input sequences including the flanking DNA sequences (5 bp) are given below the plot. No obvious trend is observed for the presence or relative position of polyA tracts in or immediately adjacent to the **PA1** binding site.

2.4.7 X-ray Crystallography Screening Matrix of Oligonucleotide with PA25

The Natrix HT matrix plate using the formate salt of **PA25** resulted in many wells with crystal growth, whereas the same conditions using the TFA salt of **PA25** was limited to 3 wells. Conversely, the matrix generated from literature conditions yielded mostly aggregates with no observable crystals. The loading scheme and screening results are shown in **Figure 2.29**.



Well #	Salt	Buffer	Precipitant	Additive
A3	0.1 M Magnesium acetate tetrahydrate	0.05 M MES monohydrate pH 5.6	20% v/v (+/-)-2-Methyl-2,4-pentanediol	
A4	0.2 M Potassium chloride, 0.01 M Magnesium sulfate heptahydrate	0.05 M MES monohydrate pH 5.6	10% v/v Polyethylene glycol 400	
B2	0.1 M Potassium chloride, 0.025 M Magnesium chloride hexahydrate	0.05 M Sodium cacodylate trihydrate pH 6.0	15% v/v 2-Propanol	
B3	0.04 M Magnesium chloride hexahydrate	0.05 M Sodium cacodylate trihydrate pH 6.0	5% v/v (+/-)-2-Methyl-2,4-pentanediol	
B9	0.2 M Potassium chloride, 0.005 M Magnesium chloride hexahydrate	0.05 M Sodium cacodylate trihydrate pH 6.5	0.9 M 1,6-Hexanediol	
D2	0.2 M Ammonium acetate, 0.15 M Magnesium acetate tetrahydrate	0.05 M HEPES Sodium pH 7.0	5% w/v Polyethylene glycol 4,000	
D3	0.1 M Ammonium acetate, 0.2 M Magnesium chloride hexahydrate	0.05 M HEPES Sodium pH 7.0	5% w/v Polyethylene glycol 8,000	
D5	0.1 M Potassium chloride, 0.015 M Magnesium chloride hexahydrate	0.05 M TRIS hydrochloride pH 7.5	10% v/v Polyethylene glycol monomethyl ether 550	
D8	0.2 M Potassium chloride, 0.05 M Magnesium chloride hexahydrate	0.05 M TRIS hydrochloride pH 7.5	10% w/v Polyethylene glycol 4,000	
D9	0.025 M Magnesium sulfate hydrate	0.05 M TRIS hydrochloride pH 8.5	1.8 M Ammonium sulfate	
D11	0.1 M Potassium chloride, 0.01 M Magnesium chloride hexahydrate	0.05 M TRIS hydrochloride pH 8.5	30% v/v Polyethylene glycol 400	
E5	0.002 M Calcium chloride dihydrate	0.05 M Sodium cacodylate trihydrate pH 6.0	1.8 M Ammonium sulfate	0.0005 M Spermine
E7	0.1 M Calcium chloride	0.05 M Sodium cacodylate trihydrate pH 6.0	10% w/v Polyethylene glycol 4,000	0.0005 M Spermine
E11	0.075 M Sodium chloride, 0.002 M Calcium chloride dihydrate	0.05 M Sodium cacodylate trihydrate pH 6.0	30% w/v 1,6-Hexanediol	0.0005 M Spermine
F1		0.05 M Sodium cacodylate trihydrate pH 6.0	30% v/v (+/-)-2-Methyl-2,4-pentanediol	
G3	0.02 M Magnesium chloride hexahydrate	0.05 M Sodium cacodylate trihydrate pH 7.0	15% v/v 2-Propanol	
G5	0.01 M Magnesium chloride hexahydrate, 0.002 M Barium chloride dihydrate	0.05 M MOPS pH 7.0	30% v/v 1,4-Dioxane	
G6	0.001 M Magnesium chloride hexahydrate, 0.002 M Calcium chloride dihydrate	0.05 M MOPS pH 7.0	15% v/v (+/-)-2-Methyl-2,4-pentanediol	
H5	0.08 M Potassium chloride	0.04 M Sodium cacodylate trihydrate pH 7.0	60% v/v (+/-)-2-Methyl-2,4-pentanediol	0.012 M Spermine tetrahydrochloride
H7	0.02 M Magnesium chloride hexahydrate	0.05 M PIPES pH 7.5	4% w/v Polyethylene glycol 8,000	0.001 M Spermine

Figure 2.29. Preliminary X-ray Crystallization Results and Reagent Formulations. A crystallization block consists of three sitting drop wells and a reservoir. The inset in the top right describes the loading scheme for the crystallization experiments; sitting drop well **a** was loaded with 1:1.3 DNA:PA, **b** with 1:2 DNA:PA and **c** with DNA alone. A red X indicates crystal growth in the particular sitting drop well for (**A**) **PA25**-formate and (**B**) **PA25**-TFA Natrix HT matrix screens. The screening conditions for those that produced crystals are given in the bottom table.

2.4.8 Circular Dichroism (CD) Titration Experiments

Figure 2.30 shows the CD spectra for the titration of CHC2050-1 DNA duplex (4.0 μM) with increasing concentrations of **PA1** (0-9.6 μM). The CHC2050-1 duplex corresponds to a **PA1** binding site (Site #14 in **Figure 2.22**) found in the HPV18 368 bp fragment. The CD spectra of the DNA peaks were characteristic of B-form DNA with a negative peak at around 250 nm and a positive peak at around 260-280 nm. Binding of **PA1** to the DNA duplex led to a strong induced CD signal at ~332 nm and ~302 nm, while **PA1** alone does not generate a CD signal as it is not optically active (achiral) (**Figure 2.30A** and **B**). Because the induced CD signal (~332 nm) is outside the CD spectrum of DNA, ligand binding to DNA can be monitored by this technique. The generation of an induced CD signal above 300 nm is consistent with **PA1** binding to the minor groove of DNA. Furthermore, the presence of multiple isodichroic points at approximately 250, 270 and 315 nm suggests a single binding mechanism.[80,81] A titration curve was generated by plotting the induced CD (mdeg) at the λ_{MAX} (~332 nm) versus the **PA1**:DNA duplex ratio (**Figure 2.30C**). **PA1** saturation of the 16-mer oligonucleotides was not reached until about 2:1 **PA1**:DNA ratio, suggesting that two molecules of the hairpin polyamide bind per DNA molecule.

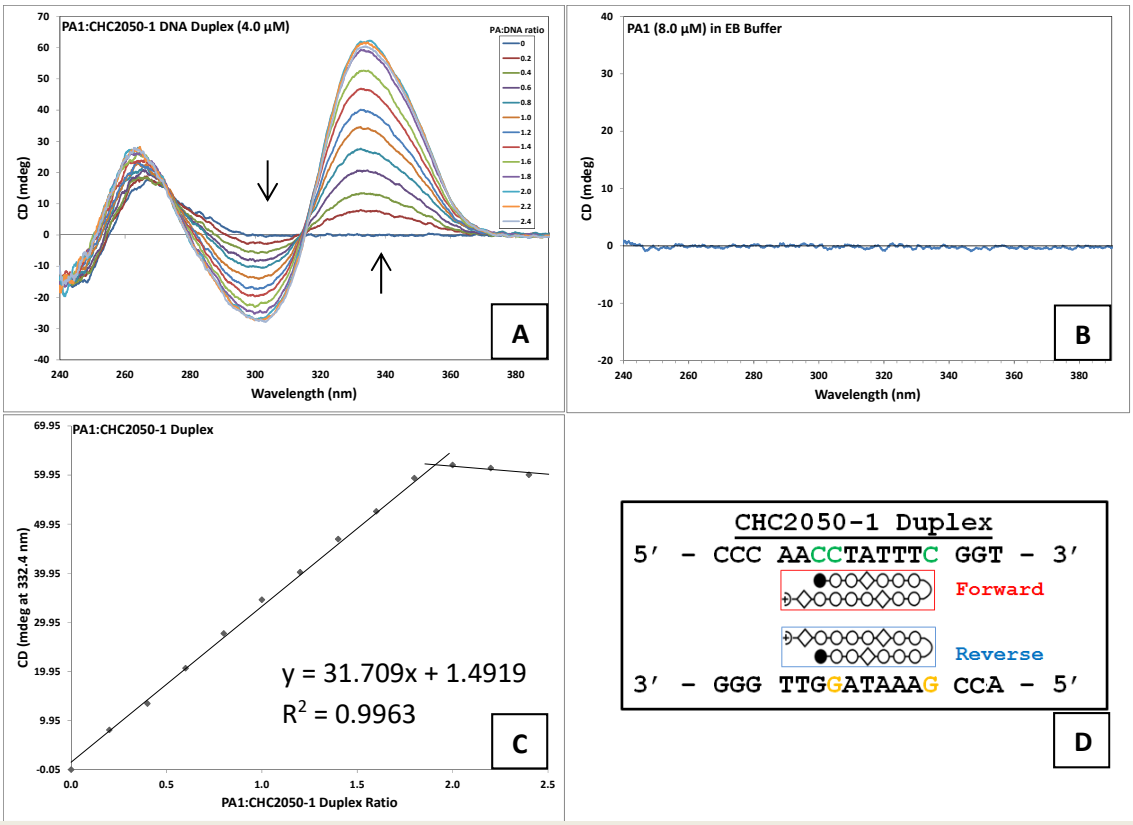


Figure 2.30. Circular Dichroism Spectra for the Titration of CHC2050-1 with PA1. (**A**) CD spectra corresponding to the titration of CHC2050-1 DNA duplex with **PA1**. Free DNA corresponds to the CD spectra with no induced CD signal at 332 nm (navy blue line). PA:DNA ratios ranged from 0.2 to 2.4. Black arrows indicate the induced CD band used to generate the titrations plots. (**B**) **PA1** is an achiral molecule and thus in the absence of DNA, it is CD inactive. (**C**) Titration curve is shown below their CD spectra and was generated by plotting the induced CD (mdeg) at the λ_{MAX} (332.4 nm) versus the ligand:DNA ratio. **PA1** saturation of the 16-mer oligonucleotides was not reached until about 2:1 **PA1**:DNA ratios. (**D**) CHC2050-1 duplex presents a forward triple-base-pair-mismatch binding site and a reverse double-pair-mismatch binding site to **PA1**. Forward and reverse binding orientations are boxed in red and blue, respectively. Mismatches for the forward binding orientation are shown on the top strand in green, while mismatches for the reverse binding orientation are highlighted on the bottom strand in orange.

Two titrations with **PA1** were also performed to determine whether binding of **PA1** to a forward single-base-pair mismatch site exhibited qualitatively different spectra than binding to a reverse single-base-pair mismatch site on a 16-mer DNA duplex. The sequences were designed with a single-base-pair mismatch at the nucleotides that interact with the γ-turn of **PA1**. The rationale was that **PA1** would prefer the least number of mismatches and would have a slight preference for the forward orientation. Consequently, the forward binding orientation of **PA1** on the CHC2050-2 duplex would lead to a single-base-pair mismatch at the γ-turn, where the reverse binding orientation results in a double-base-pair-mismatch. On the other hand, binding of **PA1** in the reverse orientation to the CHC2050-3 duplex might be more favorable because this conformation results in a single-base-pair mismatch at the γ-turn, in contrast to a double-base-pair mismatch in the forward binding orientation. **Figure 2.31** shows the CD spectra for the titrations of CHC2050-2 and CHC2050-3 DNA duplexes with increasing concentrations of **PA1**. Similar to the results observed with CHC2050-1, the CD spectra of the DNA peaks were characteristic of B-form DNA with a negative peak at around 250 nm and a positive peak at around 260-280 nm. Binding of **PA1** to the DNA duplex led to a strong

induced CD signals at ~332 nm and ~302 nm (**Figure 2.31**). The generation of induced CD signals above 300 nm is consistent with **PA1** binding to the minor groove of DNA. Furthermore, the presence of multiple isodichroic points at approximately 250, 270 and 315 nm suggests a single binding mechanism.[80,81] Titration curves were generated by plotting the induced CD (mdeg) at the λ_{MAX} (~332 nm) versus the **PA1**:DNA duplex ratio (**Figure 2.31C** and **D**). **PA1** saturation of the 16-mer oligonucleotides was not reached until approximately 2:1 **PA1**:DNA ratio, suggesting that two molecules of the hairpin polyamide bind per DNA molecule. Comparison of the obtained CD spectra for the expected forward and reverse **PA1** binding orientations displayed nearly identical bands, suggesting that either the CD technique cannot differentiate between the two polyamide binding modes or that **PA1** binds in the same manner—in a forward conformation—to the studied DNA sequences. If the latter is correct, **PA1** would bind in a forward orientation with a single-base-pair mismatch to CHC2050-2 and a forward orientation with a double-base-pair mismatch to CHC2050-3.

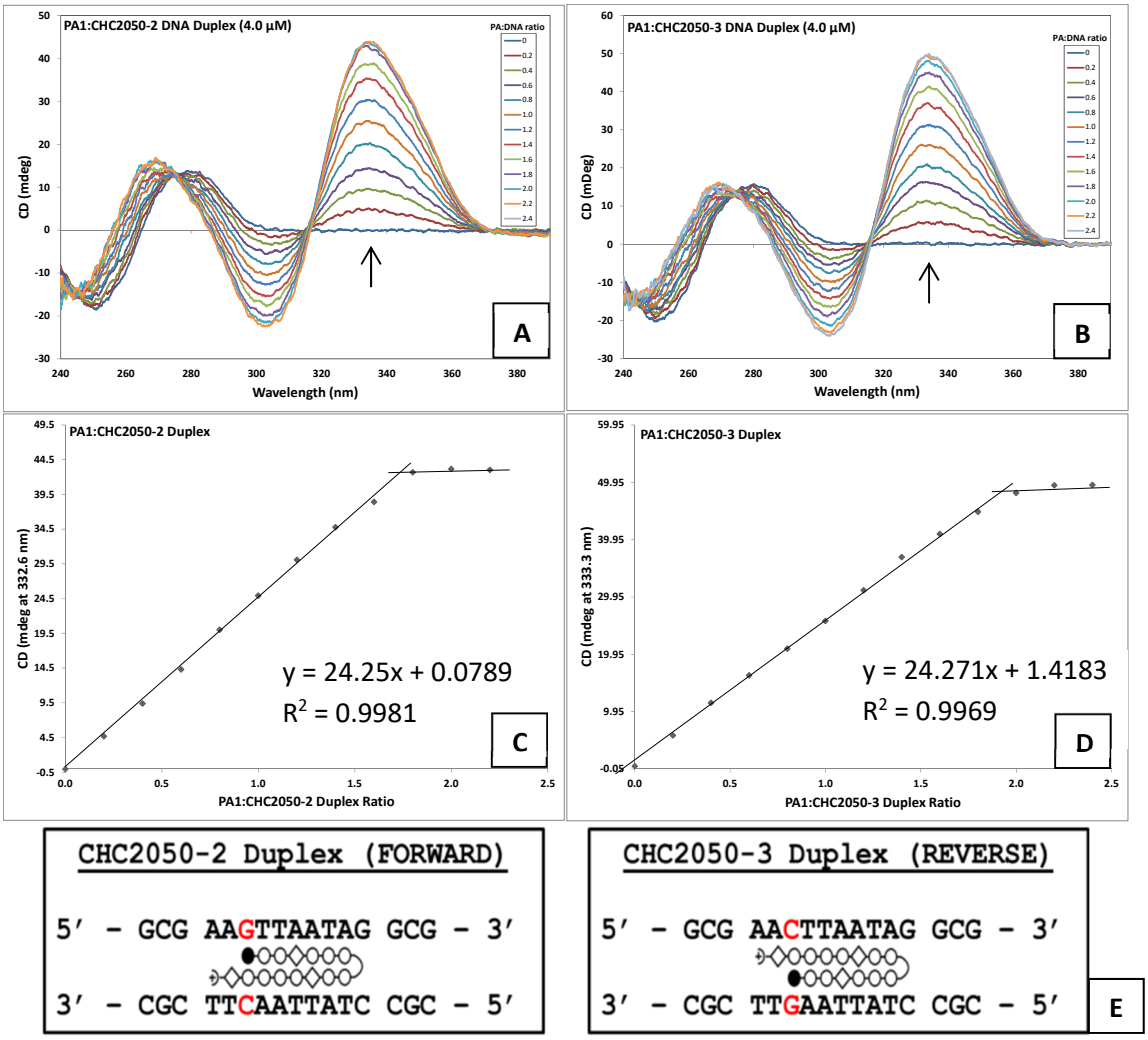


Figure 2.31. *Caption on next page.*

Figure 2.31. Circular Dichroism Spectra for the Titration of CHC2050-2 and CHC2050-3 with PA1. (**A** and **B**) CD spectra corresponding to the titrations of CHC2050-2 and CHC2050-3 DNA duplexes with **PA1**. Free DNA corresponds to the CD spectra with no induced CD signal at 332 nm (navy blue line). PA:DNA ratios ranged from 0.2 to 2.4. Black arrows indicate the induced CD band used to generate the titrations plots. (**C** and **D**) Titration curves are shown below their CD spectra and were generated by plotting the induced CD (mdeg) at the λ_{MAX} (332.6 and 333.3 nm) versus the ligand:DNA ratio. **PA1** saturation of the 16-mer oligonucleotides was not reached until about 1.5:1 and 2:1 **PA1**:DNA ratios. (**E**) CHC2050-2 and CHC2050-3 duplexes present a forward single-base-pair-mismatch binding site and a reverse single-pair-mismatch binding site to **PA1**. Mismatches are located on the nucleotides that make contact with the γ-turn from **PA1**. Red font highlights the nucleotides that are swapped in the sequences.

The same experiment was performed using **PA2** and two different match binding sequences. **PA2** shares the same sequence as **PA1**, except that it has a chiral (R)-NH_2 in the γ-turn. As a result, it recognizes the same DNA sequence (5'-W2G7-3') as **PA1**; however, the chiral (R)-substitution on the γ-turn is expected to increase the binding preference for the forward orientation.[77] The purpose of this experiment was to determine whether binding of **PA2** to a forward match site exhibited qualitatively different spectra than binding to an oligomer with both a forward single-base-pair mismatch site and a reverse match site (**Figure 2.32**). The CD spectra of the DNA peaks are similar to those observed for **PA1** experiments. However, a second binding event was observed as judged by the presence of two distinct slopes at 1.5:1 PA2:DNA ratio.

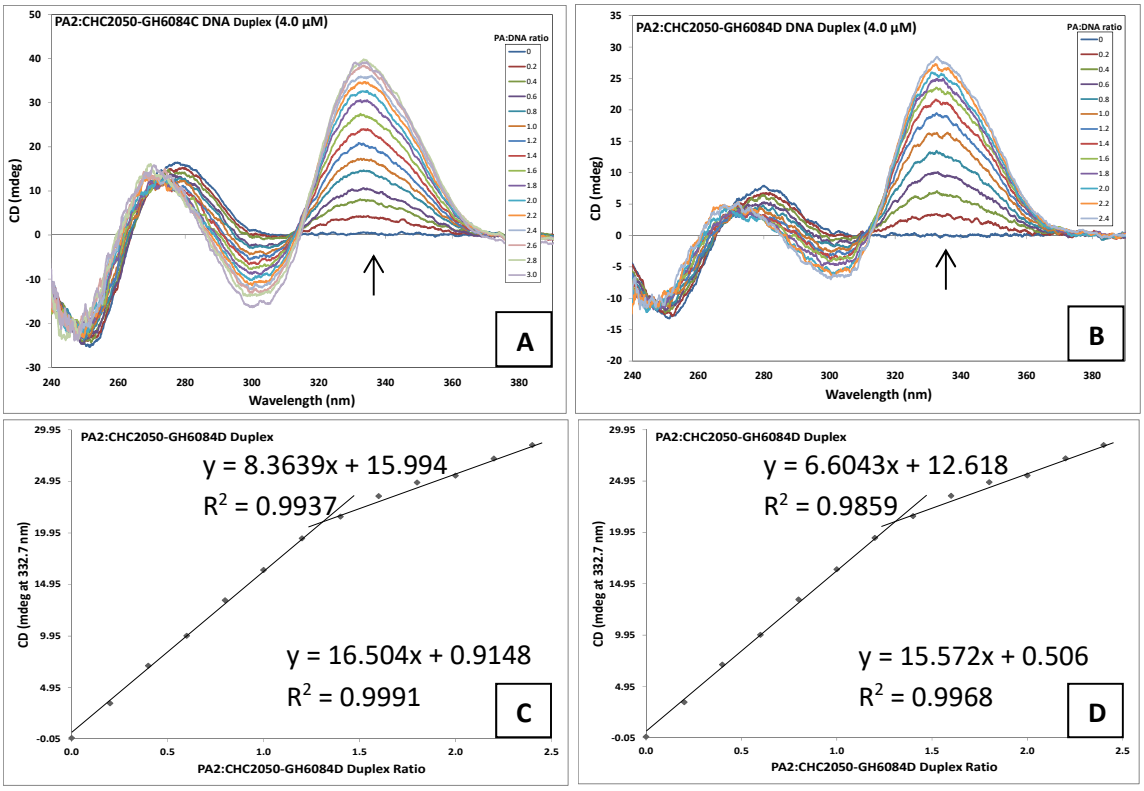


Figure 2.32. *Caption on next page.*

Figure 2.32. Circular Dichroism Spectra for the Titration of CHC2050-GH6084C and CHC2050-GH6084D with PA2. (**A** and **B**) CD spectra corresponding to the titrations of CHC2050-GH6084C and CHC2050-GH6084D DNA duplexes with **PA2**. Free DNA corresponds to the CD spectra with no induced CD signal at 332 nm (navy blue line). PA:DNA ratios ranged from 0.2 to 3.0. Black arrows indicate the induced CD band used to generate the titrations plots. (**C** and **D**) Titration curves are shown below their CD spectra and were generated by plotting the induced CD (mdeg) at the λ_{MAX} (332.6 and 332.7 nm) versus the ligand:DNA ratio. **PA2** saturation of the 16-mer oligonucleotides was not reached until about 2.5:1 PA:DNA ratios. (**E**) CHC2050-GH6084C and CHC2050-GH6084D DNA duplexes present a forward and a reverse match binding site to **PA2**, respectively. Red font highlights the nucleotides that are swapped in the sequences.

2.5 CONCLUSIONS

We have relied on quantitative DNase I footprinting and affinity cleavage experiments to determine the binding sites and equilibrium dissociation constants of **PA1** and **PA25** to PCR-amplified DNA sequences corresponding to the HPV18 LCR genomic region.

Results obtained from circular dichroism and affinity cleavage studies support **PA1** and **PA25** binding to the minor groove of DNA. The characteristic induced CD signals produced above 300 nm and the observed asymmetrical 3' shift between DNA strands in affinity cleavage experiments are consistent with **PA1** and **PA25** binding to the minor groove of DNA.

We report that **PA1** and **PA25** hairpin polyamides bind avidly to the minor groove of A/T-rich sequences in the LCR of HPV18 with dissociation constants ranging from 1.1 nM to 2.6 nM for **PA1** and 0.4 nM to 1.2 nM for **PA25**. Compared to the reported short hairpin polyamides, these PAs exhibit greater binding promiscuity.[82] Specifically, these large anti-HPV PAs bind in both forward and reverse orientations and tolerate PA-DNA mismatches without a significant decrease in binding affinity. The large nature of these antiviral agents may lead to extensive favorable interactions with the minor groove of DNA, and as a result a relatively small number of unfavorable interactions are ineffective at reducing the binding affinity. While the binding preference for A·T and T·A at Py/Py pairs and G·C at Im/Py pairs is retained in these PAs, PA-DNA base-pair mismatches are allowed in all of the positions in the PA binding motif. In contrast to **PA1**, **PA25** has the ability to tolerate a higher number of base-pair mismatches probably due to its expected larger DNA binding site. Most of the binding events observed for **PA1** can be explained with binding sites corresponding to match, single- and double-base-pair mismatch sequences; while most binding sites for **PA25** are characterized with single-, double-, and triple-base-pair mismatch sequences.

In some instances, these compounds bind well to double-base-pair mismatch binding sites, yet eschew some predicted single-base-pair mismatch DNA sequences. These observations may be due to the fact that the published PA-DNA recognition rules, while they provide a general prediction of polyamide binding, do not account for subtle DNA structural differences between seemingly equivalent DNA sequences. Furthermore, these rules fail to include the flanking sequences around the predicted binding site and other nearby binding sites that may induce a conformational change in the local DNA structure. Thus, recognition of the minor groove by large anti-HPV PAs is not solely based on the established recognition rules.

Quantitative DNase I footprinting and affinity cleavage results suggest no clear correlation of antiviral activity with binding affinity. Certainly, the dissociation constants

for **PA25** are on average 2-fold lower than those obtained for **PA1**, but this fold change is not sufficient to explain the antiviral differences. However, **PA25** exhibits an extensive DNase I protection, expected from its larger binding site, than **PA1**. Because **PA25** exhibits a more extended coverage of the HPV genome proximal to the origin of replication, its improved antiviral activity against HPV18 as compared to **PA1** may be afforded by a more effective inhibition of DNA-protein interactions.

Hairpin polyamides **PA1** and **PA25** were rationally designed to inhibit the access of viral proteins E1 and E2 to their cognate binding sites within the HPV genome. Yet, our results suggest that these bioactive compounds exhibit significant binding promiscuity. Thus, it is expected that binding to the host genome also occurs. However, we hypothesize that significant selectivity toward the HPV genome may occur because: **1)** infected keratinocytes maintain 500-1000 episomes per cell in comparison to only 2 copies of host genome. **2)** The human genome is highly compacted into chromosomes that may preclude binding of polyamides to potential binding sites, while the viral genome is condensed into chromatin-like structures that may be more accessible to polyamides.[83-88] For example, Dervan *et al*. suggested that sensitivity of cancer versus normal cells to hairpin polyamide-chlorambucil conjugates was a result of an open chromatin conformation of the target region in cancer cells, leading to a greater binding accessibility of these compounds to their cognate sites.[89] Recent findings by our collaborators at NanoVir revealed that effective episomal elimination by anti-HPV PAs is accompanied by a significant change in the expression of members from the DNA Damage Response (DDR) pathways.[44] Most notably, the authors demonstrated that this altered expression of the DDR pathways was dependent on the presence of HPV in the episomal form, as SiHa cells harboring integrated copies of HPV16 and HPV-negative cells did not elicit such a response. Likewise, the inactive eight-ring hairpin polyamide **PA11** had no effect on the DDR transcripts studied.[44] Thus, the mode of action by which antiviral PAs eliminate HPV episomes is more complex than a simple allosteric inhibition of DNA-binding proteins and it may comprise the concerted interplay of multiple cellular pathways. Furthermore, one can envisage that more PA-binding events would cause substantial topological perturbations in the DNA helix of closed-circular, negatively supercoiled episomes. These effects would in turn significantly distort the viral genome, leading to its detection as foreign or damaged DNA by the DDR response.[44,90]

2.6 BIBLIOGRAPHY

(1) Doorbar, J. *Clin Sci (Lond)* **2006**, *110*, 525.

(2) Dunne, E. F.; Markowitz, L. E.; Saraiya, M.; Stokley, S.; Middleman, A.; Unger, E. R.; Williams, A.; Iskander, J. *Morbidity and Mortality Weekly Report (MMWR)* **2014**, *63(04)*, 69.

(3) zur Hausen, H. *Cancer Research* **1989**, *49*, 4677.

(4) Giuliano, A. R.; Tortolero-Luna, G.; Ferrer, E.; Burchell, A. N.; de Sanjose, S.; Kjaer, S. K.; Muñoz, N.; Schiffman, M.; Bosch, F. X. *Vaccine* **2008**, *26*, K17.

(5) Ryser, M. D.; McGoff, K.; Herzog, D. P.; Sivakoff, D. J.; Myers, E. R. *Epidemics* **2015**, *11*, 32.

(6) Bottalico, D.; Chen, Z.; Kocjan, B. J.; Seme, K.; Poljak, M.; Burk, R. D. *J Gen Virol* **2012**, *93*, 1774.

(7) Muñoz, N.; Bosch, F. X.; Castellsagué, X.; Díaz, M.; de Sanjose, S.; Hammouda, D.; Shah, K. V.; Meijer, C. J. L. M. *International Journal of Cancer* **2004**, *111*, 278.
(8) Archambault, J.; Melendy, T. *Antivir Ther* **2013**, *18*, 271.
(9) Clifford, G.; Franceschi, S.; Diaz, M.; Munoz, N.; Villa, L. L. *Vaccine* **2006**, *24 Suppl 3*, S3/26.
(10) Stanley, M. A. P., M.M.; Coleman N. *Biochemical Society Transaction* **2007**, *35*.
(11) Wang, J. W.; Roden, R. B. *Virology* **2013**, *445*, 175.
(12) Adams, H. P.; Carnright, E. L. *Clinician Reviews* **2013**, *23*, 42.
(13) Graham, S. V. *Future microbiology* **2010**, *5*, 1493.
(14) Stanley, M. *Gynecologic Oncology* **2008**, *109*, S15.
(15) Han, K. T.; Sin, J. I. *Clin Exp Vaccine Res* **2013**, *2*, 106.
(16) Baseman, J. G.; Koutsky, L. A. *J Clin Virol* **2005**, *32 Suppl 1*, S16.
(17) Faridi, R.; Zahra, A.; Khan, K.; Idrees, M. *Virol J* **2011**, *8*, 269.
(18) Edwards, T. G.; Koeller, K. J.; Slomczynska, U.; Fok, K.; Helmus, M.; Bashkin, J. K.; Fisher, C. *Antiviral Res* **2011**, *91*, 177.
(19) Bashkin, J. K.; Koeller, K. J.; Edwards, T. G.; Fisher, C.; Google Patents: 2007.
(20) Dervan, P. B.; Edelson, B. S. *Current Opinion in Structural Biology* **2003**, *13*, 284.
(21) Chen, X. R., Boopathy; Rao, Sambhorao T.; Sundaralingam, Muttaiya *Nat Struct Mol Biol* **1994**, *1*, 7.
(22) Marky, L. A.; Breslauer, K. J. *Proc Natl Acad Sci U S A* **1987**, *84*, 4359.
(23) Finlay, A. C.; Hochstein, F. A.; Sobin, B. A.; Murphy, F. X. *Journal of the American Chemical Society* **1951**, *73*, 341.
(24) Herman, D. M.; Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1998**, *120*, 1382.
(25) Pilch, D. S.; Poklar, N.; Gelfand, C. A.; Law, S. M.; Breslauer, K. J.; Baird, E. E.; Dervan, P. B. *Proceedings of the National Academy of Sciences* **1996**, *93*, 8306.
(26) White, S.; Baird, E. E.; Dervan, P. B. *Biochemistry* **1996**, *35*, 12532.
(27) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *Journal of the American Chemical Society* **1996**, *118*, 6160.
(28) Wang, C. C. C. E., U.; Dervan, P. B. *Bioorg Med Chem* **2001**, *9*, 5.
(29) Dervan, P. B. D., R.M.; Marques, M.A. *Current Medicinal Chemistry - Anti-Cancer Agents* **2005**, *5*.
(30) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc Natl Acad Sci U S A* **1985**, *82*, 1376.
(31) Bremer, R. E.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. *Bioorg Med Chem* **2000**, *8*, 1947.
(32) Arora, P. S.; Ansari, A. Z.; Best, T. P.; Ptashne, M.; Dervan, P. B. *Journal of the American Chemical Society* **2002**, *124*, 13067.
(33) Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan, P. B. *Proc Natl Acad Sci U S A* **1998**, *95*, 12890.
(34) Mapp, A. K.; Ansari, A. Z.; Ptashne, M.; Dervan, P. B. *Proceedings of the National Academy of Sciences* **2000**, *97*, 3930.
(35) Nickols, N. G.; Szablowski, J. O.; Hargrove, A. E.; Li, B. C.; Raskatov, J. A.; Dervan, P. B. *Molecular Cancer Therapeutics* **2013**, *12*, 675.

(36) Raskatov, J. A.; Meier, J. L.; Puckett, J. W.; Yang, F.; Ramakrishnan, P.; Dervan, P. B. *Proceedings of the National Academy of Sciences* **2012**, *109*, 1023.
(37) Wang, X.; Nagase, H.; Watanabe, T.; Nobusue, H.; Suzuki, T.; Asami, Y.; Shinojima, Y.; Kawashima, H.; Takagi, K.; Mishra, R.; Igarashi, J.; Kimura, M.; Takayama, T.; Fukuda, N.; Sugiyama, H. *Cancer Sci* **2010**, *101*, 759.
(38) Nozeret, K.; Loll, F.; Escudé, C.; Boutorine, A. S. *ChemBioChem* **2015**, *16*, 549.
(39) Chenoweth, D. M.; Viger, A.; Dervan, P. B. *Journal of the American Chemical Society* **2007**, *129*, 2216.
(40) Maeshima, K.; Janssen, S.; Laemmli, U. K. *The EMBO Journal* **2001**, *20*, 3218.
(41) Janssen, S.; Durussel, T.; Laemmli, U. K. *Molecular Cell* **2000**, *6*, 999.
(42) Yasuda, A.; Noguchi, K.; Minoshima, M.; Kashiwazaki, G.; Kanda, T.; Katayama, K.; Mitsuhashi, J.; Bando, T.; Sugiyama, H.; Sugimoto, Y. *Cancer Sci* **2011**, *102*, 2221.
(43) Schaal, T. D.; Mallet, W. G.; McMinn, D. L.; Nguyen, N. V.; Sopko, M. M.; John, S.; Parekh, B. S. *Nucleic Acids Res* **2003**, *31*, 1282.
(44) Edwards, T. G.; Vidmar, T. J.; Koeller, K.; Bashkin, J. K.; Fisher, C. *PLoS One* **2013**, *8*, e75406.
(45) Chenoweth, D. M.; Dervan, P. B. *Proc Natl Acad Sci U S A* **2009**, *106*, 13175.
(46) Van Tine, B. A.; Dao, L. D.; Wu, S.-Y.; Sonbuchner, T. M.; Lin, B. Y.; Zou, N.; Chiang, C.-M.; Broker, T. R.; Chow, L. T. *Proc Natl Acad Sci U S A* **2004**, *101*, 4030.
(47) Hegde, R. S. *Annu Rev Biophys Biomol Struct* **2002**, *31*, 343.
(48) Bashkin, J. K.; EDWARDS, T. G.; Fisher, C.; HARRIS, J. R. G. D.; Koeller, K. J.; Google Patents: 2014.
(49) Edwards, T. G.; Koeller, K. J.; Slomczynska, U.; Fok, K.; Helmus, M.; Bashkin, J. K.; Fisher, C. *Antiviral Research* **2011**, *91*, 177.
(50) Galas, D. J.; Schmitz, A. *Nucleic Acids Res* **1978**, *5*, 3157.
(51) Fox, K. R. *Drug-DNA Interaction Protocols* 1997; Vol. 90, p 1.
(52) Lane, M. J.; Dabrowiak, J. C.; Vournakis, J. N. *Proc Natl Acad Sci U S A* **1983**, *80*, 3260.
(53) Fox, K. R.; Waring, M. J. *Nucleic Acids Res* **1984**, *12*, 9271.
(54) Chaires, J. B.; Fox, K. R.; Herrera, J. E.; Britt, M.; Waring, M. J. *Biochemistry* **1987**, *26*, 8227.
(55) Van Dyke, M. W.; Dervan, P. B. *Nucleic Acids Res* **1983**, *11*, 5555.
(56) Yoshinaga, S. K.; Boulanger, P. A.; Berk, A. J. *Proc Natl Acad Sci U S A* **1987**, *84*, 3585.
(57) Hesselberth, J. R.; Chen, X.; Zhang, Z.; Sabo, P. J.; Sandstrom, R.; Reynolds, A. P.; Thurman, R. E.; Neph, S.; Kuehn, M. S.; Noble, W. S.; Fields, S.; Stamatoyannopoulos, J. A. *Nat Methods* **2009**, *6*, 283.
(58) Van Dyke, M. W.; Sawadogo, M. *Mol Cell Biol* **1990**, *10*, 3415.
(59) Abu-Daya, A.; Brown, P. M.; Fox, K. R. *Nucleic Acids Res* **1995**, *23*, 3385.
(60) Mrksich, M.; Parks, M. E.; Dervan, P. B. *Journal of the American Chemical Society* **1994**, *116*, 7983.
(61) Hampshire, A. J.; Rusling, D. A.; Broughton-Head, V. J.; Fox, K. R. *Methods* **2007**, *42*, 128.

(62) Cardew, A.; Fox, K. In *Drug-DNA Interaction Protocols*; Fox, K. R., Ed.; Humana Press: 2010; Vol. 613, p 153.
(63) Leblanc, B.; Moss, T. In *DNA-Protein Interactions*; Leblanc, B., Moss, T., Eds.; Humana Press: 2009; Vol. 543, p 37.
(64) Dervan, P. B. *Science* **1986**, *232*, 464.
(65) Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1996**, *118*, 6141.
(66) Vasilieva, E.; Niederschulte, J.; Song, Y.; Harris Jr, G. D.; Koeller, K. J.; Liao, P.; Bashkin, J. K.; Dupureur, C. M. *Biochimie* **2016**, *127*, 103.
(67) Cole, S. T.; Danos, O. *J Mol Biol* **1987**, *193*, 599.
(68) Boshart, M.; Gissmann, L.; Ikenberg, H.; Kleinheinz, A.; Scheurlen, W.; zur Hausen, H. *The EMBO Journal* **1984**, *3*, 1151.
(69) Sverdrup, F.; Khan, S. A. *J Virol* **1995**, *69*, 1319.
(70) IDT. DNA Sequencing. [Online Early Access]. Published Online: 2005. https://www.idtdna.com/pages/docs/educational-resources/dna-sequencing.pdf.
(71) Karger, B. L.; Guttman, A. *Electrophoresis* **2009**, *30 Suppl 1*, S196.
(72) He, G.; Vasilieva, E.; Bashkin, J. K.; Dupureur, C. M. *Analytical Biochemistry* **2013**, *439*, 99.
(73) Maxam, A. M.; Gilbert, W. In *Methods in Enzymology*; Lawrence Grossman, K. M., Ed.; Academic Press: 1980; Vol. Volume 65, p 499.
(74) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning: A Laboratory Manual*; 2nd ed.; Cold Spring Harbor, N.Y. : Cold Spring Harbor Laboratory Press, 1989.
(75) Heiger, D. A Primer: High Performance Capillary Electrophoresis. [Online Early Access]. Published Online: 2000.
(76) Chang, Y. M.; Chen, C. K.; Hou, M. H. *Int J Mol Sci* **2012**, *13*, 3394.
(77) Meier, J. L.; Yu, A. S.; Korf, I.; Segal, D. J.; Dervan, P. B. *Journal of the American Chemical Society* **2012**, *134*, 17814.
(78) Holt, S. E.; Schuller, G.; Wilson, V. G. *J Virol* **1994**, *68*, 1094.
(79) Doerks, T.; Copley, R. R.; Schultz, J.; Ponting, C. P.; Bork, P. *Genome Res* **2002**, *12*, 47.
(80) Caesar, C. E. B.; Johnsson, R.; Ellervik, U.; Fox, K. R.; Lincoln, P.; Nordén, B. *Biophysical Journal* **2006**, *91*, 904.
(81) Lacy, E. R.; Cox, K. K.; Wilson, W. D.; Lee, M. *Nucleic Acids Res* **2002**, *30*, 1834.
(82) Koeller, K. J.; Harris, G. D.; Aston, K.; He, G.; Castaneda, C. H.; Thornton, M. A.; Edwards, T. G.; Wang, S.; Nanjunda, R.; Wilson, W. D.; Fisher, C.; Bashkin, J. K. *Medicinal Chemistry* **2014**, *4*, 338.
(83) Favre, M.; Breitburd, F.; Croissant, O.; Orth, G. *J Virol* **1977**, *21*, 1205.
(84) Lee, D.; Sohn, H.; Kalpana, G. V.; Choe, J. *Nature* **1999**, *399*, 487.
(85) Peña, L. d. M.; Laimins, L. A. *J Virol* **2001**, *75*, 10005.
(86) Stünkel, W.; Bernard, H.-U. *J Virol* **1999**, *73*, 1918.
(87) Demeret, C.; Le Moal, M.; Yaniv, M.; Thierry, F. *Nucleic Acids Res* **1995**, *23*, 4777.
(88) Badal, V.; Chuang, L. S. H.; Tan, E. H. H.; Badal, S.; Villa, L. L.; Wheeler, C. M.; Li, B. F. L.; Bernard, H. U. *J Virol* **2003**, *77*, 6227.

(89) Jespersen, C.; Soragni, E.; James Chou, C.; Arora, P. S.; Dervan, P. B.; Gottesfeld, J. M. *Bioorganic & Medicinal Chemistry Letters* **2012**, *22*, 4068.
(90) Fisher, C. *Journal of Clinical Medicine* **2015**, *4*, 204.
(91) Dupureur, C. M.; Bashkin, J. K.; Aston, K.; Koeller, K. J.; Gaston, K. R.; He, G. *Anal Biochem* **2012**, *423*, 178.

2.7 SUPPLEMENTAL INFORMATION

2.7.1 Maxam-Gilbert Sequencing Protocol

Note: Hydrazine will destroy HEX dye.

Reagents	Preparation
Piperidine Formate	Add 4 % formic acid solution to 10 mL of 1 M Piperidine until pH 2.0 is obtained
30 % Mono Hydrazine	As is
3 M NaOAc pH = 5.0	Add 1.23 g of NaOAc to 3 mL of MQ H_2O, pH to 5.0 with glacial acetic acid. Add MQ H2O to a 5 mL total volume
Stop Solution	**0.3 M NaOAc pH = 7.0, 0.1 mM EDTA:** 1000 µL of 3 M NaOAc pH = 5.0 + 2 µL of 0.5 M EDTA pH = 8.0 + 8998 µL MQ H_2O
1 M Piperidine	Add 1 mL of 10 M Piperidine to 9 mL of MQ H_2O

A+G:

1. 16 nM DNA (24 µL) + 4 µL of piperidine formate; overall volume = 28 µL.
2. Incubate for 15 min or 30 min at 37 °C to optimize fragmentation.

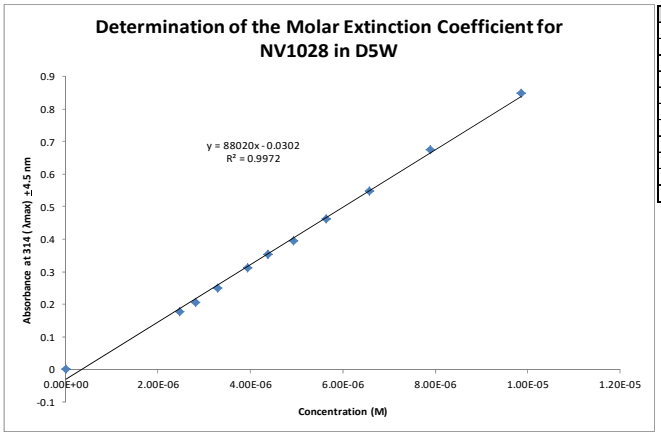
C+T:

1. 6 nM DNA (24 µL) + 40 µL of mono hydrazine or 11.5 nM DNA (24 µL) + 20 µL of mono hydrazine.
2. Incubate for 15 min or 30 min at 37 °C to optimize fragmentation.

3. Add 200 µL of stop solution at 0 °C:
   0.3 M NaOAc pH =7.0 (If you have it available, dilute from 3 M NaOAc pH = 5.0)
   0.1 mM EDTA pH = 8.0
4. Add 750 µL of EtOH (-20 °C) incubate in freezer (-20 °C) for 20 min.
5. Centrifuge for 20 min 13000*x g* at 4 °C.
6. Remove supernatant carefully and dispose of supernatant to appropriate waste bottle.
7. Add 300 µL of 0.3 M NaOAc (pH = 5.2) at 0 °C.
8. Add 900 µL of EtOH (-20 °C).
9. Store 20 min in -20 °C.
10. Centrifuge for 20 min 13000*x g* at 4 °C.
11. Remove supernatant carefully.
12. Add 1 mL of EtOH (-20 °C).
13. Centrifuge for 10 min 13000*x g* at 4 °C.
14. Remove supernatant carefully.
15. Speed vac DNA pellet for at least 4 h.
16. Add 100 µL of 1 M piperidine, mix (pure piperidine is 10 M, just dilute in water to get 1 M).
17. Incubate 30 min at 90 °C.
18. Speed vac for 2 h (good stopping point for the day, speed vac piperidine overnight).
19. Add 20 µL of H_2O, vortex for 30 sec.
20. Centrifuge for 10 sec.
21. Speed vac for 2 h.
22. Add 50 µL of H_2O.
23. Send different amounts of DNA for fragmentation analysis keeping the total volume equal to 2 µL.

2.7.2 Molar Extinction Coefficient Determination for PA1 in 5 % Dextrose / 95 % Water (D5W)

Because the molar extinction coefficient of polyamides can differ substantially in different solvents,[91] the molar extinction coefficient for **PA1** in 5 % dextrose / 95 % water (D5W) was determined as the provisional value of 88,020 $M^{-1}cm^{-1}$. This value assumes that polyamide did not aggregate in the stock and diluted solutions. Briefly, a 2.24 mM **PA1** in D5W (FT1124) was used as the stock and different concentrations of **PA1** were prepared by further dilution with D5W (9.90-2.45 μM). The absorbance at λ_{max} was determined (around 305 nm) using a Thermo Scientific Evolution 260 Bio UV-visible spectrophotometer for different **PA1** concentrations. Since Beer-Lambert Law states that A = εbc (where A is the absorbance at a particular wavelength, ε is the molar extinction coefficient, b is the cell path length and c is the concentration of the sample), a graph was constructed plotting the measured absorbance versus the polyamide concentration (M). The slope of this line corresponds to the molar extinction coefficient (**Figure SI2.1**).



λmax (nm)	Absorbance at λmax	Concentration (μM)	Concentration (M)
	0	0	0
313.80	0.847	9.9	9.86E-06
316.55	0.675	7.9	7.88E-06
314.83	0.547	6.6	6.57E-06
315.62	0.463	5.6	5.63E-06
313.92	0.394	4.9	4.93E-06
315.25	0.354	4.4	4.38E-06
314.82	0.312	3.9	3.94E-06
312.08	0.251	3.3	3.29E-06
312.18	0.205	2.8	2.82E-06
315.79	0.176	2.5	2.46E-06

Figure SI2.1. Absorbance Data and Plot for Different Concentrations of PA1 in 5 % Dextrose / 95 % Water (D5W). Experimental data used for the determination of the molar extinction coefficient for **PA1** in D5W. Plot of absorbances at 314 $\pm$ 4.5 nm versus polyamide concentrations (M). Equation of the line is y = 88020x-0.0302, R^2 =

2.7.3 X-ray Crystallography Matrix

Table SI2.1. X-ray Crystallization Screening Matrix Generated from Conditions Published in PA Literature.

						Matrix Solution Preparation							
	CaAc (mM)	MPD (%)	TRIS (mM), pH 7.5	TRIS (mM), 8.0 pH	Comments	1000 mM CaAc Stock (µL)	100% MPD Stock (µL)	1000 mM TRIS (µL), pH 7.5	1000 mM TRIS (µL), pH 8.0	MilliQ H_2O	Total (µL)	Experiment A (1:1.3 DNA:PA)	Experiment B (1:2 DNA:PA)
A1	100	40		20	CaAc titration	25	100		5	120	250	50 mM CaAc, 20% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	50 mM CaAc, 20% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
B1	90	40		20	CaAc titration	22.5	100		5	122.5	250	45 mM CaAc, 20% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	45 mM CaAc, 20% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
C1	80	40		20	CaAc titration	20	100		5	125	250	40 mM CaAc, 20% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	40 mM CaAc, 20% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
D1	70	40		20	Dervan's Conditions in well	17.5	100		5	127.5	250	35 mM CaAc, 20% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	35 mM CaAc, 20% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
E1	60	40		20	CaAc titration	15	100		5	130	250	30 mM CaAc, 20% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	30 mM CaAc, 20% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
F1	50	40		20	CaAc titration	12.5	100		5	132.5	250	25 mM CaAc, 20% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	25 mM CaAc, 20% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
G1	40	40		20	CaAc titration	10	100		5	135	250	20 mM CaAc, 20% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	20 mM CaAc, 20% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
H1	30	40		20	CaAc titration	7.5	100		5	137.5	250	15 mM CaAc, 20% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	15 mM CaAc, 20% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
A2	70	50		20	MPD titration	17.5	125		5	102.5	250	35 mM CaAc, 25% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	35 mM CaAc, 25% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
B2	70	45		20	MPD titration	17.5	112.5		5	115	250	35 mM CaAc, 22.5% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	35 mM CaAc, 22.5% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
C2	70	40	20		TRIS Low	17.5	100	5		127.5	250	35 mM CaAc, 20% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 0.65 mM PA	35 mM CaAc, 20% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 1.0 mM PA
D2	70	35		20	MPD titration	17.5	87.5		5	140	250	35 mM CaAc, 17.5% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	35 mM CaAc, 17.5% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
E2	70	30		20	MPD titration	17.5	75		5	152.5	250	35 mM CaAc, 15% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	35 mM CaAc, 15% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
F2	70	25		20	MPD titration	17.5	62.5		5	165	250	35 mM CaAc, 12.5% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	35 mM CaAc, 12.5% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
G2	70	15		20	MPD titration	17.5	37.5		5	190	250	35 mM CaAc, 7.5% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	35 mM CaAc, 7.5% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
H2	70	10		20	MPD titration	17.5	25		5	202.5	250	35 mM CaAc, 5% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	35 mM CaAc, 5% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
A3	100	50		20	High High High	25	125		5	95	250	50 mM CaAc, 25% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	50 mM CaAc, 25% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
B3	100	50	20		High High Low	25	125	5		95	250	50 mM CaAc, 25% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 0.65 mM PA	50 mM CaAc, 25% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 1.0 mM PA
C3	100	10		20	High Low High	25	25		5	195	250	50 mM CaAc, 5% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	50 mM CaAc, 5% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
D3	30	50		20	Low High High	7.5	125		5	112.5	250	15 mM CaAc, 25% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	15 mM CaAc, 25% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
E3	30	10		20	Low Low High	7.5	25		5	212.5	250	15 mM CaAc, 5% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	15 mM CaAc, 5% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
F3	30	50	20		Low High Low	7.5	125	5		112.5	250	15 mM CaAc, 25% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 0.65 mM PA	15 mM CaAc, 25% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 1.0 mM PA
G3	100	10	20		High Low Low	25	25	5		195	250	50 mM CaAc, 5% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 0.65 mM PA	50 mM CaAc, 5% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 1.0 mM PA
H3	30	10	20		Low Low Low	7.5	25	5		212.5	250	15 mM CaAc, 5% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 0.65 mM PA	15 mM CaAc, 5% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 1.0 mM PA
A4	100	40	20		High Mid Low	25	100	5		120	250	50 mM CaAc, 20% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 0.65 mM PA	50 mM CaAc, 20% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 1.0 mM PA
B4	30	40	20		Low Mid Low	7.5	100	5		137.5	250	15 mM CaAc, 20% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 0.65 mM PA	15 mM CaAc, 20% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 1.0 mM PA
C4	70	10	20		Mid Low Low	17.5	25	5		202.5	250	35 mM CaAc, 5% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 0.65 mM PA	35 mM CaAc, 5% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 1.0 mM PA
D4	70	50	20		Mid High Low	17.5	125	5		102.5	250	35 mM CaAc, 25% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 0.65 mM PA	35 mM CaAc, 25% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 1.0 mM PA
E4	90	40		20	Mid-H Mid High	22.5	100		5	122.5	250	45 mM CaAc, 20% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	45 mM CaAc, 20% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
F4	90	40	20		Mid-H Mid Low	22.5	100	5		122.5	250	45 mM CaAc, 20% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 0.65 mM PA	45 mM CaAc, 20% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 1.0 mM PA
G4	50	40		20	Mid-L Mid High	12.5	100		5	132.5	250	25 mM CaAc, 20% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	25 mM CaAc, 20% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
H4	50	40	20		Mid-L Mid Low	12.5	100	5		132.5	250	25 mM CaAc, 20% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 0.65 mM PA	25 mM CaAc, 20% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 1.0 mM PA
A5	90	50		20	Mid-H High High	22.5	125		5	97.5	250	45 mM CaAc, 25% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	45 mM CaAc, 25% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
B5	90	50	20		Mid-H High Low	22.5	125	5		97.5	250	45 mM CaAc, 25% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 0.65 mM PA	45 mM CaAc, 25% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 1.0 mM PA
C5	50	50		20	Mid-L High High	12.5	125		5	107.5	250	25 mM CaAc, 25% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	25 mM CaAc, 25% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
D5	50	50	20		Mid-L High Low	12.5	125	5		107.5	250	25 mM CaAc, 25% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 0.65 mM PA	25 mM CaAc, 25% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 1.0 mM PA
E5	90	45		20	Mid-H Mid-H High	22.5	112.5		5	110	250	45 mM CaAc, 22.5% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	45 mM CaAc, 22.5% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
F5	90	15	20		Mid-H Mid-L High	22.5	37.5		5	185	250	45 mM CaAc, 7.5% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	45 mM CaAc, 7.5% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
G5	50	45		20	Mid-L Mid-H High	12.5	112.5		5	120	250	25 mM CaAc, 22.5% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	25 mM CaAc, 22.5% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
H5	50	15	20		Mid-L Mid-L High	12.5	37.5		5	195	250	25 mM CaAc, 7.5% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 0.65 mM PA	25 mM CaAc, 7.5% MPD, 10 mM TRIS pH 8.0, 0.5 mM DNA, 1.0 mM PA
A6	90	45		20	Mid-H Mid-H Low	22.5	112.5	5		110	250	45 mM CaAc, 22.5% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 0.65 mM PA	45 mM CaAc, 22.5% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 1.0 mM PA
B6	90	15	20		Mid-H Mid-L Low	22.5	37.5	5		185	250	45 mM CaAc, 7.5% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 0.65 mM PA	45 mM CaAc, 7.5% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 1.0 mM PA
C6	50	45	20		Mid-L Mid-H Low	12.5	112.5	5		120	250	25 mM CaAc, 22.5% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 0.65 mM PA	25 mM CaAc, 22.5% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 1.0 mM PA
D6	50	15	20		Mid-L Mid-L Low	12.5	37.5	5		195	250	25 mM CaAc, 7.5% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 0.65 mM PA	25 mM CaAc, 7.5% MPD, 10 mM TRIS pH 7.5, 0.5 mM DNA, 1.0 mM PA
E6		25					62.5			187.5	250	35 mM CaAc, 20% MPD, 10 mM TRIS, 0.5 mM DNA, 0.65 mM PA	35 mM CaAc, 20% MPD, 10 mM TRIS, 0.5 mM DNA, 1.0 mM PA
F6		30			MDP alone in reservoir		75			175	250	35 mM CaAc, 20% MPD, 10 mM TRIS, 0.5 mM DNA, 0.65 mM PA	35 mM CaAc, 20% MPD, 10 mM TRIS, 0.5 mM DNA, 1.0 mM PA
G6		35					87.5			162.5	250	35 mM CaAc, 20% MPD, 10 mM TRIS, 0.5 mM DNA, 0.65 mM PA	35 mM CaAc, 20% MPD, 10 mM TRIS, 0.5 mM DNA, 1.0 mM PA
H6		40					100			150	250	35 mM CaAc, 20% MPD, 10 mM TRIS, 0.5 mM DNA, 0.65 mM PA	35 mM CaAc, 20% MPD, 10 mM TRIS, 0.5 mM DNA, 1.0 mM PA

2.7.4 CE Sanger USB and Maxam-Gilbert Indexing

HPV18 (7479-7783) 305 bp (FAM ECF160 Lanes E1, F1, G2, H9 (plate JS4313)
Correction between CE size calling to USB sequencing results

CE postion	Sequence	Real Position	Genome Position		CE postion	Sequence	Real Position	Genome Position		CE postion	Sequence	Real Position	Genome Position	
	C	1	7479		116.1	C	121	7599	-4.9	236.9	C	241	7719	-4.1
	T	2	7480		117.1	A	122	7600	-4.9	237.8	T	242	7720	-4.2
	T	3	7481		117.8	C	123	7601	-5.2	238.7	A	243	7721	-4.3
	A	4	7482		118.7	C	124	7602	-5.3	240	A	244	7722	-4
	T	5	7483		119.7	T	125	7603	-5.3	240.9	T	245	7723	-4.1
	G	6	7484		120.8	G	126	7604	-5.2	241.9	T	246	7724	-4.1
	T	7	7485		122	G	127	7605	-5	243.1	G	247	7725	-3.9
	C	8	7486		123.3	T	128	7606	-4.7	244	C	248	7726	-4
	T	9	7487		124.2	A	129	7607	-4.8	245	A	249	7727	-4
	G	10	7488		125.2	T	130	7608	-4.8	245.9	T	250	7728	-4.1
	T	11	7489		126.2	T	131	7609	-4.8	247	A	251	7729	-4
	G	12	7490		127.2	A	132	7610	-4.8	248	C	252	7730	-4
	G	13	7491		128.2	G	133	7611	-4.8	249	T	253	7731	-4
	T	14	7492		129.2	T	134	7612	-4.8	250.1	T	254	7732	-3.9
	T	15	7493		130.2	C	135	7613	-4.8	251.3	G	255	7733	-3.7
	T	16	7494		131	A	136	7614	-5	252.3	G	256	7734	-3.7
	T	17	7495		132	T	137	7615	-5	253.3	C	257	7735	-3.7
	C	18	7496		133	T	138	7616	-5	254.2	T	258	7736	-3.8
	T	19	7497		134	T	139	7617	-5	255.1	T	259	7737	-3.9
	G	20	7498		135	T	140	7618	-5	256.3	G	260	7738	-3.7
	C	21	7499		136	C	141	7619	-5	257.2	T	261	7739	-3.8
	A	22	7500		136.9	C	142	7620	-5.1	258.3	A	262	7740	-3.7
	C	23	7501		137	T	143	7621	-6	259.3	C	263	7741	-3.7
	A	24	7502		139.1	G	144	7622	-4.9	260.1	A	264	7742	-3.9
	A	25	7503		140.2	T	145	7623	-4.8	261	A	265	7743	-4
	T	26	7504		141.2	C	146	7624	-4.8	262	C	266	7744	-4
	A	27	7505		142	C	147	7625	-5	262.9	T	267	7745	-4.1
	C	28	7506		143	A	148	7626	-5	264	A	268	7746	-4
	A	29	7507		144.1	G	149	7627	-4.9	265.1	C	269	7747	-3.9
	G	30	7508		145	G	150	7628	-5	266	T	270	7748	-4
	T	31	7509		146.2	T	151	7629	-4.8	267	T	271	7749	-4
	A	32	7510		147.4	G	152	7630	-4.6	268.2	T	272	7750	-3.8
	C	33	7511		148.3	C	153	7631	-4.7	269.2	C	273	7751	-3.8
	G	34	7512			G	154	7632		270.2	A	274	7752	-3.8
	C	35	7513		150.2	C	155	7633	-4.8	271.2	T	275	7753	-3.8
	T	36	7514		151	T	156	7634	-5	272.5	G	276	7754	-3.5
	G	37	7515		152	A	157	7635	-5		T	277	7755	
	G	38	7516		153.1	C	158	7636	-4.9	274.4	C	278	7756	-3.6
	C	39	7517		154	A	159	7637	-5	275.5	C	279	7757	-3.5
	A	40	7518		155.1	A	160	7638	-4.9	276.1	A	280	7758	-3.9
	C	41	7519		156.1	C	161	7639	-4.9	277.1	A	281	7759	-3.9
	T	42	7520		157.1	A	162	7640	-4.9	278.2	C	282	7760	-3.8
	A	43	7521		158	A	163	7641	-5	279.2	A	283	7761	-3.8
	T	44	7522		159.1	T	164	7642	-4.9	280	T	284	7762	-4
	T	45	7523		160.2	T	165	7643	-4.8	281.2	T	285	7763	-3.8
	G	46	7524		161.4	G	166	7644	-4.6	282.3	C	286	7764	-3.7
	C	47	7525		162.4	C	167	7645	-4.6	283.2	T	287	7765	-3.8
	A	48	7526		163.3	T	168	7646	-4.7	284.4	G	288	7766	-3.6
	A	49	7527		164.3	T	169	7647	-4.7	285.3	T	289	7767	-3.7
	A	50	7528		165.4	G	170	7648	-4.6	286.5	C	290	7768	-3.5
40.5	C	51	7529	-10.5	166.4	C	171	7649	-4.6	287.3	T	291	7769	-3.7
42	T	52	7530	-10	167.3	A	172	7650	-4.7	288.4	A	292	7770	-3.6
43.2	T	53	7531	-9.8	168.3	T	173	7651	-4.7	289.3	C	293	7771	-3.7
44.5	T	54	7532	-9.5	169.3	A	174	7652	-4.7	290.2	C	294	7772	-3.8
	A	55	7533		170.3	A	175	7653	-4.7	291.1	C	295	7773	-3.9
	A	56	7534		171.3	C	176	7654	-4.7	292	T	296	7774	-4
48	T	57	7535	-9	172.3	T	177	7655	-4.7	293.7	T	297	7775	-3.3
49	C	58	7536	-9	173.3	A	178	7656	-4.7	294.3	A	298	7776	-3.7
	T	59	7537		174.4	T	179	7657	-4.6	295.3	A	299	7777	-3.7
	T	60	7538		175.5	A	180	7658	-4.5	296.3	C	300	7778	-3.7
54.1	T	61	7539	-6.9	176.4	T	181	7659	-4.6	297.3	A	301	7779	-3.7
55.2	T	62	7540	-6.8	177.4	C	182	7660	-4.6	298.2	T	302	7780	-3.8
	G	63	7541		178.3	C	183	7661	-4.7		G	303	7781	
	G	64	7542		179.3	A	184	7662	-4.7		A	304	7782	
	G	65	7543		180.2	C	185	7663	-4.8		A	305	7783	
59.3	C	66	7544	-6.7	181.2	T	186	7664	-4.8					
60.2	A	67	7545	-6.8	182.3	C	187	7665	-4.7					
60.5	C	68	7546	-7.5	183.2	C	188	7666	-4.8					
61.7	T	69	7547	-7.3	184.1	C	189	7667	-4.9					
62.9	G	70	7548	-7.1	185.1	T	190	7668	-4.9					
63.9	C	71	7549	-7.1	186.4	A	191	7669	-4.6					
65	T	72	7550	-7	187.4	A	192	7670	-4.6					
65.9	C	73	7551	-7.1	188.5	G	193	7671	-4.5					
66.8	C	74	7552	-7.2	189.5	T	194	7672	-4.5					
68	T	75	7553	-7	190.6	A	195	7673	-4.4					
69.3	A	76	7554	-6.7	191.4	A	196	7674	-4.6					
70.2	C	77	7555	-6.8	192.5	T	197	7675	-4.5					
71.5	A	78	7556	-6.5	193.5	A	198	7676	-4.5					
72.5	T	79	7557	-6.5	194.5	A	199	7677	-4.5					
73.7	A	80	7558	-6.3	195.4	A	200	7678	-4.6					
74.7	T	81	7559	-6.3	196.4	A	201	7679	-4.6					
75.9	T	82	7560	-6.1	197.4	C	202	7680	-4.6					
77	T	83	7561	-6	198.4	T	203	7681	-4.6					
78.2	T	84	7562	-5.8	199.5	G	204	7682	-4.5					
79.3	G	85	7563	-5.7	200.5	C	205	7683	-4.5					
	A	86	7564		201.5	T	206	7684	-4.5					
	A	87	7565		202.5	T	207	7685	-4.5					
82.2	C	88	7566	-5.8	203.5	T	208	7686	-4.5					
83.2	A	89	7567	-5.8	204.6	T	209	7687	-4.4					
	A	90	7568		205.6	A	210	7688	-4.4					
	T	91	7569		206.8	G	211	7689	-4.2					
	T	92	7570		207.9	G	212	7690	-4.1					
	G	93	7571		208.8	C	213	7691	-4.2					
	G	94	7572		209.6	A	214	7692	-4.4					
	C	95	7573		210.4	C	215	7693	-4.6					
	G	96	7574		211.5	A	216	7694	-4.5					
	C	97	7575		212.3	T	217	7695	-4.7					
	G	98	7576		213.5	A	218	7696	-4.5					
	C	99	7577		214.5	T	219	7697	-4.5					
	C	100	7578		215.6	T	220	7698	-4.4					
	T	101	7579		216.7	T	221	7699	-4.3					
96.1	C	102	7580	-5.9	217.7	T	222	7700	-4.3					
97.3	T	103	7581	-5.7	218.7	A	223	7701	-4.3					
98.5	T	104	7582	-5.5	219.8	G	224	7702	-4.2					
99.8	T	105	7583	-5.2	220.9	T	225	7703	-4.1					
101.1	G	106	7584	-4.9	221.9	T	226	7704	-4.1					
102.2	G	107	7585	-4.8	222.9	T	227	7705	-4.1					
103.3	C	108	7586	-4.7	223.9	G	228	7706	-4.1					
104.1	G	109	7587	-4.9	225	T	229	7707	-4					
104.6	C	110	7588	-5.4	226	T	230	7708	-4					
105.6	A	111	7589	-5.4	226.8	T	231	7709	-4.2					
106.7	T	112	7590	-5.3	228	T	232	7710	-4					
107.9	A	113	7591	-5.1	228.9	T	233	7711	-4.1					
109	T	114	7592	-5	229.9	A	234	7712	-4.1					
110.2	A	115	7593	-4.8	230.8	C	235	7713	-4.2					
111.2	A	116	7594	-4.8	231.9	T	236	7714	-4.1					
112.5	G	117	7595	-4.5	232.8	T	237	7715	-4.2					
113.1	G	118	7596	-4.9	233.8	A	238	7716	-4.2					
114.4	C	119	7597	-4.6	234.8	A	239	7717	-4.2					
115.3	G	120	7598	-4.7	235.9	G	240	7718	-4.1					

Figure SI2.2A. CE Sanger USB Indexing of the 305 bp HPV18 (7479-7783): Top Strand. CE position *versus* the real position of nucleotides from the top strand (5'-FAM) of the 305 bp HPV18 (7479-7783) fragment.

HPV18 (7479-7783) 305 bp (HEX Lanes ECF160 A1,B2,C2,D2)
Correction between CE size calling to USB sequencing results

CE postion	Sequence	Real Position	Genome Position		CE postion	Sequence	Real Position	Genome Position		CE postion	Sequence	Real Position	Genome Position	
	T	1	7783		114.9	G	121	7663	-6.1	234.7	C	241	7543	-6.3
	T	2	7782		115.8	T	122	7662	-6.2	235.6	C	242	7542	-6.4
	C	3	7781		116.9	G	123	7661	-6.1	236.5	C	243	7541	-6.5
	A	4	7780		117.9	G	124	7660	-6.1	237.5	A	244	7540	-6.5
	T	5	7779		118.7	A	125	7659	-6.3	238.5	A	245	7539	-6.5
	G	6	7778		119.8	T	126	7658	-6.2	239.5	A	246	7538	-6.5
	T	7	7777		120.5	A	127	7657	-6.5	240.5	A	247	7537	-6.5
	T	8	7776		121.6	T	128	7656	-6.4	241.8	G	248	7536	-6.2
	A	9	7775		122.4	A	129	7655	-6.6	242.6	A	249	7535	-6.4
	A	10	7774		123.7	G	130	7654	-6.3	244	T	250	7534	-6
	G	11	7773		124.6	T	131	7653	-6.4	244.8	T	251	7533	-6.2
	G	12	7772		125.6	T	132	7652	-6.4	245.7	A	252	7532	-6.3
	G	13	7771		126.4	A	133	7651	-6.6	246.7	A	253	7531	-6.3
	T	14	7770		127.5	T	134	7650	-6.5	247.7	A	254	7530	-6.3
	A	15	7769		128.8	G	135	7649	-6.2	248.9	G	255	7529	-6.1
	G	16	7768		129.5	C	136	7648	-6.5	249.9	T	256	7528	-6.1
	A	17	7767		130.3	A	137	7647	-6.7	250.9	T	257	7527	-6.1
	C	18	7766		131.1	A	138	7646	-6.9	251.8	T	258	7526	-6.2
	A	19	7765		132.5	G	139	7645	-6.5	252.9	G	259	7525	-6.1
	G	20	7764		133	C	140	7644	-7	253.8	C	260	7524	-6.2
	A	21	7763		134.1	A	141	7643	-6.9	254.7	A	261	7523	-6.3
	A	22	7762		135	A	142	7642	-7	255.6	A	262	7522	-6.4
	T	23	7761		136.1	T	143	7641	-6.9	256.7	T	263	7521	-6.3
	G	24	7760		137.2	T	144	7640	-6.8	257.6	A	264	7520	-6.4
	T	25	7759			G	145	7639		258.8	G	265	7519	-6.2
	T	26	7758			T	146	7638		259.7	T	266	7518	-6.3
	G	27	7757		140.4	T	147	7637	-6.6	260.9	G	267	7517	-6.1
	G	28	7756		141.8	G	148	7636	-6.2	261.6	C	268	7516	-6.4
	A	29	7755		142.6	T	149	7635	-6.4	262	C	269	7515	-7
	C	30	7754		143.5	A	150	7634	-6.5	263.4	A	270	7514	-6.6
	A	31	7753		144.8	G	151	7633	-6.2	264.6	G	271	7513	-6.4
	T	32	7752		145.5	C	152	7632	-6.5	265.4	C	272	7512	-6.6
	G	33	7751		146.8	G	153	7631	-6.2	266.6	G	273	7511	-6.4
	A	34	7750		147.4	C	154	7630	-6.6	267.4	T	274	7510	-6.6
	A	35	7749		148.3	A	155	7629	-6.7	268.4	A	275	7509	-6.6
	A	36	7748		149.2	C	156	7628	-6.8	269.4	C	276	7508	-6.6
	G	37	7747		150	C	157	7627	-7	270.5	T	277	7507	-6.5
	T	38	7746		151.2	T	158	7626	-6.8	271.6	G	278	7506	-6.4
	A	39	7745		152.5	G	159	7625	-6.5	272.7	T	279	7505	-6.3
	G	40	7744		153.7	G	160	7624	-6.3	273.6	A	280	7504	-6.4
	T	41	7743		154.7	A	161	7623	-6.3	274.7	T	281	7503	-6.3
	T	42	7742		155.7	C	162	7622	-6.3	275.8	T	282	7502	-6.2
	G	43	7741		156.6	A	163	7621	-6.4	276.9	G	283	7501	-6.1
	T	44	7740		157.8	G	164	7620	-6.2	277.9	T	284	7500	-6.1
	A	45	7739		159	G	165	7619	-6	279	G	285	7499	-6
	C	46	7738		160	A	166	7618	-6	279.8	C	286	7498	-6.2
	A	47	7737		160.8	A	167	7617	-6.2	280.7	A	287	7497	-6.3
	A	48	7736		162	A	168	7616	-6	281.8	G	288	7496	-6.2
	G	49	7735		162.6	A	169	7615	-6.4	283	A	289	7495	-6
	C	50	7734		163.8	T	170	7614	-6.2	283.6	A	290	7494	-6.4
	C	51	7733		164.9	G	171	7613	-6.1	285	A	291	7493	-6
	A	52	7732		165.7	A	172	7612	-6.3	285.5	A	292	7492	-6.5
	A	53	7731		166.7	C	173	7611	-6.3	287	C	293	7491	-6
	G	54	7730		167.8	T	174	7610	-6.2	288	C	294	7490	-6
	T	55	7729		168.6	A	175	7609	-6.4	288.3	A	295	7489	-6.7
46.4	A	56	7728	-9.6	170	A	176	7608	-6	289.4	C	296	7488	-6.6
47.9	T	57	7727	-9.1	170.8	T	177	7607	-6.2	290.2	A	297	7487	-6.8
48.8	G	58	7726	-9.2	171.7	A	178	7606	-6.3	291	G	298	7486	-7
49.6	C	59	7725	-9.4	172.7	C	179	7605	-6.3	292.3	A	299	7485	-6.7
50.5	A	60	7724	-9.5	173.5	C	180	7604	-6.5	293.4	C	300	7484	-6.6
51.2	A	61	7723	-9.8	174.4	A	181	7603	-6.6	294.3	A	301	7483	-6.7
52.7	T	62	7722	-9.3	175.7	G	182	7602	-6.3	295.3	T	302	7482	-6.7
53.9	T	63	7721	-9.1	177	G	183	7601	-6	296.3	A	303	7481	-6.7
54.7	A	64	7720	-9.3	178	T	184	7600	-6		A	304	7480	
56.4	G	65	7719	-8.6	179.1	G	185	7599	-5.9		G	305	7479	
57.1	C	66	7718	-8.9	179.7	C	186	7598	-6.3					
58.4	T	67	7717	-8.6	181	G	187	7597	-6					
59.5	T	68	7716	-8.5	181.7	C	188	7596	-6.3					
60.3	A	69	7715	-8.7	182.6	C	189	7595	-6.4					
61.4	A	70	7714	-8.6	183.6	T	190	7594	-6.4					
63.1	G	71	7713	-7.9	185	T	191	7593	-6					
64	T	72	7712	-8	185.7	A	192	7592	-6.3					
64.9	A	73	7711	-8.1	187	T	193	7591	-6					
65.7	A	74	7710	-8.3	187.8	A	194	7590	-6.2					
66.7	A	75	7709	-8.3	189	T	195	7589	-6					
67.6	A	76	7708	-8.4	190.3	G	196	7588	-5.7					
68.6	A	77	7707	-8.4	191.1	C	197	7587	-5.9					
69.5	C	78	7706	-8.5	192.3	G	198	7586	-5.7					
70.6	A	79	7705	-8.4	193	C	199	7585	-6					
71.5	A	80	7704	-8.5	193.9	C	200	7584	-6.1					
72.6	A	81	7703	-8.4	194.8	A	201	7583	-6.2					
73.5	C	82	7702	-8.5	196	A	202	7582	-6					
74.1	T	83	7701	-8.9	196.9	A	203	7581	-6.1					
74.9	A	84	7700	-9.1	198.1	G	204	7580	-5.9					
75.8	A	85	7699	-9.2	198.9	A	205	7579	-6.1					
76.9	A	86	7698	-9.1	200.2	G	206	7578	-5.8					
	A	87	7697		201.2	G	207	7577	-5.8					
	T	88	7696		202	C	208	7576	-6					
	A	89	7695		203.1	G	209	7575	-5.9					
	T	90	7694		203.6	C	210	7574	-6.4					
	G	91	7693		204.7	G	211	7573	-6.3					
	T	92	7692		205.6	C	212	7572	-6.4					
85.9	G	93	7691	-7.1	206.5	C	213	7571	-6.5					
86.5	C	94	7690	-7.5	207.5	A	214	7570	-6.5					
87.4	C	95	7689	-7.6	208.5	A	215	7569	-6.5					
87.8	T	96	7688	-8.2	210	T	216	7568	-6					
89.5	A	97	7687	-7.5	210.8	T	217	7567	-6.2					
90.5	A	98	7686	-7.5	212	G	218	7566	-6					
91.6	A	99	7685	-7.4	213	T	219	7565	-6					
92.5	A	100	7684	-7.5	214	T	220	7564	-6					
93.9	G	101	7683	-7.1	215	C	221	7563	-6					
94	C	102	7682	-8	215.8	A	222	7562	-6.2					
95.3	A	103	7681	-7.7	216.8	A	223	7561	-6.2					
96.6	G	104	7680	-7.4	217.8	A	224	7560	-6.2					
97.5	T	105	7679	-7.5	218.8	A	225	7559	-6.2					
98.6	T	106	7678	-7.4	220	T	226	7558	-6					
99.7	T	107	7677	-7.3	220.8	A	227	7557	-6.2					
100.8	T	108	7676	-7.2	222	T	228	7556	-6					
101.7	A	109	7675	-7.3	223	G	229	7555	-6					
103.1	T	110	7674	-6.9	224.1	T	230	7554	-5.9					
104.2	T	111	7673	-6.8	224.9	A	231	7553	-6.1					
105.1	A	112	7672	-6.9	226	G	232	7552	-6					
106	C	113	7671	-7	227.2	G	233	7551	-5.8					
107.2	T	114	7670	-6.8	228	A	234	7550	-6					
108	T	115	7669	-7	229.2	G	235	7549	-5.8					
109.2	A	116	7668	-6.8	229.9	C	236	7548	-6.1					
110.5	G	117	7667	-6.5	230.8	A	237	7547	-6.2					
111.8	G	118	7666	-6.2	232	G	238	7546	-6					
112.9	G	119	7665	-6.1	232.9	T	239	7545	-6.1					
113.7	A	120	7664	-6.3	234	G	240	7544	-6					

Figure SI2.2B. CE Sanger USB Indexing of the 305 bp HPV18 (7479-7783): Bottom Strand. CE position *versus* the real position of nucleotides from the bottom strand (5'-HEX) of the 305 bp HPV18 (7479-7783) fragment.

HPV18 (7647-157) 368 bp Samples A1, B1, C1 and D3

Top Strand (5'-FAM)

Correction between CE size calling to sequencing results

CE postion	Sequence	Real Position	Genome Position		CE postion	Sequence	Real Position	Genome Position		CE postion	Sequence	Real Position	Genome Position	
	T	1	7647		115.9	T	121	7767	-5.1	236.7	A	241	30	-4.3
	G	2	7648		116.8	C	122	7768	-5.2	237.8	A	242	31	-4.2
	C	3	7649		118	T	123	7769	-5	238.8	A	243	32	-4.2
	A	4	7650		118.8	A	124	7770	-5.2	240	A	244	33	-4
	T	5	7651		119.5	C	125	7771	-5.5	240.7	A	245	34	-4.3
	A	6	7652		120	C	126	7772	-6	241.7	G	246	35	-4.3
	A	7	7653		121.2	C	127	7773	-5.8	242.9	G	247	36	-4.1
	C	8	7654		122.3	T	128	7774	-5.7	243.9	G	248	37	-4.1
	T	9	7655		123.5	T	129	7775	-5.5	244.7	A	249	38	-4.3
	A	10	7656		124.7	A	130	7776	-5.3	245.9	G	250	39	-4.1
	T	11	7657		125.7	A	131	7777	-5.3	246.8	T	251	40	-4.2
	A	12	7658		126.5	C	132	7778	-5.5	247.5	A	252	41	-4.5
	T	13	7659		127.6	A	133	7779	-5.4	248.3	A	253	42	-4.7
	C	14	7660		128.5	T	134	7780	-5.5	249.6	C	254	43	-4.4
	C	15	7661		129.8	G	135	7781	-5.2	250.5	C	255	44	-4.5
	A	16	7662		130.8	A	136	7782	-5.2	251.7	G	256	45	-4.3
	C	17	7663		131.7	A	137	7783	-5.3	253.2	A	257	46	-3.8
	T	18	7664		132.5	C	138	7784	-5.5	254	A	258	47	-4
	C	19	7665		133.6	T	139	7785	-5.4	254.9	A	259	48	-4.1
	C	20	7666		134.6	A	140	7786	-5.4	255.6	A	260	49	-4.4
	C	21	7667		135.8	T	141	7787	-5.2	256.5	C	261	50	-4.5
	T	22	7668		136.7	A	142	7788	-5.3	257.6	G	262	51	-4.4
	A	23	7669		138	A	143	7789	-5	259	G	263	52	-4
	A	24	7670		138.7	T	144	7790	-5.3	259.9	T	264	53	-4.1
	G	25	7671		139.7	A	145	7791	-5.3	260.5	C	265	54	-4.5
	T	26	7672		141	T	146	7792	-5	261.6	G	266	55	-4.4
	A	27	7673		141.9	G	147	7793	-5.1	262.6	G	267	56	-4.4
	A	28	7674		142.9	A	148	7794	-5.1	263.8	G	268	57	-4.2
	T	29	7675		143.7	C	149	7795	-5.3	264	A	269	58	-5
	A	30	7676		144.8	T	150	7796	-5.2	265.6	C	270	59	-4.4
	A	31	7677		145.8	A	151	7797	-5.2	266.5	C	271	60	-4.5
	A	32	7678		146.8	A	152	7798	-5.2	267.5	G	272	61	-4.5
	A	33	7679		148	G	153	7799	-5	268.8	A	273	62	-4.2
	C	34	7680		148.7	C	154	7800	-5.3	269.6	A	274	63	-4.4
	T	35	7681		149.7	T	155	7801	-5.3	270.7	A	275	64	-4.3
	G	36	7682		151	G	156	7802	-5	271.5	A	276	65	-4.5
	C	37	7683		152.1	T	157	7803	-4.9	272.4	C	277	66	-4.6
	T	38	7684		153.2	G	158	7804	-4.8	273.6	G	278	67	-4.4
	T	39	7685		153.8	C	159	7805	-5.2	274.6	G	279	68	-4.4
	T	40	7686		154.8	A	160	7806	-5.2	275.5	T	280	69	-4.5
	T	41	7687		156	T	161	7807	-5	276.7	G	281	70	-4.3
	A	42	7688		156.9	A	162	7808	-5.1	277.4	T	282	71	-4.6
	G	43	7689		157.7	C	163	7809	-5.3	278.6	A	283	72	-4.4
	G	44	7690		158.7	A	164	7810	-5.3	279.8	T	284	73	-4.2
	C	45	7691		160	T	165	7811	-5	280.7	A	285	74	-4.3
	A	46	7692		160.9	A	166	7812	-5.1	281.5	T	286	75	-4.5
	C	47	7693		162.1	G	167	7813	-4.9	282.8	A	287	76	-4.2
	A	48	7694		163.2	T	168	7814	-4.8	283.6	A	288	77	-4.4
	T	49	7695		164.2	T	169	7815	-4.8	284.7	A	289	78	-4.3
	A	50	7696		165.2	T	170	7816	-4.8	285.7	A	290	79	-4.3
	T	51	7697		166.1	A	171	7817	-4.9	286.6	G	291	80	-4.4
	T	52	7698		167.2	T	172	7818	-4.8	287.6	A	292	81	-4.4
	T	53	7699		168.4	G	173	7819	-4.6	288.4	T	293	82	-4.6
	T	54	7700		169.2	C	174	7820	-4.8	289.6	G	294	83	-4.4
	A	55	7701		170.1	A	175	7821	-4.9	290.8	T	295	84	-4.2
47.5	G	56	7702	-8.5	171	A	176	7822	-5	291.6	G	296	85	-4.4
48.6	T	57	7703	-8.4	171.9	C	177	7823	-5.1	292.8	A	297	86	-4.2
49.5	T	58	7704	-8.5	173	C	178	7824	-5	293.7	G	298	87	-4.3
50.2	T	59	7705	-8.8	174.1	G	179	7825	-4.9	294.7	A	299	88	-4.3
51.4	G	60	7706	-8.6	175	A	180	7826	-5	295.6	A	300	89	-4.4
52.2	T	61	7707	-8.8	176	A	181	7827	-5	296.3	A	301	90	-4.7
54	T	62	7708	-8	177	A	182	7828	-5	297.2	C	302	91	-4.8
55.1	T	63	7709	-7.9	178.1	T	183	7829	-4.9	298.1	A	303	92	-4.9
56.2	T	64	7710	-7.8	179.1	A	184	7830	-4.9	299	C	304	93	-5
56.9	T	65	7711	-8.1	180.3	G	185	7831	-4.7	300	A	305	94	-5
57.6	A	66	7712	-8.4	181.4	G	186	7832	-4.6	300.9	C	306	95	-5.1
58.3	C	67	7713	-8.7	182.4	T	187	7833	-4.6	301.8	C	307	96	-5.2
59.2	T	68	7714	-8.8	183.4	T	188	7834	-4.6	302.8	A	308	97	-5.2
60.5	T	69	7715	-8.5	184.4	G	189	7835	-4.6	303.8	C	309	98	-5.2
61.8	A	70	7716	-8.2	185.5	G	190	7836	-4.5	304.8	A	310	99	-5.2
62.8	A	71	7717	-8.2	186.6	G	191	7837	-4.4	305.6	A	311	100	-5.4
64.3	G	72	7718	-7.7	188	C	192	7838	-4	306.8	T	312	101	-5.2
65	C	73	7719	-8	188.3	A	193	7839	-4.7	307.9	A	313	102	-5.1
66.1	T	74	7720	-7.9	189.5	G	194	7840	-4.5	308.8	C	314	103	-5.2
66.9	A	75	7721	-8.1	190.3	C	195	7841	-4.7	309.6	T	315	104	-5.4
68.1	A	76	7722	-7.9	191.2	A	196	7842	-4.8	310.8	A	316	105	-5.2
69.3	T	77	7723	-7.7	192.1	C	197	7843	-4.9	311.6	T	317	106	-5.4
70.5	T	78	7724	-7.5	193	A	198	7844	-5	313.1	G	318	107	-4.9
71.9	G	79	7725	-7.1	194.2	T	199	7845	-4.8	314.2	G	319	108	-4.8
72.6	C	80	7726	-7.4	195.2	A	200	7846	-4.8	315.1	C	320	109	-4.9
73.6	A	81	7727	-7.4	196.1	C	201	7847	-4.9	316.3	G	321	110	-4.7
74.3	T	82	7728	-7.7	197.6	T	202	7848	-4.4	317	C	322	111	-5
75.7	A	83	7729	-7.3	198.2	A	203	7849	-4.8	318.2	G	323	112	-4.8
76.5	C	84	7730	-7.5	199.6	T	204	7850	-4.4	319	C	324	113	-5
77.8	T	85	7731	-7.2	200.4	A	205	7851	-4.6	319.9	T	325	114	-5.1
79	T	86	7732	-7	201.3	C	206	7852	-4.7	320.7	T	326	115	-5.3
80	G	87	7733	-7	202.4	T	207	7853	-4.6	321.6	T	327	116	-5.4
81.6	G	88	7734	-6.4	203.5	T	208	7854	-4.5	323.3	G	328	117	-4.7
82.4	C	89	7735	-6.6	204.3	T	209	7855	-4.7	324.2	A	329	118	-4.8
83.5	T	90	7736	-6.5	205.6	T	210	7856	-4.4	325.4	G	330	119	-4.6
84.7	T	91	7737	-6.3	206.5	C	211	7857	-4.5	326.4	G	331	120	-4.6
85.7	G	92	7738	-6.3	207.4	A	212	1	-4.6	327.4	A	332	121	-4.6
86.4	T	93	7739	-6.6	208.4	T	213	2	-4.6	328.8	T	333	122	-4.2
87.8	A	94	7740	-6.2	209.6	T	214	3	-4.4	329.3	C	334	123	-4.7
88.6	C	95	7741	-6.4	210.6	A	215	4	-4.4	330.1	C	335	124	-4.9
89.6	A	96	7742	-6.4	211.5	A	216	5	-4.5	331	A	336	125	-5
90.5	A	97	7743	-6.5	212.8	T	217	6	-4.2	332.1	A	337	126	-4.9
91.4	C	98	7744	-6.6	213.6	A	218	7	-4.4	333	C	338	127	-5
92.5	T	99	7745	-6.5	214.5	C	219	8	-4.5	333.8	A	339	128	-5.2
93.6	A	100	7746	-6.4	215.5	T	220	9	-4.5	334.9	C	340	129	-5.1
94.5	C	101	7747	-6.5	217	T	221	10	-4	336	G	341	130	-5
96	T	102	7748	-6	217.7	T	222	11	-4.3	337.1	G	342	131	-4.9
96.8	T	103	7749	-6.2	218.7	T	223	12	-4.3	338.1	C	343	132	-4.9
98	T	104	7750	-6	219.8	A	224	13	-4.2	339.1	G	344	133	-4.9
99	C	105	7751	-6	221	A	225	14	-4	340	A	345	134	-5
100	A	106	7752	-6	221.6	C	226	15	-4.4	340.9	C	346	135	-5.1
101	T	107	7753	-6	222.6	A	227	16	-4.4	341.8	C	347	136	-5.2
102.4	G	108	7754	-5.6	223.4	A	228	17	-4.6	342.7	C	348	137	-5.3
103.6	T	109	7755	-5.4	224.2	T	229	18	-4.8	343.8	T	349	138	-5.2
104.4	C	110	7756	-5.6	225.7	T	230	19	-4.3	344.8	A	350	139	-5.2
105.2	C	111	7757	-5.8	226.8	G	231	20	-4.2	346	C	351	140	-5
106.2	A	112	7758	-5.8	227.7	T	232	21	-4.3	347.3	A	352	141	-4.7
107.2	A	113	7759	-5.8	228.6	A	233	22	-4.4	348.1	A	353	142	-4.9
108.1	C	114	7760	-5.9	229.9	G	234	23	-4.1	349	G	354	143	-5
109.2	A	115	7761	-5.8	230.5	T	235	24	-4.5	349.8	C	355	144	-5.2
110.3	T	116	7762	-5.7	231.7	A	236	25	-4.3	350.8	T	356	145	-5.2
111.5	T	117	7763	-5.5	232.9	T	237	26	-4.1	351.8	A	357	146	-5.2
112.4	C	118	7764	-5.6	233.5	A	238	27	-4.5	352.8	C	358	147	-5.2
113.6	T	119	7765	-5.4	234.7	T	239	28	-4.3	353.7	C	359	148	-5.3
114.8	G	120	7766	-5.2	235.8	A	240	29	-4.2	354.8	T	360	149	-5.2
										355.9	G	361	150	-5.1
										357	A	362	151	-5
										357.8	T	363	152	-5.2
										358.9	C	364	153	-5.1
										360.1	T	365	154	-4.9
										361.1	G	366	155	-4.9
											T	367	156	
											G	368	157	

Figure SI2.2C. CE Sanger USB Indexing of the 368 bp HPV18 (7647-157): Top Strand. CE position *versus* the real position of nucleotides from the top strand (5'-FAM) of the 368 bp HPV18 (7647-157) fragment.

HPV18 (7647-157) 368 bp Samples E1, F1, G1 and H1
Bottom Strand (5'-HEX)
Correction between CE size calling to sequencing results

CE postion	Sequence	Real Position	Genome Position		CE postion	Sequence	Real Position	Genome Position		CE postion	Sequence	Real Position	Genome Position	
	C	1	157		116.2	C	121	37	-4.8	237	A	241	7774	-4
	A	2	156		117	C	122	36	-5	238.3	G	242	7773	-3.7
	C	3	155		117.9	C	123	35	-5.1	239.3	G	243	7772	-3.7
	A	4	154		119	T	124	34	-5	240	G	244	7771	-4
	G	5	153		120.3	T	125	33	-4.7	241.4	T	245	7770	-3.6
	A	6	152		121.5	T	126	32	-4.5	242.4	A	246	7769	-3.6
	T	7	151		122.6	T	127	31	-4.4	243.4	G	247	7768	-3.6
	C	8	150		123.7	T	128	30	-4.3	244.4	A	248	7767	-3.6
	A	9	149		124.8	T	129	29	-4.2	245.2	C	249	7766	-3.8
	G	10	148		125.7	A	130	28	-4.3	246.3	A	250	7765	-3.7
	G	11	147		126.7	T	131	27	-4.3	247.3	G	251	7764	-3.7
	T	12	146		127.7	A	132	26	-4.3	248.4	A	252	7763	-3.6
	A	13	145		128.7	T	133	25	-4.3	249	A	253	7762	-4
	G	14	144		129.6	A	134	24	-4.4	250.3	T	254	7761	-3.7
	C	15	143		130.3	C	135	23	-4.7	251.3	G	255	7760	-3.7
	T	16	142		131.5	T	136	22	-4.5	252	T	256	7759	-4
	T	17	141		132.5	A	137	21	-4.5	253.3	T	257	7758	-3.7
	G	18	140		133.3	C	138	20	-4.7	254.4	G	258	7757	-3.6
	T	19	139		134.4	A	139	19	-4.6	255	G	259	7756	-4
	A	20	138		135.3	A	140	18	-4.7	256.5	A	260	7755	-3.5
	G	21	137		136.4	T	141	17	-4.6	257.3	C	261	7754	-3.7
	G	22	136		137.4	T	142	16	-4.6	258.4	A	262	7753	-3.6
	G	23	135		138.5	G	143	15	-4.5	259.3	T	263	7752	-3.7
	T	24	134		139.7	T	144	14	-4.3	260.3	G	264	7751	-3.7
	C	25	133		140.7	T	145	13	-4.3	261	A	265	7750	-4
	G	26	132		141.7	A	146	12	-4.3	262.3	A	266	7749	-3.7
	C	27	131		142.6	A	147	11	-4.4	263	A	267	7748	-4
	C	28	130		143.4	A	148	10	-4.6	264.3	G	268	7747	-3.7
	G	29	129		144.4	A	149	9	-4.6	265.3	T	269	7746	-3.7
	T	30	128		145.6	G	150	8	-4.4	266.3	A	270	7745	-3.7
	G	31	127		146.6	T	151	7	-4.4	267.3	G	271	7744	-3.7
	T	32	126		147.5	A	152	6	-4.5	268	T	272	7743	-4
	T	33	125		148.5	T	153	5	-4.5	269.3	T	273	7742	-3.7
	G	34	124		149.5	T	154	4	-4.5	270.3	G	274	7741	-3.7
	G	35	123		150.5	A	155	3	-4.5	271.4	T	275	7740	-3.6
	A	36	122		151.4	A	156	2	-4.6	272.3	A	276	7739	-3.7
	T	37	121		152.6	T	157	1	-4.4	273.2	C	277	7738	-3.8
	C	38	120		153.8	G	158	7857	-4.2	274	A	278	7737	-4
	C	39	119		154.8	A	159	7856	-4.2	275.1	A	279	7736	-3.9
	T	40	118		155.8	A	160	7855	-4.2	276.2	G	280	7735	-3.8
	C	41	117		156.5	A	161	7854	-4.5	277	C	281	7734	-4
	A	42	116		157.5	A	162	7853	-4.5	278	C	282	7733	-4
	A	43	115		158.7	G	163	7852	-4.3	279	A	283	7732	-4
	A	44	114		159.6	T	164	7851	-4.4	280	A	284	7731	-4
	G	45	113		160.6	A	165	7850	-4.4	281.1	G	285	7730	-3.9
	C	46	112		161.5	T	166	7849	-4.5	282.1	T	286	7729	-3.9
	G	47	111		162.5	A	167	7848	-4.5	283.1	A	287	7728	-3.9
	C	48	110		163.7	G	168	7847	-4.3	284	T	288	7727	-4
	G	49	109		164.7	T	169	7846	-4.3	285.1	G	289	7726	-3.9
	C	50	108		165.6	A	170	7845	-4.4	286	C	290	7725	-4
	C	51	107		166.6	T	171	7844	-4.4	287	A	291	7724	-4
	A	52	106		167.7	G	172	7843	-4.3	288	A	292	7723	-4
	T	53	105		168.7	T	173	7842	-4.3	289	T	293	7722	-4
	A	54	104		169.9	G	174	7841	-4.1	289.9	T	294	7721	-4.1
47.4	G	55	103	-7.6	170.7	C	175	7840	-4.3	290.9	A	295	7720	-4.1
48.5	T	56	102	-7.5	171.7	T	176	7839	-4.3	292	G	296	7719	-4
49.2	A	57	101	-7.8	172.8	G	177	7838	-4.2	293	C	297	7718	-4
50.1	T	58	100	-7.9	173.6	C	178	7837	-4.4	294	T	298	7717	-4
51.1	T	59	99	-7.9	174.5	C	179	7836	-4.5	295	T	299	7716	-4
52.4	G	60	98	-7.6	175.4	C	180	7835	-4.6	296	A	300	7715	-4
53.3	T	61	97	-7.7	176.4	A	181	7834	-4.6	297	A	301	7714	-4
54.8	G	62	96	-7.2	177.4	A	182	7833	-4.6	298	G	302	7713	-4
56	G	63	95	-7	178.3	C	183	7832	-4.7	299	T	303	7712	-4
57	T	64	94	-7	179.2	C	184	7831	-4.8	300.1	A	304	7711	-3.9
58.3	G	65	93	-6.7	180.4	T	185	7830	-4.6	300.9	A	305	7710	-4.1
59.2	T	66	92	-6.8	181.5	A	186	7829	-4.5	301.9	A	306	7709	-4.1
60.4	G	67	91	-6.6	182.6	T	187	7828	-4.4	302.9	A	307	7708	-4.1
61.3	T	68	90	-6.7	183.7	T	188	7827	-4.3	303.8	A	308	7707	-4.2
62.3	T	69	89	-6.7	184.8	T	189	7826	-4.2	304.6	C	309	7706	-4.4
63.4	T	70	88	-6.6	185.7	C	190	7825	-4.3	305.7	A	310	7705	-4.3
64	C	71	87	-7	187	G	191	7824	-4	306.6	A	311	7704	-4.4
65.2	T	72	86	-6.8	188	G	192	7823	-4	307.6	A	312	7703	-4.4
65.9	C	73	85	-7.1	189.1	T	193	7822	-3.9	308.4	C	313	7702	-4.6
67	A	74	84	-7	190.1	T	194	7821	-3.9	309.6	T	314	7701	-4.4
67.8	C	75	83	-7.2	191.1	G	195	7820	-3.9	310.5	A	315	7700	-4.5
69	A	76	82	-7	192	C	196	7819	-4	311.6	A	316	7699	-4.4
70.3	T	77	81	-6.7	193	A	197	7818	-4	312.6	A	317	7698	-4.4
71.1	C	78	80	-6.9	194	T	198	7817	-4	313.6	A	318	7697	-4.4
72.4	T	79	79	-6.6	195	A	199	7816	-4	314.5	T	319	7696	-4.5
73.6	T	80	78	-6.4	196	A	200	7815	-4	315.7	A	320	7695	-4.3
74.9	T	81	77	-6.1	196.9	A	201	7814	-4.1	316.7	T	321	7694	-4.3
76	T	82	76	-6	197.8	C	202	7813	-4.2	317.8	G	322	7693	-4.2
77	A	83	75	-6	198.9	T	203	7812	-4.1	318.8	T	323	7692	-4.2
78.2	T	84	74	-5.8	199.9	A	204	7811	-4.1	319.9	G	324	7691	-4.1
79.2	A	85	73	-5.8	201	T	205	7810	-4	321	C	325	7690	-4
80.3	T	86	72	-5.7	202	G	206	7809	-4	321.8	C	326	7689	-4.2
81.2	A	87	71	-5.8	203.1	T	207	7808	-3.9	322.7	T	327	7688	-4.3
82.1	C	88	70	-5.9	204	A	208	7807	-4	323.9	A	328	7687	-4.1
83.1	A	89	69	-5.9	205.1	T	209	7806	-3.9	324.8	A	329	7686	-4.2
83.8	C	90	68	-6.2	206.2	G	210	7805	-3.8	325.9	A	330	7685	-4.1
84.7	C	91	67	-6.3	207.1	C	211	7804	-3.9	326.9	A	331	7684	-4.1
86.2	G	92	66	-5.8	208	A	212	7803	-4	327.8	G	332	7683	-4.2
87.3	T	93	65	-5.7	208.8	C	213	7802	-4.2	328.7	C	333	7682	-4.3
88	T	94	64	-6	209.8	A	214	7801	-4.2	330	A	334	7681	-4
89.5	T	95	63	-5.5	210.8	G	215	7800	-4.2	330.7	G	335	7680	-4.3
91	T	96	62	-5	211.7	C	216	7799	-4.3	332	T	336	7679	-4
91.5	C	97	61	-5.5	212.8	T	217	7798	-4.2	333	T	337	7678	-4
92.8	G	98	60	-5.2	213.8	T	218	7797	-4.2	333.7	T	338	7677	-4.3
93.8	G	99	59	-5.2	214.9	A	219	7796	-4.1	335	T	339	7676	-4
94.9	T	100	58	-5.1	216	G	220	7795	-4	335.8	A	340	7675	-4.2
95.8	C	101	57	-5.2	217.2	T	221	7794	-3.8	337	T	341	7674	-4
97	C	102	56	-5	218	C	222	7793	-4	337.7	T	342	7673	-4.3
97.4	C	103	55	-5.6	219	A	223	7792	-4	338.8	A	343	7672	-4.2
98.7	G	104	54	-5.3	220.1	T	224	7791	-3.9	339.7	C	344	7671	-4.3
99.7	A	105	53	-5.3	221.1	A	225	7790	-3.9	341	T	345	7670	-4
100.5	C	106	52	-5.5	222.1	T	226	7789	-3.9	342	T	346	7669	-4
101.5	C	107	51	-5.5	223.1	T	227	7788	-3.9	343	A	347	7668	-4
102.7	G	108	50	-5.3	224.1	A	228	7787	-3.9	343.8	G	348	7667	-4.2
103.9	T	109	49	-5.1	225.2	T	229	7786	-3.8	345	G	349	7666	-4
105	T	110	48	-5	226.1	A	230	7785	-3.9	346	G	350	7665	-4
106.2	T	111	47	-4.8	227.3	G	231	7784	-3.7	347	A	351	7664	-4
107.3	T	112	46	-4.7	228	T	232	7783	-4	347.9	G	352	7663	-4.1
108.2	C	113	45	-4.8	229.3	T	233	7782	-3.7	348.9	T	353	7662	-4.1
109	G	114	44	-5	230.1	C	234	7781	-3.9	350	G	354	7661	-4
110.4	G	115	43	-4.6	231	A	235	7780	-4	351.1	G	355	7660	-3.9
111	T	116	42	-5	232.1	T	236	7779	-3.9	352	A	356	7659	-4
112.6	T	117	41	-4.4	233.2	G	237	7778	-3.8	353	T	357	7658	-4
113.5	A	118	40	-4.5	234.3	T	238	7777	-3.7	354	A	358	7657	-4
114.4	C	119	39	-4.6	235.3	T	239	7776	-3.7	355	T	359	7656	-4
115.4	T	120	38	-4.6	236.2	A	240	7775	-3.8	356	A	360	7655	-4
										357	G	361	7654	-4
										358	T	362	7653	-4
										359	T	363	7652	-4
										360	A	364	7651	-4
										361	T	365	7650	-4
										362	G	366	7649	-4
											C	367	7648	
											A	368	7647	

Figure SI2.2D. CE Sanger USB Indexing of the 368 bp HPV18 (7647-157): Bottom Strand. CE position *versus* the real position of nucleotides from the bottom strand (5'-HEX) of the 368 bp HPV18 (7647-157) fragment.

HPV18 (7479-7783) 305 bp (FAM Lanes A10, C10 and G12)
Correction between CE size calling to Maxam-Gilbert sequencing results

CE postion	Sequence	Real Position	Genome Position		CE postion	Sequence	Real Position	Genome Position		CE postion	Sequence	Real Position	Genome Position	
	C	1	7479		113.7	C	121	7599	-7.3	235	C	241	7719	-6
	T	2	7480		114.4	A	122	7600	-7.6	235.8	T	242	7720	-6.2
	T	3	7481		115.3	C	123	7601	-7.7	237	A	243	7721	-6
	A	4	7482		116.2	C	124	7602	-7.8	238	A	244	7722	-6
	T	5	7483		117.1	T	125	7603	-7.9	238.8	T	245	7723	-6.2
	G	6	7484		118.3	G	126	7604	-7.7	239.9	T	246	7724	-6.1
	T	7	7485		119.6	G	127	7605	-7.4	241.1	G	247	7725	-5.9
	C	8	7486		120.8	T	128	7606	-7.2	242.2	C	248	7726	-5.8
	T	9	7487		121.8	A	129	7607	-7.2	243.1	A	249	7727	-5.9
	G	10	7488		122.7	T	130	7608	-7.3	244	T	250	7728	-6
	T	11	7489		123.7	T	131	7609	-7.3	245.2	A	251	7729	-5.8
	G	12	7490		124.7	A	132	7610	-7.3	246.1	C	252	7730	-5.9
	G	13	7491		125.6	G	133	7611	-7.4	246.9	T	253	7731	-6.1
	T	14	7492		127	T	134	7612	-7	248.2	T	254	7732	-5.8
	T	15	7493		127.9	C	135	7613	-7.1	249.4	G	255	7733	-5.6
	T	16	7494		128.6	A	136	7614	-7.4	250.5	G	256	7734	-5.5
	T	17	7495		129.6	T	137	7615	-7.4	251.7	C	257	7735	-5.3
	C	18	7496		130.7	T	138	7616	-7.3	252.6	T	258	7736	-5.4
	T	19	7497		131.7	T	139	7617	-7.3	253.6	T	259	7737	-5.4
	G	20	7498		132.8	T	140	7618	-7.2	254.6	G	260	7738	-5.4
	C	21	7499		133.8	C	141	7619	-7.2	255.7	T	261	7739	-5.3
	A	22	7500		134.6	C	142	7620	-7.4	256.6	A	262	7740	-5.4
	C	23	7501		135.3	T	143	7621	-7.7	257.5	C	263	7741	-5.5
	A	24	7502		136.4	G	144	7622	-7.6	258.5	A	264	7742	-5.5
	A	25	7503		137.9	T	145	7623	-7.1	259.4	A	265	7743	-5.6
	T	26	7504		138.8	C	146	7624	-7.2	260.2	C	266	7744	-5.8
	A	27	7505		139.6	C	147	7625	-7.4	261.2	T	267	7745	-5.8
	C	28	7506		140.5	A	148	7626	-7.5	262.4	A	268	7746	-5.6
	A	29	7507		141.5	G	149	7627	-7.5	263.3	C	269	7747	-5.7
	G	30	7508		142.8	G	150	7628	-7.2	264.1	T	270	7748	-5.9
	T	31	7509		143.9	T	151	7629	-7.1	265.4	T	271	7749	-5.6
	A	32	7510		145	G	152	7630	-7	266.4	T	272	7750	-5.6
	C	33	7511		146.2	C	153	7631	-6.8	267.5	C	273	7751	-5.5
	G	34	7512		147	G	154	7632	-7	268.5	A	274	7752	-5.5
	C	35	7513			C	155	7633		269.2	T	275	7753	-5.8
20.5	T	36	7514	-15.5		T	156	7634		270.5	G	276	7754	-5.5
21.8	G	37	7515	-15.2	149.9	A	157	7635	-7.1	271.7	T	277	7755	-5.3
23.3	G	38	7516	-14.7	150.8	C	158	7636	-7.2	272.7	C	278	7756	-5.3
24.2	C	39	7517		151.8	A	159	7637	-7.2	273.6	C	279	7757	-5.4
25.1	A	40	7518	-14.9	152.9	A	160	7638	-7.1	274.5	A	280	7758	-5.5
26.3	C	41	7519	-14.7	153.7	C	161	7639	-7.3	275.5	A	281	7759	-5.5
26.6	T	42	7520	-15.4	154.7	A	162	7640	-7.3	276.4	C	282	7760	-5.6
27.9	A	43	7521	-15.1	155.8	A	163	7641	-7.2	277.4	A	283	7761	-5.6
29	T	44	7522	-15	156.7	T	164	7642	-7.3	278.3	T	284	7762	-5.7
30.2	T	45	7523	-14.8	157.8	T	165	7643	-7.2	279.6	T	285	7763	-5.4
31.5	G	46	7524	-14.5	159.1	G	166	7644	-6.9	280.5	C	286	7764	-5.5
	C	47	7525		160.3	C	167	7645	-6.7	281.4	T	287	7765	-5.6
34.1	A	48	7526	-13.9	161.1	T	168	7646	-6.9	282.6	G	288	7766	-5.4
34.7	A	49	7527	-14.3	162.2	T	169	7647	-6.8	283.7	T	289	7767	-5.3
35.3	A	50	7528	-14.7	163.3	G	170	7648	-6.7	284.8	C	290	7768	-5.2
36.2	C	51	7529	-14.8	164.4	C	171	7649	-6.6	285.6	T	291	7769	-5.4
36.9	T	52	7530	-15.1	165.3	A	172	7650	-6.7	286.8	A	292	7770	-5.2
38	T	53	7531	-15	166	T	173	7651	-7	287.7	C	293	7771	-5.3
39.1	T	54	7532	-14.9	167.2	A	174	7652	-6.8	288.4	C	294	7772	-5.6
40.2	A	55	7533	-14.8	168.2	A	175	7653	-6.8		C	295	7773	
41.2	A	56	7534	-14.8	169	C	176	7654	-7		T	296	7774	
42.2	T	57	7535	-14.8	169.9	T	177	7655	-7.1		T	297	7775	
43.6	C	58	7536	-14.4	171.2	A	178	7656	-6.8		A	298	7776	
44.5	T	59	7537	-14.5	172	T	179	7657	-7		A	299	7777	
46	T	60	7538	-14	173.2	A	180	7658	-6.8		C	300	7778	
47.4	T	61	7539	-13.6	174.1	T	181	7659	-6.9		A	301	7779	
48.7	T	62	7540	-13.3	175.2	C	182	7660	-6.8		T	302	7780	
49.9	G	63	7541	-13.1	176.1	C	183	7661	-6.9		G	303	7781	
51.9	G	64	7542	-12.1	177	A	184	7662	-7		A	304	7782	
53.7	G	65	7543	-11.3	177.9	C	185	7663	-7.1		A	305	7783	
54.9	C	66	7544	-11.1	178.8	T	186	7664	-7.2					
55.7	A	67	7545	-11.3	180	C	187	7665	-7					
56.4	C	68	7546	-11.6	180.9	C	188	7666	-7.1					
57.2	T	69	7547	-11.8	181.9	C	189	7667	-7.1					
58.5	G	70	7548	-11.5	182.7	T	190	7668	-7.3					
60	C	71	7549	-11	184.1	A	191	7669	-6.9					
60.7	T	72	7550	-11.3	185.1	A	192	7670	-6.9					
62	C	73	7551	-11	186	G	193	7671	-7					
62.8	C	74	7552	-11.2	187.2	T	194	7672	-6.8					
63.8	T	75	7553	-11.2	188.4	A	195	7673	-6.6					
65	A	76	7554	-11	189.3	A	196	7674	-6.7					
66.2	C	77	7555	-10.8	190.1	T	197	7675	-6.9					
67	A	78	7556	-11	191.2	A	198	7676	-6.8					
68.3	T	79	7557	-10.7	192.3	A	199	7677	-6.7					
69.4	A	80	7558	-10.6	193.2	A	200	7678	-6.8					
70.6	T	81	7559	-10.4	194.1	A	201	7679	-6.9					
71.9	T	82	7560	-10.1	195	C	202	7680	-7					
73.1	T	83	7561	-9.9	195.9	T	203	7681	-7.1					
74.2	T	84	7562	-9.8	197.2	G	204	7682	-6.8					
75.3	G	85	7563	-9.7	198.4	C	205	7683	-6.6					
76.9	A	86	7564	-9.1	199.3	T	206	7684	-6.7					
77.7	A	87	7565	-9.3	200.3	T	207	7685	-6.7					
78.6	C	88	7566	-9.4	201.3	T	208	7686	-6.7					
79.4	A	89	7567	-9.6	202.5	T	209	7687	-6.5					
80.4	A	90	7568	-9.6	203.6	A	210	7688	-6.4					
81.4	T	91	7569	-9.6	204.5	G	211	7689	-6.5					
82.5	T	92	7570	-9.5	205.6	G	212	7690	-6.4					
83.6	G	93	7571	-9.4	206.7	C	213	7691	-6.3					
85	G	94	7572	-9	207.6	A	214	7692	-6.4					
86.2	C	95	7573	-8.8	208.4	C	215	7693	-6.6					
86.9	G	96	7574	-9.1	209.3	A	216	7694	-6.7					
88	C	97	7575	-9	210.3	T	217	7695	-6.7					
89	G	98	7576	-9	211.5	A	218	7696	-6.5					
90.4	C	99	7577	-8.6	212.4	T	219	7697	-6.6					
90.9	C	100	7578	-9.1	213.5	T	220	7698	-6.5					
91.7	T	101	7579	-9.3	214.5	T	221	7699	-6.5					
93	C	102	7580	-9	215.6	T	222	7700	-6.4					
94	T	103	7581	-9	216.7	A	223	7701	-6.3					
95.7	T	104	7582	-8.3	217.6	G	224	7702	-6.4					
97.3	T	105	7583	-7.7	218.9	T	225	7703	-6.1					
98.3	G	106	7584	-7.7	219.9	T	226	7704	-6.1					
99.4	G	107	7585	-7.6	220.9	T	227	7705	-6.1					
100.5	C	108	7586	-7.5	222	G	228	7706	-6					
101.7	G	109	7587	-7.3	223.1	T	229	7707	-5.9					
102.4	C	110	7588	-7.6	224	T	230	7708	-6					
103.4	A	111	7589	-7.6	225.1	T	231	7709	-5.9					
104.4	T	112	7590	-7.6	226	T	232	7710	-6					
	A	113	7591		227	T	233	7711	-6					
	T	114	7592		228.2	A	234	7712	-5.8					
107.1	A	115	7593	-7.9	229	C	235	7713	-6					
108.1	A	116	7594	-7.9	229.7	T	236	7714	-6.3					
109.1	G	117	7595	-7.9	230.9	T	237	7715	-6.1					
110.4	G	118	7596	-7.6	232.1	A	238	7716	-5.9					
111.6	C	119	7597	-7.4	233	A	239	7717	-6					
112.5	G	120	7598	-7.5	233.9	G	240	7718	-6.1					

Figure SI2.3A. CE Maxam-Gilbert Indexing of the 305 bp HPV18 (7479-7783): Top Strand. CE position *versus* the real position of nucleotides from the top strand (5'-FAM) of the 305 bp HPV18 (7479-7783) fragment.

HPV18 (7479-7783) 305 bp (HEX Lanes A10 and G12)
Correction between CE size calling to Maxam-Gilbert sequencing results

CE postion	Sequence	Real Position	Genome Position		CE postion	Sequence	Real Position	Genome Position		CE postion	Sequence	Real Position	Genome Position	
	T	1	7783		112.6	G	121	7663	-8.4	234	C	241	7543	-7
	T	2	7782		113.8	T	122	7662	-8.2	234.9	C	242	7542	-7.1
	C	3	7781		114.9	G	123	7661	-8.1	235.9	C	243	7541	-7.1
	A	4	7780		116.1	G	124	7660	-7.9	236.9	A	244	7540	-7.1
	T	5	7779		117.2	A	125	7659	-7.8	237.8	A	245	7539	-7.2
	G	6	7778		118	T	126	7658	-8	238.8	A	246	7538	-7.2
	T	7	7777		119	A	127	7657	-8	239.8	A	247	7537	-7.2
	T	8	7776		119.9	T	128	7656	-8.1	240.8	G	248	7536	-7.2
	A	9	7775		120.8	A	129	7655	-8.2	242	A	249	7535	-7
	A	10	7774		121.7	G	130	7654	-8.3	242.8	T	250	7534	-7.2
	G	11	7773		123.1	T	131	7653	-7.9	243.9	T	251	7533	-7.1
	G	12	7772		124	T	132	7652	-8	245.1	A	252	7532	-6.9
	G	13	7771		125	A	133	7651	-8	246	A	253	7531	-7
	T	14	7770		125.9	T	134	7650	-8.1	246.9	A	254	7530	-7.1
	A	15	7769		126.9	G	135	7649	-8.1	247.9	G	255	7529	-7.1
	G	16	7768		128.2	C	136	7648	-7.8	248.9	T	256	7528	-7.1
	A	17	7767		128.9	A	137	7647	-8.1	250	T	257	7527	-7
	C	18	7766		129.8	A	138	7646	-8.2	251.1	T	258	7526	-6.9
	A	19	7765		130.7	G	139	7645	-8.3	252.2	G	259	7525	-6.8
	G	20	7764		131.9	C	140	7644	-8.1	253.2	C	260	7524	-6.8
	A	21	7763		132.7	A	141	7643	-8.3	254.1	A	261	7523	-6.9
	A	22	7762		133.6	A	142	7642	-8.4	255	A	262	7522	-7
	T	23	7761		134.5	T	143	7641	-8.5	256	T	263	7521	-7
	G	24	7760		135.8	T	144	7640	-8.2	257	A	264	7520	-7
	T	25	7759		136.8	G	145	7639	-8.2	257.9	G	265	7519	-7.1
	T	26	7758		137.9	T	146	7638	-8.1	259	T	266	7518	-7
	G	27	7757		139.1	T	147	7637	-7.9	260.2	G	267	7517	-6.8
	G	28	7756		140.1	G	148	7636	-7.9	261.3	C	268	7516	-6.7
	A	29	7755		141.3	T	149	7635	-7.7	261.7	C	269	7515	-7.3
	C	30	7754		142.3	A	150	7634	-7.7	262.9	A	270	7514	-7.1
	A	31	7753		143.3	G	151	7633	-7.7	263.9	G	271	7513	-7.1
	T	32	7752		144.4	C	152	7632	-7.6	264.9	C	272	7512	-7.1
	G	33	7751		145.2	G	153	7631	-7.8	265.9	G	273	7511	-7.1
	A	34	7750		146.6	C	154	7630	-7.4	266.9	T	274	7510	-7.1
	A	35	7749		147.3	A	155	7629	-7.7	268.1	A	275	7509	-6.9
	A	36	7748		148.2	C	156	7628	-7.8	268.9	C	276	7508	-7.1
	G	37	7747		149.1	C	157	7627	-7.9	269.7	T	277	7507	-7.3
	T	38	7746		150	T	158	7626	-8	271	G	278	7506	-7
	A	39	7745		151.2	G	159	7625	-7.8	272	T	279	7505	-7
	G	40	7744		152.5	G	160	7624	-7.5	273.2	A	280	7504	-6.8
	T	41	7743		153.7	A	161	7623	-7.3	274	T	281	7503	-7
	T	42	7742		154.6	C	162	7622	-7.4	275.1	T	282	7502	-6.9
	G	43	7741		155.6	A	163	7621	-7.4	276.2	G	283	7501	-6.8
	T	44	7740		156.6	G	164	7620	-7.4	277.3	T	284	7500	-6.7
	A	45	7739		157.8	G	165	7619	-7.2	278.4	G	285	7499	-6.6
	C	46	7738		158.9	A	166	7618	-7.1	279.3	C	286	7498	-6.7
32	A	47	7737	-15	159.9	A	167	7617	-7.1	280.4	A	287	7497	-6.6
33.5	A	48	7736	-14.5	160.8	A	168	7616	-7.2	281.2	G	288	7496	-6.8
34.2	G	49	7735	-14.8	161.7	A	169	7615	-7.3	282.3	A	289	7495	-6.7
35.4	C	50	7734	-14.6	162.5	T	170	7614	-7.5	283.2	A	290	7494	-6.8
36	C	51	7733	-15	163.7	G	171	7613	-7.3	284.1	A	291	7493	-6.9
36.6	A	52	7732	-15.4	165	A	172	7612	-7	285	A	292	7492	-7
37.5	A	53	7731	-15.5	165.8	C	173	7611	-7.2	286.1	C	293	7491	-6.9
38.3	G	54	7730	-15.7	166.7	T	174	7610	-7.3		C	294	7490	-294
39.7	T	55	7729	-15.3	167.9	A	175	7609	-7.1	287.8	A	295	7489	-7.2
40.9	A	56	7728	-15.1	168.8	A	176	7608	-7.2		C	296	7488	-296
42	T	57	7727	-15	169.7	T	177	7607	-7.3	289.7	A	297	7487	-7.3
43.3	G	58	7726	-14.7	170.9	A	178	7606	-7.1	290.7	G	298	7486	-7.3
45.1	C	59	7725	-13.9	171.8	C	179	7605	-7.2	292	A	299	7485	-7
45.7	A	60	7724	-14.3	172.6	C	180	7604	-7.4	292.8	C	300	7484	-7.2
46.7	A	61	7723	-14.3	173.6	A	181	7603	-7.4	293.1	A	301	7483	-7.9
47.7	T	62	7722	-14.3	174.6	G	182	7602	-7.4		T	302	7482	-302
49.1	T	63	7721	-13.9	175.9	G	183	7601	-7.1		A	303	7481	-303
50.4	A	64	7720	-13.6	177.1	T	184	7600	-6.9		A	304	7480	-304
51.7	G	65	7719	-13.3	178.1	G	185	7599	-6.9		G	305	7479	-305
53.3	C	66	7718	-12.7	179.2	C	186	7598	-6.8					
54.3	T	67	7717	-12.7	180.1	G	187	7597	-6.9					
55.5	T	68	7716	-12.5	181.3	C	188	7596	-6.7					
56.8	A	69	7715	-12.2	182.5	C	189	7595	-6.5					
57.9	A	70	7714	-12.1	183.8	T	190	7594	-6.2					
58.7	G	71	7713	-12.3	184.1	T	191	7593	-6.9					
59.9	T	72	7712	-12.1	185.1	A	192	7592	-6.9					
61.3	A	73	7711	-11.7	186	T	193	7591	-7					
62.3	A	74	7710	-11.7	187.2	A	194	7590	-6.8					
63.2	A	75	7709	-11.8	188.1	T	195	7589	-6.9					
64.1	A	76	7708	-11.9	189.3	G	196	7588	-6.7					
65	A	77	7707	-12	190.4	C	197	7587	-6.6					
66.1	C	78	7706	-11.9	191.3	G	198	7586	-6.7					
66.9	A	79	7705	-12.1	192.4	C	199	7585	-6.6					
68	A	80	7704	-12	193.1	C	200	7584	-6.9					
69	A	81	7703	-12	194.1	A	201	7583	-6.9					
70.1	C	82	7702	-11.9	195.1	A	202	7582	-6.9					
71	T	83	7701	-12	196.1	A	203	7581	-6.9					
72.4	A	84	7700	-11.6	197.1	G	204	7580	-6.9					
73.5	A	85	7699	-11.5	198.3	A	205	7579	-6.7					
74.5	A	86	7698	-11.5	199.2	G	206	7578	-6.8					
75.5	A	87	7697	-11.5	200.4	G	207	7577	-6.6					
76.6	T	88	7696	-11.4	201.5	C	208	7576	-6.5					
77.9	A	89	7695	-11.1	202.3	G	209	7575	-6.7					
79	T	90	7694	-11		C	210	7574						
80.1	G	91	7693	-10.9		G	211	7573						
81.5	T	92	7692	-10.5	204.4	C	212	7572	-7.6					
82.4	G	93	7691	-10.6	205.4	C	213	7571	-7.6					
83.6	C	94	7690	-10.4	206.5	A	214	7570	-7.5					
84.5	C	95	7689	-10.5	207.5	A	215	7569	-7.5					
85.4	T	96	7688	-10.6	208.5	T	216	7568	-7.5					
86.4	A	97	7687	-10.6	209.7	T	217	7567	-7.3					
87.4	A	98	7686	-10.6	210.9	G	218	7566	-7.1					
88.3	A	99	7685	-10.7	212.1	T	219	7565	-6.9					
89.2	A	100	7684	-10.8	213.2	T	220	7564	-6.8					
90.1	G	101	7683	-10.9	214.3	C	221	7563	-6.7					
91.5	C	102	7682	-10.5	215.1	A	222	7562	-6.9					
92.3	A	103	7681	-10.7	216.1	A	223	7561	-6.9					
93.3	G	104	7680	-10.7	217	A	224	7560	-7					
94.9	T	105	7679	-10.1	217.9	A	225	7559	-7.1					
95.9	T	106	7678	-10.1	218.8	T	226	7558	-7.2					
96.6	T	107	7677	-10.4	220	A	227	7557	-7					
98.3	T	108	7676	-9.7	220.9	T	228	7556	-7.1					
99.5	A	109	7675	-9.5	222	G	229	7555	-7					
100.4	T	110	7674	-9.6	223.2	T	230	7554	-6.8					
101.4	T	111	7673	-9.6	224.3	A	231	7553	-6.7					
102.7	A	112	7672	-9.3	225.1	G	232	7552	-6.9					
103.6	C	113	7671	-9.4	226.3	G	233	7551	-6.7					
104.5	T	114	7670	-9.5	227.4	A	234	7550	-6.6					
105.7	T	115	7669	-9.3	228.3	G	235	7549	-6.7					
106.9	A	116	7668	-9.1	229.3	C	236	7548	-6.7					
107.9	G	117	7667	-9.1	230.2	A	237	7547	-6.8					
109.2	G	118	7666	-8.8	231.1	G	238	7546	-6.9					
110.5	G	119	7665	-8.5	232.2	T	239	7545	-6.8					
111.7	A	120	7664	-8.3	233.3	G	240	7544	-6.7					

Figure SI2.3B. CE Maxam-Gilbert Indexing of the 305 bp HPV18 (7479-7783): Bottom Strand. CE position *versus* the real position of nucleotides from the bottom strand (5'-HEX) of the 305 bp HPV18 (7479-7783) fragment.

HPV18 (7647-157) 368 bp
Correction between CE size calling to Maxam-Gilbert sequencing results

CE postion	Sequence	Real Position	Genome Position		CE postion	Sequence	Real Position	Genome Position		CE postion	Sequence	Real Position	Genome Position	
	T	1	7647		113.4	T	121	7767	-7.6	236.3	A	241	30	-4.7
	G	2	7648		114.4	C	122	7768	-7.6	237.3	A	242	31	-4.7
	C	3	7649		115.2	T	123	7769	-7.8	238	A	243	32	-5
	A	4	7650		116.2	A	124	7770	-7.8	239.2	A	244	33	-4.8
	T	5	7651		117.2	C	125	7771	-7.8	240.2	A	245	34	-4.8
	A	6	7652		118	C	126	7772	-8	241.4	G	246	35	-4.6
	A	7	7653		118.9	C	127	7773	-8.1	242.6	G	247	36	-4.4
	C	8	7654		119.8	T	128	7774	-8.2	243.7	G	248	37	-4.3
	T	9	7655		121.1	T	129	7775	-7.9	244.7	A	249	38	-4.3
	A	10	7656		122.1	A	130	7776	-7.9	244.7	G	250	39	-5.3
	T	11	7657		123.1	A	131	7777	-7.9	245.8	T	251	40	-5.2
	A	12	7658		124.1	C	132	7778	-7.9	246.8	A	252	41	-5.2
	T	13	7659		124.9	A	133	7779	-8.1	247.7	A	253	42	-5.3
	C	14	7660		126	T	134	7780	-8	248.7	C	254	43	-5.3
	C	15	7661		127.1	G	135	7781	-7.9	249.6	C	255	44	-5.4
	A	16	7662		128.4	A	136	7782	-7.6	250.4	G	256	45	-5.6
	C	17	7663		129.3	A	137	7783	-7.7	251.7	A	257	46	-5.3
	T	18	7664		130.2	C	138	7784	-7.8	252.6	A	258	47	-5.4
	C	19	7665		131.1	T	139	7785	-7.9	254	A	259	48	-5
	C	20	7666		132.2	A	140	7786	-7.8	254.5	A	260	49	-5.5
	C	21	7667		133.2	T	141	7787	-7.8	255.6	C	261	50	-5.4
	T	22	7668		134.3	A	142	7788	-7.7	256.4	G	262	51	-5.6
	A	23	7669		135.3	A	143	7789	-7.7	257.6	G	263	52	-5.4
	A	24	7670		136.2	T	144	7790	-7.8	258.8	T	264	53	-5.2
	G	25	7671		137.3	A	145	7791	-7.7	259.8	C	265	54	-5.2
	T	26	7672		138.3	T	146	7792	-7.7	260.6	G	266	55	-5.4
8.4	A	27	7673	-18.6	139.4	G	147	7793	-7.6	261.8	G	267	56	-5.2
9.2	A	28	7674	-18.8	140.6	A	148	7794	-7.4	262.9	G	268	57	-5.1
10.3	T	29	7675	-18.7	141.5	C	149	7795	-7.5	264	A	269	58	-5
11.3	A	30	7676	-18.7	142.4	T	150	7796	-7.6	265	C	270	59	-5
12.2	A	31	7677	-18.8	143.5	A	151	7797	-7.5	265.9	C	271	60	-5.1
13.1	A	32	7678	-18.9	144.5	A	152	7798	-7.5	266.6	G	272	61	-5.4
13.9	A	33	7679	-19.1	145.5	G	153	7799	-7.5	267.6	A	273	62	-5.4
15.2	C	34	7680	-18.8	146.7	C	154	7800	-7.3	268.6	A	274	63	-5.4
16.4	T	35	7681	-18.6	147.6	T	155	7801	-7.4	269.6	A	275	64	-5.4
18.3	G	36	7682	-17.7	148.7	G	156	7802	-7.3	270.6	A	276	65	-5.4
	C	37	7683		150.1	T	157	7803	-6.9	271.7	C	277	66	-5.3
20.5	T	38	7684	-17.5	151.1	G	158	7804	-6.9	272.5	G	278	67	-5.5
21.9	T	39	7685	-17.1	152.3	C	159	7805	-6.7	273.7	G	279	68	-5.3
22.9	T	40	7686	-17.1	152.8	A	160	7806	-7.2	275	T	280	69	-5
24	T	41	7687	-17	153.9	T	161	7807	-7.1	275.9	G	281	70	-5.1
25	A	42	7688	-17	155	A	162	7808	-7	277.1	T	282	71	-4.9
25.8	G	43	7689	-17.2	156.1	C	163	7809	-6.9	278.1	A	283	72	-4.9
27.2	G	44	7690	-16.8	156.8	A	164	7810	-7.2	279	T	284	73	-5
	C	45	7691		157.9	T	165	7811	-7.1	280	A	285	74	-5
29.1	A	46	7692	-16.9	159	A	166	7812	-7	281	T	286	75	-5
	C	47	7693		160	G	167	7813	-7	281.9	A	287	76	-5.1
30.5	A	48	7694	-17.5	161.4	T	168	7814	-6.6	282.9	A	288	77	-5.1
31.7	T	49	7695	-17.3	162.4	T	169	7815	-6.6	283.8	A	289	78	-5.2
32.7	A	50	7696	-17.3	163.5	T	170	7816	-6.5	284.7	A	290	79	-5.3
33.8	T	51	7697	-17.2	164.4	A	171	7817	-6.6	285.7	G	291	80	-5.3
34.9	T	52	7698	-17.1	165.4	T	172	7818	-6.6	286.8	A	292	81	-5.2
36.1	T	53	7699	-16.9	166.4	G	173	7819	-6.6	287.8	T	293	82	-5.2
37.4	T	54	7700	-16.6	167.7	C	174	7820	-6.3	288.8	G	294	83	-5.2
38.8	A	55	7701	-16.2	168.4	A	175	7821	-6.6	290	T	295	84	-5
39.7	G	56	7702	-16.3	169.3	A	176	7822	-6.7	291	G	296	85	-5
41.2	T	57	7703	-15.8	170.3	C	177	7823	-6.7	292.1	A	297	86	-4.9
42.3	T	58	7704	-15.7	171.2	C	178	7824	-6.8	292.9	G	298	87	-5.1
43.5	T	59	7705	-15.5	172	G	179	7825	-7	294.1	A	299	88	-4.9
44.5	G	60	7706	-15.5	173.4	A	180	7826	-6.6	294.9	A	300	89	-5.1
46.3	T	61	7707	-14.7	174.3	A	181	7827	-6.7	295.8	A	301	90	-5.2
47.5	T	62	7708	-14.5	175.2	A	182	7828	-6.8	296.8	C	302	91	-5.2
48.7	T	63	7709	-14.3	176.3	T	183	7829	-6.7	297.6	A	303	92	-5.4
49.9	T	64	7710	-14.1	177.3	A	184	7830	-6.7	298.6	C	304	93	-5.4
51.1	T	65	7711	-13.9	178.3	G	185	7831	-6.7	299.5	A	305	94	-5.5
52	A	66	7712	-14	179.5	G	186	7832	-6.5	300.5	C	306	95	-5.5
53.2	C	67	7713	-13.8	180.8	T	187	7833	-6.2	301.4	C	307	96	-5.6
54.1	T	68	7714	-13.9	181.8	T	188	7834	-6.2	302.3	A	308	97	-5.7
55.6	T	69	7715	-13.4	182.8	G	189	7835	-6.2	303.3	C	309	98	-5.7
56.7	A	70	7716	-13.3	183.9	G	190	7836	-6.1	304.3	A	310	99	-5.7
58	A	71	7717	-13	185.1	G	191	7837	-5.9	305.2	A	311	100	-5.8
59.3	G	72	7718	-12.7	186.3	C	192	7838	-5.7	306.2	T	312	101	-5.8
60.9	C	73	7719	-12.1	187	A	193	7839	-6	307.3	A	313	102	-5.7
61.7	T	74	7720	-12.3	187.9	G	194	7840	-6.1	308.4	C	314	103	-5.6
62.7	A	75	7721	-12.3	189.2	C	195	7841	-5.8	309.2	T	315	104	-5.8
63.7	A	76	7722	-12.3	189.9	A	196	7842	-6.1	310.3	A	316	105	-5.7
64.9	T	77	7723	-12.1	190.9	C	197	7843	-6.1	311.4	T	317	106	-5.6
66.2	T	78	7724	-11.8	191.7	A	198	7844	-6.3	312.4	G	318	107	-5.6
67.3	G	79	7725	-11.7	192.8	T	199	7845	-6.2	313.6	G	319	108	-5.4
69	C	80	7726	-11	193.8	A	200	7846	-6.2	314.8	C	320	109	-5.2
69.5	A	81	7727	-11.5	194.9	C	201	7847	-6.1	315.6	G	321	110	-5.4
70.8	T	82	7728	-11.2	195.8	T	202	7848	-6.2	316.9	C	322	111	-5.1
71.7	A	83	7729	-11.3	196.9	A	203	7849	-6.1	317.7	G	323	112	-5.3
73	C	84	7730	-11	197.9	T	204	7850	-6.1	319	C	324	113	-5
73.9	T	85	7731	-11.1	199	A	205	7851	-6	319.7	T	325	114	-5.3
75.3	T	86	7732	-10.7	200	C	206	7852	-6	320.9	T	326	115	-5.1
76.3	G	87	7733	-10.7	200.9	T	207	7853	-6.1	321.8	T	327	116	-5.2
77.7	G	88	7734	-10.3	202	T	208	7854	-6	322.9	G	328	117	-5.1
79.3	C	89	7735	-9.7	203.1	T	209	7855	-5.9	324	A	329	118	-5
80.2	T	90	7736	-9.8	204.2	T	210	7856	-5.8	324.9	G	330	119	-5.1
81.4	T	91	7737	-9.6	205.2	C	211	7857	-5.8	326.1	G	331	120	-4.9
82.5	G	92	7738	-9.5	206	A	212	1	-6	327.2	A	332	121	-4.8
84.1	T	93	7739	-8.9	207	T	213	2	-6	328.1	T	333	122	-4.9
85.1	A	94	7740	-8.9	208.1	T	214	3	-5.9	329.1	C	334	123	-4.9
85.9	C	95	7741	-9.1	209.2	A	215	4	-5.8	329.9	C	335	124	-5.1
86.6	A	96	7742	-9.4	210.2	A	216	5	-5.8	331	A	336	125	-5
87.6	A	97	7743	-9.4	211	T	217	6	-6	331.7	A	337	126	-5.3
88.5	C	98	7744	-9.5	212.1	A	218	7	-5.9	332.7	C	338	127	-5.3
89.4	T	99	7745	-9.6	213.1	C	219	8	-5.9	333.6	A	339	128	-5.4
90.6	A	100	7746	-9.4	214	T	220	9	-6	334.7	C	340	129	-5.3
91.6	C	101	7747	-9.4	215.2	T	221	10	-5.8	335.6	G	341	130	-5.4
92.4	T	102	7748	-9.6	216.3	T	222	11	-5.7	336.8	G	342	131	-5.2
93.7	T	103	7749	-9.3	217.3	T	223	12	-5.7	338	C	343	132	-5
94.9	T	104	7750	-9.1	218.4	A	224	13	-5.6	338.8	G	344	133	-5.2
96	C	105	7751	-9	219.3	A	225	14	-5.7	340	A	345	134	-5
96.8	A	106	7752	-9.2	220.3	C	226	15	-5.7	341	C	346	135	-5
97.8	T	107	7753	-9.2	221.1	A	227	16	-5.9	341.8	C	347	136	-5.2
99.1	G	108	7754	-8.9	222.1	A	228	17	-5.9	342.6	C	348	137	-5.4
101.2	T	109	7755	-7.8	223.2	T	229	18	-5.8	343.5	T	349	138	-5.5
102.6	C	110	7756	-7.4	224.3	T	230	19	-5.7	344.7	A	350	139	-5.3
103.7	C	111	7757	-7.3	225.2	G	231	20	-5.8	345.8	C	351	140	-5.2
104.6	A	112	7758	-7.4	226.6	T	232	21	-5.4	346.7	A	352	141	-5.3
105.8	A	113	7759	-7.2	227.5	A	233	22	-5.5	347.7	A	353	142	-5.3
106.6	C	114	7760	-7.4	228.4	G	234	23	-5.6	348.7	G	354	143	-5.3
107.1	A	115	7761	-7.9	229.6	T	235	24	-5.4	350	C	355	144	-5
108.2	T	116	7762	-7.8	230.6	A	236	25	-5.4	350.6	T	356	145	-5.4
109.2	T	117	7763	-7.8	231.6	T	237	26	-5.4	351.9	A	357	146	-5.1
110.2	C	118	7764	-7.8	232.5	A	238	27	-5.5	353.1	C	358	147	-4.9
111	T	119	7765	-8	233.5	T	239	28	-5.5	353.8	C	359	148	-5.2
111.9	G	120	7766	-8.1	234.5	A	240	29	-5.5	354.6	T	360	149	-5.4
											G	361	150	
											A	362	151	
											T	363	152	
											C	364	153	
											T	365	154	
											G	366	155	
											T	367	156	
											G	368	157	

Figure SI2.3C. CE Maxam-Gilbert Indexing of the 368 bp HPV18 (7647-157): Top Strand. CE position *versus* the real position of nucleotides from the top strand (5'-FAM) of the 368 bp HPV18 (7647-157) fragment.

HPV18 (7647-157) 368 bp
Correction between CE size calling to Maxam-Gilbert sequencing results

CE postion	Sequence	Real Position	Genome Position		CE postion	Sequence	Real Position	Genome Position		CE postion	Sequence	Real Position	Genome Position	
	C	1	157		114.7	C	121	37	-6.3	236.7	A	241	7774	-4.3
	A	2	156		116	C	122	36	-6	237.5	G	242	7773	-4.5
	C	3	155		116.4	C	123	35	-6.6	238.7	G	243	7772	-4.3
	A	4	154		117.3	T	124	34	-6.7	239.8	G	244	7771	-4.2
	G	5	153		118.5	T	125	33	-6.5	241.1	T	245	7770	-3.9
	A	6	152		119.7	T	126	32	-6.3	242	A	246	7769	-4
	T	7	151		120.8	T	127	31	-6.2	242.9	G	247	7768	-4.1
	C	8	150		121.9	T	128	30	-6.1	244	A	248	7767	-4
	A	9	149		123	T	129	29	-6	245	C	249	7766	-4
	G	10	148		123.9	A	130	28	-6.1	245.8	A	250	7765	-4.2
	G	11	147		124.8	T	131	27	-6.2	246.8	G	251	7764	-4.2
	T	12	146		125.8	A	132	26	-6.2	248	A	252	7763	-4
	A	13	145		126.8	T	133	25	-6.2	248.9	A	253	7762	-4.1
	G	14	144		127.8	A	134	24	-6.2	249.9	T	254	7761	-4.1
	C	15	143		128.8	C	135	23	-6.2	250.9	G	255	7760	-4.1
	T	16	142		129.6	T	136	22	-6.4	252.2	T	256	7759	-3.8
	T	17	141		130.7	A	137	21	-6.3	253.1	T	257	7758	-3.9
	G	18	140		131.7	C	138	20	-6.3	254.1	G	258	7757	-3.9
	T	19	139		132.5	A	139	19	-6.5	255.2	G	259	7756	-3.8
	A	20	138		133.5	A	140	18	-6.5	256.3	A	260	7755	-3.7
	G	21	137		134.5	T	141	17	-6.5	257.2	C	261	7754	-3.8
	G	22	136		135.6	T	142	16	-6.4	258	A	262	7753	-4
	G	23	135		136.7	G	143	15	-6.3	259.1	T	263	7752	-3.9
	T	24	134		138	T	144	14	-6	260	G	264	7751	-4
	C	25	133		139	T	145	13	-6	261.2	A	265	7750	-3.8
	G	26	132		140	A	146	12	-6	262	A	266	7749	-4
	C	27	131		140.9	A	147	11	-6.1	263	A	267	7748	-4
	C	28	130		141.8	A	148	10	-6.2	263.9	G	268	7747	-4.1
	G	29	129		142.7	A	149	9	-6.3	265.2	T	269	7746	-3.8
	T	30	128		143.6	G	150	8	-6.4	266.1	A	270	7745	-3.9
16.1	G	31	127	-14.9	145	T	151	7	-6	267	G	271	7744	-4
	T	32	126		146	A	152	6	-6	268.2	T	272	7743	-3.8
	T	33	125		147	T	153	5	-6	269.3	T	273	7742	-3.7
20.6	G	34	124	-13.4	148	T	154	4	-6	270.2	G	274	7741	-3.8
21.8	G	35	123	-13.2	149	A	155	3	-6	271.4	T	275	7740	-3.6
22.9	A	36	122	-13.1	150	A	156	2	-6	272.3	A	276	7739	-3.7
	T	37	121		150.9	T	157	1	-6.1	273.2	C	277	7738	-3.8
	C	38	120		152	G	158	7857	-6	274.1	A	278	7737	-3.9
	C	39	119		153.3	A	159	7856	-5.7	275	A	279	7736	-4
	T	40	118		154.2	A	160	7855	-5.8	276	G	280	7735	-4
26.7	C	41	117	-14.3	155.1	A	161	7854	-5.9	277.2	C	281	7734	-3.8
27.2	A	42	116	-14.8	156	A	162	7853	-6	278	C	282	7733	-4
28.2	A	43	115	-14.8	157	G	163	7852	-6	278.9	A	283	7732	-4.1
29	A	44	114	-15	158.3	T	164	7851	-5.7	279.9	A	284	7731	-4.1
29.8	G	45	113	-15.2	159.3	A	165	7850	-5.7	280.8	G	285	7730	-4.2
31.3	C	46	112	-14.7	160.3	T	166	7849	-5.7	282.1	T	286	7729	-3.9
31.8	G	47	111	-15.2	161.2	A	167	7848	-5.8	283.1	A	287	7728	-3.9
33.5	C	48	110	-14.5	162.2	G	168	7847	-5.8	284.1	T	288	7727	-3.9
34	G	49	109	-15	163.5	T	169	7846	-5.5	285.1	G	289	7726	-3.9
35.7	C	50	108	-14.3	164.4	A	170	7845	-5.6	286.3	C	290	7725	-3.7
36.2	C	51	107	-14.8	165.5	T	171	7844	-5.5	287.1	A	291	7724	-3.9
36.8	A	52	106	-15.2	166.4	G	172	7843	-5.6	288	A	292	7723	-4
38.3	T	53	105	-14.7	167.7	T	173	7842	-5.3	289	T	293	7722	-4
39.7	A	54	104	-14.3	168.6	G	174	7841	-5.4	290	T	294	7721	-4
40.6	G	55	103	-14.4	169.9	C	175	7840	-5.1	291.1	A	295	7720	-3.9
42.1	T	56	102	-13.9	170.7	T	176	7839	-5.3	292	G	296	7719	-4
43.2	A	57	101	-13.8	171.7	G	177	7838	-5.3	293.3	C	297	7718	-3.7
44.3	T	58	100	-13.7	173	C	178	7837	-5	294.1	T	298	7717	-3.9
45.6	T	59	99	-13.4	173.8	C	179	7836	-5.2	295	T	299	7716	-4
46.6	G	60	98	-13.4	174.6	C	180	7835	-5.4	296.2	A	300	7715	-3.8
48.5	T	61	97	-12.5	175.4	A	181	7834	-5.6	297.1	A	301	7714	-3.9
49.4	G	62	96	-12.6	176.3	A	182	7833	-5.7	298	G	302	7713	-4
50.9	G	63	95	-12.1	177.4	C	183	7832	-5.6	299.2	T	303	7712	-3.8
52.6	T	64	94	-11.4	178.3	C	184	7831	-5.7	300.2	A	304	7711	-3.8
53.5	G	65	93	-11.5	179.2	T	185	7830	-5.8	301.1	A	305	7710	-3.9
55.3	T	66	92	-10.7	180.3	A	186	7829	-5.7	302	A	306	7709	-4
56.1	G	67	91	-10.9	181.5	T	187	7828	-5.5	302.9	A	307	7708	-4.1
58.2	T	68	90	-9.8	182.8	T	188	7827	-5.2	303.8	A	308	7707	-4.2
59.4	T	69	89	-9.6	183.8	T	189	7826	-5.2	304.7	C	309	7706	-4.3
60.4	T	70	88	-9.6	184.8	C	190	7825	-5.2	305.7	A	310	7705	-4.3
61.5	C	71	87	-9.5	185.6	G	191	7824	-5.4	306.7	A	311	7704	-4.3
62	T	72	86	-10	186.9	G	192	7823	-5.1	307.6	A	312	7703	-4.4
63.3	C	73	85	-9.7	188.2	T	193	7822	-4.8	308.6	C	313	7702	-4.4
63.8	A	74	84	-10.2	189.2	T	194	7821	-4.8	309.5	T	314	7701	-4.5
65.1	C	75	83	-9.9	190.1	G	195	7820	-4.9	310.7	A	315	7700	-4.3
65.6	A	76	82	-10.4	191.4	C	196	7819	-4.6	311.6	A	316	7699	-4.4
67.1	T	77	81	-9.9	192.1	A	197	7818	-4.9	312.6	A	317	7698	-4.4
68.3	C	78	80	-9.7	193.1	T	198	7817	-4.9	313.6	A	318	7697	-4.4
69.2	T	79	79	-9.8	194.1	A	199	7816	-4.9	314.5	T	319	7696	-4.5
70.5	T	80	78	-9.5	195	A	200	7815	-5	315.7	A	320	7695	-4.3
71.8	T	81	77	-9.2	196	A	201	7814	-5	316.8	T	321	7694	-4.2
73	T	82	76	-9	196.9	C	202	7813	-5.1	317.8	G	322	7693	-4.2
73.9	A	83	75	-9.1	197.8	T	203	7812	-5.2	319.1	T	323	7692	-3.9
75.2	T	84	74	-8.8	199	A	204	7811	-5	320	G	324	7691	-4
76.1	A	85	73	-8.9	200	T	205	7810	-5	321.3	C	325	7690	-3.7
77.3	T	86	72	-8.7	201.1	G	206	7809	-4.9	322	C	326	7689	-4
78.3	A	87	71	-8.7	202.3	T	207	7808	-4.7	322.8	T	327	7688	-4.2
79.5	C	88	70	-8.5	203.3	A	208	7807	-4.7	324	A	328	7687	-4
80.1	A	89	69	-8.9	204.2	T	209	7806	-4.8	325	A	329	7686	-4
81.4	C	90	68	-8.6	205.3	G	210	7805	-4.7	325.9	A	330	7685	-4.1
82.3	C	91	67	-8.7	206.5	C	211	7804	-4.5	326.9	A	331	7684	-4.1
83.2	G	92	66	-8.8	207.3	A	212	7803	-4.7	327.8	G	332	7683	-4.2
84.9	T	93	65	-8.1	208.2	C	213	7802	-4.8	329.1	C	333	7682	-3.9
86	T	94	64	-8	209	A	214	7801	-5	329.9	A	334	7681	-4.1
87.1	T	95	63	-7.9	210	G	215	7800	-5	330.8	G	335	7680	-4.2
88.2	T	96	62	-7.8	211.3	C	216	7799	-4.7	332.2	T	336	7679	-3.8
89.2	C	97	61	-7.8	212	T	217	7798	-5	333.1	T	337	7678	-3.9
89.9	G	98	60	-8.1	213.1	T	218	7797	-4.9	334.1	T	338	7677	-3.9
91.3	G	99	59	-7.7	214.2	A	219	7796	-4.8	335	T	339	7676	-4
92.5	T	100	58	-7.5	215.2	G	220	7795	-4.8	336.1	A	340	7675	-3.9
93.5	C	101	57	-7.5	216.5	T	221	7794	-4.5	337	T	341	7674	-4
94.3	C	102	56	-7.7	217.5	C	222	7793	-4.5	338.1	T	342	7673	-3.9
95.2	C	103	55	-7.8	218.3	A	223	7792	-4.7	339.1	A	343	7672	-3.9
96	G	104	54	-8	219.3	T	224	7791	-4.7	340.2	C	344	7671	-3.8
97.4	A	105	53	-7.6	220.4	A	225	7790	-4.6	340.9	T	345	7670	-4.1
98.4	C	106	52	-7.6	221.4	T	226	7789	-4.6	342	T	346	7669	-4
99.3	C	107	51	-7.7	222.5	T	227	7788	-4.5	343.1	A	347	7668	-3.9
100.3	G	108	50	-7.7	223.5	A	228	7787	-4.5	344	G	348	7667	-4
102.1	T	109	49	-6.9	224.5	T	229	7786	-4.5	345.2	G	349	7666	-3.8
104.8	T	110	48	-5.2	225.5	A	230	7785	-4.5	346.4	G	350	7665	-3.6
105.7	T	111	47	-5.3	226.4	G	231	7784	-4.6	347.5	A	351	7664	-3.5
106.5	T	112	46	-5.5	227.7	T	232	7783	-4.3	348.3	G	352	7663	-3.7
107.4	C	113	45	-5.6	228.8	T	233	7782	-4.2	349.6	T	353	7662	-3.4
107.8	G	114	44	-6.2	229.8	C	234	7781	-4.2	350.4	G	354	7661	-3.6
108.9	G	115	43	-6.1	230.5	A	235	7780	-4.5	351.5	G	355	7660	-3.5
110.1	T	116	42	-5.9	231.5	T	236	7779	-4.5	352.6	A	356	7659	-3.4
111.1	T	117	41	-5.9	232.5	G	237	7778	-4.5	353.5	T	357	7658	-3.5
112	A	118	40	-6	233.8	T	238	7777	-4.2	354.6	A	358	7657	-3.4
112.9	C	119	39	-6.1	234.8	T	239	7776	-4.2	355.4	T	359	7656	-3.6
113.7	T	120	38	-6.3	235.7	A	240	7775	-4.3		A	360	7655	
											G	361	7654	
											T	362	7653	
											T	363	7652	
											A	364	7651	
											T	365	7650	
											G	366	7649	
											C	367	7648	
											A	368	7647	

Figure SI2.3D. CE Maxam-Gilbert Indexing of the 368 bp HPV18 (7647-157): Bottom Strand. CE position *versus* the real position of nucleotides from the bottom strand (5'-HEX) of the 368 bp HPV18 (7647-157) fragment.

2.7.5 Putative PA1 Binding Sites on HPV18 7647-157 bp

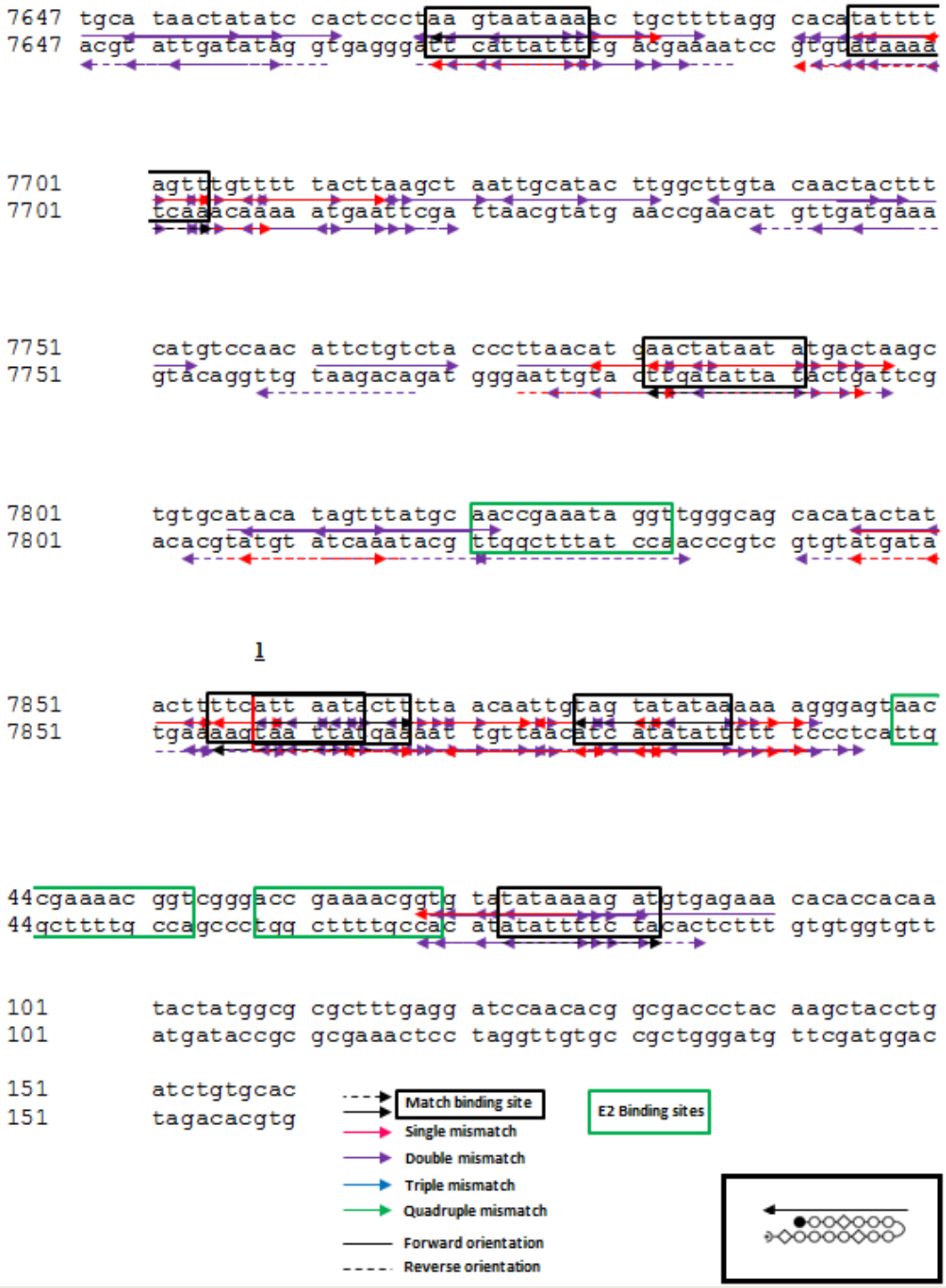


Figure SI2.4. Putative PA1 Binding Sites on the HPV18 Region from 7647-157 bp. According to the literature, **PA1** recognizes the DNA sequence 5'-W_2GW_7-3'. **PA1** binding sites are depicted with either a solid arrow (forward orientation) or a dashed arrow (reverse orientation). The head of the arrow indicates the polyamide end where the Ta is located (see inset). HPV E2 binding sites are depicted in green boxes. HPV E1 binds in the region (HPV18 3-19 bp) flanked by the first and second E2 binding sites. The rationale for the project was that the polyamide would interfere with viral protein binding leading to repression of DNA replication and/or transcription. The origin of replication is marked with 1. Numbers on the left correspond to the nucleotide position in HPV18 genome.

2.7.6 PA1 Single-Base-Pair Mismatch Sites Predicted, But Not Observed on HPV18 7647-157 bp

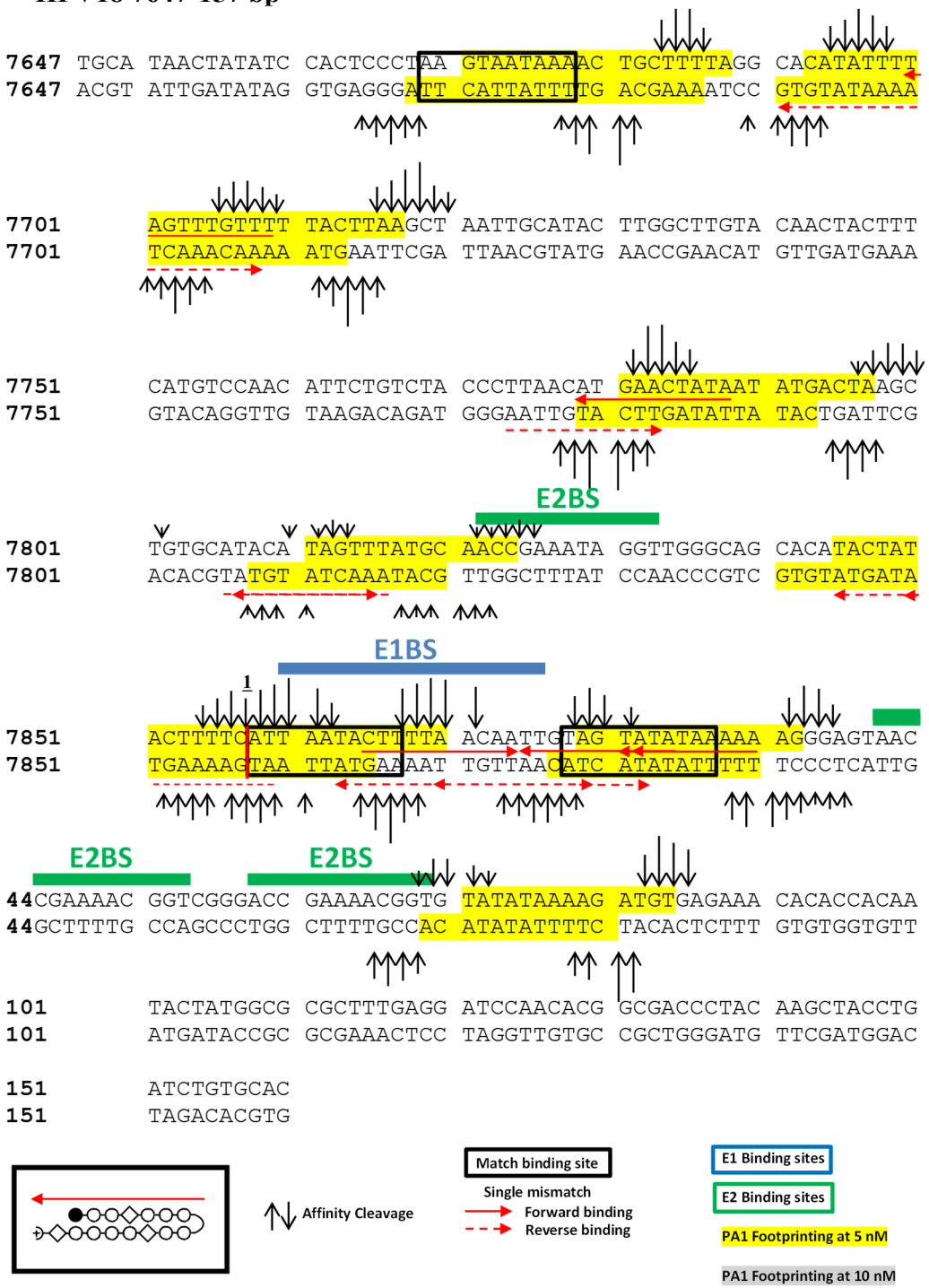


Figure SI2.5. PA1 Single-base-pair Mismatch Sites Predicted by PA-DNA Recognition Rules, Yet Not Observed on the 368 bp (7647-157) HPV18 DNA Fragment. Predicted single-base-pair mismatch binding sites for **PA1** are depicted with a horizontal solid arrow (forward orientation) or a horizontal dashed arrow (reverse orientation) between DNA duplex. **PA1** sites shown in this figure either do not agree with affinity cleavage patterns or lack DNase I footprints. The head of the arrow indicates the polyamide end where the Ta is located (see inset). HPV E2 binding sites are depicted with green lines above the sequence duplex. HPV E1 binds in the region (HPV18 3-19 bp) flanked by the first and second E2 binding sites (blue line above sequence duplex). The origin of replication is marked with 1. Numbers on the left correspond to the nucleotide position in HPV18 genome.

Table SI2.2. PA1 Predicted Single-Base-Pair Mismatch Sites Which Were Not Observed on 7647-157 Fragment of HPV18.

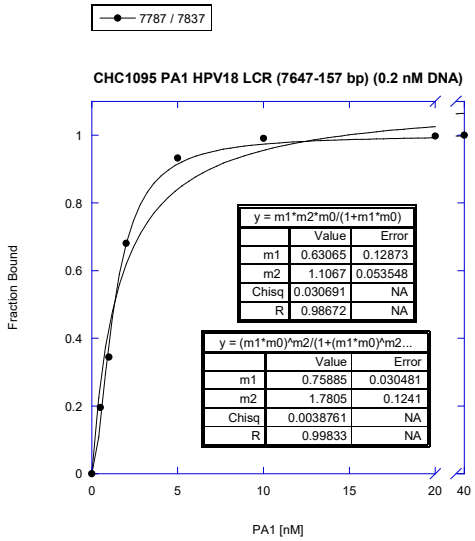
Sequence	Position	Binding orientation	Comment
TTAGG CA**C**ATATTTT AGTTT	7691-7700	Reverse	No AC
TATTT TA**G**TTTGTTT TTACT	7699-7708	Forward	No AC
GTAAA AA**C**AAACTAA AATAT	7708-7699	Reverse	No AC
TATAG TT**C**ATGTTAA GGGTA	7783-7774	Reverse	No AC or FP
TTAAC AT**G**AACTATA ATATG	7779-7788	Forward	No AC
GCATA AA**C**TATGTAT GCACA	7815-7806	Reverse	No AC or FP
GTGCA TA**C**ATAGTTT ATGCA	7807-7816	Reverse	No AC or FP
GCACA TA**C**TATACTT TTCAT	7845-7854	Reverse	No AC
TACTA TA**C**TTTTCAT TAATA	7850-2	Reverse	No AC
ATTAA TA**C**TTTTAAC AATTG	6-15	Reverse	No AC or FP
TACAA TT**G**TTAAAAG TATTA	17-8	Forward	No AC or FP
CTTTT AA**C**AATTGTA GTATA	13-22	Reverse	No FP
TATAC TA**C**AATTGTT AAAAG	22-13	Reverse	No AC or FP
AACAA TT**G**TAGTATA TAAAA	18-27	Forward	No AC or FP
TTATA TA**C**TACAATT GTTAA	25-16	Reverse	No AC or FP
TGTAG TA**T**ATAAAAA AGGGA	24-33	Forward	No AC
GTAGT AT**A**TAAAAAA GGGAG	25-34	Forward	No AC or FP

PA1 was predicted to bind these sequences with single-base-pair mismatches; however, these sequences did not generate affinity cleavage (AC) patterns and/or footprints (FP).

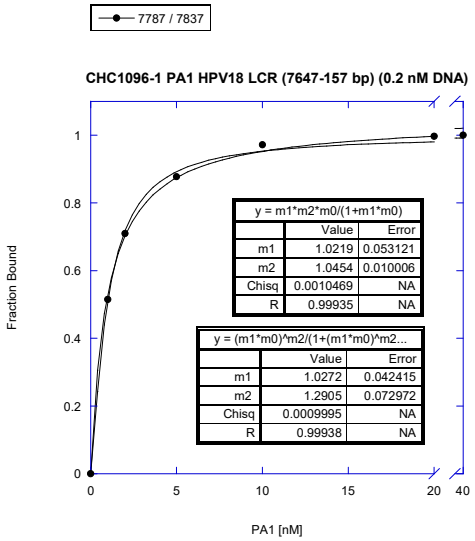
2.7.7 Representative PA1 Binding Isotherms Obtained from Quantitative DNase I Footprinting Sites Along the DNA Fragment Corresponding to 7647-157 bp of the HPV18 LCR

SITE #9 (Integration nt: 7787; Reference nt: 7837 in HEX Channel)

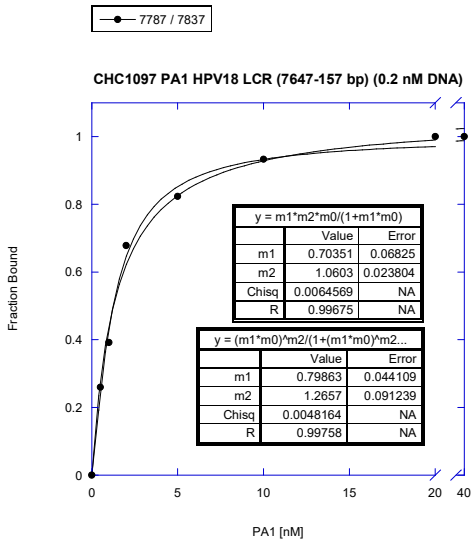
Trial #1



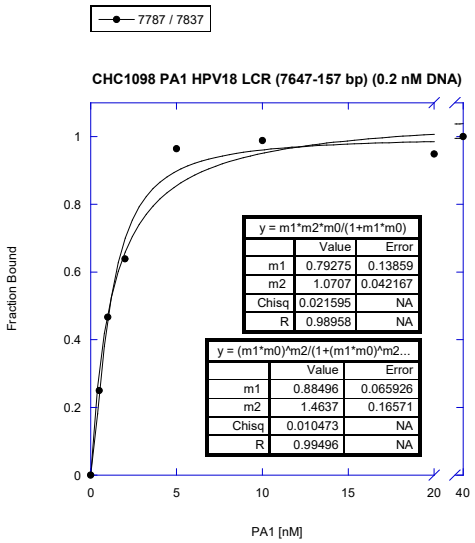
Trial #2



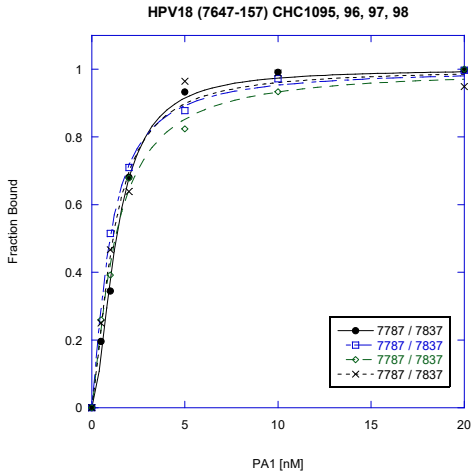
Trial #3



Trial #4

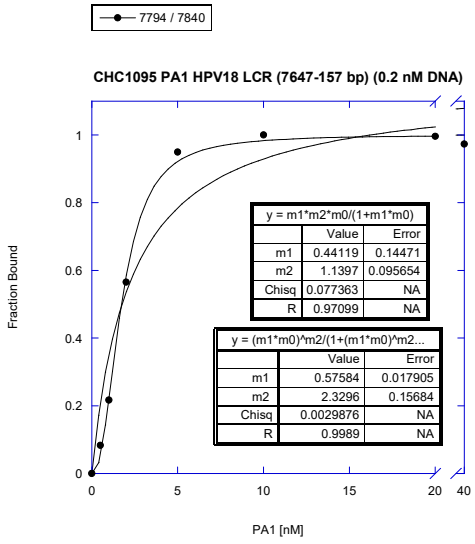


SITE #9 HEX: All Trials (Hill fit)



SITE #10 (Integration nt: 7794; Reference nt: 7840 in FAM Channel)

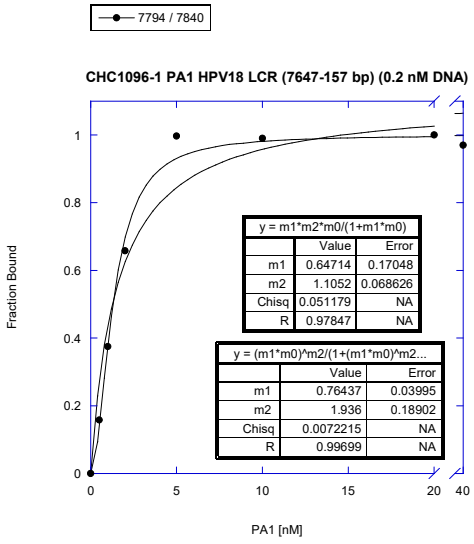
Trial #1



y = m1*m2*m0/(1+m1*m0)	Value	Error
m1	0.44119	0.14471
m2	1.1397	0.095654
Chisq	0.077363	NA
R	0.97099	NA

y = (m1*m0)^m2/(1+(m1*m0)^m2...	Value	Error
m1	0.57584	0.017905
m2	2.3296	0.15684
Chisq	0.0029876	NA
R	0.9989	NA

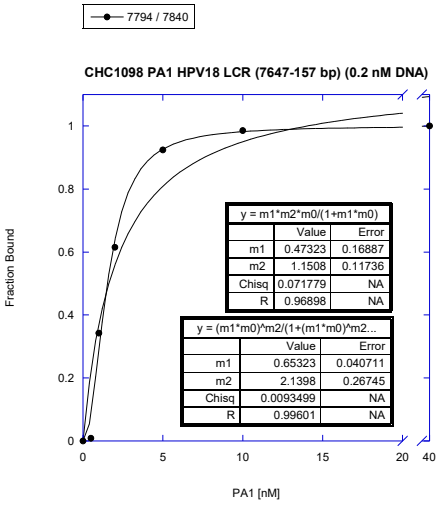
Trial #2



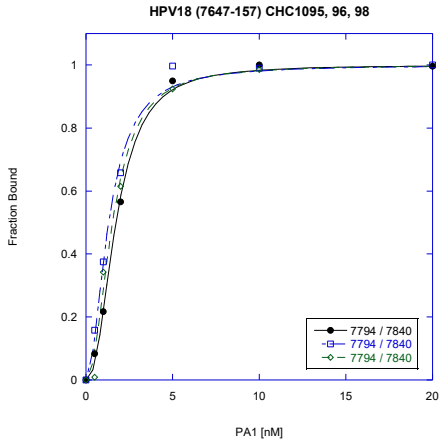
y = m1*m2*m0/(1+m1*m0)	Value	Error
m1	0.64714	0.17048
m2	1.1052	0.068626
Chisq	0.051179	NA
R	0.97847	NA

y = (m1*m0)^m2/(1+(m1*m0)^m2...	Value	Error
m1	0.76437	0.03995
m2	1.936	0.18902
Chisq	0.0072215	NA
R	0.99699	NA

Trial #3

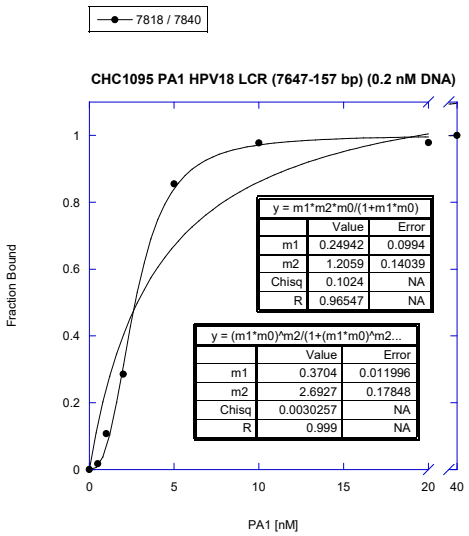


SITE #10 FAM: All Trials (Hill fit)

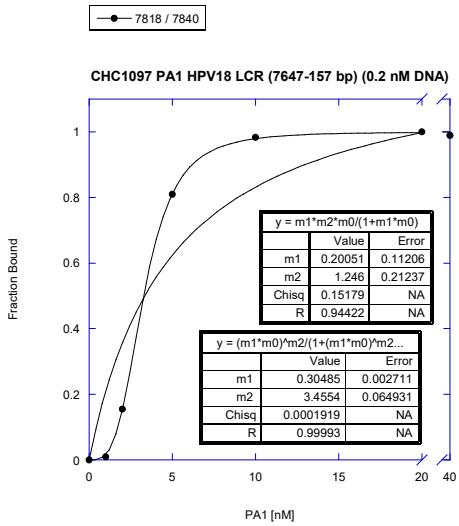


SITE #12 & 13 (Integration nt: 7818; Reference nt: 7840 in FAM Channel)

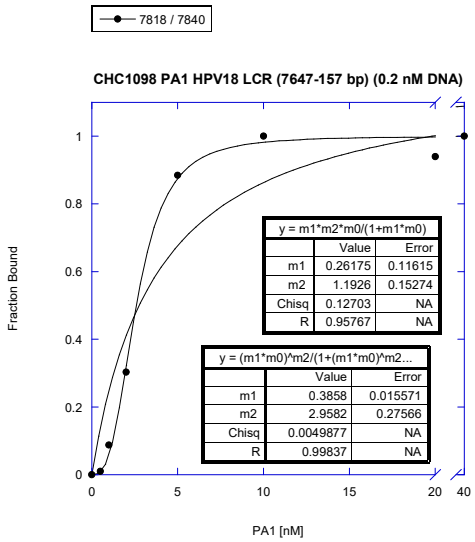
Trial #1



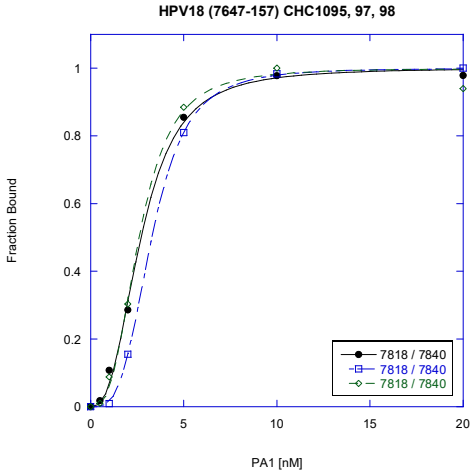
Trial #2



Trial #3

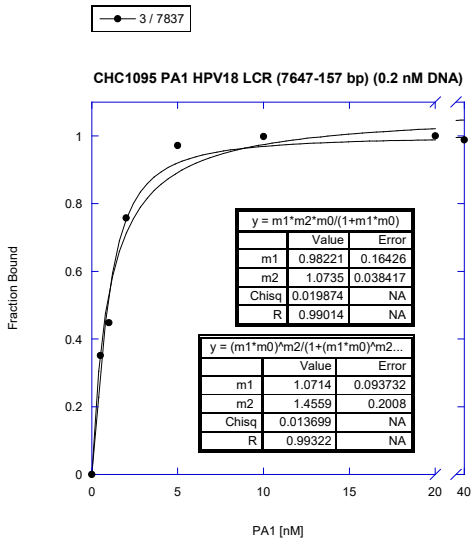


SITE #12 & 13 FAM: All Trials (Hill fit)

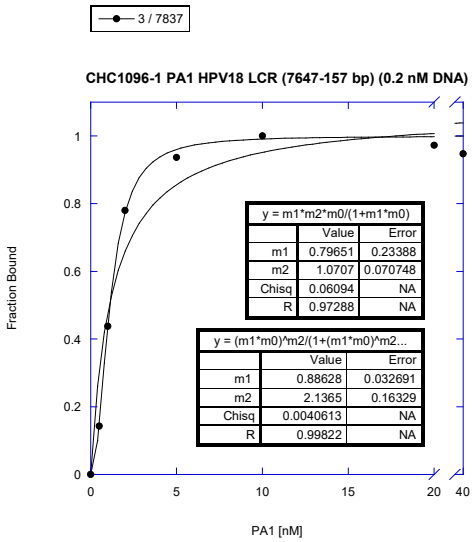


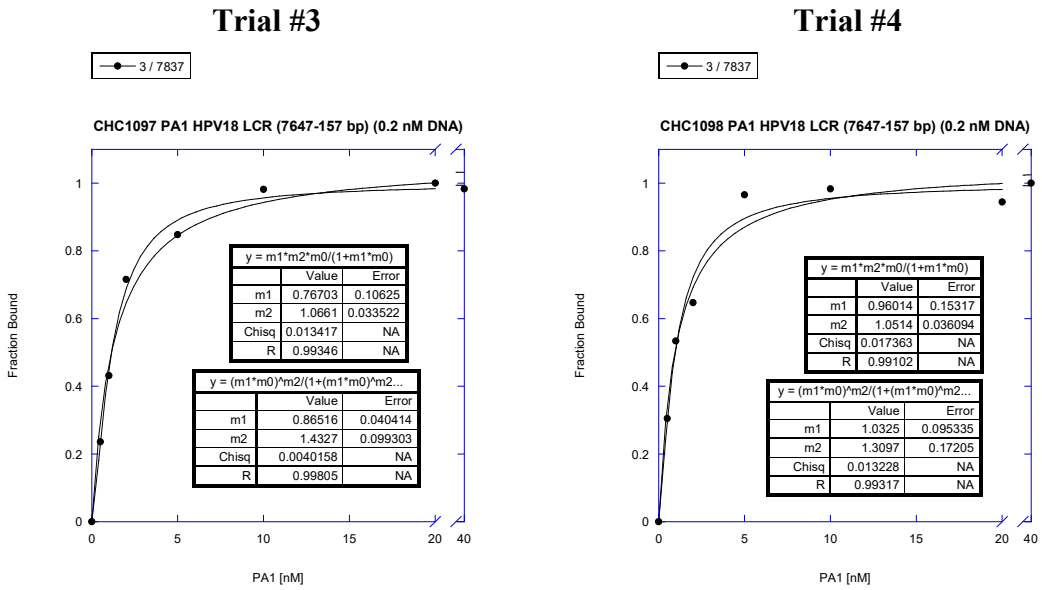
SITE #15 & #16 (Integration nt: 3; Reference nt: 7837 in HEX Channel)

Trial #1



Trial #2





SITE #15 & 16 HEX: All Trials (Hill fit)

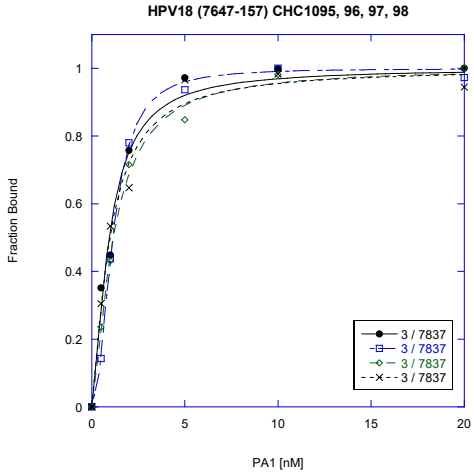


Figure SI2.6. Representative PA1 Binding Isotherms Obtained from Quantitative DNase I Footprinting Sites Along the DNA Fragment Corresponding to 7647-157 bp of the HPV18 LCR. Binding isotherms for each trial of **PA1** are plotted. Calculated parameters for the Langmuir (*top inset table*) and the Hill (*bottom inset table*) equations (KaleidaGraph 4.1 software). Summary plots are fitted using the Hill equation.

2.7.8 Putative PA25 Binding Sites on HPV18 Fragments corresponding to 7479-7783 bp and 7647-157 bp

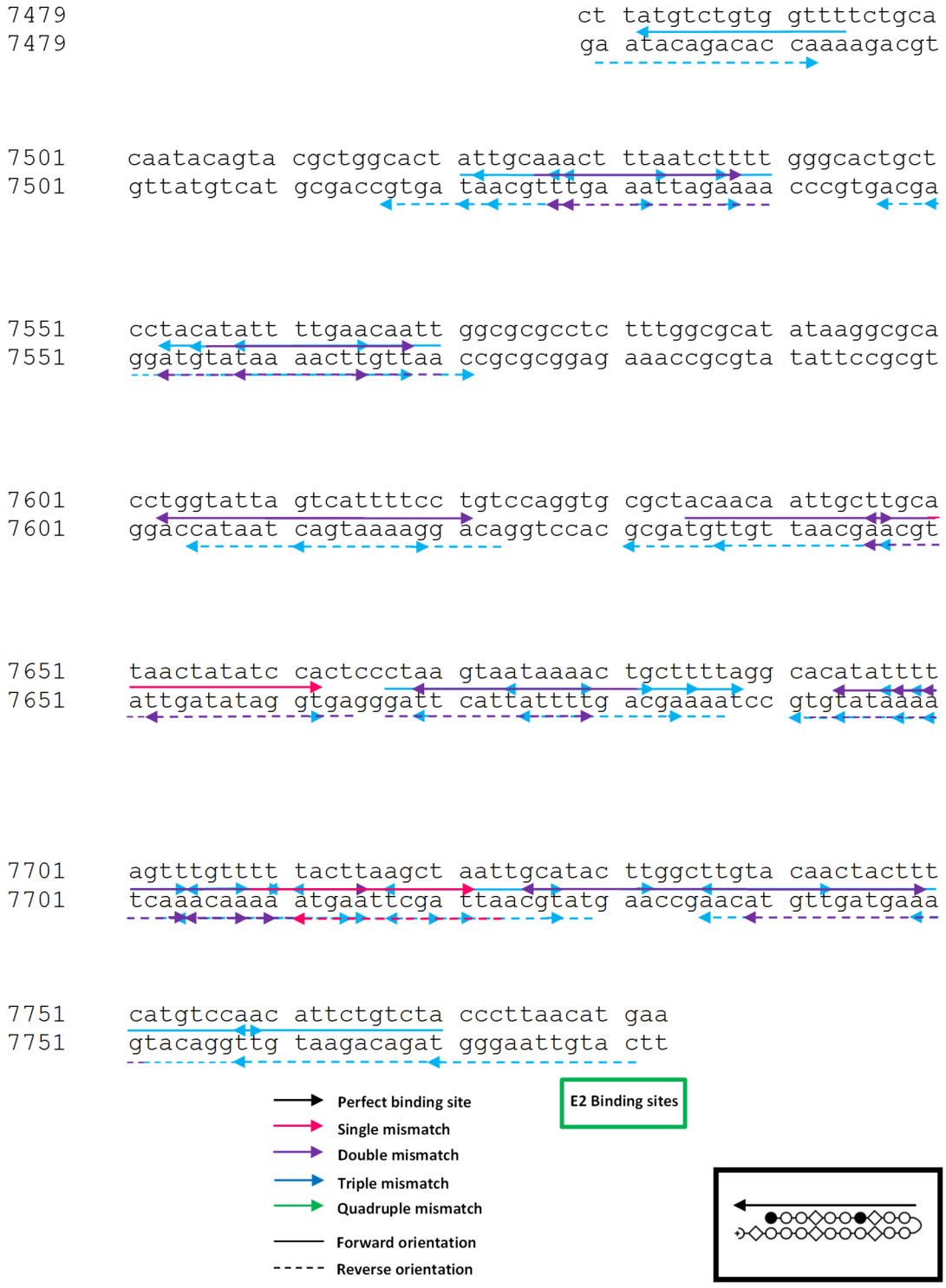


Figure SI2.7. Putative PA25 Binding Sites on the HPV18 Region from 7479-7783 bp. According to the literature, **PA25** recognizes the DNA sequence 5'-$W_2GW_5GW_4$-3'. **PA25** binding sites are depicted with either a solid arrow (forward orientation) or a dashed arrow (reverse orientation). The head of the arrow indicates the polyamide end where the Ta is located (see inset). This fragment does not harbor any HPV E2 binding sites. Numbers on the left correspond to the nucleotide position in HPV18 genome.

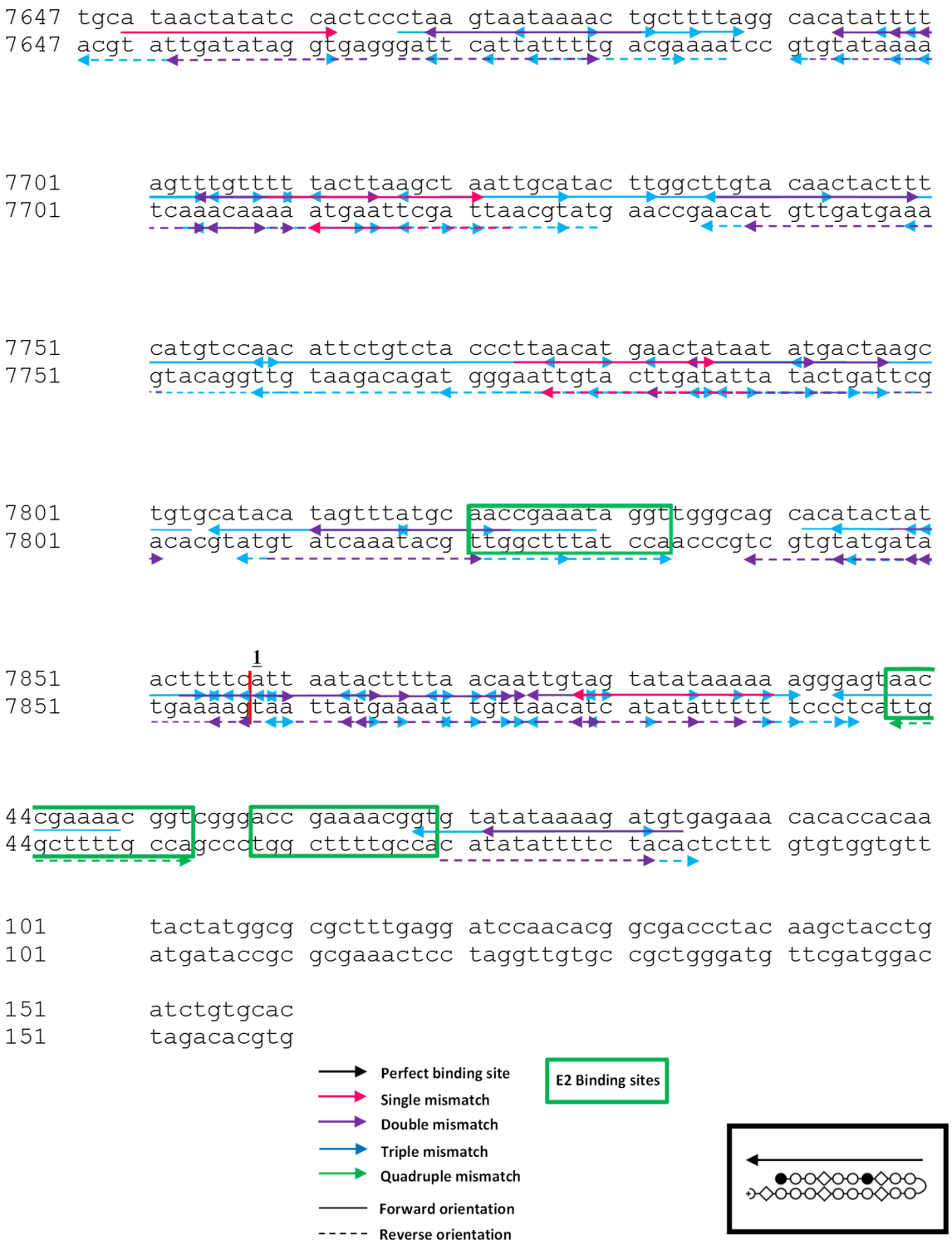


Figure SI2.8. Putative PA25 Binding Sites on the HPV18 Region from 7647-157 bp. According to the literature, **PA25** recognizes the DNA sequence 5'-$W_2GW_5GW_4$-3'. **PA25** binding sites are depicted with either a solid arrow (forward orientation) or a dashed arrow (reverse orientation). The head of the arrow indicates the polyamide end where the Ta is located (see inset). HPV E2 binding sites are depicted in green boxes. HPV E1 binds in the region (HPV18 3-19 bp) flanked by the first and second E2 binding sites. The rationale for the project was that the polyamide would interfere with viral protein binding leading to repression of DNA replication and/or transcription. The origin of replication is marked with 1. Numbers on the left correspond to the nucleotide position in HPV18 genome.

2.7.9 PA25 Single-base-pair Mismatch Sites Predicted, But Not Observed on HPV18 7479-7783 bp

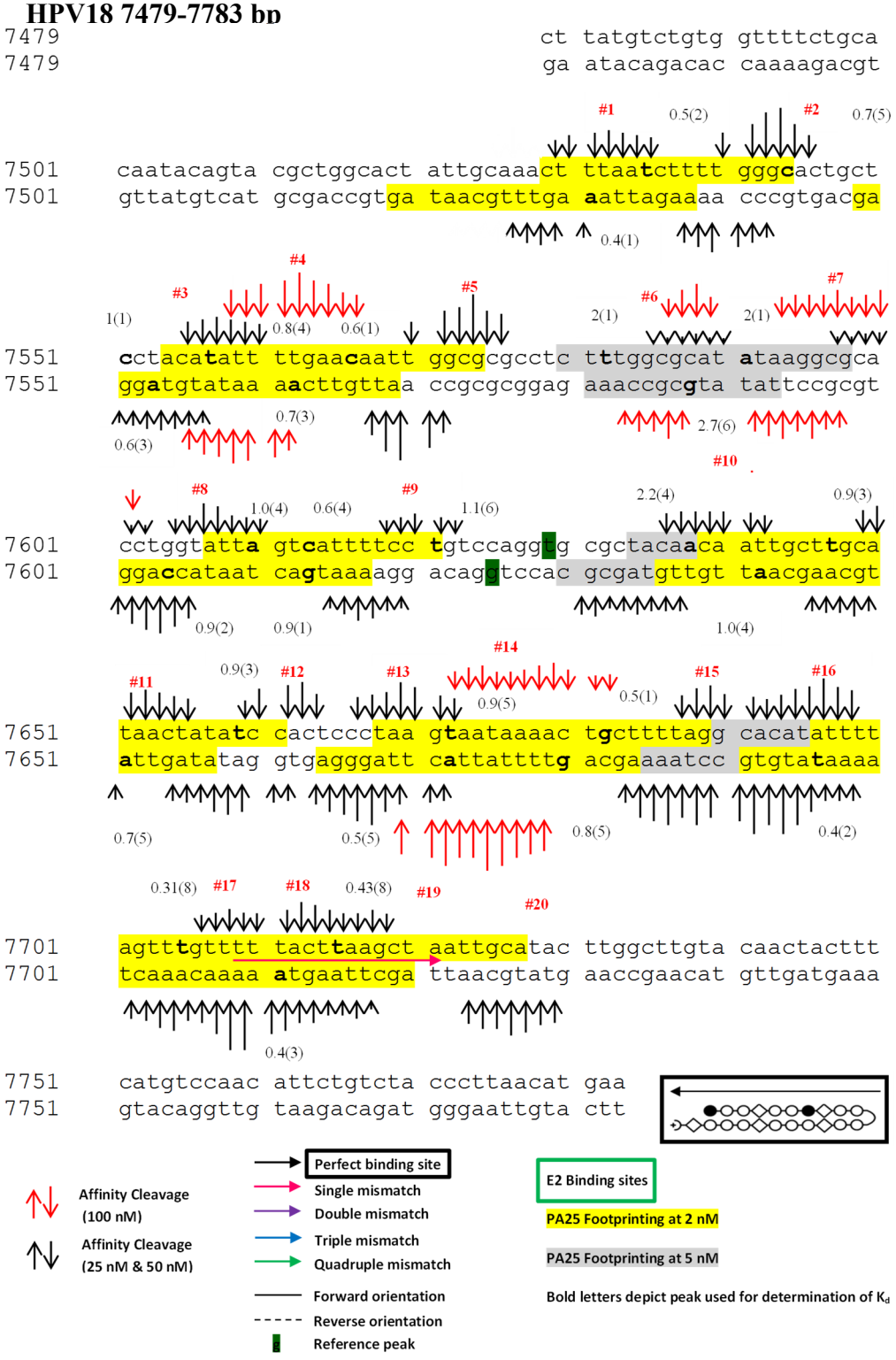


Figure SI2.9. PA25 Single-base-pair Mismatch Sites Predicted by PA-DNA Recognition Rules, But Not Observed on the 305 bp (7479-7783) HPV18 DNA Fragment. Predicted single-base-pair mismatch binding site for **PA25** at 7709-7721 bp is depicted with a horizontal solid arrow (forward orientation) between DNA duplex. The head of the arrow indicates the polyamide end where the Ta is located (see inset). This fragment does not harbor any HPV E2 binding sites. Numbers on the left correspond to the nucleotide position in HPV18 genome.

Table SI2.3. PA25 Predicted Single-Base-Pair Mismatch Sites Which Were Not Observed on 7479-7783 Fragment of HPV18 LCR.

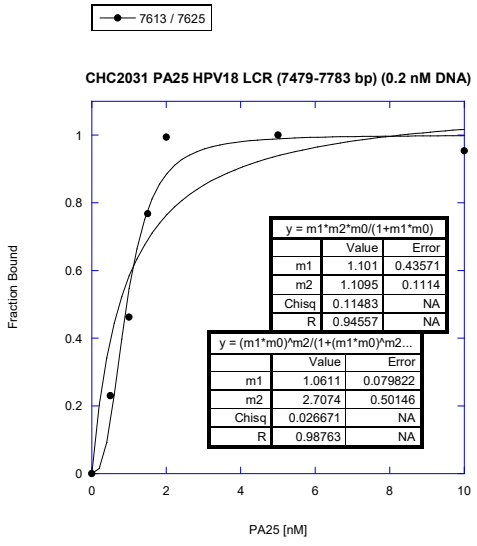
Sequence	Position	Binding orientation	Comment
GCAAT TAG**C**TTAA**G**TAAA AACAA	7709-7721	Forward	No AC

PA25 was predicted to bind this sequence with a single-base-pair mismatch; however, this sequence did not generate an affinity cleavage (AC) pattern.

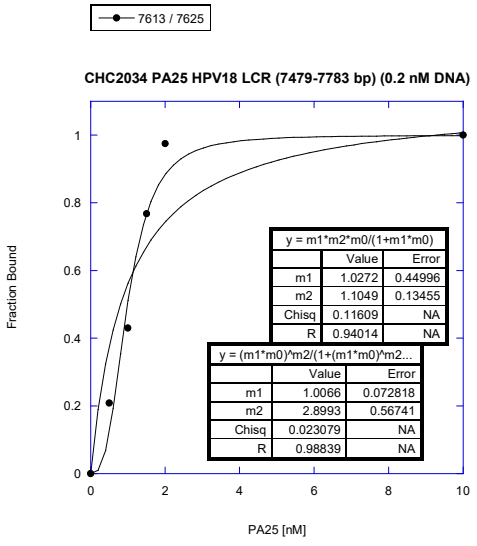
2.7.10 Representative PA25 Binding Isotherms Obtained from Quantitative DNase I Footprinting Sites Along the DNA Fragment Corresponding to 7479-7783 bp of the HPV18 LCR

SITE #8 and #9 (Integration nt: 7613; Reference nt: 7625 in HEX Channel)

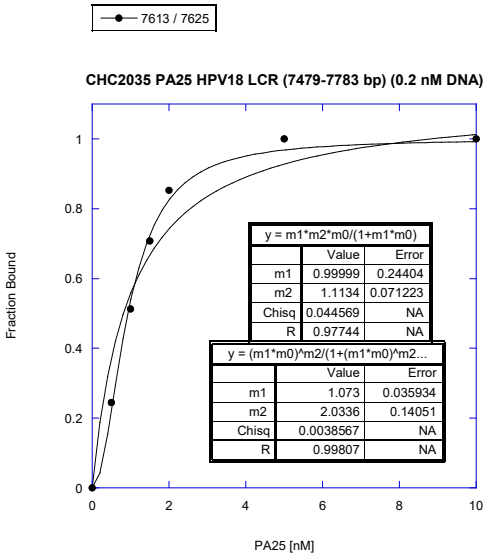
Trial #1



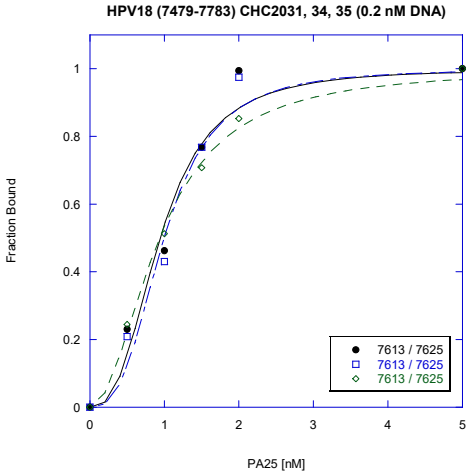
Trial #2



Trial #3

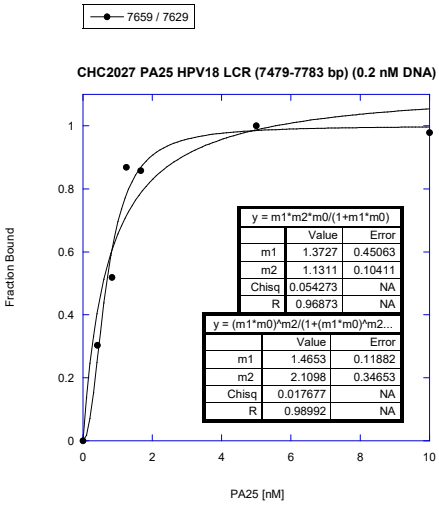


SITE #8 & 9 HEX: All Trials (Hill fit)

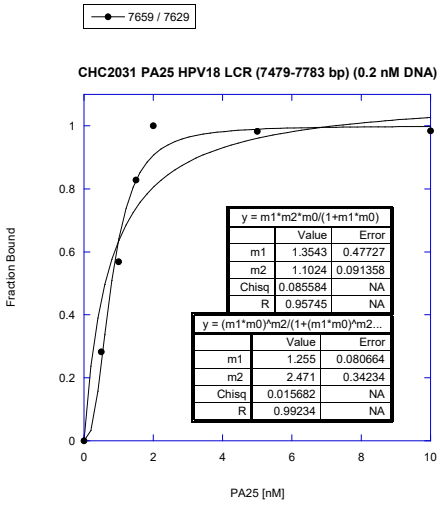


SITE #11 and #12 (Integration nt: 7659; Reference nt: 7629 in FAM Channel)

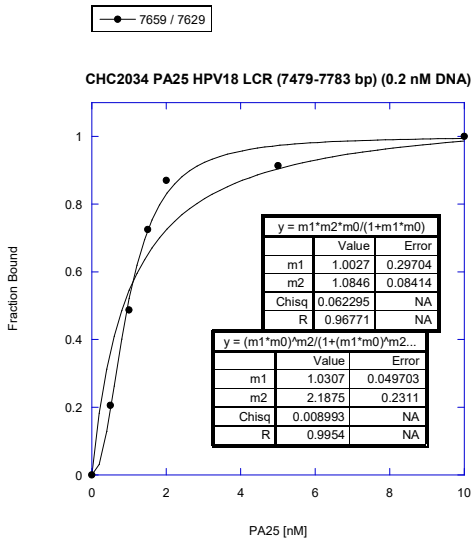
Trial #1



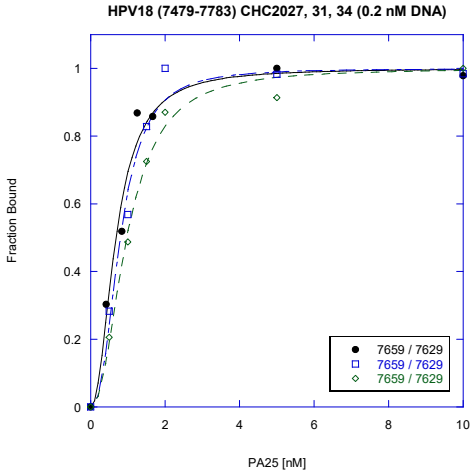
Trial #2



Trial #3

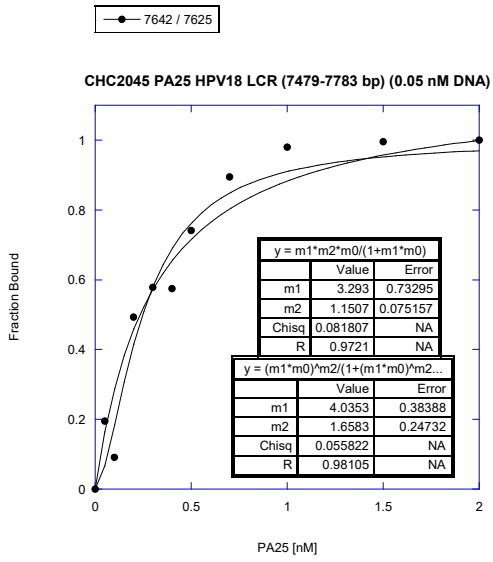


SITE #11 & 12 FAM: All Trials (Hill fit)

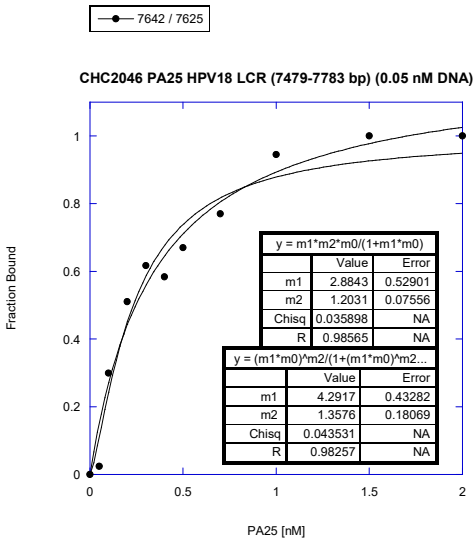


SITE #10 and #11 (Integration nt: 7642; Reference nt: 7625 in HEX Channel)

Trial #1



Trial #2



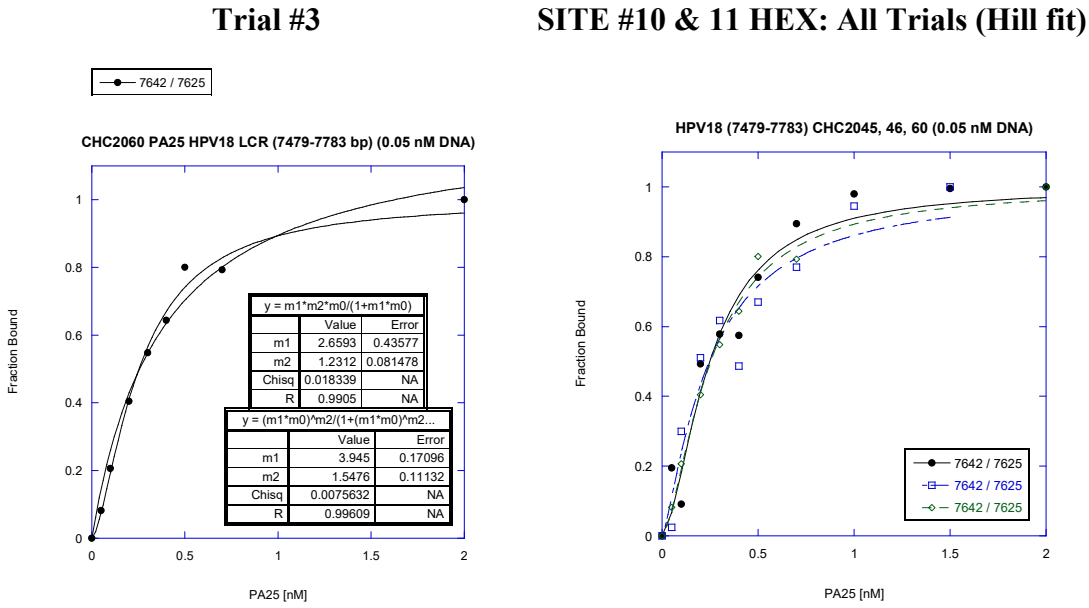


Figure SI2.10. Representative PA25 Binding Isotherms Obtained from Quantitative DNase I Footprinting Sites Along the DNA Fragment Corresponding to 7479-7783 bp of the HPV18 LCR. Binding isotherms for each trial of **PA25** are plotted. Calculated parameters for the Langmuir (*top inset table*) and the Hill (*bottom inset table*) equations (KaleidaGraph 4.1 software). Summary plots are fitted using the Hill equation.

2.7.11 PA25 Single-base-pair Mismatch Sites Predicted, But Not Observed on HPV18 7647-157 bp

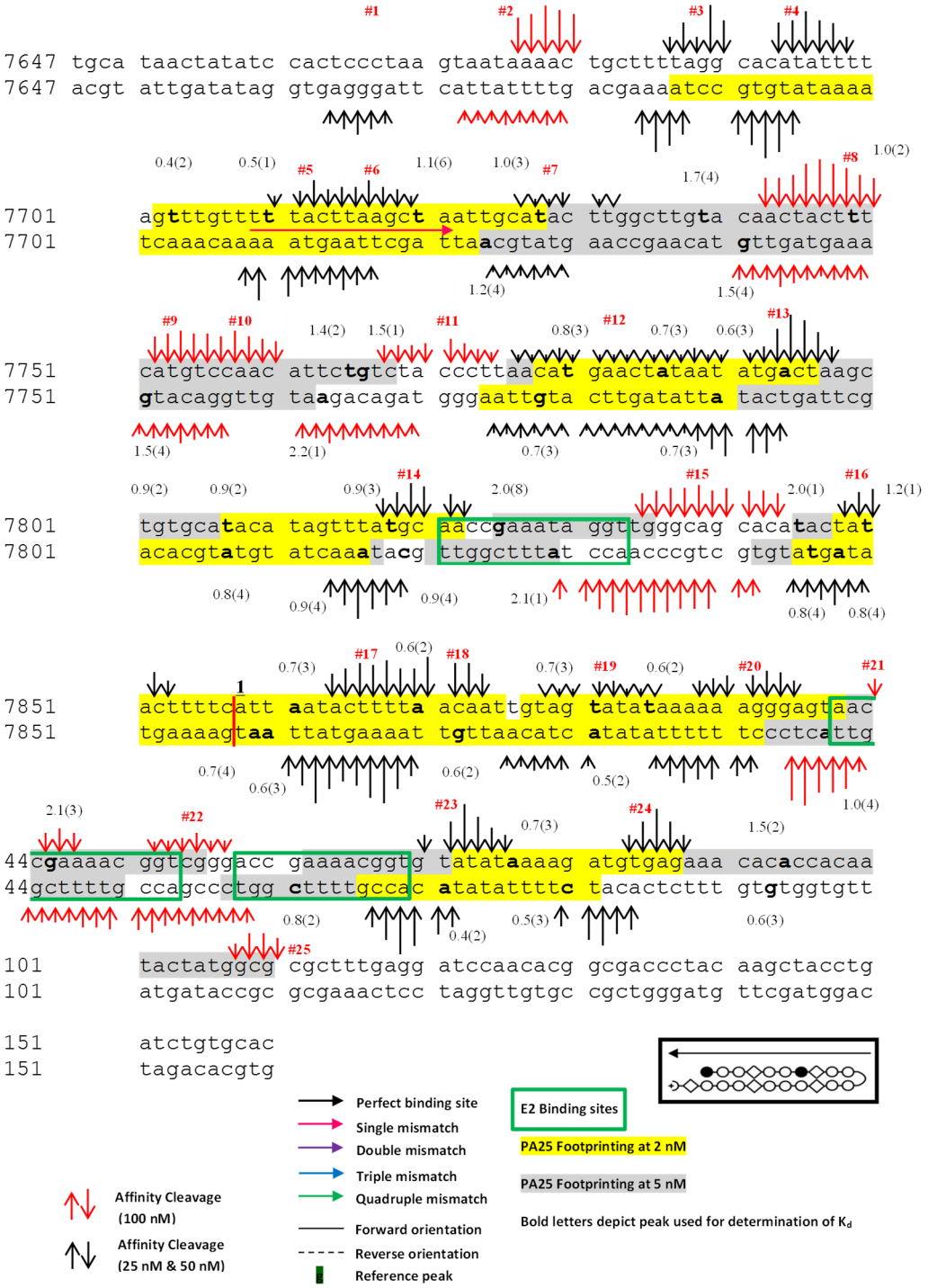


Figure SI2.11. PA25 Single-base-pair Mismatch Sites Predicted by PA-DNA Recognition Rules, But Not Observed on the 368 bp (7647-157) HPV18 DNA Fragment. A predicted single-base-pair mismatch binding site for **PA25** at 7709-7721 bp is depicted with a horizontal solid arrow (forward orientation) between DNA duplex. The head of the arrow indicates the polyamide end where the Ta is located (see inset). HPV E2 binding sites are depicted in green boxes. HPV E1 binds in the region (HPV18 3-19 bp) flanked by the first and second E2 binding sites. The rationale for the project was that the polyamide would interfere with viral protein binding leading to repression of DNA replication and/or transcription. The origin of replication is marked with 1. Numbers on the left correspond to the nucleotide position in HPV18 genome.

Table SI2.4. PA25 predicted single-base-pair mismatch sites which were not observed on 7647-157 fragment of HPV18 LCR.

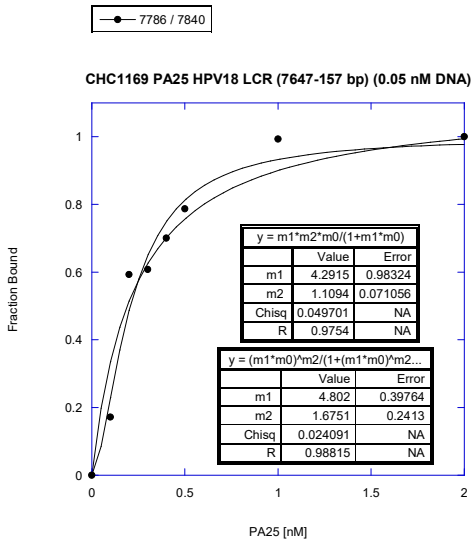
Sequence	Position	Binding orientation	Comment
GCAAT TAG**C**TTAA**G**TAAA AACAA	7709-7721	Forward	No AC

PA25 was predicted to bind this sequence with a single-base-pair mismatch; however, this sequence did not generate an affinity cleavage (AC) pattern.

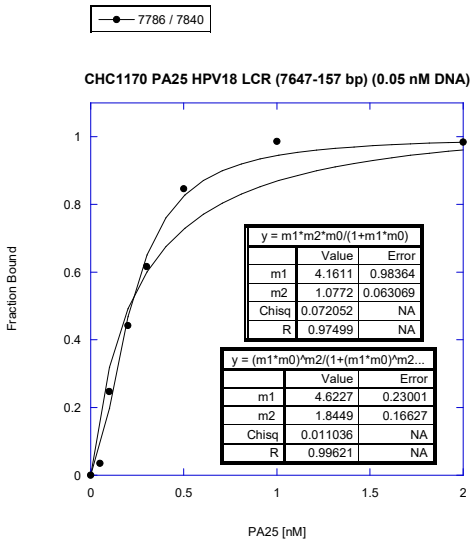
2.7.12 Representative PA25 Binding Isotherms Obtained from Quantitative DNase I Footprinting Sites Along the DNA Fragment Corresponding to 7647-157 bp of the HPV18 LCR

SITE #12 and #13 (Integration nt: 7786; Reference nt: 7840 in FAM Channel)

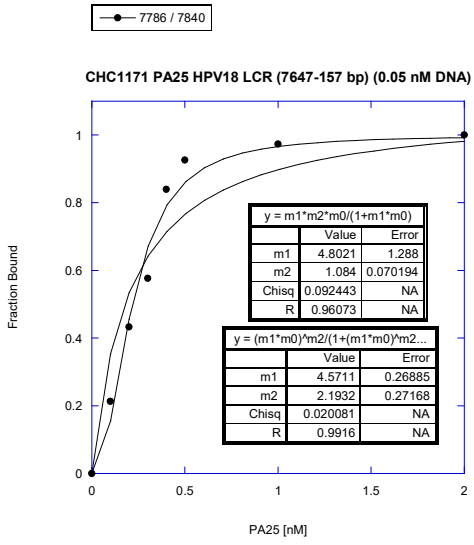
Trial #1



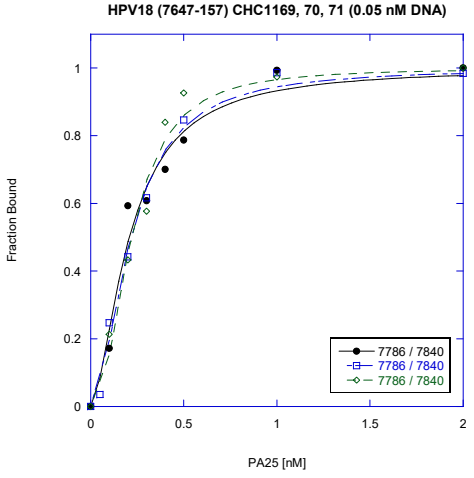
Trial #2



Trial #3

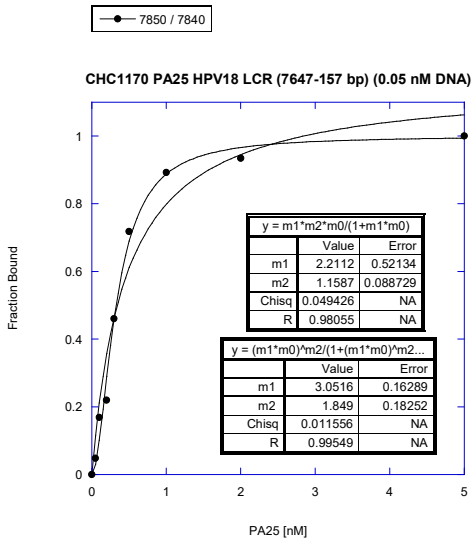


SITE #10 & 11 FAM: All Trials (Hill fit)

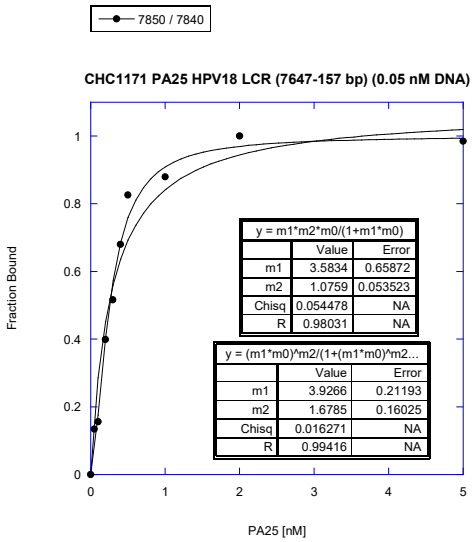


SITE #16 (Integration nt: 7850; Reference nt: 7840 in FAM Channel)

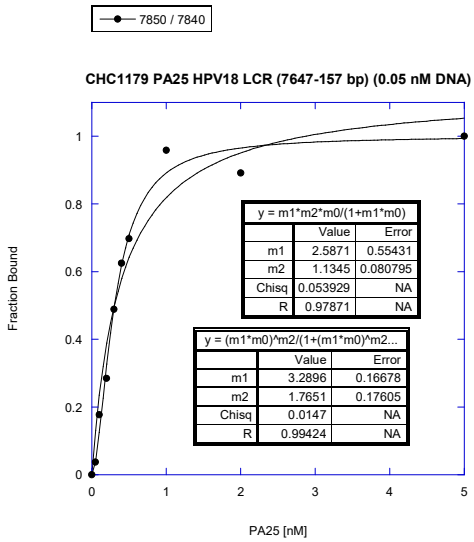
Trial #1



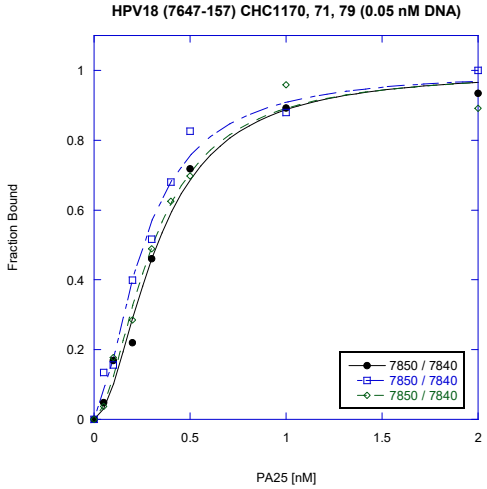
Trial #2



Trial #3



SITE #16 FAM: All Trials (Hill fit)



SITE #20 and #21 (Integration nt: 40; Reference nt: 7837 in HEX Channel)

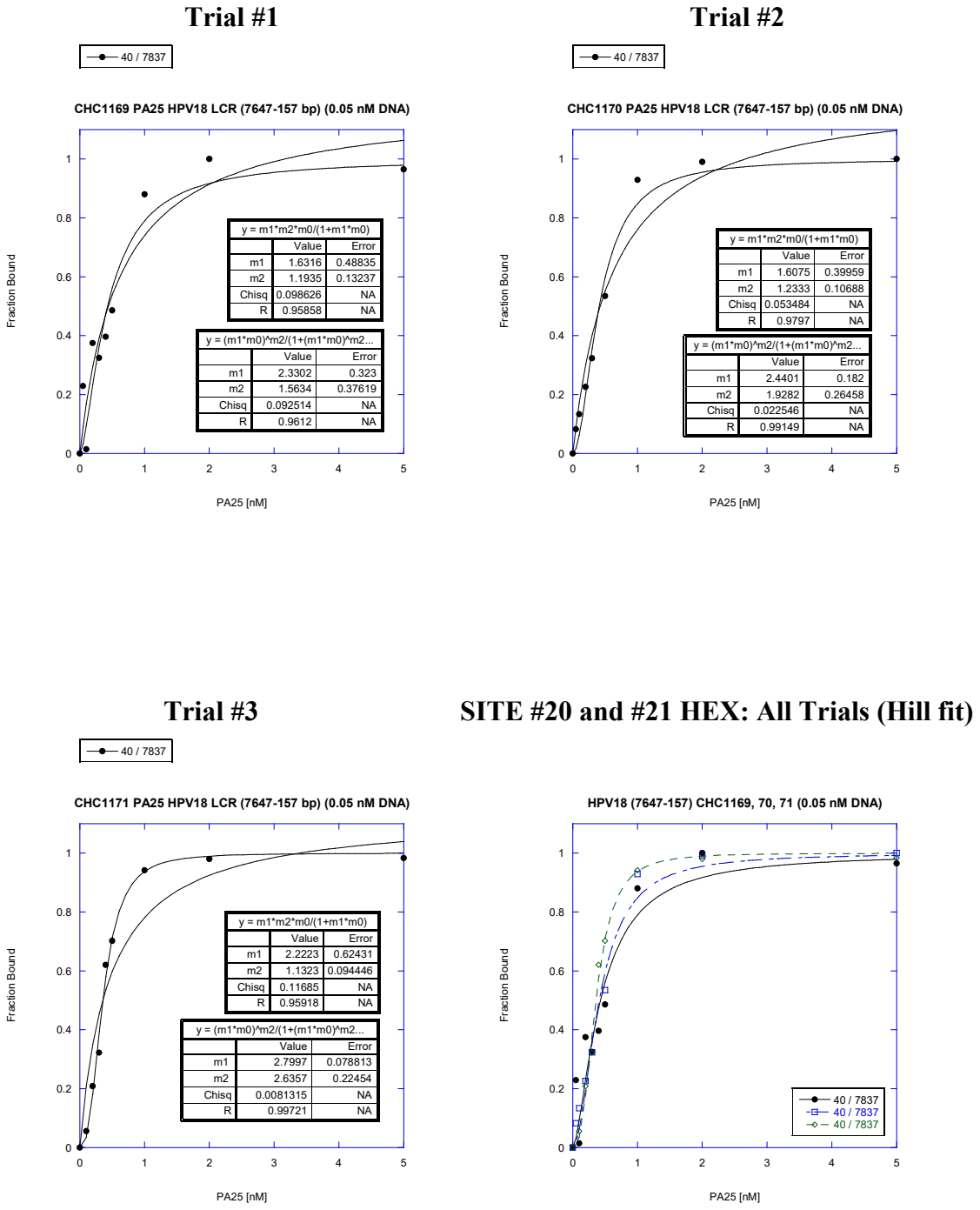


Figure SI2.12. Representative PA25 Binding Isotherms Obtained from Quantitative DNase I Footprinting Sites Along the DNA Fragment Corresponding to 7647-157 bp of the HPV18 LCR. Binding isotherms for each trial of **PA25** are plotted. Calculated parameters for the Langmuir (*top inset table*) and the Hill (*bottom inset table*) equations (KaleidaGraph 4.1 software). Summary plots are fitted using the Hill equation.

Chapter 3

Improved Antiviral Activity of a Polyamide Against High-Risk Human Papillomavirus via N-terminal Guanidinium Substitution

The text of this chapter was taken in part from a publication co-author with Maria J. Scuderi, Terri G. Edwards, George D. Harris, Cynthia M. Dupureur, Kevin J. Koeller, Chris Fisher and James K. Bashkin (Department of Chemistry & Biochemistry, University of Missouri-St. Louis, St. Louis, MO 63121, USA; NanoVir, LLC, Kalamazoo, MI 49008, USA)

Castaneda, C. H.; Scuderi, M. J.; Edwards, T. G.; Harris, G. D.; Dupureur, C. M.; Koeller, K. J.; Fisher, C.; Bashkin, J. K. (2016) Improved antiviral activity of a polyamide against high-risk human papillomavirus via N-terminal guanidinium substitution. *MedChemComm,* doi:10.1039/c6md00371k.

3.1 ABSTRACT

We report the synthesis of two novel pyrrole–imidazole polyamides with N-terminal guanidinium or tetramethylguanidinium groups and evaluate their antiviral activity against three cancer-causing human papillomavirus strains. Introduction of guanidinium improves antiviral activity when compared to an unsubstituted analog, especially in IC_{90} values. These substitutions change DNA-binding preferences, while binding affinity remains unchanged.

3.2 INTRODUCTION

N-methylpyrrole (Py) and *N*-methylimidazole (Im) hairpin polyamides (PAs) are synthetic homologs of the natural antibiotics distamycin A and netropsin which bind AT-rich regions of DNA in the minor groove.[1-4] The antiparallel nature of hairpin PAs allows for the side-by-side pairing of building blocks and recognition of DNA base pairs (bp) within a double strand (dsDNA). It has been extensively reported that Py/Py and Py/β-alanine (β) pairs recognize A/T and T/A base pairs, while an Im/Py pair distinguishes G/C from C/G.[5,6] Furthermore, introduction of the flexible β linker in place of heterocycles resets the heterocycle-base pair register and allows targeting of longer DNA sequences.[7,8]

In numerous reports hairpin PAs distinguish DNA sequences by binding to patterns of hydrogen-bond donors and acceptors in the minor groove of DNA,[6,9-11] so synthetic PAs are attractive for many biomedical applications. Thus, for polyamides themselves and related minor groove binders, reports include the regulation of gene expression,[12-14] biological imaging,[15] and use as antimicrobial/antiviral agents.[16-22] Antiviral polyamides include the long polyamides reported here which, as a class, access a new biological mechanism: they lead to the degradation of the genome of dsDNA tumour viruses in cells while avoiding acute cytotoxicity. The DNA degradation is accomplished *via* the DNA Damage Response (DDR), for which expression is altered significantly only in infected cells carrying small circular, dsDNA genomes of the virus, known as episomes.[23,24] As discussed below, another characteristic of this new class of antiviral polyamides is their failure to follow the reported DNA recognition rules described above.

The process by which the DDR works in concert with the polyamides is not completely established at the molecular level, but many details have emerged as a result of a variety of experiments, a portion of which are summarized here. For example, 224 DDR genes were knocked down individually using siRNA, and we checked to see if the knockdowns affected polyamide activity in any way; 21 genes were found to affect PA activity, and the results were validated by additional experiments using QPCR (rather than relying on the PCR array data alone).[23,24] Some gene knockdowns enhanced PA activity and others opposed PA activity. Additional experiments used Mirin, a small molecule inhibitor of DDR enzyme Mre11. Mre11 has many activities alone and in protein complexes. For example, Mre11 is a member of the MRN complex, which participates in double-strand DNA break repair and telomere maintenance as a member of the ATM branch of the DDR.[25] As a standalone enzyme, one activity associated with MRE11 is that of an endonuclease. MRE11 inhibition experiments were carried out in the presence and absence of polyamides.[23] 100 μM Mirin had no effect on its own, but it sensitized cells to PA25, making the compound 4× more active in terms of lower IC_{50}

and IC_{90} values. Part of the activation of the DDR was also followed by phosphorylation of Rad9, a protein that is phosphorylated upon DNA damage as part of the ATR pathway of the DDR, one of the two major DDR branches (the other being ATM).[23,24,26] Treatment of HPV16 episome bearing W12E cells with active polyamide causes Rad9 phosphorylation within 2–4 hours, but HPV-free cervical cancer cell line C33A shows no phosphorylation of Rad9 upon treatment with polyamide.[23,24]

We previously reported a library of long hairpin PAs, including **PA1** and **PA25**, which exhibits antiviral activity against three high-risk (oncogenic) HPVs.[27-29] HPV is an AT-rich dsDNA virus[30,31] that causes cervical cancer, among other diseases, and subtypes HPV16 and 18 are responsible for 70 % of cervical cancer in most of the world.[32,33] Although specific HPV vaccines are available,[34,35] they are effective prophylactics, not therapies. **PA1**, our first preclinical lead, exhibits anti-HPV activity with no cytotoxicity up to its solubility limit in cell culture media of 200 μM.[27-29] A characteristic of these active anti-HPV polyamides is that they bind at least ten bp, or one full turn of DNA.

Here, we report the antiviral activities and DNA-binding properties of two PAs of a new structural class, **PA30** and **PA31**, where the des-amino N-terminus of the parent compound **PA1** has been substituted with either tetramethylguanidinium (TMG) or guanidinium (Guan), respectively. **PA30** and **PA31** are biomimetic in that netropsin has an N-terminal guanidinium, though **PA30** and **31** are not exact structural analogues of netropsin. While others have explored guanidinylated DNA-binding ligands, no biophysical studies on guanidinylated hairpin PAs have been undertaken, to our knowledge.[36-38] We found that these substitutions at the N-terminus, which are small relative to the overall molecular size, can significantly improve anti-HPV activity. However, improvement in antiviral efficacy cannot be attributed to altered DNA binding strength, as will be shown. Nevertheless, **PA30** and **PA31** display noticeable differences in binding distribution on a fragment of native HPV18 DNA in comparison with parent **PA1**.

3.3 MATERIALS AND METHODS

3.3.1 Polyamide Synthesis and Characterization

Polyamides were synthesized by Boc solid-phase methods[39,40] as reported[27] (**PA1**) or followed by TMG (**PA30**) or Guan (**PA31**) conjugation. After cleavage from the resin, polyamides were purified by reverse-phase HPLC and characterized with ^{1}H and ^{13}C NMR and HR mass spectrometry.

PA1 and its EDTA conjugate were synthesized and characterized as previously described.[27,41,42] **PA30** and **31** were synthesized by virtually identical solid-phase methods up to the Im group (4-amino-N-methyl imidazole-2-carboxamide) and its 4-N-attached guanidinyl substituent. For **PA30**, HATU[43] was utilized as the final building block to incorporate the TMG group onto the H_2N-Im moiety. For **PA31**, in parallel to the literature,[44] treatment with N,N'-Bis (tert-butoxycarbonyl)-1H-pyrazole-1-carboxamidine and diisopropylethylamine for 24 h, followed by TFA treatment to remove the BOC groups produced the desired, unsubstituted Guan group. The EDTA conjugate of **PA30** was prepared via the same method as previously described for **PA1**'s conjugate.[41,42] Briefly, to a vigorously-stirred mixture of EDTA dianhydride (15.8 mg, 0.060 mmol, 22

eq) in N,N-diisopropylethylamine (DIEA) (0.5 mL), DMF (0.25 mL) and DMSO (0.25 mL) at 55 °C was added a mixture of **PA30** (5.9 mg, 0.0027 mmol, 1 eq) in DIEA (0.5 mL) and DMF (0.5 mL) dropwise over a period of 30 min. The reaction mixture was stirred at 55 °C for 30 min. Following addition of 0.1 N NaOH (1.03 mL), the mixture was stirred at 55 °C for an additional 20 min. The bottom layer of the biphasic mixture was removed from the reaction vessel and neutralized with formic acid (0.013 mL). The mixture was diluted with 0.5 mL DMSO, filtered through a 20 μm polyethylene filter, and purified by reversed-phase HPLC using a Phenomenex Luna 250 x 30 mm, 5 μm, 100 Å, C_{18} (version 2) column maintained at 25 ˚C. Mobile phases were 0.2 % formic acid in water (A) and 100 % methanol (B). The gradient was 35 % B for 5 min followed by a ramp to 70 % B over 40 min at 50 mL/min. Concentration of pooled fractions, followed by lyophilization, gave the **PA30-EDTA** conjugate (1.0 mg, 16 % yield) as a white, fluffy solid: exact mass $[M+H]^+$= 2282.0657, experimental (ESI) $[M+H]^+$ = 2282.1257. The EDTA conjugate of **PA31** was prepared *via* a virtually identical procedure to the one described for **PA30**, but with the following changes. **PA31** (4.0 mg) was substituted for **PA30**. The preparative HPLC purification parameters were identical to those described above except mobile phase A consisted of 0.2 % trifluoroacetic acid in water. Concentration of pooled fractions, followed by lyophilization gave the **PA31-EDTA** conjugate (2.3 mg, 55 % yield) as an off-white fluffy solid: exact mass $[M+2H+Na]^+$ = 2250.0007, experimental (ESI) $[M+2H+Na]^+$ = 2249.9562. Analytical HPLC characterization was performed using a Phenomenex Jupiter Proteo (C12), 4.6 x 50 mm, 4 μm, 90 Å maintained at 40˚C. Mobile phases consisted of 0.1 % formic acid in water (A) and 100 % acetonitrile (B). The gradient consisted of 5 % B for 0.75 min followed by a ramp to 60 % B over 3.25 min at 2 mL/min. Retention times were 2.641 min, 2.486 min and 2.437 min for **PA1**, **PA30** and **PA31**, respectively **(Figure 3.1)**.

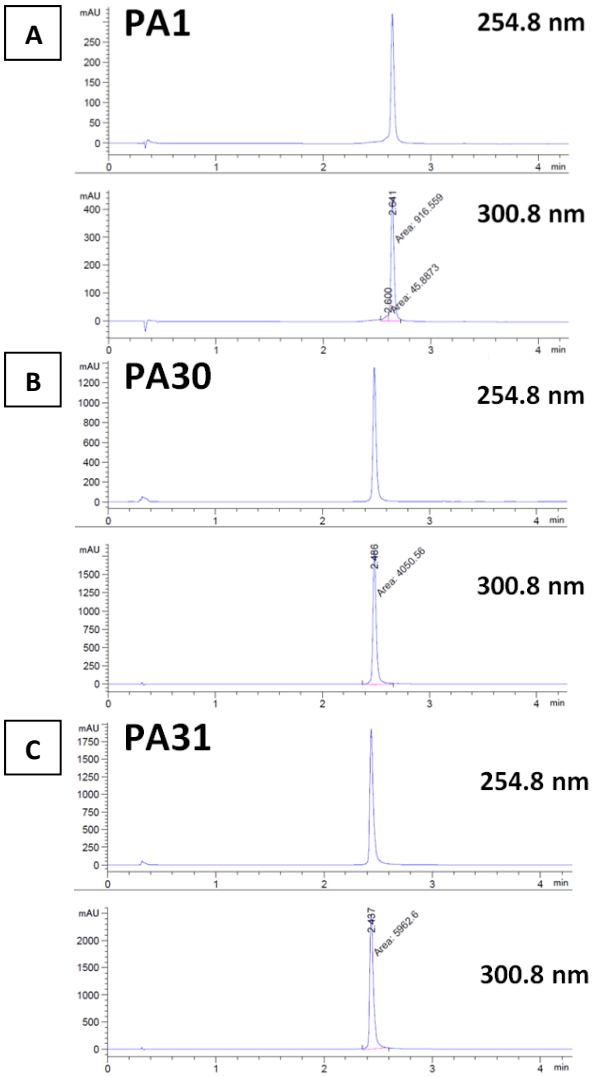


Figure 3.1. HPLC Purity of Synthesized Polyamides with λ Monitored. (A) PA1 (98 % purity, retention time = 2.64 min), **(B) PA30** (99 % purity, retention time = 2.49 min) and **(C) PA31** (98 % purity, retention time = 2.44 min)

3.3.1.1 PA1 Im-Py-Py-β-Py-Py-Py-β-Py-Py-β-Py-Py-Py-Py-β-Ta (3 TFA)

^{1}H NMR (600 MHz, DMSO-d$_6$) δ = 10.46 (s, 1 H), 9.93 (s, 1 H), 9.91 (s, 4 H), 9.90 - 9.87 (m, 2 H), 9.86 (s, 1 H), 9.83 (s, 1 H), 9.63 (br. s., 2 H), 8.12 - 8.02 (m, 5 H), 7.98 - 7.92 (m, 1 H), 7.88 (br. s., 3 H), 7.40 (s, 1 H), 7.28 (d, *J* = 1.8 Hz, 1 H), 7.24 - 7.21 (m, 3 H), 7.21 - 7.18 (m, 4 H), 7.18 - 7.16 (m, 3 H), 7.15 (d, *J* = 1.8 Hz, 1 H), 7.08 (d, *J* = 1.8 Hz, 1 H), 7.07 (d, *J* = 1.8 Hz, 1 H), 7.06 (s, 1 H), 7.04 (d, *J* = 1.8 Hz, 1 H), 6.90 (d, *J* = 1.8 Hz, 1 H), 6.89 (d, *J* = 1.8 Hz, 1 H), 6.89 (s, 2 H), 6.85 (s, 2 H), 6.84 (d, *J* = 1.2 Hz, 1 H), 3.99 (s, 3 H), 3.85 (s, 3 H), 3.85 (s, 3 H), 3.84 (s, 9 H), 3.82 (s, 6 H), 3.81 (s, 3 H), 3.81 (s, 9 H), 3.49 - 3.42 (m, 4 H), 3.41 - 3.35 (m, 4 H), 3.25 - 3.16 (m, 4 H), 3.16 - 3.10 (m, 2 H), 3.10 - 2.98 (m, 4 H), 2.87 (br. s, 2 H), 2.74 (d, *J* = 4.7 Hz, 3 H), 2.55 - 2.51 (m, 2 H), 2.36 (t, *J* = 7.0 Hz, 2 H), 2.28 (t, *J* = 7.3 Hz, 2 H), 1.96 - 1.87 (m, 2 H), 1.82 - 1.74 (m, 2 H)

^{13}C NMR (151 MHz, DMSO-d_6) δ = 171.1, 169.3, 167.9, 167.8, 161.3, 158.6, 158.5, 158.5, 158.4, 158.4, 158.4, 158.1, 157.9, 155.9, 138.7, 126.8, 126.3, 123.0, 122.9, 122.8, 122.8, 122.8, 122.7, 122.7, 122.2, 122.2, 122.1, 122.1, 122.1, 122.0, 121.9, 121.4, 118.6, 118.5, 118.4, 118.2, 118.1, 118.0, 117.8, 104.9, 104.8, 104.8, 104.7, 104.3, 104.3, 104.2, 104.0, 104.0, 103.9, 53.3, 52.1, 40.0, 39.3, 38.2, 36.2, 36.1, 36.1, 36.0, 35.8, 35.8, 35.6, 35.6, 35.5, 35.4, 35.1, 33.3, 25.7, 24.0, 21.8

HRMS (ESI) calculated for $C_{91}H_{111}N_{31}O_{16}$ $[MH]^+$, 1894.8908, found, 1894.8933.

HPLC purity: 98 %. (see **Figure 3.1**).

3.3.1.2 PA30 TMG-Im-Py-Py-β-Py-Py-Py-β-Py-Py-β-Py-Py-Py-Py-β-Ta (4 TFA)

^{1}H NMR (600 MHz, DMSO-d_6) δ = 10.19 (s, 1 H), 10.08 (s, 1 H), 9.94 (s, 1 H), 9.92 (s, 3 H), 9.91 (s, 1 H), 9.90 (s, 1 H), 9.87 (s, 1 H), 9.84 (s, 1 H), 9.45 (br. s., 1 H), 8.13 - 8.03 (m, 5 H), 7.86 (br. s., 1 H), 7.80 (br. s., 3 H), 7.29 (d, *J* = 1.8 Hz, 1 H), 7.23 (d, *J* = 1.2 Hz, 2 H), 7.22 - 7.21 (m, 2 H), 7.21 - 7.18 (m, 4 H), 7.16 (s, 3 H), 7.15 (d, *J* = 1.8 Hz, 1 H), 7.08 (d, *J* = 1.8 Hz, 1 H), 7.07 (d, *J* = 1.8 Hz, 1 H), 7.05 (d, *J* = 1.2 Hz, 1 H), 6.91 (d, *J* = 1.8 Hz, 1 H), 6.89 (d, *J* = 1.8 Hz, 2 H), 6.88 (s, 1 H), 6.85 (s, 2 H), 6.84 (d, *J* = 1.2 Hz, 1 H), 6.56 (br. s., 2 H), 3.99 (s, 3 H), 3.85 (s, 3 H), 3.85 (s, 3 H), 3.84 (s, 3 H), 3.84 (s, 3 H), 3.84 (s, 6 H), 3.81 (s, 6 H), 3.81 (s, 3 H), 3.80 (s, 6 H), 3.50 - 3.26 (m, 6 H), 3.24 - 3.16 (m, 3 H), 3.16 - 3.09 (m, 2 H), 3.09 - 2.98 (m, 3 H), 2.98 - 2.80 (m, 14 H), 2.74 (d, *J* = 5.3 Hz, 3 H), 2.55 - 2.51 (m, 4 H), 2.36 (t, *J* = 7.3 Hz, 2 H), 2.27 (t, *J* = 7.3 Hz, 2 H), 1.94 - 1.85 (m, 2 H), 1.82 - 1.73 (m, 4 H)

^{13}C NMR (151 MHz, DMSO-d_6) δ = 171.0, 169.3, 167.9, 167.8, 161.3, 158.5, 158.5, 158.4, 158.4, 158.4, 158.2, 157.9, 157.8, 155.4, 135.9, 134.8, 123.1, 122.9, 122.8, 122.8, 122.8, 122.7, 122.7, 122.2, 122.2, 122.1, 122.1, 122.1, 122.1, 122.0, 121.9, 121.1, 118.9, 118.5, 118.5, 118.2, 118.1, 118.0, 117.8, 115.8, 105.0, 104.8, 104.8, 104.7, 104.3, 104.3, 104.0, 104.0, 103.9, 53.3, 52.1, 40.0, 39.3, 38.2, 36.2, 36.1, 36.1, 36.0, 35.8, 35.8, 35.6, 35.6, 35.5, 35.4, 33.3, 25.7, 24.0, 22.5, 21.8

HRMS (ESI) calculated for $C_{96}H_{122}N_{34}O_{16}$ $[M]^+$, 2006.9778, found, 2006.9653.

HPLC purity: 99 %.(see **Figure 3.1**).

3.3.1.3 PA31 Guan-Im-Py-Py-β-Py-Py-Py-β-Py-Py-β-Py-Py-Py-Py-β-Ta (4 TFA)

^{1}H NMR (600 MHz, DMSO-d_6) δ = 10.46 (s, 1 H), 10.42 (s, 1 H), 9.94 (s, 1 H), 9.93 (s, 1 H), 9.92 (s, 3 H), 9.91 (s, 1 H), 9.90 (s, 1 H), 9.87 (s, 1 H), 9.84 (s, 1 H), 9.46 (br. s., 1 H), 8.13 - 8.07 (m, 4 H), 8.06 (t, *J* = 5.6 Hz, 2 H), 7.86 (br. s., 1 H), 7.81 (br. s., 3 H), 7.28 (d, *J* = 1.8 Hz, 1 H), 7.23 (d, *J* = 1.2 Hz, 2 H), 7.22 (d, *J* = 1.2 Hz, 1 H), 7.21 - 7.18 (m, 5 H), 7.17 (s, 3 H), 7.14 (d, *J* = 1.8 Hz, 1 H), 7.08 (d, *J* = 1.2 Hz, 1 H), 7.07 (d, *J* = 1.8 Hz, 1 H), 7.04 (d, *J* = 1.8 Hz, 1 H), 6.91 (d, *J* = 1.8 Hz, 1 H), 6.89 (s, 2 H), 6.88 (d, *J* = 1.8 Hz, 1 H), 6.87 - 6.85 (m, 2 H), 6.84 (d, *J* = 1.8 Hz, 1 H), 6.69 (s, 1 H), 6.56 (br. s., 2 H), 4.00 (s, 3 H), 3.86 (s, 3 H), 3.85 (s, 3 H), 3.84 (s, 3 H), 3.84 (s, 3 H), 3.84 (s, 6 H), 3.81 (s, 6 H), 3.81 (s, 3 H), 3.80 (s, 6 H), 3.51 - 3.32 (m, 6 H), 3.24 - 3.16 (m, 3 H), 3.16 - 3.09 (m, *J* = 6.2, 12.5, 12.5 Hz, 2 H), 3.09 - 2.98 (m, 3 H), 2.90 - 2.81 (m, 2 H), 2.74 (d, *J* = 4.7 Hz, 3 H), 2.55 - 2.51 (m, 4 H), 2.36 (t, *J* = 7.3 Hz, 2 H), 2.27 (t, *J* = 7.3 Hz, 2 H), 1.94 - 1.85 (m, 2 H), 1.82 - 1.75 (m, 4 H)

^{13}C NMR (151 MHz, DMSO-d_6) δ = 171.1, 169.3, 167.9, 161.3, 161.3, 158.6, 158.5, 158.5, 158.4, 158.4, 158.3, 158.2, 158.0, 155.2, 154.6, 134.7, 134.4, 123.1, 122.9, 122.8, 122.8, 122.7, 122.7, 122.2, 122.2, 122.1, 122.1, 122.1, 122.0, 121.9, 120.9, 118.9, 118.5, 118.4, 118.2, 118.1, 118.0, 118.0, 117.8, 116.1, 113.7, 105.1, 104.8, 104.8, 104.7, 104.3, 104.3, 104.2, 104.0, 104.0, 103.9, 53.3, 52.1, 40.0, 39.3, 38.2, 36.2, 36.2, 36.1, 36.0, 36.0, 35.8, 35.8, 35.6, 35.6, 35.5, 35.4, 35.4, 33.3, 25.7, 24.1, 22.5, 21.8

HRMS (ESI) calculated for $C_{92}H_{114}N_{34}O_{16}$ $[M]^+$, 1950.9152, found, 1950.9034.

HPLC purity: 98 %. (see **Figure 3.1**).

3.3.2 Compound Efficacy Testing

The antiviral activities were calculated by measuring the ability of PAs to decrease the viral DNA (or episomal) load in monolayer keratinocyte cultures. The episomal load was determined by Q-PCR as previously described.[27] **PA1** was already reported, along with extensive discussion of how antiviral parameters are defined for this new mechanism of antiviral action in which the viral genome is degraded while the host cells are unharmed as measured by standard MTT and LDH assays and observations of numbers of floating *vs.* adherent cells, etc.[23,27-30] Typical assays were run for 48 h though much longer assays have also been run, and **ref. 27** also describes organotypic, differentiated keratinocyte tissue culture assays as well as cell monolayer assays. Because **PA1** has been central to our work for over a decade, it has been used as an internal standard to benchmark new compound series, so even though we cite the prior study, **PA1** was studied side-by-side with **PA30** and **PA31** for the current work. The new **PA1** results were indistinguishable from the dose-response curves previously reported. Dose-response curves for **PA30** and **PA31** are given below in **Figure 3.4A** and **3.4B.**

3.3.3 Molar Extinction Coefficient Determination

The molar extinction coefficient for **PA1** was determined as ε = 88,235 $M^{-1}cm^{-1}$ by fluorescence as previously described.[45] The molar extinction coefficients for **PA30** and **PA31** were determined as ε = 91,164 $M^{-1}cm^{-1}$, ε = 89,263 $M^{-1}cm^{-1}$ by measuring the absorbance at 305 nm of solutions of known polyamide concentration. Briefly, stock solutions of 8.44 x 10^{-4} M of **PA30** and 9.22 x 10^{-4} M of **PA31** were prepared by separately diluting 1.04 mg of **PA30** (4TFA, MW = 2464.3 g/mol, Lot KJK6169) and 1.11 mg of **PA31** (4TFA, MW = 2408.2 g/mol, Lot KJK6170) in 500 μL of DMSO. These stocks were further diluted in DMSO to known polyamide concentrations in the range 5 x 10^{-7} M to 7 x 10^{-5} M using cuvettes of various path lengths (0.2, 1, 5 cm). Since Beer-Lambert Law states that A = εbc (where A is the absorbance at a particular wavelength, ε is the molar extinction coefficient, b is the cell path length, and c is the concentration of the sample), a graph was constructed plotting the measured absorbance versus the polyamide concentration (M). The slope of this line corresponds to the molar extinction coefficient **(Figure 3.2)**.

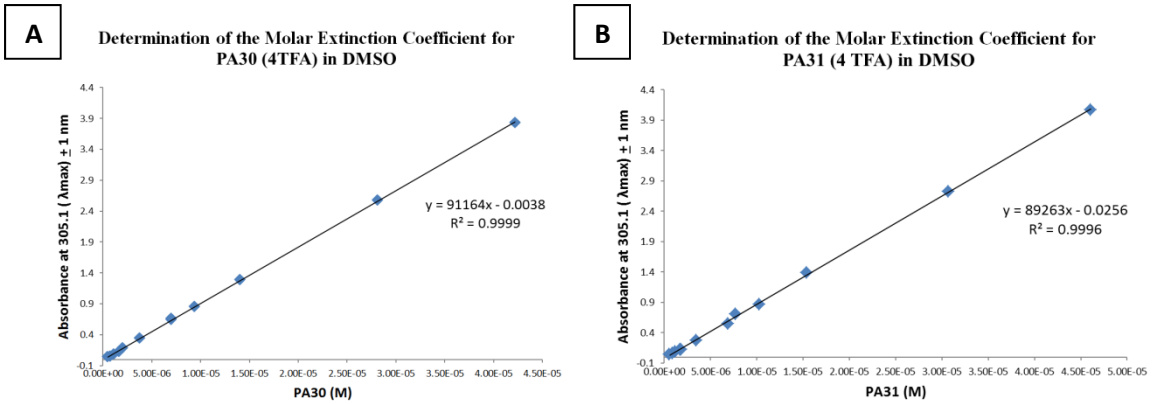


Figure 3.2. Plots of Measured Absorbance at 305 nm for Different Concentrations of PA30 and PA31. Experimental data used for the determination of the molar extinction coefficients for **(A) PA30** and **(B) PA31** in DMSO.

3.3.4 Quantitative DNase I Footprinting

Quantitative DNase I footprinting experiments were performed as described previously.[41,42] Briefly, a 305 bp DNA fragment corresponding to the nucleotides 7479-7783 in the Long Control Region (LCR) of the HPV18 genome (Genbank accession number X05015) was PCR amplified and fluorescently-labeled at both 5' ends using the following primers: (top strand) 5'-FAM-CT TAT GTC TGT GGT TTT CTG and (bottom strand) 5'-HEX-TT CAT GTT AAG GGT AGA CAG. The DNA fragment was diluted to 200 pM in TKMC buffer with 10 mM CHAPS and incubated with different concentrations of each polyamide (0.5-40 nM) for at least 4 h at 37 °C. The polyamide solutions were always added to DNA solutions not *vice versa*. At the end of this incubation period, each solution was subjected to ~0.08 U of DNase I for 5 min. The reactions were then quenched with EDTA and the DNA fragments were purified using a QIAquick PCR purification kit. Each experiment was performed at least three times.

3.3.5 Affinity Cleavage

Affinity cleavage experiments were performed as previously described.[29,46] Briefly, PA-EDTA conjugates were complexed with 0.8 eq of Fe^{2+} in the form of ammonium iron(II) sulfate hexahydrate. 1 nM DNA fragment in 10 mM Tris, 10 mM CHAPS, pH 7.5 was incubated with different concentrations of PA-EDTA-conjugates (5-200 nM) for at least 4 h at 37 °C. The reaction was initiated with 5 µL of 100 mM DTT and incubated at room temperature for 1-2 h. Each cleavage reaction was then quenched and fragments were purified using a QIAquick PCR purification kit.

The samples from quantitative DNase I footprinting and affinity cleavage experiments were analyzed using capillary electrophoresis and indexed using Sanger and Maxam-Gilbert sequencing, respectively.[46] Dissociation constant (K_d) values were calculated by normalizing the area from a peak in the footprint region to the area of a peak not sensitive to PA concentration. These values were plotted and fitted to Hill and Langmuir binding isotherms (as appropriate) with KaleidaGraph 4.1 software in methods reported in detail elsewhere.[29,41,42,46]

3.4 RESULTS

Structures of **PA1**, **PA30** and **PA31** and their EDTA-conjugates used in affinity cleavage maps are shown in **Figure 3.3**. Anti-HPV activities and MTT cell toxicity assays (shown for **PA1**, **Figure SI3.1**) were determined as previously described.[27] New side-by-side, concentration-dependent assays were performed for previously reported **PA1**.[27] **Table 3.1** provides the resulting pseudo IC_{50} and IC_{90} values for **PA1**, **PA30** and **PA31** corresponding to the PA concentration that causes a 50 % or 90 % decrease in viral DNA concentration.

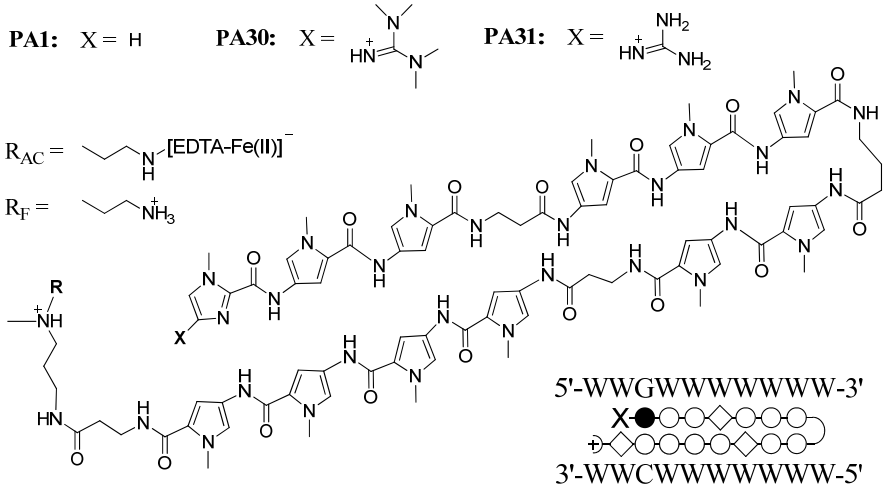


Figure 3.3. Chemical Structure of PA1, PA30 and PA31. R_F refers to the substituent used for DNase I footprinting experiments, whereas R_{AC} indicates the EDTA-Fe conjugate for affinity cleavage experiments. Inset: cartoon representation of the PA with building blocks represented by open circles for N-methylpyrrole, filled circles for N-methylimidazole, diamonds for β, + for the Ta tail and ⊃ for γ-aminobutyric acid. **PA1**, **PA30**, and **PA31** share the same predicted DNA binding site (5'-WWGWWWWWWW-3', where W = A or T) and a general PA representing all three is depicted between two complementary DNA strands.

Table 3.1 IC_{50} and IC_{90} Values of **PA1**,[a] **PA30,** and **PA31** against High-Risk HPV16, HPV18 and HPV31 Types.

	HPV16			HPV18			HPV31		
	IC_{50} (μM)	**IC_{90} (μM)**	**n**	**IC_{50} (μM)**	**IC_{90} (μM)**	**n**	**IC_{50} (μM)**	**IC_{90} (μM)**	**n**
PA1	0.1(1)	1.1(8)	4	0.7(4)	>10	3	0.1(1)	1.0(5)	4
PA30	0.3(1)	>10	3	0.2(2)	>10	3	0.13(4)	0.9(7)	3
PA31	0.10(1)	0.38(8)	3	0.2(2)	0.9(6)	3	0.14(4)	0.8(2)	3

The numbers in parentheses are sample standard deviations; n is the number of independent measurements. [a] IC_{50} values for **PA1** were first reported in reference 27; we re-measured them and found them unchanged.

Figure 3.4 gives dose-response curves for **PA30** and **PA31** while data are given for **PA1** against HPV16 and HPV31 in **Figure SI3.1**; data for **PA1** were the same as previously reported.[27,47] Although IC_{50} values for the compounds are comparable against HPV16 and HPV31, **PA30** and **PA31** afforded somewhat improved IC_{50} values against HPV18. In contrast, significant improvement in IC_{90} for HPV16 and HPV18 was observed with **PA31**. However, **PA30** activity plateaus just below a 90 % decrease in viral DNA and, in two days, does not quite reach an IC_{90}, an important measure of antiviral activity. We will explore longer time intervals, but higher concentrations do not

seem promising based on the dose-response curves, and would make **PA30** a poor competitor with the more-active **PA31**.

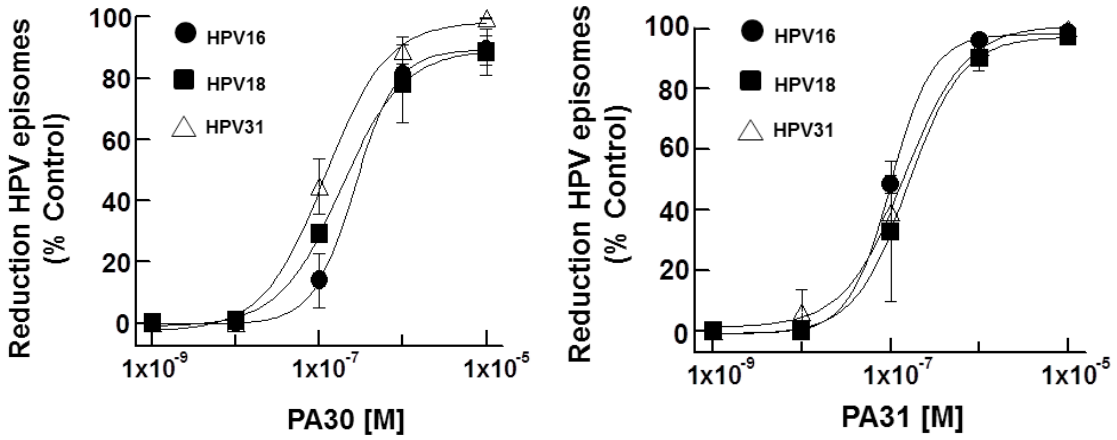


Figure 3.4. Dose-response Curves for PA30 and PA31 Against HPV16, 18 and 31. IC_{50} and IC_{90} values were determined for **(A) PA30** and for **(B) PA31**.

Because only **PA31** showed an improved IC_{90} against HPV18, we investigated the DNA-binding properties of these hairpin PAs on a natural HPV18 sequence and compared them to the parent compound, **PA1**. Based on well-established PA-DNA pairing rules,[6] **PA1**, **PA30**, and **PA31** were expected to target the ten base-pair DNA sequence WWGWWWWWWW, where W=A or T (**Figure 3.3 inset**). A PA-DNA mismatch refers to any deviation from this recognition motif, *i.e.* where a PA building block pair binds to a noncanonical Watson-Crick base pair: one not predicted by the pairing rules.[1] We used quantitative DNase I footprinting[48,49] to obtain dissociation constants (K_d) and affinity cleavage experiments to assign binding orientations of the compounds on a 305 bp DNA fragment corresponding to nucleotides (nt) 7479 to 7783 in the Long Control Region (LCR) of HPV18, a viral genomic region containing regulatory elements for replication.[50] We determined K_d values and orientation for all PA binding events along the entire 305 bp DNA fragment (see **Figure SI3.2** for **PA30** binding sites in this fragment), and representative dissociation constants at three binding sites are given in **Table 3.2**. Similar results were observed in the rest of the DNA fragment (see **Figure SI3.2** and **Table SI3.1** for **PA30** dissociation constants for entire fragment). K_d was determined by the Hill or Langmuir equation, based on the magnitude of the derived Hill coefficient and overall fit (see **Table 3.3 and 3.4**).[41,42]

Table 3.2. Representative **PA1**, **PA30** and **PA31** Binding Sites on nt 7479-7783 of the HPV18 LCR as Determined by Quantitative DNase I Footprinting and Affinity Cleavage.

				K_d (nM)		
Site	**Sequence**	**Position**	**Site Type**	**PA1**	**PA30**	**PA31**
1	CCTGG TATTA**G**T<u>C</u>AT TTTCC	7606-7615	Single-mismatch	0.9 ± 0.1	1.2 ± 0.2	1.0 ± 0.1
2	ACATA TTTT**G**AA<u>C</u>AA TTGGC	7559-7568	Single-mismatch	0.9 ± 0.2	1.1 ± 0.2	1.0 ± 0.1
3	CTTTG **GC**<u>G</u>**C**ATATAA GGCGC	7585-7594	Triple-mismatch	2.7 ± 0.5	3.2 ± 0.3	2.2 ± 0.2

DNA sequences are illustrated so that the middle block represents the PA binding site along with 5 flanking nt. In each sequence, the underlined base is bound by imidazole/pyrrole (or pyrrole/imidazole) pair of the polyamide. The bold bases in column 2 correspond to mismatches according to recognition "rules." K_d values were calculated from DNase I experiments performed at least in triplicate (**Table 3.3 and 3.4**).

Table 3.3. K_d Values Determined by Hill Equation for Sites 1-3 with Hill Coefficients and Overall Fit.

	PA1			**PA30**			**PA31**		
	***Kd* (Hill)**	**Hill Coeff.**	**R**	***Kd* (Hill)**	**Hill Coeff.**	**R**	***Kd* (Hill)**	**Hill Coeff.**	**R**
Site 1	0.9 ± 0.1	1.6 ± 0.4	>0.99	1.2 ± 0.2	1.8 ± 0.2	>0.99	1.0 ± 0.1	2.0 ± 0.3	>0.99
Site 2	0.9 ± 0.2	1.8 ± 0.6	1	1.1 ± 0.2	2.0 ± 0.3	>0.98	1.0 ± 0.1	1.6 ± 0.2	>0.99
Site 3	2.7 ± 0.5	3.8 ± 0.8	>0.99	3.2 ± 0.3	3.4 ± 0.5	>0.98	2.2 ± 0.2	3.8 ± 0.8	1

Table 3.4. K_d Values Determined by Langmuir Equation for Sites 1-3 with Overall Fit.

	PA1		**PA30**		**PA31**	
	***Kd* (Langmuir)**	**R**	***Kd* (Langmuir)**	**R**	***Kd* (Langmuir)**	**R**
Site 1	1.0 ± 0.1	>0.98	1.4 ± 0.3	>0.98	1.3 ± 0.4	>0.97
Site 2	1.0 ± 0.2	>0.97	1.3 ± 0.3	>0.96	1.3 ± 0.3	>0.99
Site 3	3.5 ± 0.5	>0.92	6 ± 1	>0.95	2.9 ± 1.5	>0.95

There is no significant difference in DNA-binding affinities of **PA1**, **PA30**, and **PA31**, indicating that the addition of a potential H-bond donor and a positive charge for **PA30** and **PA31** *vs.* **PA1** does not explain enhanced antiviral activity. It is noteworthy that since we are sampling an ensemble of different binding conformations in multiple linear, cell-free DNA molecules, binding affinities are the sum of these binding events at a particular binding site. Furthermore, the negative supercoiling of the viral genome, and bound viral and host proteins *in vivo*, probably affect binding (see Conclusions). We are currently examining these matters experimentally. All three PAs bind to sites 1 and 2 (single-mismatch sites) with K_d values ranging from 0.9-1.2 nM, and it is noteworthy that they all tolerate a triple base-pair mismatch at Site 3 with K_d values ranging from 2.2-3.2 nM (**Table 3.2**). The high Hill coefficients at this site are explained by previously-reported polyamide binding cooperativity,[11] in which the binding of one polyamide can preorganize the minor groove for the binding of a second (and more) polyamides, for example by widening the minor groove. This cooperativity would be most expected in areas of high polyamide density on the DNA, which is what we found in the Long Control Region. We need to point out that we used a single-site binding model to analyze binding thermodynamics. If we had parameterized binding to allow multiple Hill coefficients to float simultaneously, we would more accurately have accounted for the

statistics of multi-site binding. The results would likely have been that some Hill *Kd* values and coefficients would have decreased, and others would have increased. However, we were not convinced that our data would hold up to the introduction of so many additional parameters (a Hill coefficient per binding site and orientation), and did not believe that the outcome would change significantly in a scientific sense with a more precise handling of the statistical question. We will revisit this with more complete data sets for multiple compounds to test whether our approximation, which is normal in polyamide literature,[1-11] is suitable or not.

Thus, large antiviral hairpin PAs tolerate reportedly unfavourable DNA interactions without a significant decrease in binding affinity, which agrees with our prior work on HPV16.[41,42] In fact, in that prior work we found for 14-ring **PA1** that numerous "mismatch" sites were bound with higher affinity than so-called canonical sites. Also, up to four mismatches were tolerated with 2 nM dissociation constants identical to numerous "perfect-match" sites. For a larger, 20-ring antiviral polyamide (**PA25**), four mismatches were even more common than with **PA1**, leading to a complete blanketing of the HPV16 Long Control Region when only sparse binding, at best, was predicted.[42]

Data for HPV18-polyamide binding to a large portion of the LCR are reported here (also see **Figure SI3.2**) allow us to expand our understanding (polyamide-HPV18 binding was previously reported for the E2-protein binding sites).[16] We hypothesize that a deviation from established binding rules occurs for our unusually large hairpin PAs because they form many H-bonds with DNA and significant internal π-stacking compared to the smaller PAs typically found in the literature.[51] Thus, a few unfavourable enthalpic interactions are likely insignificant. The entropy of binding is likely very favourable for long PAs from the loss of many water molecules previously resident in the minor groove and from loss of exposed nonpolar surface area upon π-stacking. Entropy and enthalpy are being determined by the CM Dupureur group. Since we measured polyamide-DNA binding properties in the context of a large, natural DNA sequences instead of a small artificial, DNA sequences designed to examine sequence preferences at a single location, our results are much more complex than those found in the literature. Published binding rules are not obeyed in our system with significant fidelity, and "mismatches" are very well tolerated (see also He et al.[41] and Qiao et al.[52]).

Binding sites for **PA1**, **PA30**, and **PA31** in a region of HPV18 DNA spanning nt 7601-7620 are shown in **Figure 3.4A**. In this part of the viral genome, there are two predicted single-mismatch binding sites, sequence (i) 5'-TATTAGTCAT-3' (7606-7615) and sequence (ii) 5'-TAGTCATTTT-3' (7609-7618). Our data indicate that **PA1** and **PA31** bind both of these sites, while **PA30** only binds to (i).

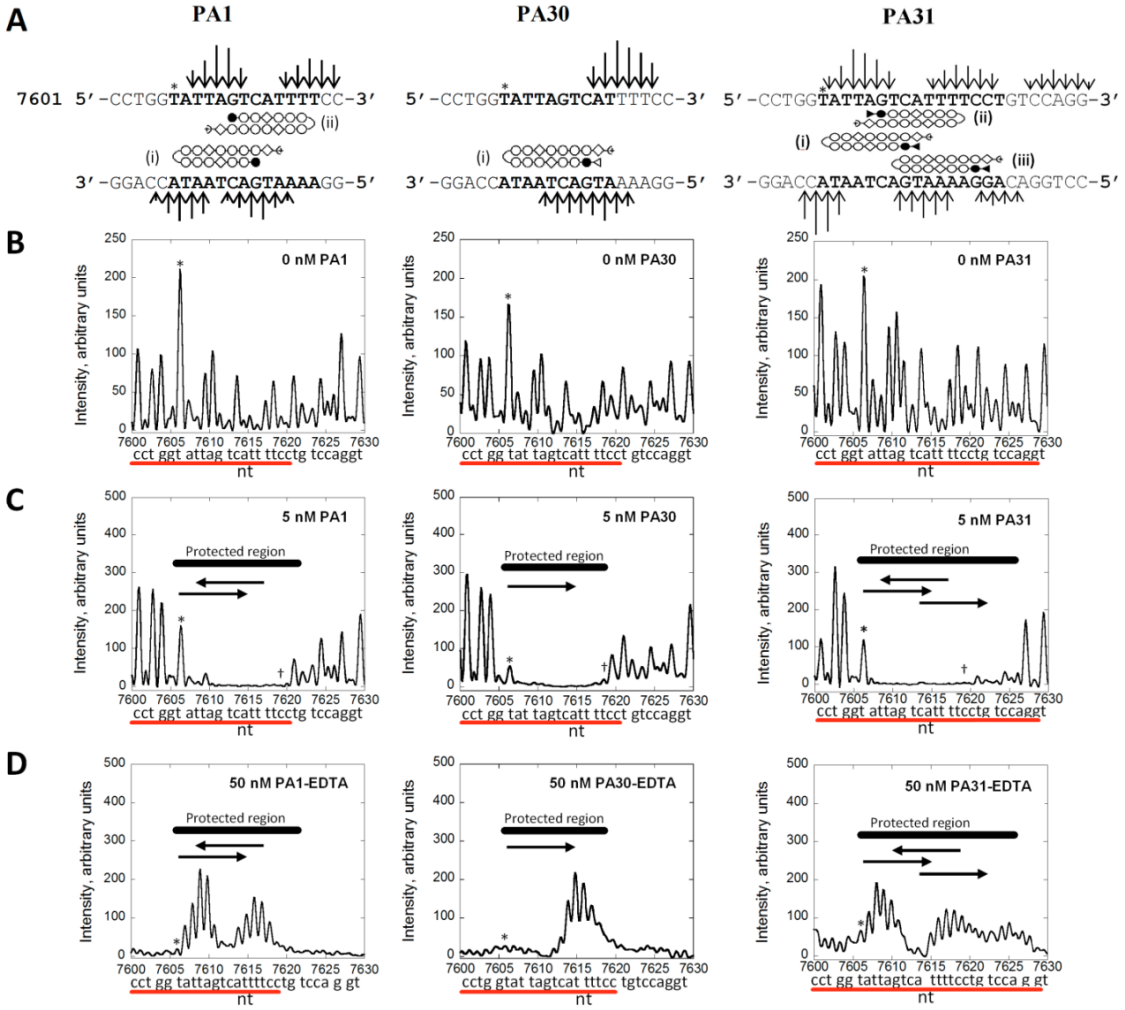


Figure 3.4. PA1, PA30 and PA31 Binding Sites on HPV18 (7600-7620 bp). (A) Binding sites for **PA1**, **PA30**, and **PA31** from a section of HPV18 LCR corresponding to 7600-7620 as determined by capillary electrophoresis, DNase I footprinting and AC. The relative heights above the DNA sequence correspond to observed AC intensities. The remaining panels show the top strand of ten more base pairs than panel **A**, 7600–7630. Representative electropherograms showing the top strand for **(B)** DNase I footprinting controls (0 nM PA), **(C)** DNase I-protected regions upon the addition of 5 nM of PA and **(D)** AC patterns at 50 nM PA. The red line below each plot corresponds to the DNA sequence shown in panel **(A)**. The * and † symbols mark nt 7606 and 7619. Arrows inside the plots show the PA binding direction, with the arrowhead in the position of the PA tail. PA building blocks as defined in **Figure 3.1**; TMG and Guan groups are open and filled triangles.

Raw DNase I footprinting data of dsDNA (top strand) is shown in **Figures 3.4B** (control) and **3.4C** (5 nM PA). Inspection of the footprints generated by the three PAs reveals subtle differences in the protected regions. For instance, the peaks at the left-hand edge of the protected region corresponding to nt 7606-09 (**Figure 3.4*** shows 7606) display increased nuclease accessibility in the case of **PA1** *vs.* **PA30** and **PA31**. At the right-hand edge of the protected region, the peak corresponding to nt 7619 (**Figure 3.4C†**) is less protected from DNase I by **PA30** (**Figure 3.4C middle panel**) than by **PA1** (**Figure 3.4C left panel**) at equivalent PA concentrations. In contrast, **PA31** affords more-extended nuclease protection than either **PA1** or **PA30**, with a footprint including extra nt 7619-7626; this extra nuclease protection corresponds to the double-mismatch

site (iii) at 7612-7621. For steric reasons, the TMG group of **PA30** may not be bound to DNA, making it less sequence-specific than **PA1** and **PA31**.

Footprinting differences are supported by affinity cleavage (AC) results.[53] Patterns generated by **PA1-**, **PA30-** and **PA31-EDTA** conjugates are given in **Figure 3.4D**. For **PA1** and **PA31** there are two binding events, each corresponding to a distinct AC pattern, agreeing with DNase I data that both **PAs** bind at both available single-mismatch sites, (i) and (ii). However, an extra AC pattern was observed for **PA31** corresponding to the double mismatch site (iii). Conversely, only one of the predicted single-mismatch sites is occupied by **PA30**: a single AC pattern corresponding to site (i) was observed. The AC pattern for site (ii) did not appear even upon increasing the concentration of **PA30** from 50 nM to 200 nM (data not shown). Our results with these 14-ring PAs agree with previous 8-ring PAs studies reporting that substitution at the N-terminal imidazole can modulate orientation and DNA-binding preferences of hairpin polyamides.[54,55]

3.5 CONCLUSIONS

In summary, we show that N-terminal substitution of anti-HPV polyamide **PA1** improves antiviral activity against the two high-risk HPV types that are most prevalent (HPV16 and HPV18). N-terminal substitution with TMG gives **PA30** and leads to improvement in anti-HPV activity (IC_{50}) against HPV18. N-terminal substitution with guanidinium itself gives **PA31** with enhanced anti-HPV activity even further relative to **PA1** by decreasing IC_{50} against HPV18 and IC_{90} for both HPV16 and HPV18. For antiviral agents in general, a dosage above IC_{90} decreases viral load, magnitude, and frequency of viral rebound,[56] and the probability of viral mutation, so improvements in IC_{90} are significant for therapy, as they are for EC_{90}.[57]

Although the dissociation constants for the reported PAs are similar and cannot account for the differences in antiviral activity, differences in PA-DNA binding distribution were observed: **PA31** binds more sites on the LCR of HPV18 than **PA1** or **PA30**. Binding kinetics may be more important for activity[58,59] and remain under investigation. Polyamide uptake is also under investigation. We note that **PA1** and its analogues **PA30** and **PA31** are among the smallest molecules in a library of 75+ active compounds that reach nearly twice the molecular weight of **PA1**. We have also prepared a series of compounds based on the highly-active **PA25** and other active series, and they will be described subsequently. **PA25** was our second preclinical lead because of its IC_{50}/IC_{90} data (μM) as follows: HPV16, 0.036/0.35; HPV18, 0.056/1.5; HPV31, 0.030/0.51.[27,47] The first biophysical studies of **PA25** were reported recently.[42] However, **PA25** is a more difficult synthetic and purification challenge than **PA1** and its derivatives, though we have scaled up both **PA1** and **PA25** to 10 g, and the cost-of-goods is much higher for the 20-ring **PA25** than the 14-ring **PA1**, so we were interested to see if modifications to **PA1** could increase its activity significantly. That goal was successful with **PA31**.

Long hairpin polyamides of this general class function by a novel polyamide mechanism that invokes the host DNA Damage Response to destroy the viral genome.[23,24,30] Their activity is not limited to HPV but is also found for other small, double-stranded DNA tumour viruses, including polyomaviruses.[47] This mechanism is unrelated to any prior polyamide literature that we can determine, even though our work was initially inspired by early PA research. **PA30/31** represent an improved class of long

polyamide that invokes this new mechanism to combat cancer-causing HPV. More recently, Dervan has reported a similar, non-sequence-specific, polyamide-induced DNA Damage Response.[60]

Why might this nonspecific binding, or poorly-specific binding, be so relevant? We have come to hypothesize and are performing experiments to test the idea, that the interaction of large polyamides with small, circular, double-stranded DNA genomes is biologically important more from physical than chemical reasons. Hydrogen bonding does go on in the minor groove, albeit at moderate selectivity, but of great importance is the binding of many hundreds or more of rigid, crescent-shaped molecules to supercoiled DNA targets. This must result in considerable strain on the supercoiled viral genomes, and could in the most extreme case result in single-stranded DNA regions appearing, which would be rapidly noticed and acted upon by the DNA Damage Response machinery. However, the process may be simpler: the dsDNA viruses evolved to use parts of the DDR for their life cycles but otherwise to hide, to avoid having their genomes recognized as foreign or damaged in any way. When the supercoiled genomes are distorted by the effect of many bound polyamides, and 50 % of HPV DNA is bound by active polyamides, at least for the linearized pieces we have reported on to date,[29,41,42] the viral genome may no longer remain hidden. The strain imposed by all of the bound polyamides may change the shape enough so that the DDR naturally notices unwanted additions to the chromatin and is automatically activated. Support for this hypothesis is found in our recently published patent application, where, as mentioned above, we showed that the same polyamides active against HPV are also active against three polyomaviruses having little in common with HPV except a small, dsDNA genome that is supercoiled.[47] The DNA Damage Response is also altered for these disparate viruses.

3.6 NOTES AND ACKNOWLEDGEMENTS

Declaration: JKB and CF hold significant equity positions in NanoVir.

We thank Jacqui Niederschulte for extinction coefficient determination of **PA1**, the DNA Core at the University of Missouri for DNA CE fragment analysis, the Danforth Plant Science Center (NSF DBI 0922879], Prof. B. Bythell for HRMS and Prof. P. Lambert for providing W12E cells. The Agilent 600 MHz NMR spectrometer was obtained using funds from NSF (#0959360). This work was supported by NIH-NIAID: AI083803, AI062182, and AI068159.

3.7 BIBLIOGRAPHY

(1) Dervan, P. B.; Edelson, B. S. *Current Opinion in Structural Biology* **2003**, *13*, 284.
(2) Chen, X. R., Boopathy; Rao, Sambhorao T.; Sundaralingam, Muttaiya *Nat Struct Mol Biol* **1994**, *1*, 7.
(3) Marky, L. A.; Breslauer, K. J. *Proc Natl Acad Sci U S A* **1987**, *84*, 4359.
(4) Finlay, A. C.; Hochstein, F. A.; Sobin, B. A.; Murphy, F. X. *Journal of the American Chemical Society* **1951**, *73*, 341.
(5) White, S.; Baird, E. E.; Dervan, P. B. *Biochemistry* **1996**, *35*, 12532.

(6) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468.
(7) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *Journal of the American Chemical Society* **1996**, *118*, 6160.
(8) Wang, C. C. C. E., U.; Dervan, P. B. *Bioorg Med Chem* **2001**, *9*, 5.
(9) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proceedings of the National Academy of Sciences* **1992**, *89*, 7586.
(10) Mrksich, M.; Parks, M. E.; Dervan, P. B. *Journal of the American Chemical Society* **1994**, *116*, 7983.
(11) Hawkins, C. A.; Baird, E. E.; Dervan, P. B.; Wemmer, D. E. *Journal of the American Chemical Society* **2002**, *124*, 12689.
(12) Mapp, A. K.; Ansari, A. Z.; Ptashne, M.; Dervan, P. B. *Proceedings of the National Academy of Sciences* **2000**, *97*, 3930.
(13) Wang, X.; Nagase, H.; Watanabe, T.; Nobusue, H.; Suzuki, T.; Asami, Y.; Shinojima, Y.; Kawashima, H.; Takagi, K.; Mishra, R.; Igarashi, J.; Kimura, M.; Takayama, T.; Fukuda, N.; Sugiyama, H. *Cancer Sci* **2010**, *101*, 759.
(14) Raskatov, J. A.; Meier, J. L.; Puckett, J. W.; Yang, F.; Ramakrishnan, P.; Dervan, P. B. *Proceedings of the National Academy of Sciences* **2012**, *109*, 1023.
(15) Maeshima, K.; Janssen, S.; Laemmli, U. K. *The EMBO Journal* **2001**, *20*, 3218.
(16) Schaal, T. D.; Mallet, W. G.; McMinn, D. L.; Nguyen, N. V.; Sopko, M. M.; John, S.; Parekh, B. S. *Nucleic Acids Res* **2003**, *31*, 1282.
(17) Yasuda, A.; Noguchi, K.; Minoshima, M.; Kashiwazaki, G.; Kanda, T.; Katayama, K.; Mitsuhashi, J.; Bando, T.; Sugiyama, H.; Sugimoto, Y. *Cancer Sci* **2011**, *102*, 2221.
(18) Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan, P. B. *Proc Natl Acad Sci U S A* **1998**, *95*, 12890.
(19) Bürli, R. W.; McMinn, D.; Kaizerman, J. A.; Hu, W.; Ge, Y.; Pack, Q.; Jiang, V.; Gross, M.; Garcia, M.; Tanaka, R.; Moser, H. E. *Bioorganic & Medicinal Chemistry Letters* **2004**, *14*, 1253.
(20) Bürli, R. W.; Jones, P.; McMinn, D.; Le, Q.; Duan, J.-X.; Kaizerman, J. A.; Difuntorum, S.; Moser, H. E. *Bioorganic & Medicinal Chemistry Letters* **2004**, *14*, 1259.
(21) Barrett, M. P.; Gemmell, C. G.; Suckling, C. J. *Pharmacology & Therapeutics* **2013**, *139*, 12.
(22) Soeiro, M. N. C.; Werbovetz, K.; Boykin, D. W.; Wilson, W. D.; Wang, M. Z.; Hemphill, A. *Parasitology* **2013**, *140*, 10.1017/S0031182013000292.
(23) Edwards, T. G.; Vidmar, T. J.; Koeller, K.; Bashkin, J. K.; Fisher, C. *PLoS One* **2013**, *8*, e75406.
(24) Fisher, C. *Journal of Clinical Medicine* **2015**, *4*, 204.
(25) Lamarche, B. J.; Orazio, N. I.; Weitzman, M. D. *FEBS Letters* **2010**, *584*, 3682.
(26) Lin, J. J.; Dutta, A. *Journal of Biological Chemistry* **2007**, *282*, 30357.
(27) Edwards, T. G.; Koeller, K. J.; Slomczynska, U.; Fok, K.; Helmus, M.; Bashkin, J. K.; Fisher, C. *Antiviral Res* **2011**, *91*, 177.
(28) He, G.; Bashkin, J. K. *Future Medicinal Chemistry* **2015**, *7*, 1953.

(29) Koeller, K. J.; Harris, G. D.; Aston, K.; He, G.; Castaneda, C. H.; Thornton, M. A.; Edwards, T. G.; Wang, S.; Nanjunda, R.; Wilson, W. D.; Fisher, C.; Bashkin, J. K. *Medicinal Chemistry* **2014**, *4*, 338.
(30) Edwards, T. G.; Helmus, M. J.; Koeller, K.; Bashkin, J. K.; Fisher, C. *J Virol* **2013**, *87*, 3979.
(31) Schwartz, S. *Virology* **2013**, *445*, 187.
(32) Walboomers, J. M. M.; Jacobs, M. V.; Manos, M. M.; Bosch, F. X.; Kummer, J. A.; Shah, K. V.; Snijders, P. J. F.; Peto, J.; Meijer, C. J. L. M.; Muñoz, N. *J Pathol* **1999**, *189*, 12.
(33) Clifford, G.; Franceschi, S.; Diaz, M.; Munoz, N.; Villa, L. L. *Vaccine* **2006**, *24 Suppl 3*, S3/26.
(34) Dochez, C.; Bogers, J. J.; Verhelst, R.; Rees, H. *Vaccine* **2014**, *32*, 1595.
(35) Joura, E. A.; Giuliano, A. R.; Iversen, O.-E.; Bouchard, C.; Mao, C.; Mehlsen, J.; Moreira, E. D.; Ngan, Y.; Petersen, L. K.; Lazcano-Ponce, E.; Pitisuttithum, P.; Restrepo, J. A.; Stuart, G.; Woelber, L.; Yang, Y. C.; Cuzick, J.; Garland, S. M.; Huh, W.; Kjaer, S. K.; Bautista, O. M.; Chan, I. S. F.; Chen, J.; Gesser, R.; Moeller, E.; Ritter, M.; Vuocolo, S.; Luxembourg, A. *New England Journal of Medicine* **2015**, *372*, 711.
(36) Nagle, P. S.; McKeever, C.; Rodriguez, F.; Nguyen, B.; Wilson, W. D.; Rozas, I. *J Med Chem* **2014**, *57*, 7663.
(37) Arafa, R. K.; Ismail, M. A.; Munde, M.; Wilson, W. D.; Wenzler, T.; Brun, R.; Boykin, D. W. *European journal of medicinal chemistry* **2008**, *43*, 10.1016/j.ejmech.2008.02.008.
(38) Rodríguez, F.; Rozas, I.; Kaiser, M.; Brun, R.; Nguyen, B.; Wilson, W. D.; García, R. N.; Dardonville, C. *J Med Chem* **2008**, *51*, 909.
(39) Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1996**, *118*, 6141.
(40) Puckett, J. W.; Green, J. T.; Dervan, P. B. *Organic Letters* **2012**, *14*, 2774.
(41) He, G.; Vasilieva, E.; Harris Jr, G. D.; Koeller, K. J.; Bashkin, J. K.; Dupureur, C. M. *Biochimie* **2014**, *102*, 83.
(42) Vasilieva, E.; Niederschulte, J.; Song, Y.; Harris Jr, G. D.; Koeller, K. J.; Liao, P.; Bashkin, J. K.; Dupureur, C. M. *Biochimie* **2016**, *127*, 103.
(43) del Fresno, M.; El-Faham, A.; Carpino, L. A.; Royo, M.; Albericio, F. *Organic Letters* **2000**, *2*, 3539.
(44) Robinson, S.; Roskamp, E. J. *Tetrahedron* **1997**, *53*, 6697.
(45) Dupureur, C. M.; Bashkin, J. K.; Aston, K.; Koeller, K. J.; Gaston, K. R.; He, G. *Anal Biochem* **2012**, *423*, 178.
(46) He, G.; Vasilieva, E.; Bashkin, J. K.; Dupureur, C. M. *Anal Biochem* **2013**, *439*, 99.
(47) Bashkin, J. K.; Edwards, T. G.; Fisher, C.; Harris, G. D.; Koeller, K. J. *US Patent Office* **November 19, 2015**, *Patent number: 20150329596*, Appln. Number: 14/818881.
(48) Galas, D. J.; Schmitz, A. *Nucleic Acids Res* **1978**, *5*, 3157.
(49) Yindeeyoungyeon, W.; Schell, M. A. *BioTechniques* **2000**, *29*, 1034.
(50) Stünkel, W.; Bernard, H.-U. *J Virol* **1999**, *73*, 1918.

(51) Turner, J. M.; Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1997**, *119*, 7636.
(52) Qiao, H.; Ma, C.; Zhang, X.; Jing, X.; Li, C.; Zhao, Y. *Bioconjugate Chemistry* **2015**, *26*, 2054.
(53) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1996**, *118*, 8198.
(54) White, S.; Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1997**, *119*, 8756.
(55) Hawkins, C. A.; de Clairac, R. P.; Dominey, R. N.; Baird, E. E.; White, S.; Dervan, P. B.; Wemmer, D. E. *Journal of the American Chemical Society* **2000**, *122*, 5235.
(56) Pawlotsky, J.-M. *Hepatology* **2011**, *53*, 1742.
(57) Duffy, S.; Shackelton, L. A.; Holmes, E. C. *Nat Rev Genet* **2008**, *9*, 267.
(58) Goyal, M.; Rizzo, M.; Schumacher, F.; Wong, C. F. *J Med Chem* **2009**, *52*, 5582.
(59) Copeland, R. A.; Pompliano, D. L.; Meek, T. D. *Nat Rev Drug Discov* **2006**, *5*, 730.
(60) Martínez, T. F.; Phillips, J. W.; Karanja, K. K.; Polaczek, P.; Wang, C.-M.; Li, B. C.; Campbell, J. L.; Dervan, P. B. *Nucleic Acids Res* **2014**, *42*, 11546.

3.8 SUPPLEMENTAL INFORMATION

3.8.1 Dose-Response Curves for PA1 against HPV16 and HPV31

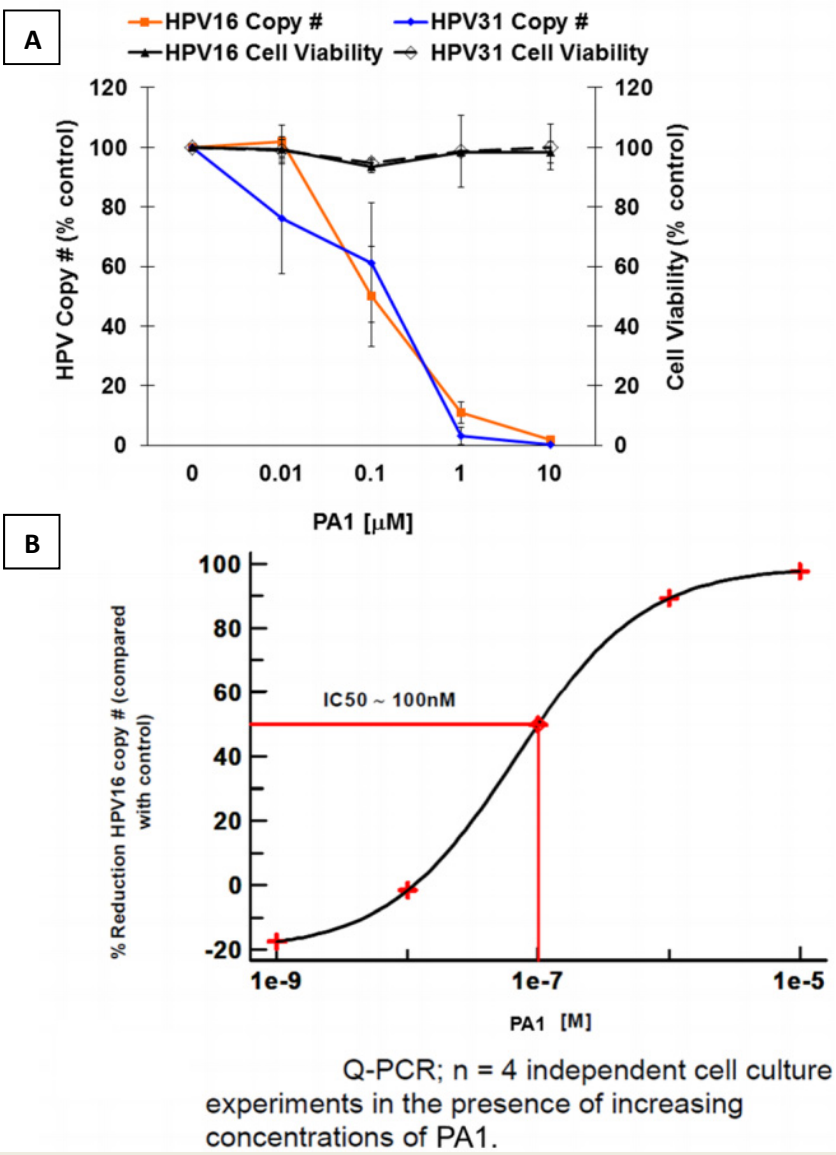


Figure SI3.1. Dose-Response Curves for PA1 against HPV16 and 31 from which IC_{50} and IC_{90} Values were Determined. (A) Results for HPV16 and 31 with cell toxicity data. **(B)** Expanded detail of results for HPV16, plotted as the inverse of the other results (as the percent decrease of viral DNA *vs.* control, reaching 100 % decrease at maximum efficacy, instead of the simple decrease in DNA *vs.* control).

3.8.2 Sequence Map of PA30 Affinity Cleavage and DNase I Footprinting in the 7479-7783 bp LCR Region of HPV18

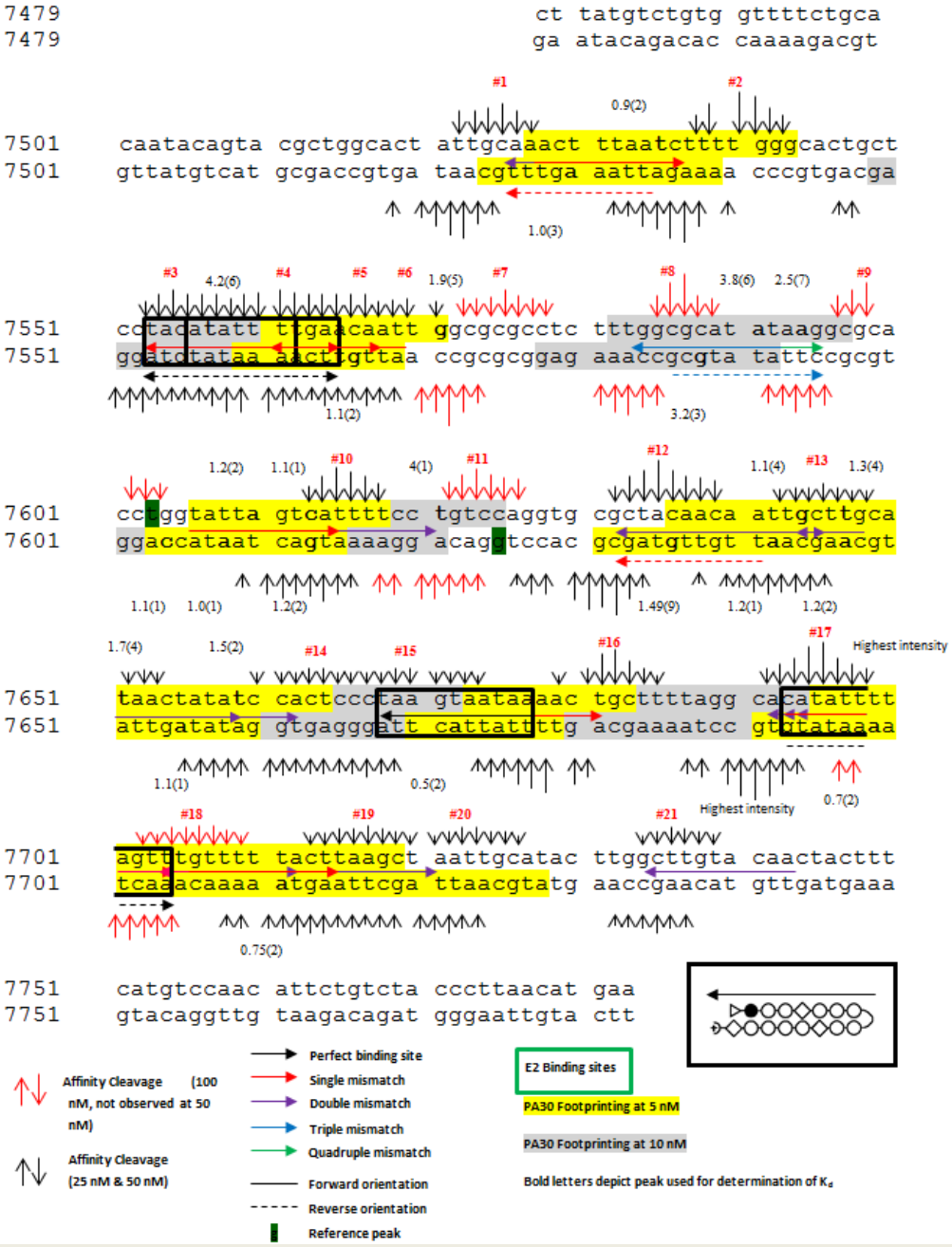


Figure SI3.2. PA30 Binding Sites and Hill-Calculated Equilibrium Dissociation Constants at 200 pM of the 305 bp (7479-7783) HPV18 DNA Fragment. The DNA duplex for the 305 bp HPV18 DNA fragment is shown along with **PA30** hairpin polyamide represented as arrows at its respective binding sites (see *inset)*. **PA30** is depicted with either a solid arrow (forward orientation) or a dashed arrow (reverse orientation). HPV E2 binding sites are depicted in green boxes. HPV E1 binds in the region flanked by the first and second E2 binding sites. The origin of replication is marked with 1. Numbers on the left correspond to the nucleotide position in HPV18 genome. Equilibrium dissociation constants calculated using the Hill equation from the FAM (blue) and HEX (green) channel of CE at the particular nucleotide are given. The sample standard deviations are in parentheses. **PA30** binding orientations were determined by affinity cleavage. Affinity cleavage patterns observed at 50 nM and 100 nM PA (not observed at 50 nM) are shown as black and red vertical arrows (n = 3). Yellow highlighted sequences represent footprinting regions observed at 5 nM **PA30**; whereas, gray highlighted sequences represent footprinting regions observed at 10 nM **PA30** (n = 5).

3.8.3 PA30 Equilibrium Dissociation Constants on HPV18 LCR 7479-7783 bp

Table SI3.1 summarizes only the higher affinity sites (**PA30** DNase I footprinting at 5 nM PA / Affinity cleavage patterns at 50 nM PA) in the HPV18 LCR region corresponding to 7479-7783 bp. DNase I footprinting experiments conducted at 200 pM DNA and affinity cleavage experiments at 1 nM DNA.

Table SI3.1. PA30 Binding Sites on HPV18 LCR Fragment Corresponding to 7479-7783 bp.

AC site #	Sequence	Position	Site type	Binding orientation	Integration nt	*Kd* (nM) DNase I
1	TTGCA AA**C**TTTAATC TTTTG	7527-7536	Double	Forward	7535	0.9 ± 0.2
			Single	Forward		
2	CCCAA AA**G**ATTAAAG TTTGC	7538-7529	Single	Forward		
3	GCTCC TA**C**ATATTTT GAACA	7553-7562	Single	Forward	7557	4.2 ± 0.6
			Perfect	Reverse		
4	ATATT TT**G**AACAATT GGCGC	7561-7570	Double	Forward	7566	1.1 ± 0.2
5	AATTG TT**C**AAAATAT GTAGG	7565-7556	Single	Forward		
			Perfect	Reverse		
6	GCCAA TT**G**TTCAAAA TATGT	7568-7559	Single	Forward		
10	GGAAA AT**G**ACTAATA CCAGG	7615-7606	Single	Forward	7613	1.1 ± 0.1
12	TGCGC TA**C**AACAATT GCTTG	7634-7643	Double	Forward	7642	1.2 ± 0.1
			Perfect	Reverse		
13	TATGC AA**G**CAATTGT TGTAG	7647-7638	Double	Forward		
13	ATTGC TT**G**CATAACT ATATC	7646-7655	Double	Forward	7655	1.1 ± 0.1
14	AGTGG AT**A**TAGTTAT GCAAG	7659-7650	Double	Forward		
14	GGGAG TG**G**ATATAGT TATGC	7822-7813	Double	Forward		
15	TCCCT AA**G**TAATAAA ACTGC	7669-7678	Perfect	Forward	7672	0.5 ± 0.2
16	AAAAG CA**G**TTTTATT ACTTA	7682-7673	Single	Forward		
17[a]	GGCAC AT**A**TTTTAGT TTGTT	7694-7703	Double	Forward	7699	0.7 ± 0.2
17[a]	GCACA TA**T**TTTAGTT TGTTT	7695-7704	Double	Forward		
17[a]	CACAT AT**T**TTAGTTT GTTTT	7696-7705	Double	Forward		
19	TTAAG TA**A**AAACAAA CTAAA	7712-7703	Double	Forward	7711	0.8 ± 0.1
19	AGCTT AA**G**TAAAAAC AAACT	7715-7706	Single	Forward		
20	GCAAT TA**G**CTTAAGT AAAAA	7721-7712	Double	Forward		
21	TTGGC TT**G**TACAACT ACTTT	7736-7745	Double	Forward	*ND*	*ND*

The affinity cleavage (AC) site # corresponds to the AC patterns in the map provided in **Figure SI3.2**. The **PA30** binding sequences are given in a 5' to 3' direction and include 5 nucleotide (nt) flanking regions. The bold nt highlights the position where the polyamide imidazole/pyrrole pair binds. The position corresponds to the HPV18 genomic nucleotides. The forward binding orientation refers to alignment of the polyamide N-terminus to C-terminus vector with the 5'-3' direction of the DNA, whereas a reverse orientation describes the polyamide N→C vector alignment with the 3'→5' direction of the DNA.

ND, Not determined. Site 21 is found at the end of the fragment.

[a] A single binding event cannot be assigned due to significant overlap of the predicted **PA30** recognition sequences.

3.8.4 PA30 Predicted, But Not Observed Single-base-pair Mismatch Sites on HPV18 LCR 7479-7783 bp

Table SI3.2. PA30 Predicted Single-Base-Pair Mismatch Sites Which Were Not Observed on HPV18 LCR Fragment Corresponding to 7479-7783 bp.

Sequence	Position	Binding orientation	Comment
GGTAT TA**G**TCATTTT CCTGT	7609-7618	Forward	No AC
TTAGG CA**C**ATATTTT AGTTT	7691-770	Reverse	No AC
AAACA AA**C**TAAAATA TGTGC	7704-7695	Forward	No AC[a]
TATTT TA**G**TTTGTTT TTACT	7700-7709	Forward	No AC
TTAGT TT**G**TTTTTAC TTAAG	7704-7713	Forward	No AC[a]

PA30 was predicted to bind these sequences with single-base-pair mismatches; however, these sequences did not generate affinity cleavage (AC) patterns and/or footprints (FP).

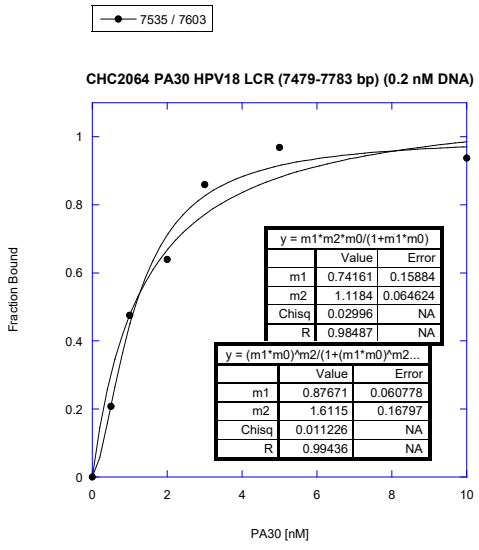
[a] The affinity cleavage pattern for Site 18, a predicted **PA30** reverse match (perfect) sequence and a single-base-pair mismatch, was only observed at 100 nM of PA-EDTA conjugate.

The differences that we observe between **PA1** and **PA30** may be due to a steric clash between TMG and the Ta tail, leading to a 180° rotation of the last C-terminal pyrrole ring (where the β-ala / tail is attached) and exclusion of the tail from the minor groove. Consequently, a mixed population of conformers with the normal pyrrole ring conformation and the rotated form may be present during the **PA30** experiments. In the case of rotated C-terminal pyrrole conformers, the favorable contacts between the minor groove and the Ta tail are not present, leading to both a decrease in affinity (may explain the reduced AC pattern reduction/disappearance at some sites) and an increased preference for reversed binding orientations.[55] Further analysis of the data is required to validate this hypothesis.

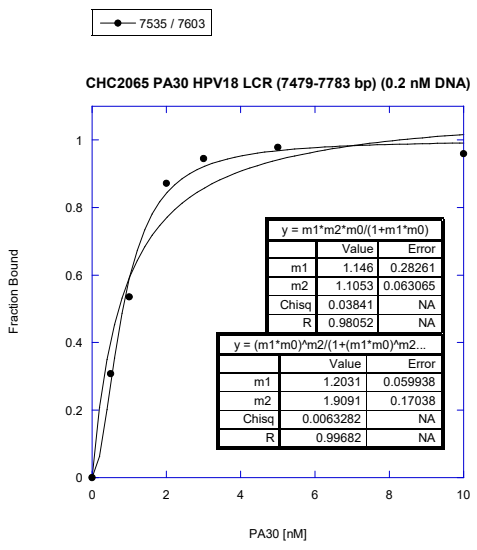
3.8.5 Representative PA30 Isotherms Obtained from Quantitative DNase I Footprinting Sites Along the DNA Fragment Corresponding to 7479-7783 bp of the HPV18 LCR

SITE #1 & 2 (Integration nt: 7535; Reference nt: 7603 in FAM Channel)

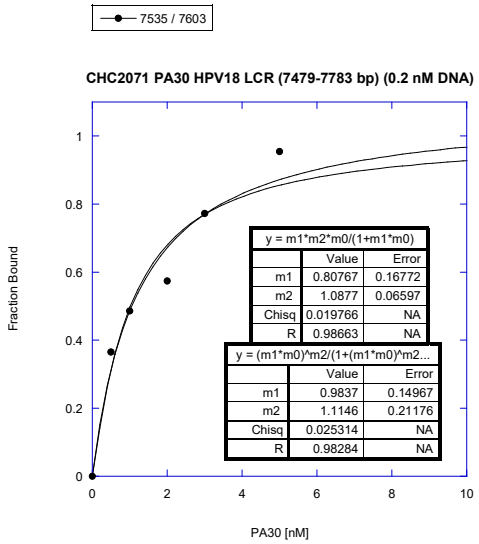
Trial #1



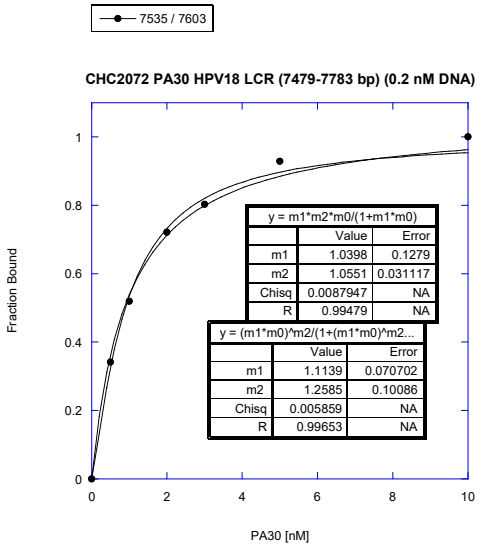
Trial #2



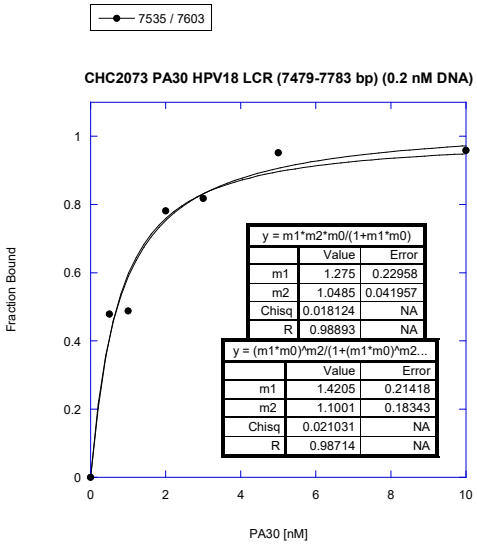
Trial #3



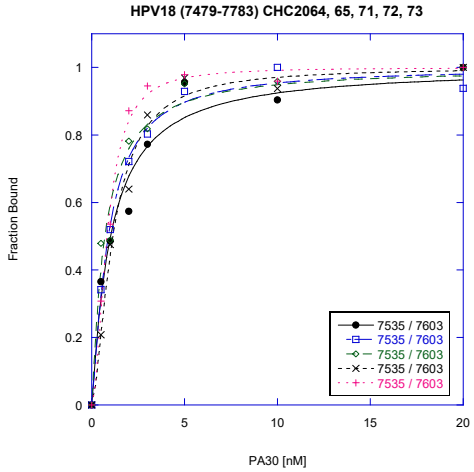
Trial #4



Trial #5

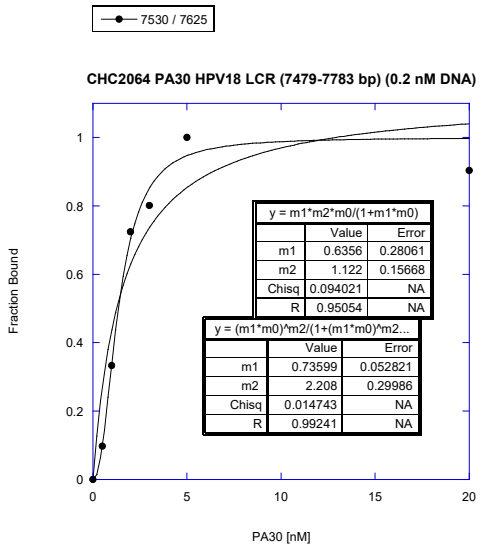


SITE #1 & 2 FAM: All Trials (Hill fit)

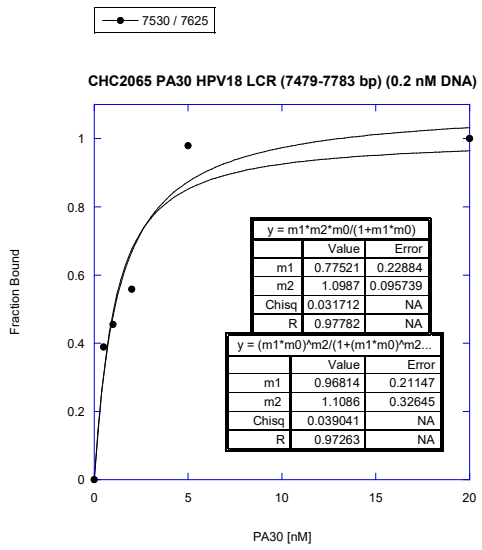


SITE #1 & 2 (Integration nt: 7530; Reference nt: 7625 in HEX Channel)

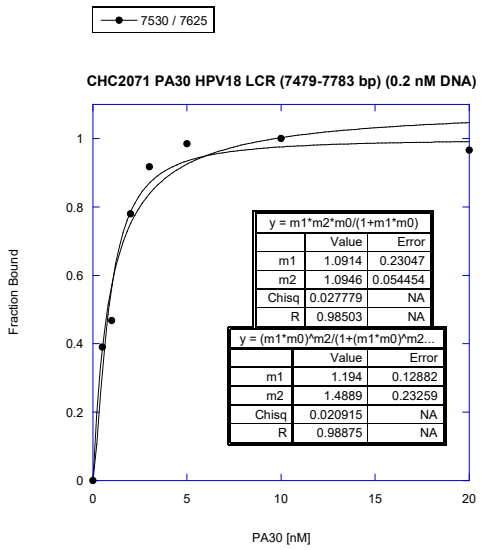
Trial #1



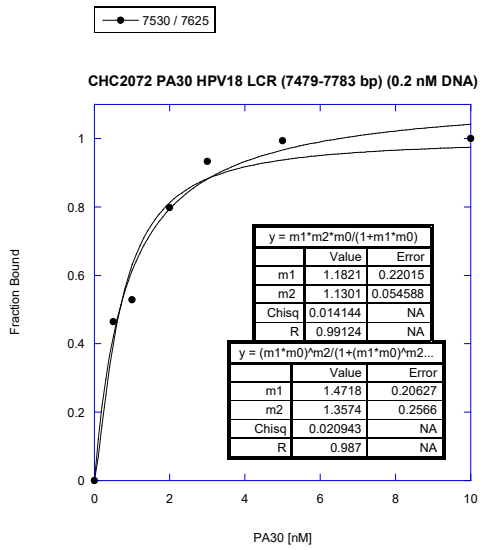
Trial #2



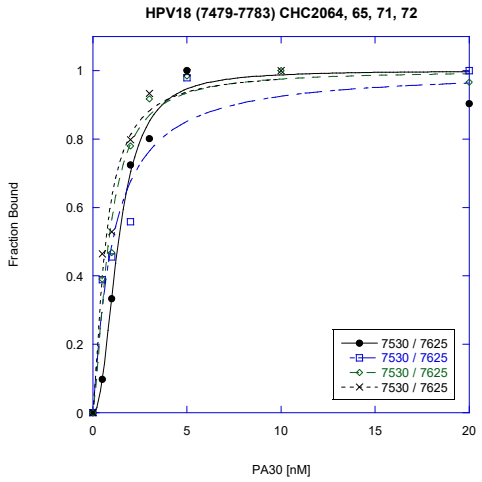
Trial #3



Trial #4

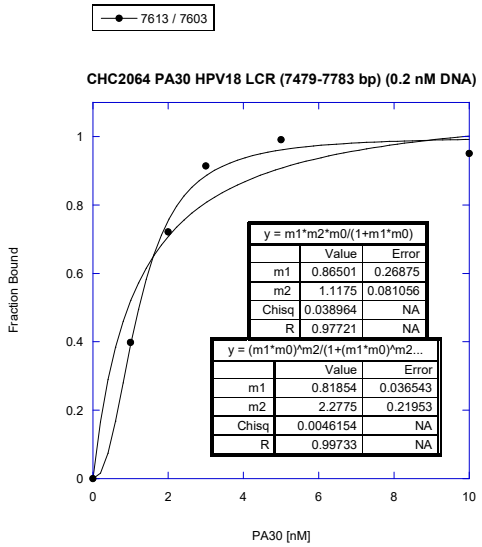


SITE #1 & 2 HEX: All Trials (Hill fit)



SITE #10 (Integration nt: 7613; Reference nt: 7603 in FAM Channel)

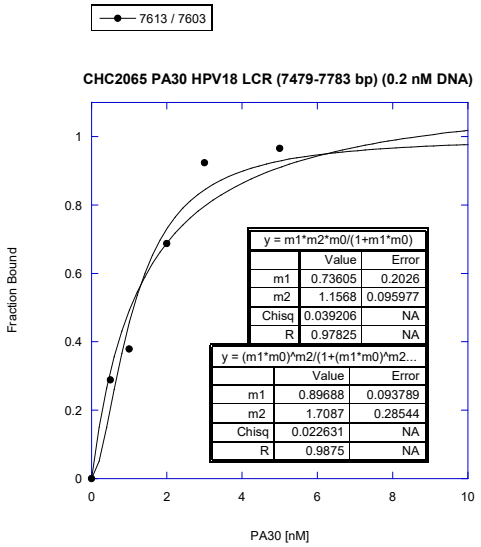
Trial #1



y = m1*m2*m0/(1+m1*m0)		
	Value	Error
m1	0.86501	0.26875
m2	1.1175	0.081056
Chisq	0.038964	NA
R	0.97721	NA

y = (m1*m0)^m2/(1+(m1*m0)^m2...		
	Value	Error
m1	0.81854	0.036543
m2	2.2775	0.21953
Chisq	0.0046154	NA
R	0.99733	NA

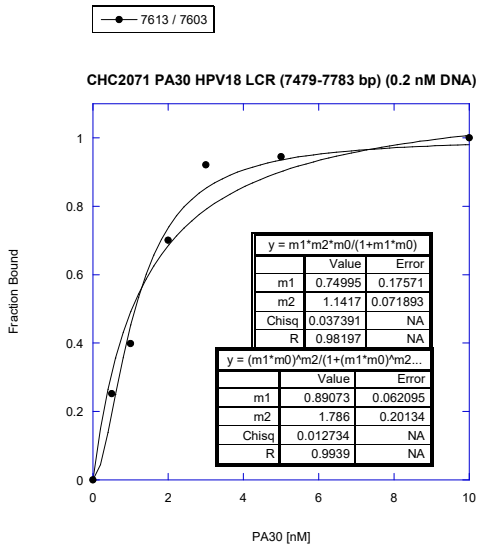
Trial #2



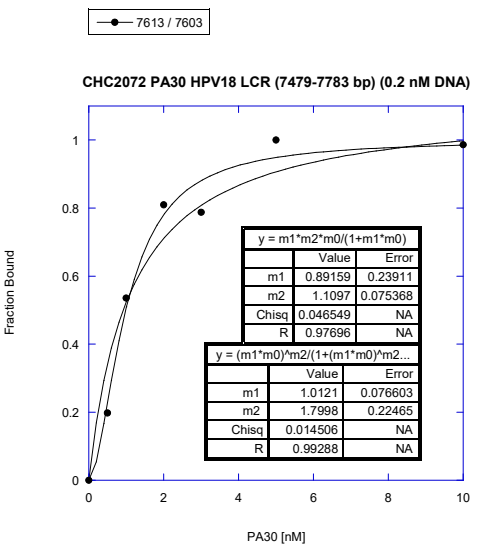
y = m1*m2*m0/(1+m1*m0)		
	Value	Error
m1	0.73605	0.2026
m2	1.1568	0.095977
Chisq	0.039206	NA
R	0.97825	NA

y = (m1*m0)^m2/(1+(m1*m0)^m2...		
	Value	Error
m1	0.89688	0.093789
m2	1.7087	0.28544
Chisq	0.022631	NA
R	0.9875	NA

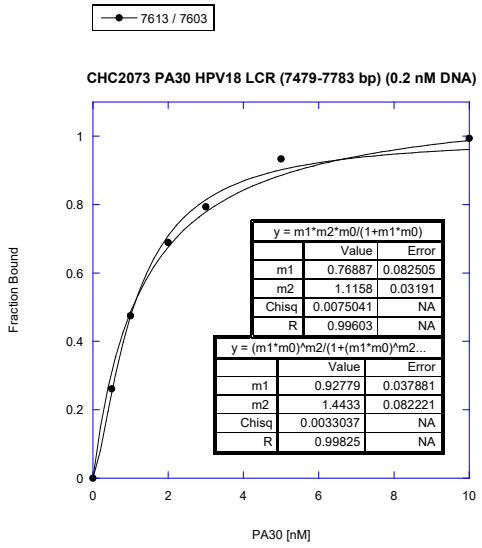
Trial #3



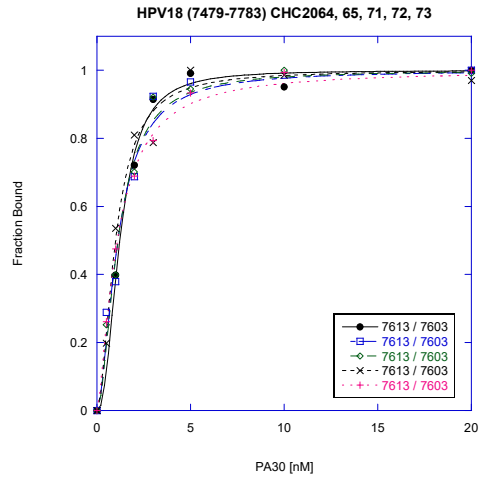
Trial #4



Trial #5

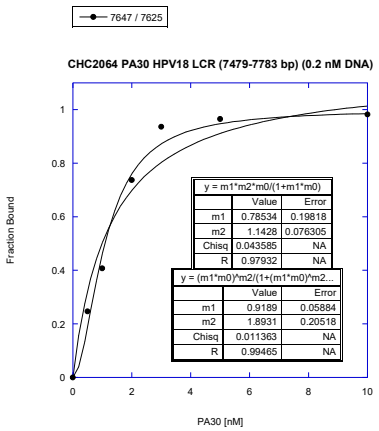


SITE #10 FAM: All Trials (Hill fit)

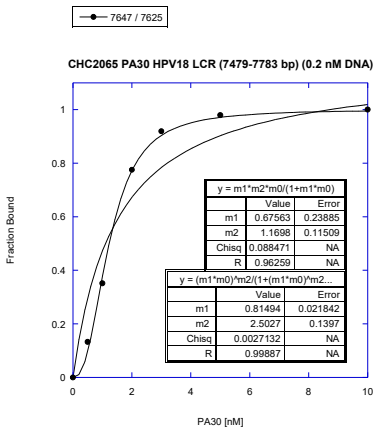


SITE #12 & 13 (Integration nt: 7647; Reference nt: 7625 in HEX Channel)

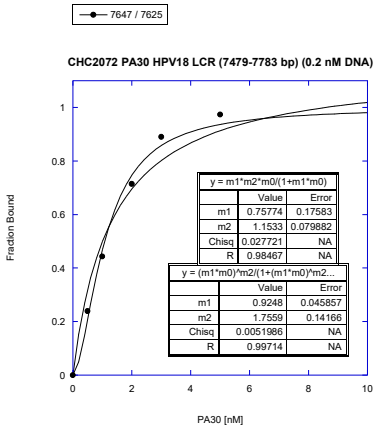
Trial #1



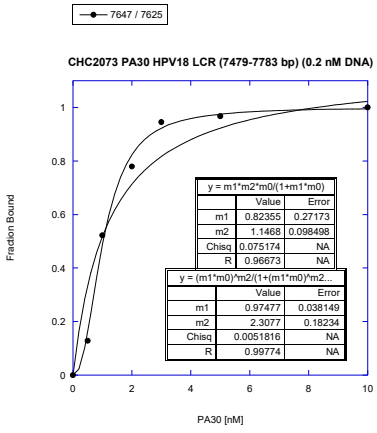
Trial #2



Trial #3



Trial #4



SITE #12 & 13 HEX: All Trials (Hill fit)

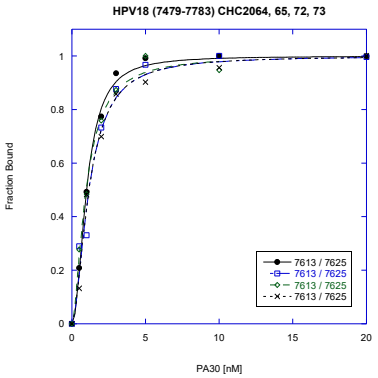


Figure SI3.3. Representative PA30 Binding Isotherms Obtained from Quantitative DNase I Footprinting Sites Along the DNA Fragment Corresponding to 7479-7783 bp of the HPV18 LCR. Binding isotherms for each trial of **PA30** are plotted. Calculated parameters for the Langmuir (*top inset table*) and the Hill (*bottom inset table*) equation (KaleidaGraph 4.1 software). Summary plots are fitted using the Hill equation.

Chapter 4

In vitro Genome-wide Mapping of a Potent Antiviral Hairpin Polyamide (PA1) to Human Papillomavirus 16 using Hydroxyl Radical Footprinting Coupled to Massively Parallel DNA Sequencing (·OH-Seq)

4.1 ABSTRACT

PA1 is a cell-permeable, minor groove binder that belongs to the hairpin polyamide class of compounds. PA1 targets the DNA sequence 5'-W_2GW_7 (where W = A or T) according to the established polyamide-DNA recognition rules. **PA1** was rationally designed to target A/T-rich DNA sequences near the HPV16 (human papillomavirus 16) origin of replication essential for both viral transcription and replication. **PA1** exhibits apparent IC_{50}/IC_{90} values of 0.1/1.1 μM against HPV16 in monolayer keratinocyte cultures without measurable cytotoxicity. Here, we have employed hydroxyl radical footprinting coupled to massively parallel DNA sequencing (·OH-Seq) to map the genome-wide binding events of **PA1** onto the HPV16 genome *in vitro* (cell-free) as a preliminary stage prior to our cell-based studies.

4.2 INTRODUCTION

4.2.1 Deciphering Specific Interactions in DNA-Ligand Complexes by Footprinting Methods

Gene expression and DNA replication are tightly regulated by the specific formation of DNA and protein complexes. Thus, deciphering the occupancy of DNA-binding proteins to their cognate binding sites is of great importance toward the understanding of these fundamental cellular processes. Footprinting methods are widely used for the interrogation of specific DNA-protein interactions.[1,2] In these techniques, a DNA fragment harboring the sequence of interest is end-labelled, either radioactively with ^{32}P or fluorescently for visualization purposes. The DNA fragment is then subjected to mild digestion by an enzymatic or chemical cleavage agent, both in the presence and absence of a DNA-ligand. The fragmented DNA is separated by electrophoresis and the cleavage patterns are compared between the naked and ligand-bound DNA samples. Binding of the ligand to its recognition site, or cognate sequence, leads to increased protection of DNA at that particular location, providing a 'footprint'.[1,2] DNase I footprinting pioneered by Galas and Schmitz[3] provides a powerful yet simple technique for deciphering the specific contacts in DNA-protein complexes. This method has been successfully employed to determine the sequence selectivity of numerous DNA-ligands including intercalators,[4-7] transcription factors,[8-10] and minor groove DNA binders.[4,5,7,11,12]

DNase I footprinting relies on the enzymatic cleavage of DNA by the endonuclease DNase I.[2,3] Although DNase I is often preferred as a cleavage agent due to its ease of use, the nature of its DNA-binding properties presents some disadvantages. For instance, the cleavage pattern generated by DNase I often exhibits uneven cleavage of the phosphodiester backbone as a result of local sequence-dependent structural motifs. Thus, regions of DNA that are poor substrates for DNase I fail to provide information about ligand binding. In addition, DNase I is relatively large, with a DNA-binding surface of approximately 10 bp, which causes overestimation of ligand binding site sizes and poor resolution of multiple ligands bound in close proximity of each other.[1,2,13]

4.2.2 Hydroxyl Radical Footprinting

Due to the disadvantages of DNase I footprinting, various smaller cleavage reagents[14-17] have been introduced for the determination of DNA structure and DNA-ligand interactions. The development of hydroxyl radicals as a high-resolution chemical probe has allowed the elucidation of nucleic acid structure[18-21] and nucleic acid-protein complexes at nucleotide-resolution.[16,22-26] In particular, hydroxyl radicals are small, charge-neutral, and highly reactive species that attack deoxyribose hydrogen atoms in a sequence-independent manner.[27,28] These qualities cause a more intimate attack of DNA providing high-resolution single-nucleotide information of the ligand-nucleic acid interactions. Furthermore, the flexibility and relative ease of hydroxyl radical generation allows for footprinting of small molecules and proteins in solution *via* Fenton chemistry[16,22] or ionizing radiation.[23,24]

The highly-reactive hydroxyl radical moiety is produced by the Fenton reaction between ferrous iron (Fe^{2+}) and hydrogen peroxide (H_2O_2) (**Equation 4.1**). Reduction of H_2O_2 by the $[Fe(II)(EDTA)]^{2-}$ complex leads to the production of a hydroxide ion (OH^-), $[Fe(III)(EDTA)]^-$ and the hydroxyl radical ($\cdot OH$). Addition of sodium ascorbate reduces Fe(III) to regenerate Fe(II), affording a reaction catalytic in iron. Furthermore, the negatively charged $[Fe(II)(EDTA)]^{2-}$ complex prevents direct electrostatic interactions between the metal ions and the negatively charged DNA backbone.[25]

$$[Fe(II)(EDTA)]^{2-} + H_2O_2 \xrightarrow[\text{ascorbate}]{} [Fe(III)(EDTA)]^- + OH^- + \cdot\mathbf{OH} \qquad \mathbf{(4.1)}$$

Once the hydroxyl radical is generated, it quickly reacts with the DNA backbone in a sequence-independent fashion by abstracting a solvent accessible hydrogen atom from the deoxyribose sugar, leading to cleavage of the DNA backbone at this position.[28] The extent of cleavage in a given position relates to either steric factors or the conformation of the deoxyribose along with the stability of the radical generated upon hydroxyl radical attack. Specifically, kinetic isotope effect studies of nucleotides deuterated separately at each deoxyribose position revealed that the degree of reactivity of hydroxyl radical toward the different hydrogen atoms correlated with their solvent accessibility within the DNA backbone (5' H > 4' H > 3' H ~ 2' H ~ 1' H)[27,28] (**Figure 4.1**).

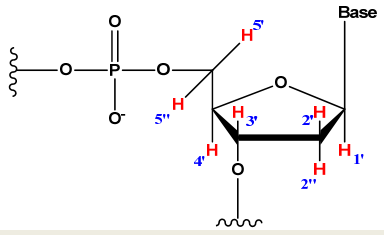


Figure 4.1. Chemical Structure of a Single DNA Nucleotide. Theoretically, hydroxyl radical attack can occur at any of the seven deoxyribose hydrogen atoms. However, their respective reactivity toward the oxidizing agent is correlated to their relative accessibilities (5' H > 4' H > 3' H ~ 2'H ~ 1'H).

The reaction products of the oxidative damage of DNA by hydroxyl radicals are comprised of single-stranded breaks along the DNA duplex. Thus far, three DNA lesions generated by a hydroxyl radical attack have been characterized: **1)** a single-nucleoside

gap flanked by two phosphates (**Figure 4.2 Top**), **2)** a single-nucleoside gap flanked by a 5' phosphate and a 3' phosphoglycolate (**Figure 4.2 Middle**) and **3)** a single-stranded nick flanked by a 5' aldehyde and a 3' phosphate (**Figure 4.2 Bottom**).[20,25,27]

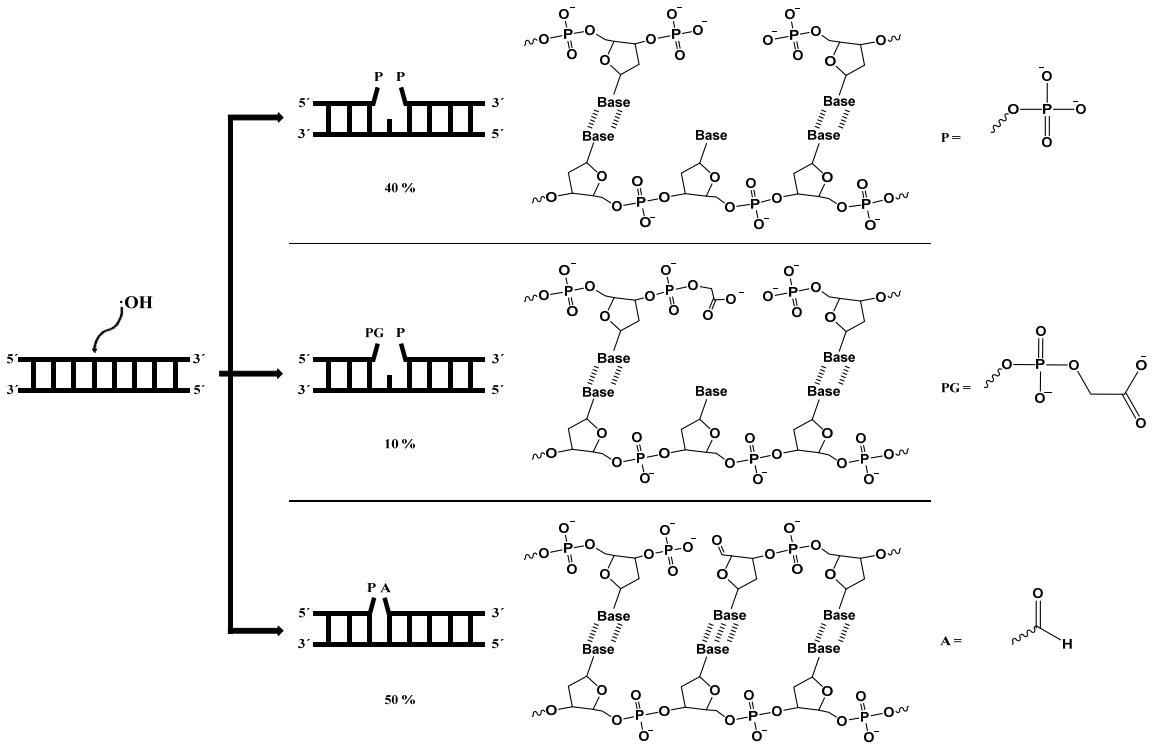


Figure 4.2. Chemical Structures of the Lesions Generated by Hydroxyl Radical DNA Cleavage. Abstraction of a deoxyribose hydrogen atom produces single-stranded nicks (~50 %) and gaps (~50 %) along the DNA duplex. Depending on the site of hydroxyl radical attack, the resulting reaction products consist of a single-nucleoside gap flanked by two phosphates (40 %) (*Top*), a single-nucleoside gap flanked by a 5' phosphate and a 3' phosphoglycolate (~10 %) (*Middle*), or a single-stranded nick flanked by a 5' aldehyde and a 3' phosphate (*Bottom*). P = phosphate, PG = phosphoglycolate, A = aldehyde.

Analysis of the cleavage reaction products has typically been conducted by high-resolution polyacrylamide gel electrophoresis (PAGE). Alternate approaches relying on fragment analysis by automated capillary electrophoresis (CE) have provided several advantages over PAGE.[29,30] Importantly, the high-throughput nature of CE allows multiple samples to be analyzed concurrently. Furthermore, detection of the fragmented DNA by CE relies on fluorescent dyes, eliminating the use of radioisotopes. However, both of these analyses are severely limited by the length of DNA that can be interrogated and require prior sequence indexing in order to accurately map DNA-ligand binding regions.[31] Here we present the use of massively parallel DNA sequencing as an attractive alternative to analyze hydroxyl radical footprinting cleavage products.

4.2.3 Next-Generation Sequencing

With the advent of next-generation sequencing (NGS), it is now possible to sequence millions of nucleotides in a high-throughput and cost-effective manner. Consequently, NGS has allowed for novel approaches to answer scientific inquiries in the areas of high-

throughput genome-wide sequencing,[32] transcriptome profiling,[33] transcription factor discovery[9,34] and genome-wide mapping of small molecules.[35,36]

The typical next-generation sequencing workflow consists of four steps: **1)** library preparation from randomly fragmented DNA, **2)** cluster generation on a solid surface, **3)** sequencing by synthesis and **4)** alignment and data analysis **(Figure 4.3)**.[37]

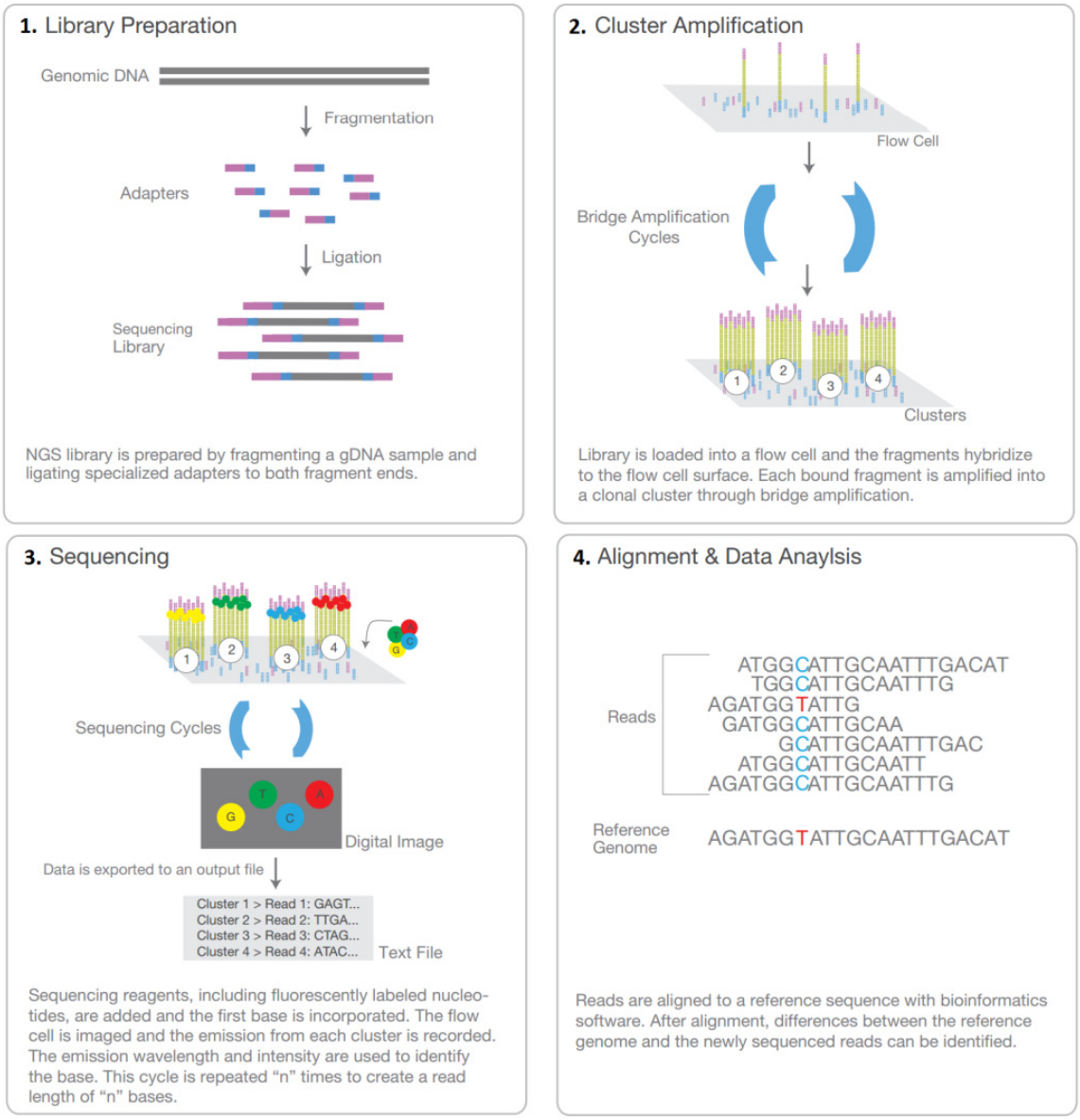


Figure 4.3. Illumina Next-generation Sequencing Technology Workflow. Illumina's sequencing by synthesis (SBS) chemistry affords for the sequencing of millions of DNA fragments in a massively parallel manner. **(1)** Library preparation - DNA is randomly fragmented and modified at the 5' and 3' ends by covalent attachment of adapters. **(2)** Cluster amplification – Constructed libraries hybridize to covalently bound oligos on the surface of the flow cell. Clonal clusters are generated throughout the flow cell, in which each cluster is derived from a single fragment through bridge amplification. **(3)** Sequencing by synthesis – SBS consists of the stepwise incorporation of fluorescently labeled nucleotides by DNA polymerase, followed by identification of the incorporated nucleotide based on the emission wavelength. The dye is then removed along with the protecting group on the 3'-OH so that the next round of SBS can be carried out. **(4)** Data analysis – Sequence reads are mapped to a reference genome using an alignment program. The aligned reads can then be subjected to other analyses determined by the basis of the scientific inquiry.
(Figure adapted from www.illumina.com/technology/next-generation-sequencing.html)

Library construction entails the random fragmentation of high-molecular DNA, followed by ligation of distinct oligonucleotide adapters to both ends of these fragments. Adapters are synthetic oligonucleotides that contain specialized indexing sequences for multiplexing samples, as well as sequences complementary to single-stranded oligonucleotides covalently attached on the flow cell surface **(Figure 4.4)**.

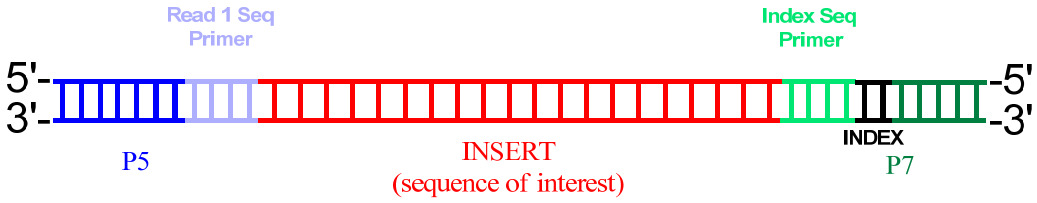


Figure 4.4. Structural Details of Adapter-Ligated Fragment. The constructed adapter-ligated fragment consists of 1) DNA insert (sequence of interest) flanked by adapter constructs that contain primer binding sites, 2) unique index sequences, and 3) sequences complementary to oligos on the flow cell surface (P5 and P7).

After ligation, the ligated fragments are enriched by primer-specific PCR amplification. These adapter-ligated fragments are then denatured into single-stranded DNA molecules, allowing these fragments to hybridize onto the flow cell and undergo cluster amplification. After generation of the spatially separated clusters, DNA polymerase and all four nucleotides with labile-fluorescently labelled reversible terminators are added to initiate the first cycle of sequencing by synthesis. Because these nucleotides have a 3'-OH blocking group, only a single, complementary nucleotide is incorporated into the nascent strand. After washing of unincorporated nucleotides, the instrument optics scans and records the emitted fluorescence at each cluster, allowing for the identification of the incorporated nucleotide. At this point, chemical cleavage of the 3'-OH blocking group in the growing strand permits the incorporation of a single nucleotide in the next round. The cycle is then repeated for a specific number of times, resulting in a "read" with known length.[37,38]

4.2.4 Hydroxyl Radical Footprinting Coupled to Massively Parallel DNA Sequencing

Tullius *et al.* introduced hydroxyl radical footprinting in 1986.[16] It has evolved into a powerful technique to decipher the structural details of DNA-ligand interactions. Recent efforts by the Tullius lab have aimed at developing a modified hydroxyl radical footprinting method using next-generation sequencing as the readout of the cleavage patterns. Xu *et al.*[18] demonstrated the feasibility of this method by studying hydroxyl radical cleavage patterns of genomic DNA *in vitro*. Using pUC18 plasmid as a model DNA genome, hydroxyl radicals generated by Fenton chemistry were used to fragment the plasmid DNA to low-molecular weight species. Because most of the commercially available NGS library preparation kits require intact double-stranded DNA fragments, the resulting single-stranded DNA breaks were subsequently processed with T7 endonuclease I to produce double-stranded DNA breaks at hydroxyl radical attack sites. Using NGS as a digital readout, they obtained "genome-wide" cleavage intensities for plasmid pUC18 that were in agreement with the cleavage pattern predicted by ORChID

(·OH Radical Cleavage Intensity) based on the pUC18 DNA sequence. As a result, the use of massively parallel DNA sequencing as a digital readout for hydroxyl radical footprinting is highly promising.

Here, we employ hydroxyl radical footprinting coupled to massively parallel DNA sequencing (·OH-Seq) to map the genome-wide binding events of **PA1** onto the human papillomavirus 16 (HPV16) genome *in vitro*. **PA1** is a cell-permeable, minor groove binder belonging to the hairpin polyamide (PA) class of compounds (**Figure 4.5A**). *N*-methylpyrrole/*N*-methylimidazole (Py/Im) hairpin polyamides are synthetic heteroaromatic compounds inspired by the chemical structures of the natural antibiotics netropsin and distamycin A.[12,39-43] These crescent-shaped molecules, collectively known as polyamides, bind with high affinity (nano- to picomolar dissociation constants) to the minor groove of dsDNA in a sequence-selective fashion.[44-46] According to the established polyamide-DNA recognition rules (*vide infra*), **PA1** targets the DNA sequence 5'-W_2GW_7 (where W = A or T) by recognizing the sequence-dependent H-bonds presented by the Watson-Crick bases in the minor groove of B-DNA (**Figure 4.5B**). In this manner, a Py/Im pair H-bonds to a G·C (but not a C·G), whereas a Py/Py H-bonds to the degenerate A·T and T·A base pairs. Notably, substitution of Py heterocycles with the flexible β-alanine (β) substituent relieves the inherent overcurvature observed in polyamides with four or more contiguous heterocycles and resets the H-bond register between the ligand and DNA.[47,48] The gamma-aminobutyric turn recognizes the degenerate A·T and T·A base pairs preferentially due to steric factors.[49] The C-terminal positively-charged tail recognizes A·T and T·A base pairs and mimics the cationic moieties found in distamycin A and netropsin.[39,49] This reported specificity is achieved by the optimal positioning of the H-bond acceptors and donors, van der Waals interactions with the walls of the minor groove, and appropriate curvature of the polyamide with respect to the minor groove of B-DNA.[50]

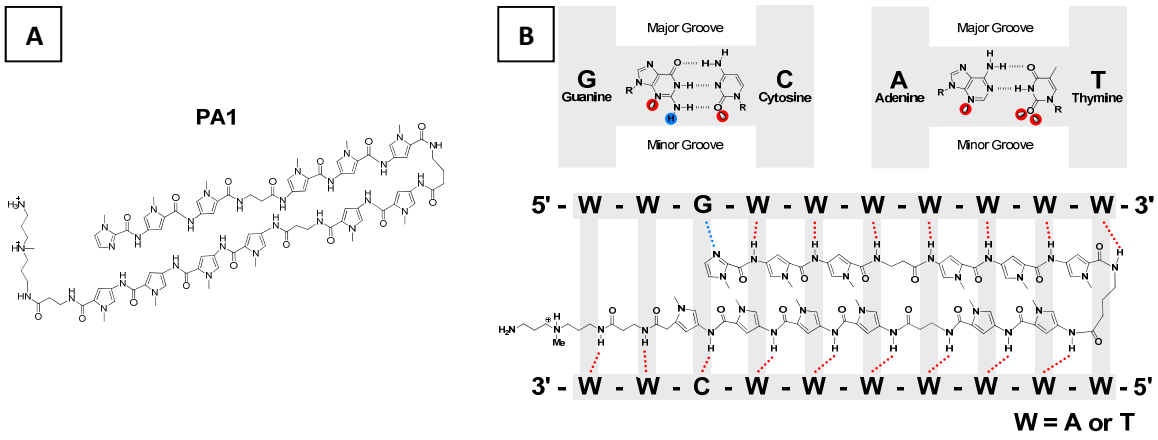


Figure 4.5. Chemical Structure of PA1 and Watson-Crick Base Pairs. (A) Hairpin polyamide 1 (**PA1**) has the sequence dIm-Py-Py-β-Py-Py-Py-γ-Py-Py-β-Py-Py-Py-Py-β-Ta (dIm = des-amino-*N*-methylimidazole, Py = *N*-methylpyrrole, β = beta-alanine, γ = gamma-aminobutyric acid, Ta = bis(aminopropyl)*N*-methylamine. **(B)** Chemical structure of the Watson-Crick base pairs showing the available H-bond acceptors (red electron cloud) and H-bond donor (blue circled H) in the DNA minor groove (*Top*). Hydrogen boding (dashed lines) scheme for **PA1**-DNA binding depicting **PA1** in the minor groove of a double-stranded DNA molecule with the sequence 5'-W_2GW_7, where W = A or T. Blue and red dashed lines represents **PA1** H-bond acceptor and donors, respectively (*Bottom*).

Our lab and its collaborators have previously demonstrated the efficacy of **PA1** as an antiviral agent against oncogenic human papillomavirus (HPV) types maintained in organotypic raft tissues and monolayer keratinocytes.[51,52] HPV is a circular, double-stranded DNA virus that infects basal keratinocytes of cutaneous and mucosal tissues.[53] To date, over 180 types have been isolated of which approximately 15 oncogenic types have been identified.[54-56] The development of cervical cancer, the second most prevalent cancer affecting women worldwide, is closely associated with persistent infections of oncogenic HPV types.[57-59] Specifically, oncogenic HPV DNA has been detected in 99.7 % of cervical cancers, as well as in ~90 % of anal cancers and ~60 % of oropharyngeal (mouth and throat) cancers.[60] **PA1** was rationally designed to target A/T-rich DNA sequences near the HPV16 origin of replication essential for both viral transcription and replication.[51] HPV16 was used as the target because it alone is responsible for approximately 50 % of cervical cancers worldwide.[61] Treatment of monolayer keratinocyte cultures with **PA1** for 48 h followed by quantification of viral DNA levels using qPCR (quantitative PCR) revealed a dose-dependent response in the decrease of viral DNA. Based on these results, **PA1** exhibits apparent IC_{50}/IC_{90} values of 0.1/1.1 μM against HPV16 without measurable cytotoxicity *via* MTT assays.[51,52]

Our lab has mapped **PA1** binding interactions and their respective binding affinities extensively by CE-based quantitative DNase I footprinting and affinity cleavage experiments on a number of DNA sequences derived from the natural sequence of different oncogenic HPV types.[52,62-64] We found that **PA1** binds avidly to the minor groove of A/T-rich sequences with nanomolar dissociation constants. Furthermore, **PA1** can tolerate PA-DNA mismatches (single- and double-base-pair) without a significant decrease in binding affinity. However, the technologies that we initially relied on, while powerful, were limited due to their scalability. Specifically, CE-based fragment analysis is severely limited by the length (maximum of ~350 bp) of DNA that can be interrogated. This fragment length is much smaller than the read-lengths amenable to CE-based DNA sequencing because of the stricter requirements for signal-to-noise ratios in CE data that integrated for quantitative DNase I footprinting than for data analyzed in a relatively digital manner for the Yes/No appearance of the A, C, G, or T-associated fluorescent dyes used in sequencing. Because the HPV16 genome consists of 7.904 kilobases, a minimum of 23 fragments of 350 base pairs must be amplified using specific fluorescently labelled primers in order to cover the entire genome. This estimate does not account the requirement for the fragments to overlap, since binding sites occurring at the beginning and end of the fragment suffer from low signal-to-noise ratios in the CE experiment. Additionally, the raw data obtained from the electropherograms require prior base pair indexing in order to accurately map the polyamide binding regions. Therefore, our prior approach is both labor-intensive and slow, though it was successful in answering fundamental questions about in binding of drug candidates to viral DNA.

Thus, hydroxyl radical footprinting coupled to massively parallel DNA sequencing is an attractive approach to map the genome-wide binding events of **PA1** onto the HPV16 genome *in vitro*. **Figure 4.6** summarizes our experimental workflow. Briefly, HPV16/pUC18 DNA was incubated in the presence and absence of **PA1**, and subsequently subjected to hydroxyl radicals generated by Fenton chemistry (**Figure 4.6 Top**). As described above, cleavage of the DNA phosphodiester backbone by hydroxyl radicals leads to single-stranded breaks with various chemistries on the 5' and 3' ends. Of

these, only the fraction of fragments containing phosphate groups on both ends can be ligated to the NGS adapters in succeeding steps (**Figure 4.6 Middle**). Furthermore, a library preparation kit compatible with nicked samples must be employed for Illumina library preparation of the fragmented DNA. Therefore, an Accel-NGS 1S Plus DNA Library Kit (Swift Biosciences) was used for library preparation of hydroxyl radical footprint samples. This kit is compatible with single-stranded, denatured and nicked input DNA making it ideal for hydroxyl radical cleavage products. The technology employed by this kit relies on the ligation of a truncated adapter to the 3' end of single-stranded DNA (ssDNA) fragments, followed by an extension step to facilitate the ligation of the second truncated adapter during the ligation step. The resulting insert flanked by truncated adapters is then subjected to PCR to amplify these fragments and to insert the indexed adapters for subsequent sequencing (**Figure 4.6 Bottom Left**). NGS sequencing was performed on these libraries using single-end, 51 base pair read length on an Illumina HiSeq 2500. Unique sequencing reads were then aligned to the HPV16/pUC18 reference genome. A detailed scheme of Illumina's NGS technology is also provided below (**Figure 4.6 Bottom Right**). Computational analysis of the aligned reads revealed the cleavage patterns at single-nucleotide resolution, allowing us to determine the genome-wide binding events of **PA1** on HPV16 *in vitro*.

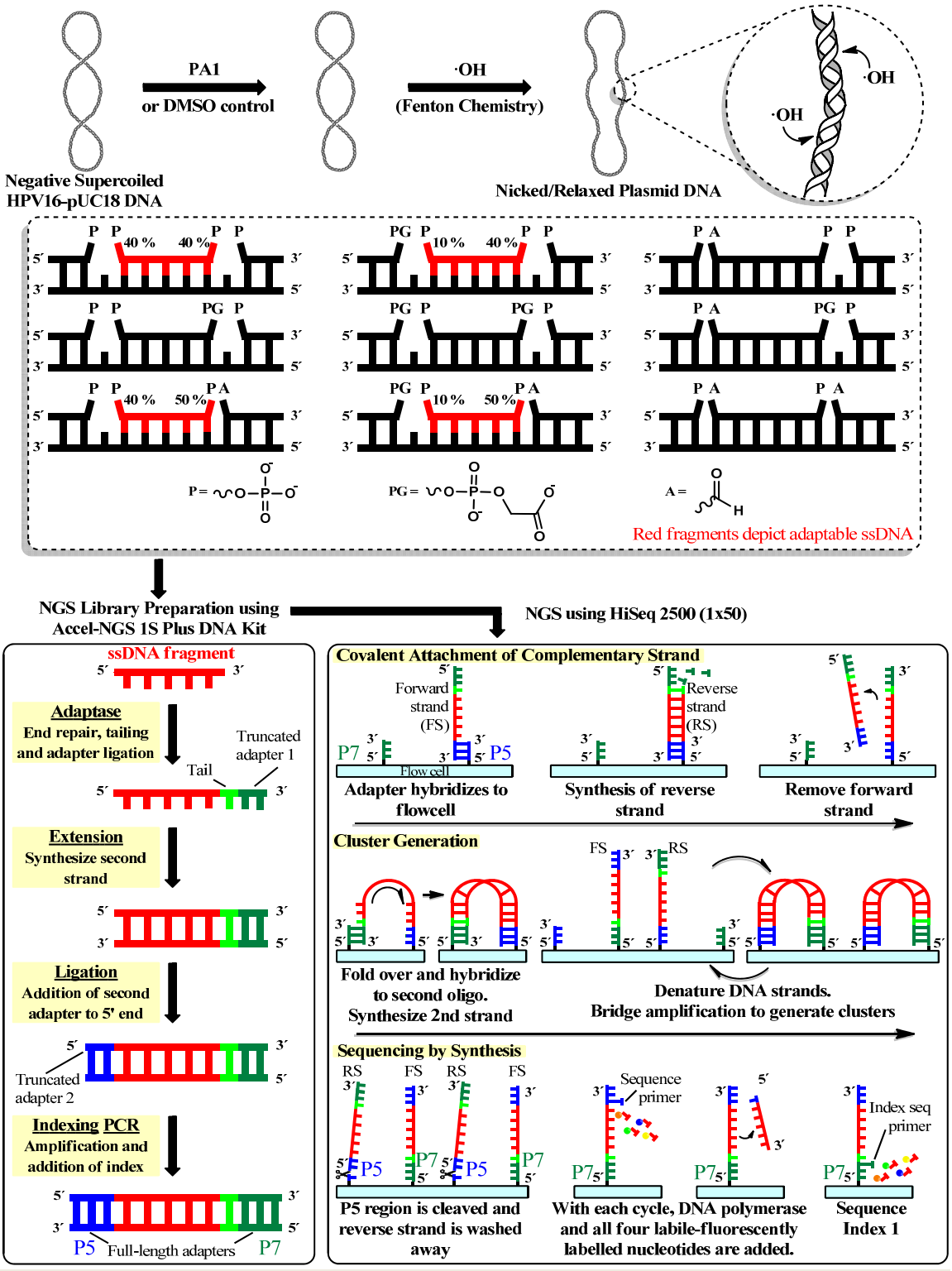


Figure 4.6. *Caption on next page.*

Figure 4.6. Hydroxyl Radical Footprinting Coupled to Illumina Massively Parallel DNA Sequencing Workflow. HPV16/pUC18 DNA was incubated in the absence (DMSO) and in the presence of **PA1**. Plasmid DNA was then fragmented by hydroxyl radicals generated by Fenton chemistry (*Top*). The chemical structures of the lesions generated by hydroxyl radical DNA cleavage are provided and the adaptable fragments are represented in red fragments (*Middle*). The cleavage products were then used to construct Illumina next-generation sequencing libraries using an Accel-NGS 1S Plus DNA Library Kit (*Bottom Left*). Sequencing was performed using an Illumina HiSeq 2500 (*Bottom Right*). Sequencing is accomplished by hybridizing single-stranded library molecules to complementary oligos found on the flow cell. After synthesis of the reverse strand, covalently attached fragments undergo cluster amplification by bridge amplification, generating millions of the spatially separated clusters throughout the flow cell surface. At this point, the P5 regions are cleaved, resulting in clusters composed of fragments (forward strand) attached covalently to the flow cell *via* their P7 region. To initiate the sequencing cycle, the sequencing primer is added and it anneals to the P5 region. In each cycle of SBS, nucleotides with labile-fluorescently labelled reversible terminators are added, followed by a detection step and a deblocking step to license the incorporation of the next nucleotide. Sequencing of the index barcode ensues by the removal of Read 1, followed by annealing of the Index sequencing primer and extension.

4.3 MATERIALS AND METHODS

4.3.1 Buffers and Reagents used for *in vitro* ·OH Footprinting Experiments

Starting reagents were used without further purification, unless specifically noted. Autoclaved MilliQ H_2O (18.2 MΩ·cm at 25 °C, Millipore Integral 10) was used in the preparation of all reagents. HPV16/pUC18 plasmid was obtained from Professor Paul Lambert's lab. Lysogenic Broth (LB) Agar was obtained from Sigma, Catalog # L7533-6X500ML. Glucose was purchased from Aldrich, Catalog # 158968-1KG. Ampicillin was purchased from Fisher Scientific, Catalog #BP1760-25. QIAGEN Plasmid Maxi kit was purchased from QIAGEN, Catalog # 12162. *Bam*HI restriction enzyme (Catalog # R0136S, 20,000 U/mL), 1X NEBuffer 3.1 (Catalog # R7203S, 10X concentration), Gel Loading Dye, Blue (6X, Catalog # B7021S), 1 kb DNA ladder (Catalog # N3232S, 500 μg/mL) were purchased from New England Biolabs. Ethylenediaminetetraacetic acid (EDTA) was purchased from Sigma, Catalog # E5134-100G. 1X TAE buffer consisted of 40 mM Tris-acetate, 1 M EDTA, pH 8.0. Ethidium bromide was obtained from Spectrum (Catalog # E1031, 1 % solution). Low EDTA TE buffer (10 mM Tris·Cl, 0.1 mM EDTA, pH 8.0) was purchased from Teknova, Catalog # T0227. 1 M Tris HCl, pH 7.5 was obtained from Invitrogen, Catalog # 15567-027. Dimethyl sulfoxide was purchased from Sigma Life Science, Catalog # D2650. Ammonium iron(II) sulfate hexahydrate was obtained from Aldrich, Catalog # 203505-25G. Sodium ascorbate was obtained from (Sigma, Catalog # 11140-50G). H_2O_2 was purchased from Fisher Chemical, Catalog # H325-500, 30 % wt/vol. 0.22 μm MILLEX-GP filters were purchased from Merck Millipore Ltd. 100 % ethanol was obtained from Decon Laboratories, Inc., Catalog # 2716. Agilent Bioanalyzer RNA 6000 Nanochip kit was purchased from Agilent Technologies, Catalog # 5067-1511. Agarose was obtained from Sigma, Catalog # A9539-100G. Qubit dsDNA BR (Broad-Range) Assay kit (Catalog # Q32853) and Qubit assay tubes (Catalog # Q32856) were obtained from Life technologies. Accel-NGS 1S Plus DNA Library Kit was purchased from Swift Biosciences, Catalog # DL-IL1SP-12/48. SPRIselect reagent kit was obtained from Beckman Coulter, Catalog # B23317. Agilent DNA 1000 kit was obtained from Agilent Technologies, Catalog # 5067-1505.

4.3.2 Polyamide Synthesis

Hairpin polyamide synthesis was performed by Dr. K. J. Koeller and Dr. G. D. Harris, Jr. **PA1** was synthesized by Boc solid-phase methods[65] as previously reported in the literature.[51]

4.3.3 Preparation of HPV16/pUC18 Plasmid DNA

4.3.3.1 *Amplification and Purification of HPV16/pUC18 Plasmid DNA*

The HPV16/pUC18 plasmid was obtained from Professor Paul Lambert's lab.[66] The plasmid consists of the full HPV16 genome from W12E keratinocyte cells (GenBank Accession No. AF125673) that has been cloned into the *Bam*HI restriction site of the pUC18 vector (total length = 10589 bp). Because the *Bam*HI restriction site (which cleaves the 5' guanine on each DNA strand of the sequence 5'-GGATCC-3') occurs within the open reading frame of L1, cloning of the HPV16 genome into the pUC18 vector leads to the disruption of the L1 coding region (**Figure 4.7**). HPV16/pUC18 plasmid DNA[66] was amplified in transformed JM109 *Escherichia coli* cells in 1 % glucose/LB media (Aldrich, Catalog # 158968-1KG / Sigma, Catalog # L7533-6X500ML) containing ampicillin (Fisher Scientific, Catalog #BP1760-25). Plasmid DNA was then purified from the harvested cells using a QIAGEN Plasmid Maxi kit (QIAGEN, Catalog # 12162) and further concentrated by ethanol precipitation. Plasmid DNA concentration was determined by measuring the absorbance at 260 nm using a Thermo Scientific Evolution 260 Bio UV-visible spectrophotometer.

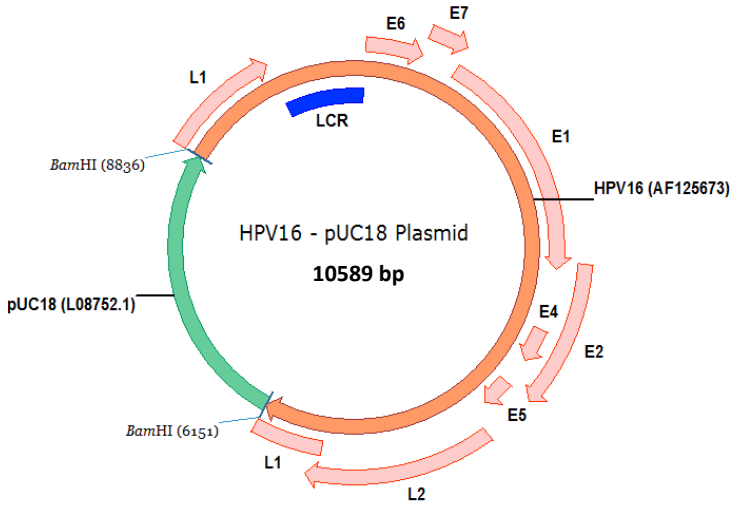


Figure 4.7. HPV16/pUC18 Plasmid Map. Plasmid construct used for the *in vitro* ·OH-Seq experiments. The HPV16 coding regions are given with their respective positions within the construct. Cloning of the full genome of HPV16 into pUC18 disrupts the L1 gene coding region.

4.3.3.2 *Restriction Digests of HPV16/pUC18 Clones*

To confirm that the correct HPV16/pUC18 DNA plasmid was amplified, a restriction digest using *Bam*HI restriction enzyme (NEB, Catalog # R0136S, 20,000 U/mL) was

performed on the purified DNA for an hour at 37 °C in 1X NEBuffer 3.1 (NEB, Catalog # R7203S, 10X concentration). The reaction products produced by *Bam*HI treatment were mixed with Gel Loading Dye, Blue (6X, NEB, Catalog # B7021S) and electrophoresed on 1 % agarose gel in 1X TAE buffer (40 mM Tris-acetate, 1 M EDTA, pH 8.0) at 110 V, room temperature for 2 h. After electrophoresis, the agarose gel was stained with 0.5 µg/mL ethidium bromide (Spectrum E1031, 1 % solution) for 20 min and washed with autoclaved MilliQ H_2O for 5 min in a staining box. The fragmented DNA and 1 kb DNA ladder (NEB, Catalog # N3232S, 500 µg/mL) were visualized with a FOTO/Convertible UV-light box equipped with an ethidium bromide filter (FOTODYNE Incorporated). The agarose gel was imaged and analyzed with FOTO/Analyst PC Image version 5.00.

4.3.4 Hydroxyl Radical (·OH) Cleavage of HPV16/pUC18 Plasmid DNA

4.3.4.1 *Optimization of Hydroxyl Radical (·OH) Reagent Concentration*

For *in vitro* ·OH-Seq footprinting experiments, hydroxyl radicals were generated by the Fenton reaction (**Equation 4.1**). To determine the appropriate working concentrations of the hydroxyl radical reagents required to generate a significant population of DNA fragments with lengths <1000 bp, 5 µg of HPV16/pUC18 plasmid DNA (6,541,200 g/mol; OligoCalc)[67] were diluted to a final concentration of 38.2 nM DNA, 5 mM Tris HCl, pH 7.5 (Invitrogen, Catalog # 15567-027) and 2 % v/v DMSO (Sigma Life Science, Catalog # D2650). This solution was incubated for 10 min at 37 °C. Stock solutions of 10 mM Fe(II)-EDTA and 0.3 % H_2O_2 (30 % wt/vol, Fisher Chemical, Catalog # H325-500) were prepared immediately before use and wrapped in aluminum foil to protect them from light exposure. The 10 mM Fe(II)-EDTA solution was prepared by mixing equal volumes of 20 mM aqueous Fe(II) solution in the form of ammonium iron(II) sulfate hexahydrate (Aldrich, Catalog # 203505-25G) and 40 mM aqueous EDTA (Sigma, Catalog # E5134-100G). The excess EDTA ensures complete chelation of the metal. A stock of 200 mM sodium ascorbate (Sigma, Catalog # 11140-50G) in autoclaved MilliQ H_2O was prepared and wrapped in aluminum foil to protect from light exposure. These stocks were then filtered through a 0.22 µm MILLEX-GP filter before use (Merck Millipore Ltd). The HPV16/pUC18 plasmid DNA solution (38.2 nM) was then treated with various concentrations of the hydroxyl radical reagents at 20 °C for a fixed reaction time of 5 min. The final concentrations of the cleavage reagents ranged from 0.43 mM Fe(II)-EDTA, 8.7 mM sodium ascorbate and 0.013 mM H_2O_2 to 2.50 mM Fe(II)-EDTA, 50 mM sodium ascorbate and 0.075 mM H_2O_2. The cleavage reaction was quenched by the addition of 3X volume of cold (4 °C) 100 % ethanol (Decon Laboratories, Inc., Catalog # 2716). Ethanol precipitation of the fragmented DNA was then performed. The DNA pellet was resuspended in 20 µL Low EDTA TE buffer (Teknova, Catalog # T0227, 10 mM Tris·Cl, 0.1 mM EDTA, pH 8.0) and stored at -20 °C. Fragment size distribution of the reaction products were assessed by 1 % agarose gel electrophoresis and Agilent Bioanalyzer RNA 6000 Nanochip kit (Agilent Technologies, Catalog # 5067-1511).

4.3.4.2 *Hydroxyl Radical (·OH) Footprinting*

HPV16/pUC18 DNA (38.2 nM) was incubated with either 2 μM or 20 μM **PA1** for at least 4 h at 37 °C before subjecting the samples to the hydroxyl radical cleavage reagents. Hydroxyl radical cleavage reactions were performed on the plasmid DNA in the absence and in the presence of **PA1** for 5 min at 20 °C in triplicate. DNA cleavage by ·OH was initiated by the addition of 20 μL each of 10 mM Fe(II)-EDTA, 200 mM sodium ascorbate and 0.3 % H_2O_2 followed by rapid mixing (2.5 mM Fe(II)-EDTA, 50 mM sodium ascorbate and 0.075 % H_2O_2 final concentration). After 5 min, the cleavage reaction was quenched by the addition of 3X volume of cold (4 °C) 100 % ethanol and vortexing. Ethanol precipitation of the fragmented DNA was then performed. The DNA pellet was resuspended in 20 μL Low EDTA TE buffer and stored at -20 °C. Fragment size distribution of the reaction products were estimated with a 1 % agarose gel electrophoresis and Agilent Bioanalyzer RNA 6000 Nanochip kit. DNA concentration was determined using a Qubit dsDNA BR (Broad-Range) Assay kit (Life technologies, Catalog # Q32853).

4.3.4.3 *Assessment of Fragment Size Distribution by 1 % Agarose Gel Electrophoresis and Agilent Bioanalyzer RNA 6000 Nanochip Kit*

The DNA cleavage products generated by hydroxyl radicals were resolved by agarose gel electrophoresis. 1.250 μg of fragmented HPV16/pUC18 was mixed with Gel Loading Dye, Blue and loaded in a 1 % agarose gel. The gel was run in a 1X TAE buffer at a voltage of 147 V for 4-5 h at room temperature. After electrophoresis, the agarose gel was stained with 0.5 μg/mL ethidium bromide for 20 min and washed with autoclaved MilliQ H_2O for 5 min in a staining box. The fragmented DNA and 1 kb DNA ladder were visualized with a FOTO/Convertible UV-light box equipped with an ethidium bromide filter (FOTODYNE Incorporated). The agarose gel was imaged and analyzed with FOTO/Analyst PC Image version 5.00.

For the assessment of ssDNA fragment size distribution produced by the cleavage reaction, an Agilent Bioanalyzer RNA 6000 Nanochip kit was used following the manufacturer's instructions. Briefly, 2 μL of 62.5 ng/μL fragmented DNA was heat denatured at 95 °C for 3 min using a Mastercycler Nexus Thermal Cycler (Eppendorf) and immediately placed on ice until ready to load on the Nanochip. Then, 1 μL of ssDNA was loaded onto a sample well containing 5 μL of Agilent RNA marker solution. The RNA 6000 Nanochip was then analyzed in an Agilent 2100 Bioanalyzer using the Eukaryote Total RNA Nano Series II assay conditions. The results were analyzed using Agilent Technologies 2100 Bioanalyzer 2100 Expert Version B.02.08.SI648 (SR3) software. The Agilent 2100 Bioanalyzer is an analytical system that allows for the determination of fragment size and concentration of nucleic acid fragments. This chip-based technology relies on the separation of nucleic acid oligos based on their size by electrophoresis that are detected, in real-time, by a specific (but proprietary) fluorescent dye that binds with nucleic acids. Using a ladder of known fragment sizes and concentrations, the size and concentration of the sample can be estimated based on its relative mobility and fluorescence intensity.

4.3.4.4 *Quantification of dsDNA using Qubit dsDNA BR Assay Kit*

DNA concentration of the ethanol-precipitated, cleavage reaction products and prepared libraries were quantified using a Qubit dsDNA BR assay kit and a Qubit 2.0 Fluorometer (Life technologies) following the manufacturer's instructions. Briefly, Qubit dsDNA BR Standards #1 and #2 were prepared by separately diluting 10 µL of either Standard #1 (0 ng/ µL) or Standard #2 (100 ng/µL) in 190 µL of Qubit working solution (1:200 Qubit dsDNA BR Reagent:Qubit dsDNA BR buffer) in separate Qubit assay tubes (Life technologies, Catalog # Q32856). Unknowns were prepared by diluting 2 µL of cleavage reaction products or 4 µL of Illumina NGS library sample to a total volume of 200 µL with Qubit working solution in separate Qubit assay tubes. Standards and unknown solutions were vigorously vortexed and allowed to incubate at room temperature for at least 2 min. After the incubation period, the fluorescence signal of each of these solutions was quantified with a Qubit 2.0 Fluorometer (Life technologies). The fluorescence signal emitted by the proprietary Qubit dsDNA BR Reagent fluorescent dye upon intercalation into dsDNA is directly proportional to the amount of dsDNA in the sample, which allows for the calculation of dsDNA concentration based on a two-point standard calibration curve. The Qubit 2.0 Fluorometer provides a dsDNA concentration in µg/mL of the diluted sample. Thus, **Equation 4.2** is used to calculate dsDNA concentration of the sample.

$$Concentration\ of\ unknown = Qubit\ 2.0\ Fluorometer\ value\ (\mu g/mL)\ \mathrm{x}\ \frac{200}{x} \qquad \textbf{(4.2)}$$

$$where\ x\ is\ the\ volume\ of\ sample\ used\ in\ microliters$$

4.3.5 Illumina Next-Generation Sequencing

4.3.5.1 *Preparation of Illumina Next-Generation Sequencing Libraries*

Illumina next-generation sequencing libraries were constructed using an Accel-NGS 1S Plus DNA Library Kit (Swift Biosciences, Catalog # SI-IL1SP-12A) following the manufacturer's protocol (refer to **Section 4.8.1** for detailed protocol). Thermal cycling steps were performed with a Mastercycler Nexus Thermal Cycler. Briefly, a 15 µL (10 ng/µL) aliquot of ·OH double-hit fragments in Low EDTA TE buffer was heat denatured in a thermal cycler at 95 °C for 2 min and immediately placed on ice for 2 min before proceeding to the Adaptase step. The Adaptase step is a proprietary reaction that performs fragment end repair, 3' end tailing and 3' ligation of a truncated adapter. 25 µL of the Adaptase Reaction Mix was added to the 15 µL sample aliquot containing the denature ssDNA and cycled with the following thermal profile: 37 °C for 15 min, 95 °C for 2 min and 4 °C hold. Following 3' end repair and truncated adapter ligation, adapter-ligated DNA molecules were amplified by adding 47 µL of the Extension Reaction Mix and cycling with the following thermal profile: 98 °C for 30 s, 63 °C for 15 s, 68 °C for 5 min and 4 °C hold. The newly synthesized strand allows for the ligation of the second truncated adapter to the 5' end of the DNA fragment. This synthesized strand is not sequenced. In order to remove oligonucleotides and small fragments, paramagnetic SPRI-based (Solid Phase Reversible Immobilization) bead purification of the Extension step

was performed with a SPRIselect reagent kit (Beckman Coulter, Catalog # B23317). Using a 1.2 ratio of SPRIselect reagent to sample, DNA fragments greater than or equal to 200 bp were size-selected and eluted with 20 µL of Low EDTA TE buffer. Although the composition of the SPRIselect reagent kit is proprietary, SPRI-based bead size-selection relies on the use of carboxyl-coated paramagnetic beads to reversibly and selectively bind dsDNA of desired lengths as a function of the concentration of a crowding agent (polyethylene glycol, PEG) and salt (NaCl or $MgCl_2$).[68-70] Specifically, increasing the ratio of the SPRIselect reagent to sample volume results in an increasing amount of fragments with shorter lengths bound to the beads.[71] Addition of the second truncated adapter were achieved by adding 20 µL of the Ligation Reaction Mix and cycled with the following thermal profile: 25 °C for 15 min and 4 °C hold. Clean-up of the Ligation step was then performed with a 1.0 ratio of SPRIselect reagent:sample to size-select for dsDNA fragments of sizes ≥ 200 bp using a 12-tube magnetic separation rack (NEB, Catalog # S1509S), followed by elution from the beads with 20 µL of Low EDTA TE buffer. Addition of the full-length indexed adapters was performed by adding 5 µL of Reagent R1 (index primer, SI-IL1SP-12A) and 25 µL of the PCR Reaction Mix. **Table 4.1** provides the specific indexing adapters used for each sample. The reaction mixture was then amplified by cycling with the following thermal profile: 98 °C for 30 s, 8 cycles of 98 °C for 10 s, 60 °C for 30 s and 68 °C for 60 s, followed by a 4 °C hold. The amplified NGS libraries were then size-selected with a 0.85 ratio of SPRIselect reagent:sample and eluted with 20 µL of Low EDTA TE buffer. Fragment size distribution was assessed by Agilent Bioanalyzer 2100 analysis using an Agilent DNA 1000 kit (Agilent Technologies, Catalog # 5067-1505). DNA concentration was determined using a Qubit 2.0 Fluorometer with a Qubit dsDNA BR Assay kit.

Table 4.1. Summary of NGS Library Preparation Parameters

Sample	Input DNA (ng)	Adapter (SI-IL1SP-12A)	Index Sequence	Number of PCR Cycles
Control-1	150	I6	GCCAAT(A)	8
PA1-1 (2 µM)	150	I15	ATGTCA(G)	8
PA1-1 (20 µM)	150	I14	AGTTCC(G)	8
Control-2	150	I5	ACAGTG(A)	8
PA1-2 (20 µM)	150	I16	CCGTCC(C)	8
Control-3	150	I18	GTCCGC(A)	8
PA1-3 (20 µM)	150	I19	GTGAAA(C)	8

Accel-NGS 1S Plus DNA Library Kit was used to prepare NGS libraries. Index sequence refers to a unique DNA sequence identifier in the adapter that allows for multiple libraries to be sequenced simultaneously in a sample lane.

4.3.5.2 *Quality Control of NGS Libraries*

Constructed libraries were quantified by Qubit 2.0 Fluorometer using a Qubit dsDNA BR assay kit and the fragment size distribution was determined by Agilent Bioanalyzer 2100 analysis using an Agilent DNA 1000 kit per manufacturer's directions. Briefly, 1 µL of the library sample was loaded onto a sample well containing 5 µL of Agilent DNA marker solution. The DNA 1000 Nanochip was then analyzed in an Agilent 2100 Bioanalyzer using the DNA 1000 Series II assay conditions. The results were analyzed using Agilent Technologies 2100 Bioanalyzer 2100 Expert Version B.02.08.SI648 (SR3) software. **Equation 4.3** was used to calculate the molar concentration of each NGS library.

$$Molar\ concentration\ of\ NGS\ library = [dsDNA] by\ Qubit\ 2.0\ (\text{ng}/\mu\text{L})\ x\ 10^6\ \mu\text{L}/\text{L}\ x \frac{\text{nmol}}{660\ \text{ng}} x \frac{1}{N} \quad \textbf{(4.3)}$$

where N is the average fragment size of the library determined by Agilent Qubit 1000 *kit*

The molar concentration of the constructed NGS libraries was also verified by the DNA Core Facility at the University of Missouri using a Qubit 3.0 Fluorometer (Life technologies) prior to high-throughput Illumina HiSeq 2500 sequencing.

4.3.5.3 *Illumina Next-generation Sequencing*

High-throughput Illumina sequencing of ·OH-treated HPV16/pUC18 plasmid samples in the presence or absence of **PA1** were performed by the DNA Core Facility at the University of Missouri using single-end, 51 base pair read length on an Illumina HiSeq 2500. Because the libraries derived from different ·OH-treated samples were constructed using different indexing adapters, three or six libraries were multiplexed per lane to achieve an appropriate signal-to-noise ratio (and cost) for the determination of **PA1**-protected regions in the HPV16/pUC18 genome.

4.3.6 Bioinformatics Analysis of NGS Results

4.3.6.1 *Sequence Alignment and Analysis of ·OH Cleavage Intensities*

Demultiplexing was subsequently performed on the sequencing reads to segregate the sequencing results into separate data sets. Unique sequencing reads were aligned to the HPV16/pUC18 reference genome using Bowtie 2 (version 2.2.9).[72] The reference genome consisted of HPV16 (GenBank Accession No. AF125673) cloned into the *Bam*HI cloning site of pUC18 (GenBank Accession No. L08752) with minor modifications in the pUC18 sequence. The aligned, or mapped reads in the resulting SAM file were then processed with CountNicks.pm, a Perl script written by Dr. Christopher Bottoms and Dr. Scott Givan of the MU Informatics Core (in consultation with our group) to determine the ·OH cleavage intensity at each nucleotide (refer to **Section 4.8.2** for detailed bioinformatics instructions). Because attack of hydroxyl radical results in the chemical destruction of a nucleotide base (refer to **Section 4.2.2**), the CountNicks.pm script assumes that the start of the sequence "read" for a nicked region occurs just 3' of the cleaved nucleotide.

4.4 RESULTS

We relied on hydroxyl radicals as a footprinting/cleavage reagent to generate short DNA fragments that were subsequently analyzed using massively parallel sequencing to determine the binding sites at single-nucleotide resolution of **PA1** to the negatively supercoiled HPV16/pUC18 plasmid. Although nicking of the negatively supercoiled DNA results in the relaxation of the supercoiled DNA, **PA1** binding sites that are observed may correspond to events that occurred under the supercoiled topology of the HPV16/pUC18 molecule.

4.4.1 Restriction Digests of HPV16/pUC18 Clones

In order to confirm that the correct plasmid DNA was amplified in transformed JM109 *E. coli* cells, a restriction digest using *Bam*HI restriction enzyme was performed on the purified DNA. *Bam*HI restriction digest of HPV16/pUC18 is expected to generate two DNA fragments with sizes of 2.7 kb and 7.9 kb, respectively. The restriction enzyme profile of the fragmented plasmid DNA separated by agarose gel electrophoresis is presented in **Figure 4.8**. The observed restriction products of approximately 2.7 kb and 7.9 kb suggest that the plasmid DNA purified from JM109 *E. coli* cells is HPV16/pUC18.

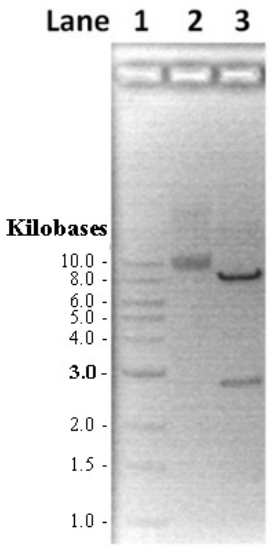


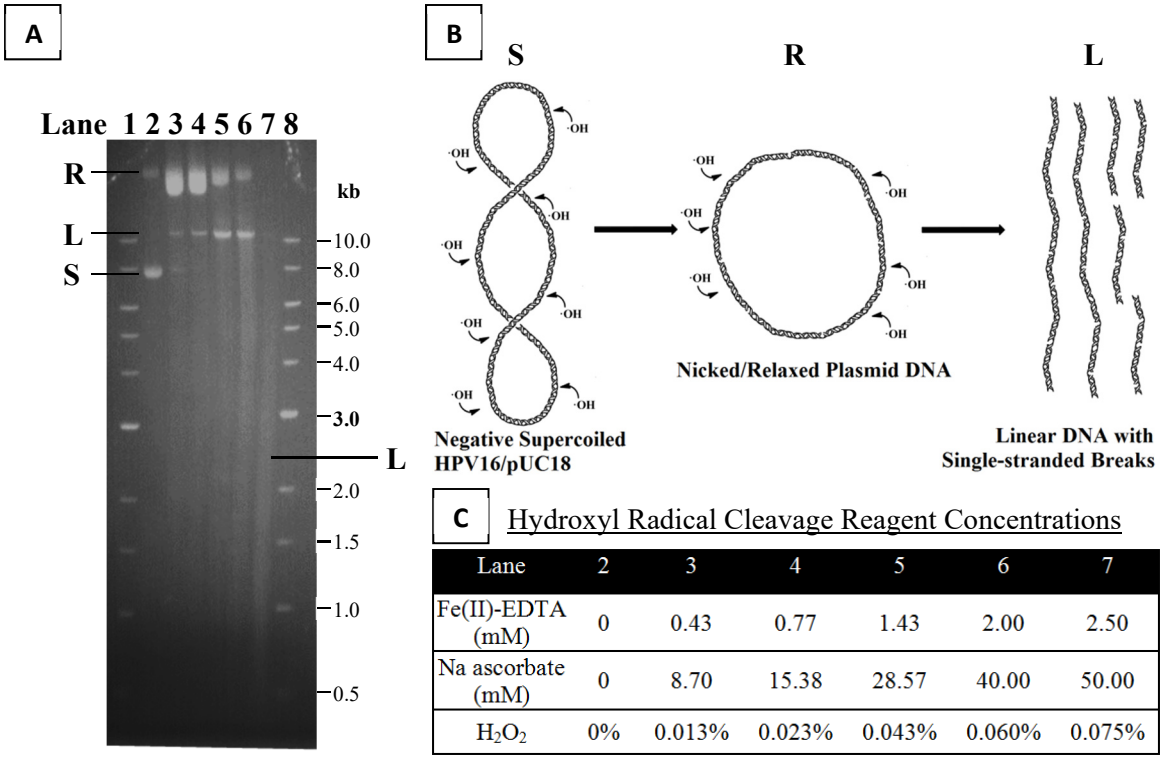
Figure 4.8. Agarose Gel Electrophoresis of Restriction Digest for HPV16/pUC18. *Bam*HI restriction digest performed on HPV16/pUC18 plasmid DNA purified from transformed JM109 *E. coli* cells. Lane 1 corresponds to 1 kb DNA molecular weight ladder with the expected size in kilobases given on the left of the picture. Lane 2 corresponds to uncut plasmid DNA (10.591 kb). Lane 3 corresponds to *Bam*HI restriction digest products (2.7 kb and 7.9 kb). Treatment of HPV16/pUC18 with *Bam*HI leads to the release of HPV16.

4.4.2 Optimization of Hydroxyl Radical Reagent Concentrations

Conventional library preparation for Illumina next-generation sequencing platform requires random fragmentation of genomic DNA into double-stranded fragment lengths of $\leq$ 600 bp by either enzymatic digestion or mechanical methods.[73] Here we have used hydroxyl radicals to produce short single-stranded fragments derived from two ·OH nicks in close proximity within the same DNA strand, allowing us to determine relative ligand-mediated protection in comparison to a control in addition to the specific location within the genome by Illumina next-generation sequencing. Since oxidative damage to DNA by hydroxyl radicals leads to single-stranded breaks along the DNA molecule, we employed the Accel-NGS 1S Plus DNA Library Kit for library preparation due to its compatibility with denatured, single-stranded, and nicked samples.

We first aimed at defining the appropriate working concentrations of the hydroxyl radical reagents required to generate a significant population of DNA fragments with lengths between 200-600 bp. To this aim, 5 μg of HPV16/pUC18 plasmid DNA were treated with various concentrations of the hydroxyl radical reagents at 20 °C for a fixed

reaction time of 5 min. The final concentrations of the cleavage reagents are tabulated in the inset of **Figure 4.9**. Agarose gel electrophoresis was used to evaluate the concentration dependence on the conversion of supercoiled plasmid DNA to nicked/relaxed DNA and linear DNA species. As the concentration of the hydroxyl radical cleavage reagents are increased, the supercoiled plasmid population decreases, giving rise to higher amounts of relaxed nicked/relaxed DNA which in turn is converted to linear DNA derived from double-stranded breaks (**Figure 4.9**). With higher concentrations of the cleavage reagents, the average fragment size of these linear DNA molecules decreases due to an increase in the population of fragments with shorter lengths. Agarose gel analysis also suggested that at hydroxyl radical reagent concentrations of 2.50 mM Fe(II)-EDTA, 50.00 mM sodium ascorbate and 0.075% H_2O_2, the supercoiled and relaxed plasmid species were completely depleted leading to a population of linear DNA molecules bearing fragment sizes in the range of approximately 0.7-5.0 kb.



C Hydroxyl Radical Cleavage Reagent Concentrations

Lane	2	3	4	5	6	7
Fe(II)-EDTA (mM)	0	0.43	0.77	1.43	2.00	2.50
Na ascorbate (mM)	0	8.70	15.38	28.57	40.00	50.00
H_2O_2	0%	0.013%	0.023%	0.043%	0.060%	0.075%

Figure 4.9. Concentration Dependence of Hydroxyl Radicals on Cleavage Reaction Products. Agarose gel electrophoresis analysis of hydroxyl radical concentration dependence on the conversion of plasmid DNA molecules from the intact, supercoiled form to the nicked/relaxed and linear forms. **(A)** Lanes 1 and 8 correspond to the 1 kb DNA ladder. Lane 2 corresponds to untreated HPV16/pUC18 plasmid control. Two bands represent the intact, supercoiled and relaxed forms of the plasmid DNA. Lanes 3-7 correspond to the ·OH-treated HPV16/pUC18 samples to different hydroxyl radical reagent concentrations (see Panel **C**). **(B)** Conversion scheme of plasmid DNA to linear DNA molecules; **R**: relaxed, **L**: linear and **S**: supercoiled.

These double-stranded fragments also contain single-stranded breaks produced by hydroxyl radical treatment. While agarose gel electrophoresis provides an estimate of dsDNA fragment size distribution, it fails to provide the fragment size distribution of ssDNA molecules that can be generated upon heat denaturation of the hydroxyl radical

products. These ssDNA fragments will ultimately serve as the template for Illumina library preparation using the Accel-NGS 1S Plus DNA Library Kit. As a result, fragment size distribution of ssDNA generated from ·OH-treated HPV16/pUC18 samples were estimated using a RNA 6000 Nanochip kit on an Agilent Bioanalyzer 2100. Although this kit is designed for sizing RNA fragments, the proprietary dye will also interact with nucleic acids from heat denatured DNA samples. We used this assay because it does not require radioactively labeled nucleic acids and it allows rapid analysis of small sample volumes. To validate the use of the RNA 6000 Nanochip kit for estimating ssDNA fragment size distribution, 1 µL (44 ng/µL) of a heat denatured, PCR amplified linear DNA fragment (368 nucleotides per strand) was analyzed. The calculated average size of the ssDNA fragment was estimated as 387 nucleotides with a range of 339 to 439 nucleotides. Furthermore, two smaller peaks of 463 and 493 nucleotides were observed, respectively (**Figure 4.10A**). While this assay overestimated the fragment size of heat denatured ssDNA, it provided an appropriate estimate of the fragment size length for the purposes of library construction. Heat denatured DNA cleavage products derived from various concentrations of the hydroxyl radical reagents were then analyzed using the RNA 6000 Nanochip assay. Sample electropherograms (**Figure 4.10B**) demonstrate that increasing the hydroxyl radical cleavage reagents led to an increase in the population of shorter length ssDNA fragments. Of interest, hydroxyl radical reagent concentrations of 2.50 mM Fe(II)-EDTA, 50.00 mM sodium ascorbate and 0.075 % H_2O_2 afforded the production of a significant population of ssDNA fragments with lengths between 200-600 bp, compatible with downstream NGS library preparation. These cleavage reaction conditions were used for ·OH footprinting experiments.

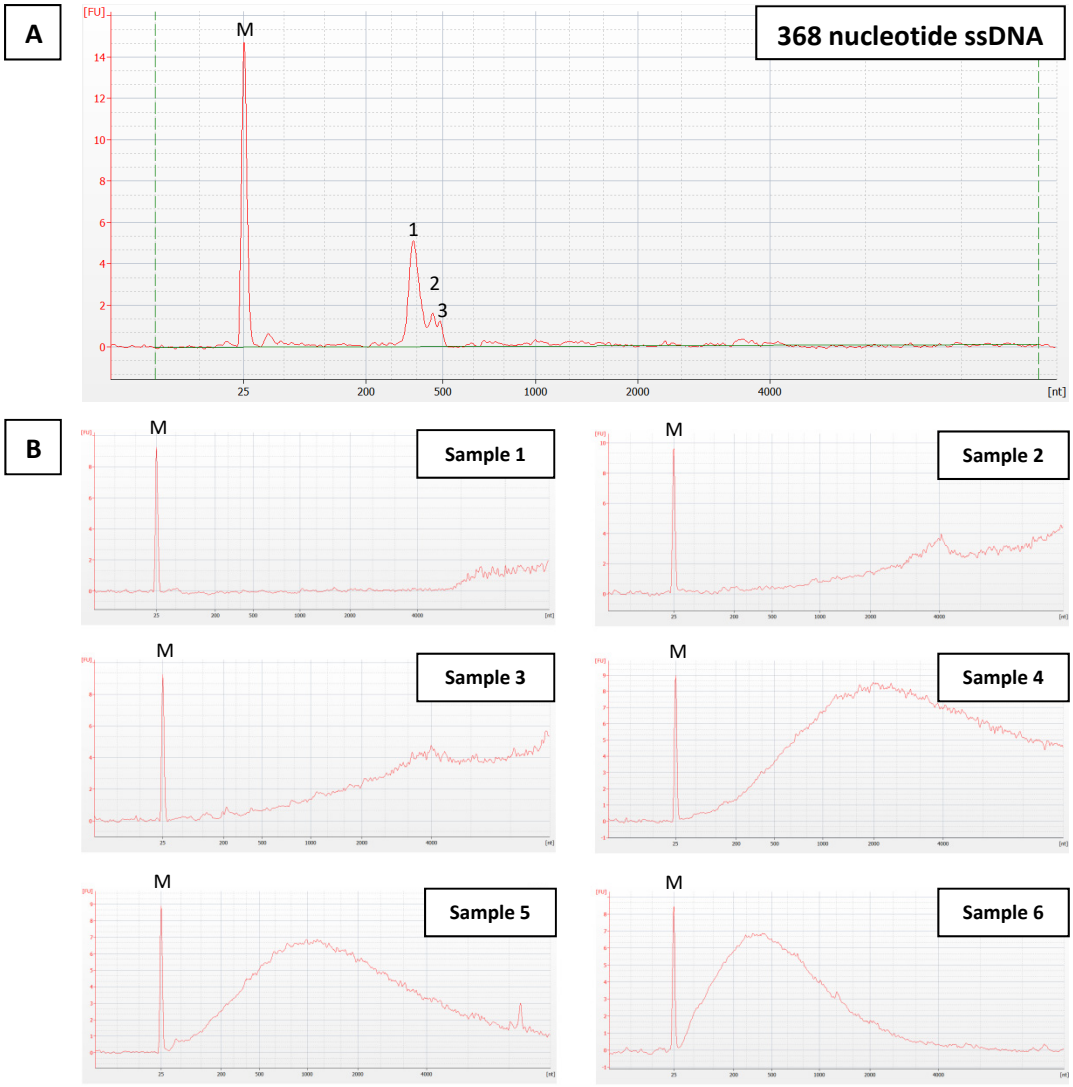


Figure 4.10. Single-Stranded DNA Fragment Size Distribution as a Function of Hydroxyl Radical Reagent Concentrations. RNA 6000 Nanochip assay analysis of heat denatured ssDNA fragment size distribution from HPV16/pUC18 DNA treated to various cleavage reagent concentrations. **(A)** A PCR-amplified dsDNA fragment was heat denatured to produce ssDNA fragments of 368 nucleotides. Analysis of the fragment size distribution revealed an overestimation of the fragment size. The calculated sizes were 387, 463 and 493 nucleotides for peak 1, 2 and 3, respectively. **(B)** Samples 1-6 correspond to HPV16/pUC18 DNA samples treated to increasing hydroxyl radical reagent concentrations. Sample 1: control; Sample 2: 0.43 mM Fe(II)-EDTA, 8.70 mM sodium ascorbate, 0.013 % H_2O_2; Sample 3: 0.77 mM Fe(II)-EDTA, 15.38 mM sodium ascorbate, 0.023 % H_2O_2; Sample 4: 1.43 mM Fe(II)-EDTA, 28.57 mM sodium ascorbate, 0.043 % H_2O_2; Sample 5: 2.00 mM Fe(II)-EDTA, 40.00 mM sodium ascorbate, 0.060 % H_2O_2; Sample 6: 2.50 mM Fe(II)-EDTA, 50.00 mM Na ascorbate, 0.075 % H_2O_2. M: RNA

4.4.3 Hydroxyl Radical (·OH) Footprinting

Samples of 5 µg of HPV16/pUC18 plasmid DNA were incubated in the presence and absence of **PA1** (2 and 20 µM) for at least 4 h at 37 °C in triplicate. After the incubation period, control and **PA1**-treated samples were subjected to hydroxyl radical cleavage reagents (2.5 mM Fe(II)-EDTA, 50 mM sodium ascorbate and 0.075 % H_2O_2) for 5 min at 20 °C. Fragmented DNA products were ethanol precipitated, followed by DNA concentration and sizing analysis.

4.4.3.1 *Determination of Fragment Size Distribution and DNA Concentration of ·OH-treated HPV16/pUC18 DNA*

Fragment size distribution for each ·OH footprinting reaction sample was estimated using the RNA 6000 Nanochip assay and their respective DNA concentrations were quantified by Qubit dsDNA BR assay. The estimated fragment size distribution for the nine ·OH footprinting samples is provided in **Figure 4.11**. The hydroxyl radical cleavage conditions used for the ·OH footprinting experiments generated a significant number of fragments of length (500 bp) compatible with subsequent NGS library construction. **Table 4.2** summarizes the DNA concentrations of the ·OH footprinting samples.

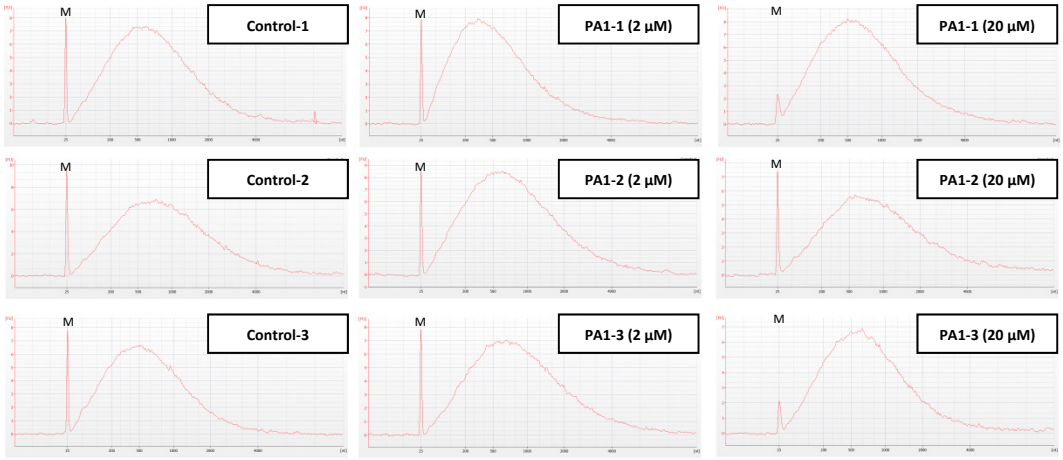


Figure 4.11. Single-Stranded DNA Fragment Size Distribution of ·OH Footprinting Samples. RNA 6000 Nanochip assay analysis of heat denatured ssDNA fragment size distribution for ·OH footprinting samples of **PA1** on HPV16/pUC18 DNA. Control and **PA1**-treated samples were each subjected to 2.5 mM Fe(II)-EDTA, 50 mM sodium ascorbate and 0.075 % H_2O_2 for 5 min at 20 °C. M: RNA marker.

Table 4.2. DNA Concentrations of ·OH Footprinting Samples by Qubit dsDNA BR Assay

Trial 1		Trial 2		Trial 3	
Sample	**[dsDNA] (ng/µL)**	**Sample**	**[dsDNA] (ng/µL)**	**Sample**	**[dsDNA] (ng/µL)**
Control-1	185	**Control-2**	214	**Control-3**	190
PA1-1 (2 µM)	122	**PA1-2 (2 µM)**	243	**PA1-3 (2 µM)**	237
PA1-1 (20 µM)	157	**PA1-2 (20 µM)**	174	**PA1-3 (20 µM)**	258

4.4.4 Illumina Next-generation Sequencing

4.4.4.1 *Quality Control of NGS Libraries*

DNA concentration was measured by the Qubit dsDNA BR assay and fragment size distribution was determined by the Agilent DNA 1000 assay for the constructed libraries. The molar concentrations of the constructed libraries were calculated according to **Equation 4.3** using the DNA concentration measured by Qubit and the average fragment size of the library determined by Bioanalyzer (**Table 4.3**). Fragment size distribution profiles of the prepared NGS libraries are provided in **Figure 4.12**.

Table 4.3. ·OH Footprinting of **PA1** on HPV16/pUC18 - Summary of NGS Library Information.

Sample	Adapter (SI-IL1SP-12A)	Index Sequence	Qubit 2.0 (ng/µL)	MU Qubit 3.0 (ng/µL)	Bioanalyzer Average Size (bp)	Molar Concentration (nM)
Control-1	I6	GCCAAT(A)	18.90	*ND*	458	62.52
PA1-1 (2 µM)	I15	ATGTCA(G)	23.30	*ND*	449	78.63
PA1-1 (20 µM)	I14	AGTTCC(G)	22.40	*ND*	467	72.68
Control-2	I5	ACAGTG(A)	19.50	34.65	456	64.79
PA1-2 (20 µM)	I16	CCGTCC(C)	12.40	20.50	451	41.66
Control-3	I18	GTCCGC(A)	23.20	39.40	428	82.13
PA1-3 (20 µM)	I19	GTGAAA(C)	5.50	8.615	450	18.52

Accel-NGS 1S Plus DNA Library Kit was used to prepare NGS libraries. Index sequence refers to a unique DNA sequence identifier in the adapter that allows for multiple libraries to be sequenced simultaneously in a single sample lane and be identified and grouped together properly *via* bioinformatics processing of the sequence data. Molar concentrations were calculated according to **Equation 4.3** using the Qubit 2.0 data. *ND* – not determined.

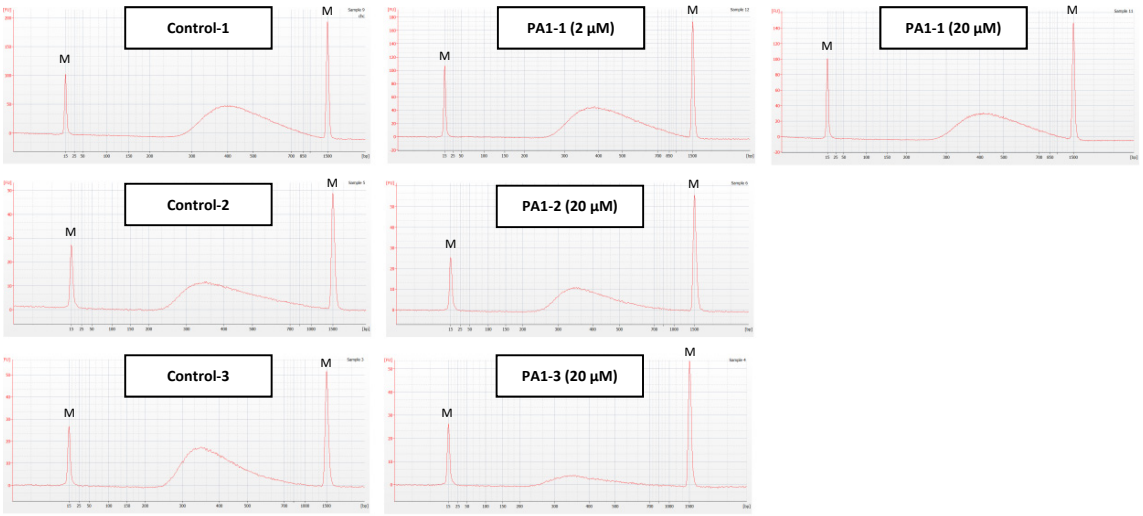


Figure 4.12. Fragment Size Distribution of Illumina NGS Libraries. Agilent DNA1000 assay analysis prepared NGS libraries.

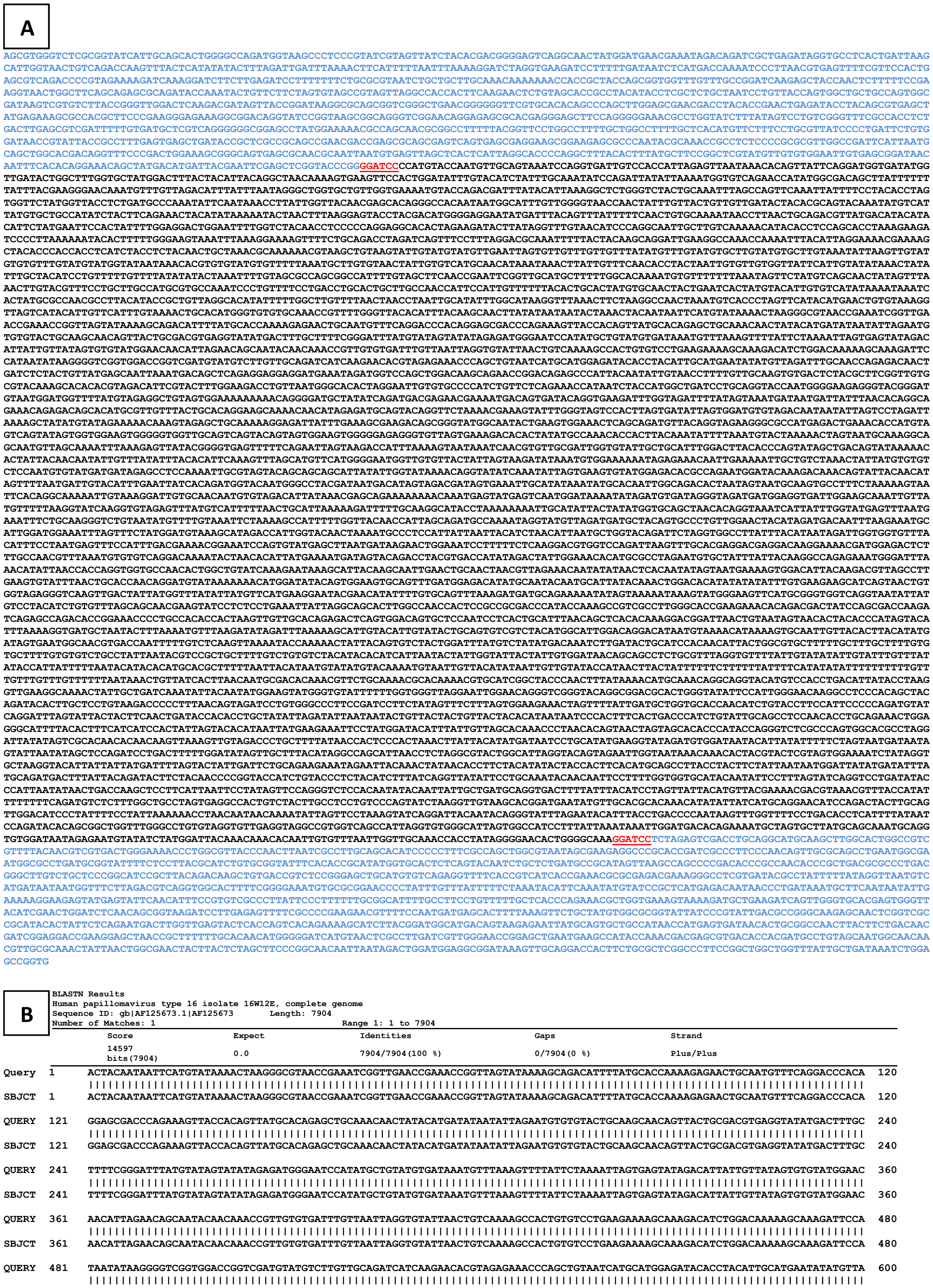
4.4.4.2 *Illumina HiSeq2500 NGS*

Illumina NGS libraries of ·OH footprinting samples were sequenced using single-end, 51 base pair read length on an Illumina HiSeq 2500. Samples were multiplexed by pooling three or six libraries per single lane for a theoretical 50 or 25 million reads per library, respectively. **Table 4.4** tabulates the number of reads obtained for each library.

4.4.4.3 *DNA Sequence Validation of HPV16/pUC18 Plasmid using Illumina HiSeq 2500 Sequencing Data*

De novo DNA sequencing of the HPV16/pUC18 plasmid further verified the identity of the amplified plasmid. The raw .fastq file from the ·OH-treated HPV16/pUC18 plasmid sample in the absence of **PA1** was used to manually aligned the sequence reads into contiguous sequences. The plasmid sequence determined matched the HPV16 sequence (GenBank Accession No. AF125673) cloned into the *Bam*HI cloning site of pUC18 (GenBank Accession No. L08752). BLASTN[74] alignment of determined HPV16 and pUC18 sequences are shown in **Figure 4.13B & C**. However, a deletion of a single cytidine at nucleotide position 220 and a mutation from cytosine to thymine at nucleotide

position 2063 were observed in the pUC18 sequence. The determined sequence for this HPV18 variant was used as the reference genome to map the sequencing reads from ·OH-treated HPV16/pUC18 samples (**Figure 4.13A**).



```
QUERY  601   GATTTGCAACCAGAGACAACTGATCTCTACTGTTATGAGCAATTAAATGACAGCTCAGAGGAGGAGGATGAAATAGATGGTCCAGCTGGACAAGCAGAACCGGACAGAGCCCATTACAAT  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  601   GATTTGCAACCAGAGACAACTGATCTCTACTGTTATGAGCAATTAAATGACAGCTCAGAGGAGGAGGATGAAATAGATGGTCCAGCTGGACAAGCAGAACCGGACAGAGCCCATTACAAT  720

QUERY  721   ATTGTAACCTTTTGTTGCAAGTGTGACTCTACGCTTCGGTTGTGCGTACAAAGCACACACGTAGACATTCGTACTTTGGAAGACCTGTTAATGGGCACACTAGGAATTGTGTGCCCCATC  840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  721   ATTGTAACCTTTTGTTGCAAGTGTGACTCTACGCTTCGGTTGTGCGTACAAAGCACACACGTAGACATTCGTACTTTGGAAGACCTGTTAATGGGCACACTAGGAATTGTGTGCCCCATC  840

QUERY  841   TGTTCTCAGAAACCATAATCTACCATGGCTGATCCTGCAGGTACCAATGGGGAAGAGGGTACGGGATGTAATGGATGGTTTTATGTAGAGGCTGTAGTGGAAAAAAAAACAGGGGATGCT  960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  841   TGTTCTCAGAAACCATAATCTACCATGGCTGATCCTGCAGGTACCAATGGGGAAGAGGGTACGGGATGTAATGGATGGTTTTATGTAGAGGCTGTAGTGGAAAAAAAAACAGGGGATGCT  960

QUERY  961   ATATCAGATGACGAGAACGAAAATGACAGTGATACAGGTGAAGATTTGGTAGATTTTATAGTAAATGATAATGATTATTTAACACAGGCAGAAACAGAGACAGCACATGCGTTGTTTACT 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  961   ATATCAGATGACGAGAACGAAAATGACAGTGATACAGGTGAAGATTTGGTAGATTTTATAGTAAATGATAATGATTATTTAACACAGGCAGAAACAGAGACAGCACATGCGTTGTTTACT 1080

QUERY 1081   GCACAGGAAGCAAAACAACATAGAGATGCAGTACAGGTTCTAAAACGAAAGTATTTGGGTAGTCCACTTAGTGATATTAGTGGATGTGTAGACAATAATATTAGTCCTAGATTAAAAGCT 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 1081   GCACAGGAAGCAAAACAACATAGAGATGCAGTACAGGTTCTAAAACGAAAGTATTTGGGTAGTCCACTTAGTGATATTAGTGGATGTGTAGACAATAATATTAGTCCTAGATTAAAAGCT 1200

QUERY 1201   ATATGTATAGAAAAACAAAGTAGAGCTGCAAAAAGGAGATTATTTGAAAGCGAAGACAGCGGGTATGGCAATACTGAAGTGGAAACTCAGCAGATGTTACAGGTAGAAGGGCGCCATGAG 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 1201   ATATGTATAGAAAAACAAAGTAGAGCTGCAAAAAGGAGATTATTTGAAAGCGAAGACAGCGGGTATGGCAATACTGAAGTGGAAACTCAGCAGATGTTACAGGTAGAAGGGCGCCATGAG 1320

QUERY 1321   ACTGAAACACCATGTAGTCAGTATAGTGGTGGAAGTGGGGGTGGTTGCAGTCAGTACAGTAGTGGAAGTGGGGGAGAGGGTGTTAGTGAAAGACACACTATATGCCAAACACCACTTACA 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 1321   ACTGAAACACCATGTAGTCAGTATAGTGGTGGAAGTGGGGGTGGTTGCAGTCAGTACAGTAGTGGAAGTGGGGGAGAGGGTGTTAGTGAAAGACACACTATATGCCAAACACCACTTACA 1440

QUERY 1441   AATATTTTAAATGTACTAAAAACTAGTAATGCAAAGGCAGCAATGTTAGCAAAATTTAAAGAGTTATACGGGGTGAGTTTTTCAGAATTAGTAAGACCATTTAAAAGTAATAAATCAACG 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 1441   AATATTTTAAATGTACTAAAAACTAGTAATGCAAAGGCAGCAATGTTAGCAAAATTTAAAGAGTTATACGGGGTGAGTTTTTCAGAATTAGTAAGACCATTTAAAAGTAATAAATCAACG 1560

QUERY 1561   TGTTGCGATTGGTGTATTGCTGCATTTGGACTTACACCCAGTATAGCTGACAGTATAAAAACACTATTACAACAATATTGTTTATATTTACACATTCAAAGTTTAGCATGTTCATGGGGA 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 1561   TGTTGCGATTGGTGTATTGCTGCATTTGGACTTACACCCAGTATAGCTGACAGTATAAAAACACTATTACAACAATATTGTTTATATTTACACATTCAAAGTTTAGCATGTTCATGGGGA 1680

QUERY 1681   ATGGTTGTGTTACTATTAGTAAGATATAAATGTGGAAAAAATAGAGAAACAATTGAAAAATTGCTGTCTAAACTATTATGTGTGTCTCCAATGTGTATGATGATAGAGCCTCCAAAATTG 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 1681   ATGGTTGTGTTACTATTAGTAAGATATAAATGTGGAAAAAATAGAGAAACAATTGAAAAATTGCTGTCTAAACTATTATGTGTGTCTCCAATGTGTATGATGATAGAGCCTCCAAAATTG 1800

QUERY 1801   CGTAGTACAGCAGCAGCATTATATTGGTATAAAACAGGTATATCAAATATTAGTGAAGTGTATGGAGACACGCCAGAATGGATACAAAGACAAACAGTATTACAACATAGTTTTAATGAT 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 1801   CGTAGTACAGCAGCAGCATTATATTGGTATAAAACAGGTATATCAAATATTAGTGAAGTGTATGGAGACACGCCAGAATGGATACAAAGACAAACAGTATTACAACATAGTTTTAATGAT 1920

QUERY 1921   TGTACATTTGAATTATCACAGATGGTACAATGGGCCTACGATAATGACATAGTAGACGATAGTGAAATTGCATATAAATATGCACAATTGGCAGACACTAATAGTAATGCAAGTGCCTTT 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 1921   TGTACATTTGAATTATCACAGATGGTACAATGGGCCTACGATAATGACATAGTAGACGATAGTGAAATTGCATATAAATATGCACAATTGGCAGACACTAATAGTAATGCAAGTGCCTTT 2040

QUERY 2041   CTAAAAAGTAATTCACAGGCAAAAATTGTAAAGGATTGTGCAACAATGTGTAGACATTATAAACGAGCAGAAAAAAAACAAATGAGTATGAGTCAATGGATAAAATATAGATGTGATAGG 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 2041   CTAAAAAGTAATTCACAGGCAAAAATTGTAAAGGATTGTGCAACAATGTGTAGACATTATAAACGAGCAGAAAAAAAACAAATGAGTATGAGTCAATGGATAAAATATAGATGTGATAGG 2160

QUERY 2161   GTAGATGATGGAGGTGATTGGAAGCAAATTGTTATGTTTTTAAGGTATCAAGGTGTAGAGTTTATGTCATTTTTAACTGCATTAAAAAGATTTTTGCAAGGCATACCTAAAAAAAATTGC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 2161   GTAGATGATGGAGGTGATTGGAAGCAAATTGTTATGTTTTTAAGGTATCAAGGTGTAGAGTTTATGTCATTTTTAACTGCATTAAAAAGATTTTTGCAAGGCATACCTAAAAAAAATTGC 2280

QUERY 2281   ATATTACTATATGGTGCAGCTAACACAGGTAAATCATTATTTGGTATGAGTTTAATGAAATTTCTGCAAGGGTCTGTAATATGTTTTGTAAATTCTAAAAGCCATTTTTGGTTACAACCA 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 2281   ATATTACTATATGGTGCAGCTAACACAGGTAAATCATTATTTGGTATGAGTTTAATGAAATTTCTGCAAGGGTCTGTAATATGTTTTGTAAATTCTAAAAGCCATTTTTGGTTACAACCA 2400

QUERY 2401   TTAGCAGATGCCAAAATAGGTATGTTAGATGATGCTACAGTGCCCTGTTGGAACTACATAGATGACAATTTAAGAAATGCATTGGATGGAAATTTAGTTTCTATGGATGTAAAGCATAGA 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 2401   TTAGCAGATGCCAAAATAGGTATGTTAGATGATGCTACAGTGCCCTGTTGGAACTACATAGATGACAATTTAAGAAATGCATTGGATGGAAATTTAGTTTCTATGGATGTAAAGCATAGA 2520

QUERY 2521   CCATTGGTACAACTAAAATGCCCTCCATTATTAATTACATCTAACATTAATGCTGGTACAGATTCTAGGTGGCCTTATTTACATAATAGATTGGTGGTGTTTACATTTCCTAATGAGTTT 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 2521   CCATTGGTACAACTAAAATGCCCTCCATTATTAATTACATCTAACATTAATGCTGGTACAGATTCTAGGTGGCCTTATTTACATAATAGATTGGTGGTGTTTACATTTCCTAATGAGTTT 2640

QUERY 2641   CCATTTGACGAAAACGGAAATCCAGTGTATGAGCTTAATGATAAGAACTGGAAATCCTTTTTCTCAAGGACGTGGTCCAGATTAAGTTTGCACGAGGACGAGGACAAGGAAAACGATGGA 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 2641   CCATTTGACGAAAACGGAAATCCAGTGTATGAGCTTAATGATAAGAACTGGAAATCCTTTTTCTCAAGGACGTGGTCCAGATTAAGTTTGCACGAGGACGAGGACAAGGAAAACGATGGA 2760

QUERY 2761   GACTCTTTGCCAACGTTTAAATGTGTGTCAGGACAAAATACTAACACATTATGAAAATGATAGTACAGACCTACGTGACCATATAGACTATTGGAAACACATGCGCCTAGAATGTGCTAT 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 2761   GACTCTTTGCCAACGTTTAAATGTGTGTCAGGACAAAATACTAACACATTATGAAAATGATAGTACAGACCTACGTGACCATATAGACTATTGGAAACACATGCGCCTAGAATGTGCTAT 2880

QUERY 2881   TTATTACAAGGCCAGAGAAATGGGATTTAAACATATTAACCACCAGGTGGTGCCAACACTGGCTGTATCAAAGAATAAAGCATTACAAGCAATTGAACTGCAACTAACGTTAGAAACAAT 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 2881   TTATTACAAGGCCAGAGAAATGGGATTTAAACATATTAACCACCAGGTGGTGCCAACACTGGCTGTATCAAAGAATAAAGCATTACAAGCAATTGAACTGCAACTAACGTTAGAAACAAT 3000

QUERY 3001   ATATAACTCACAATATAGTAATGAAAAGTGGACATTACAAGACGTTAGCCTTGAAGTGTATTTAACTGCACCAACAGGATGTATAAAAAAACATGGATATACAGTGGAAGTGCAGTTTGA 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 3001   ATATAACTCACAATATAGTAATGAAAAGTGGACATTACAAGACGTTAGCCTTGAAGTGTATTTAACTGCACCAACAGGATGTATAAAAAAACATGGATATACAGTGGAAGTGCAGTTTGA 3120

QUERY 3121   TGGAGACATATGCAATACAATGCATTATACAAACTGGACACATATATATATTTGTGAAGAAGCATCAGTAACTGTGGTAGAGGGTCAAGTTGACTATTATGGTTTATATTATGTTCATGA 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 3121   TGGAGACATATGCAATACAATGCATTATACAAACTGGACACATATATATATTTGTGAAGAAGCATCAGTAACTGTGGTAGAGGGTCAAGTTGACTATTATGGTTTATATTATGTTCATGA 3240

QUERY 3241   AGGAATACGAACATATTTTGTGCAGTTTAAAGATGATGCAGAAAAATATAGTAAAAATAAAGTATGGGAAGTTCATGCGGGTGGTCAGGTAATATTATGTCCTACATCTGTGTTTAGCAG 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 3241   AGGAATACGAACATATTTTGTGCAGTTTAAAGATGATGCAGAAAAATATAGTAAAAATAAAGTATGGGAAGTTCATGCGGGTGGTCAGGTAATATTATGTCCTACATCTGTGTTTAGCAG 3360

QUERY 3361   CAACGAAGTATCCTCTCCTGAAATTATTAGGCAGCACTTGGCCAACCACTCCGCCGCGACCCATACCAAAGCCGTCGCCTTGGGCACCGAAGAAACACAGACGACTATCCAGCGACCAAG 3480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 3361   CAACGAAGTATCCTCTCCTGAAATTATTAGGCAGCACTTGGCCAACCACTCCGCCGCGACCCATACCAAAGCCGTCGCCTTGGGCACCGAAGAAACACAGACGACTATCCAGCGACCAAG 3480

QUERY 3481   ATCAGAGCCAGACACCGGAAACCCCTGCCACACCACCTAAGTTGTTGCACAGAGACTCAGTGGACAGTGCTCCAATCCTCACTGCATTTAACAGCTCACACAAAGGACGGATTAACTGTAA 3600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 3481   ATCAGAGCCAGACACCGGAAACCCCTGCCACACCACCTAAGTTGTTGCACAGAGACTCAGTGGACAGTGCTCCAATCCTCACTGCATTTAACAGCTCACACAAAGGACGGATTAACTGTAA 3600

QUERY 3601   TAGTAACACTACACCCATAGTACATTTAAAAGGTGATGCTAATACTTTAAAATGTTTAAGATATAGATTTAAAAAGCATTGTACATTGTATACTGCAGTGTCGTCTACATGGCATTGGAC 3720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 3601   TAGTAACACTACACCCATAGTACATTTAAAAGGTGATGCTAATACTTTAAAATGTTTAAGATATAGATTTAAAAAGCATTGTACATTGTATACTGCAGTGTCGTCTACATGGCATTGGAC 3720

QUERY 3721   AGGACATAATGTAAAACATAAAAGTGCAATTGTTACACTTACATATGATAGTGAATGGCAACGTGACCAATTTTTGTCTCAAGTTAAAATACCAAAAACTATTACAGTGTCTACTGGATT 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 3721   AGGACATAATGTAAAACATAAAAGTGCAATTGTTACACTTACATATGATAGTGAATGGCAACGTGACCAATTTTTGTCTCAAGTTAAAATACCAAAAACTATTACAGTGTCTACTGGATT 3840

QUERY 3841   TATGTCTATATGACAAATCTTGATACTGCATCCACAACATTACTGGCGTGCTTTTTGCTTTGCTTTGTGTGCTTTTGTGTGTCTGCCTATTAATACGTCCGCTGCTTTTGTCTGTGTCT 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT 3841   TATGTCTATATGACAAATCTTGATACTGCATCCACAACATTACTGGCGTGCTTTTTGCTTTGCTTTGTGTGCTTTTGTGTGTCTGCCTATTAATACGTCCGCTGCTTTTGTCTGTGTCT 3960
```

```
QUERY  3961  ACATACACATCATTAATACTATTGGTATTACTATTGTGGATAACAGCAGCCTCTGCGTTTAGGTGTTTTATTGTATATATTGTATTTGTTTATATACCATTATTTTTAATACATACACAT  4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  3961  ACATACACATCATTAATACTATTGGTATTACTATTGTGGATAACAGCAGCCTCTGCGTTTAGGTGTTTTATTGTATATATTGTATTTGTTTATATACCATTATTTTTAATACATACACAT  4080

QUERY  4081  GCACGCTTTTTAATTACATAATGTATATGTACAAAATGTAATTGTTACATATAATTGTTGTATACCATAACTTACTATTTTTTCTTTTTTATTTTCATATATATTTTTTTTTTGTTTGTT  4200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  4081  GCACGCTTTTTAATTACATAATGTATATGTACAAAATGTAATTGTTACATATAATTGTTGTATACCATAACTTACTATTTTTTCTTTTTTATTTTCATATATATTTTTTTTTTGTTTGTT  4200

QUERY  4201  TGTTTGTTTTTTAATAAACTGTTATCACTTAACAATGCGACACAAACGTTCTGCAAAACGCACAAAACGTGCATCGGCTACCCAACTTTATAAAACATGCAAACAGGCAGGTACATGTCC  4320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  4201  TGTTTGTTTTTTAATAAACTGTTATCACTTAACAATGCGACACAAACGTTCTGCAAAACGCACAAAACGTGCATCGGCTACCCAACTTTATAAAACATGCAAACAGGCAGGTACATGTCC  4320

QUERY  4321  ACCTGACATTATACCTAAGGTTGAAGGCAAAACTATTGCTGATCAAATATTACAATATGGAAGTATGGGTGTATTTTTTGGTGGGTTAGGAATTGGAACAGGGTCGGGTACAGGCGGACG  4440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  4321  ACCTGACATTATACCTAAGGTTGAAGGCAAAACTATTGCTGATCAAATATTACAATATGGAAGTATGGGTGTATTTTTTGGTGGGTTAGGAATTGGAACAGGGTCGGGTACAGGCGGACG  4440

QUERY  4441  CACTGGGTATATTCCATTGGGAACAAGGCCTCCCACAGCTACAGATACACTTGCTCCTGTAAGACCCCCTTTAACAGTAGATCCTGTGGGCCCTTCCGATCCTTCTATAGTTTCTTTAGT  4560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  4441  CACTGGGTATATTCCATTGGGAACAAGGCCTCCCACAGCTACAGATACACTTGCTCCTGTAAGACCCCCTTTAACAGTAGATCCTGTGGGCCCTTCCGATCCTTCTATAGTTTCTTTAGT  4560

QUERY  4561  GGAAGAAACTAGTTTTATTGATGCTGGTGCACCAACATCTGTACCTTCCATTCCCCCAGATGTATCAGGATTTAGTATTACTACTTCAACTGATACCACACCTGCTATATTAGATATTAA  4680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  4561  GGAAGAAACTAGTTTTATTGATGCTGGTGCACCAACATCTGTACCTTCCATTCCCCCAGATGTATCAGGATTTAGTATTACTACTTCAACTGATACCACACCTGCTATATTAGATATTAA  4680

QUERY  4681  TAATACTGTTACTACTGTTACTACACATAATAATCCCACTTTCACTGACCCATCTGTATTGCAGCCTCCAACACCTGCAGAAACTGGAGGGCATTTTACACTTTCATCATCCACTATTAG  4800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  4681  TAATACTGTTACTACTGTTACTACACATAATAATCCCACTTTCACTGACCCATCTGTATTGCAGCCTCCAACACCTGCAGAAACTGGAGGGCATTTTACACTTTCATCATCCACTATTAG  4800

QUERY  4801  TACACATAATTATGAAGAAATTCCTATGGATACATTTATTGTTAGCACAAACCCTAACACAGTAACTAGTAGCACACCCATACCAGGGTCTCGCCCAGTGGCACGCCTAGGATTATATAG  4920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  4801  TACACATAATTATGAAGAAATTCCTATGGATACATTTATTGTTAGCACAAACCCTAACACAGTAACTAGTAGCACACCCATACCAGGGTCTCGCCCAGTGGCACGCCTAGGATTATATAG  4920

QUERY  4921  TCGCACAACACAACAAGTTAAAGTTGTAGACCCTGCTTTTATAACCACTCCCACTAAACTTATTACATATGATAATCCTGCATATGAAGGTATAGATGTGGATAATACATTATATTTTTC  5040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  4921  TCGCACAACACAACAAGTTAAAGTTGTAGACCCTGCTTTTATAACCACTCCCACTAAACTTATTACATATGATAATCCTGCATATGAAGGTATAGATGTGGATAATACATTATATTTTTC  5040

QUERY  5041  TAGTAATGATAATAGTATTAATATAGCTCCAGATCCTGACTTTTTGGATATAGTTGCTTTACATAGGCCAGCATTAACCTCTAGGCGTACTGGCATTAGGTACAGTAGAATTGGTAATAA  5160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  5041  TAGTAATGATAATAGTATTAATATAGCTCCAGATCCTGACTTTTTGGATATAGTTGCTTTACATAGGCCAGCATTAACCTCTAGGCGTACTGGCATTAGGTACAGTAGAATTGGTAATAA  5160

QUERY  5161  ACAAACACTACGTACTCGTAGTGGAAAATCTATAGGTGCTAAGGTACATTATTATTATGATTTTAGTACTATTGATTCTGCAGAAGAAATAGAATTACAAACTATAACACCTTCTACATA  5280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  5161  ACAAACACTACGTACTCGTAGTGGAAAATCTATAGGTGCTAAGGTACATTATTATTATGATTTTAGTACTATTGATTCTGCAGAAGAAATAGAATTACAAACTATAACACCTTCTACATA  5280

QUERY  5281  TACTACCACTTCACATGCAGCCTTACCTACTTCTATTAATAATGGATTATATGATATTTATGCAGATGACTTTATTACAGATACTTCTACAACCCCGGTACCATCTGTACCCTCTACATC  5400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  5281  TACTACCACTTCACATGCAGCCTTACCTACTTCTATTAATAATGGATTATATGATATTTATGCAGATGACTTTATTACAGATACTTCTACAACCCCGGTACCATCTGTACCCTCTACATC  5400

QUERY  5401  TTTATCAGGTTATATTCCTGCAAATACAACAATTCCTTTTGGTGGTGCATACAATATTCCTTTAGTATCAGGTCCTGATATACCCATTAATATAACTGACCAAGCTCCTTCATTAATTCC  5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  5401  TTTATCAGGTTATATTCCTGCAAATACAACAATTCCTTTTGGTGGTGCATACAATATTCCTTTAGTATCAGGTCCTGATATACCCATTAATATAACTGACCAAGCTCCTTCATTAATTCC  5520

QUERY  5521  TATAGTTCCAGGGTCTCCACAATATACAATTATTGCTGATGCAGGTGACTTTTATTTACATCCTAGTTATTACATGTTACGAAAACGACGTAAACGTTTACCATATTTTTTTTCAGATGT  5640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  5521  TATAGTTCCAGGGTCTCCACAATATACAATTATTGCTGATGCAGGTGACTTTTATTTACATCCTAGTTATTACATGTTACGAAAACGACGTAAACGTTTACCATATTTTTTTTCAGATGT  5640

QUERY  5641  CTCTTTGGCTGCCTAGTGAGGCCACTGTCTACTTGCCTCCTGTCCCAGTATCTAAGGTTGTAAGCACGGATGAATATGTTGCACGCACAAACATATATTATCATGCAGGAACATCCAGAC  5760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  5641  CTCTTTGGCTGCCTAGTGAGGCCACTGTCTACTTGCCTCCTGTCCCAGTATCTAAGGTTGTAAGCACGGATGAATATGTTGCACGCACAAACATATATTATCATGCAGGAACATCCAGAC  5760

QUERY  5761  TACTTGCAGTTGGACATCCCTATTTTCCTATTAAAAAACCTAACAATAACAAAATATTAGTTCCTAAAGTATCAGGATTACAATACAGGGTATTTAGAATACATTTACCTGACCCCAATA  5880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  5761  TACTTGCAGTTGGACATCCCTATTTTCCTATTAAAAAACCTAACAATAACAAAATATTAGTTCCTAAAGTATCAGGATTACAATACAGGGTATTTAGAATACATTTACCTGACCCCAATA  5880

QUERY  5881  AGTTTGGTTTTCCTGACACCTCATTTTATAATCCAGATACACAGCGGCTGGTTTGGGCCTGTGTAGGTGTTGAGGTAGGCCGTGGTCAGCCATTAGGTGTGGGCATTAGTGGCCATCCTT  6000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  5881  AGTTTGGTTTTCCTGACACCTCATTTTATAATCCAGATACACAGCGGCTGGTTTGGGCCTGTGTAGGTGTTGAGGTAGGCCGTGGTCAGCCATTAGGTGTGGGCATTAGTGGCCATCCTT  6000

QUERY  6001  TATTAAATAAATTGGATGACACAGAAAATGCTAGTGCTTATGCAGCAAATGCAGGTGTGGATAATAGAGAATGTATATCTATGGATTACAAACAAACACAATTGTGTTTAATTGGTTGCA  6120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  6001  TATTAAATAAATTGGATGACACAGAAAATGCTAGTGCTTATGCAGCAAATGCAGGTGTGGATAATAGAGAATGTATATCTATGGATTACAAACAAACACAATTGTGTTTAATTGGTTGCA  6120

QUERY  6121  AACCACCTATAGGGGAACACTGGGGCAAAGGATCCCCATGTACCAATGTTGCAGTAAATCCAGGTGATTGTCCACCATTAGAGTTAATAAACACAGTTATTCAGGATGGTGATATGGTTG  6240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  6121  AACCACCTATAGGGGAACACTGGGGCAAAGGATCCCCATGTACCAATGTTGCAGTAAATCCAGGTGATTGTCCACCATTAGAGTTAATAAACACAGTTATTCAGGATGGTGATATGGTTG  6240

QUERY  6241  ATACTGGCTTTGGTGCTATGGACTTTACTACATTACAGGCTAACAAAAGTGAAGTTCCACTGGATATTTGTACATCTATTTGCAAATATCCAGATTATATTAAAATGGTGTCAGAACCAT  6360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  6241  ATACTGGCTTTGGTGCTATGGACTTTACTACATTACAGGCTAACAAAAGTGAAGTTCCACTGGATATTTGTACATCTATTTGCAAATATCCAGATTATATTAAAATGGTGTCAGAACCAT  6360

QUERY  6361  ATGGCGACAGCTTATTTTTTTATTTACGAAGGGAACAAATGTTTGTTAGACATTTATTTAATAGGGCTGGTGCTGTTGGTGAAAATGTACCAGACGATTTATACATTAAAGGCTCTGGGT  6480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  6361  ATGGCGACAGCTTATTTTTTTATTTACGAAGGGAACAAATGTTTGTTAGACATTTATTTAATAGGGCTGGTGCTGTTGGTGAAAATGTACCAGACGATTTATACATTAAAGGCTCTGGGT  6480

QUERY  6481  CTACTGCAAATTTAGCCAGTTCAAATTATTTTCCTACACCTAGTGGTTCTATGGTTACCTCTGATGCCCAAATATTCAATAAACCTTATTGGTTACAACGAGCACAGGGCCACAATAATG  6600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  6481  CTACTGCAAATTTAGCCAGTTCAAATTATTTTCCTACACCTAGTGGTTCTATGGTTACCTCTGATGCCCAAATATTCAATAAACCTTATTGGTTACAACGAGCACAGGGCCACAATAATG  6600

QUERY  6601  GCATTTGTTGGGGTAACCAACTATTTGTTACTGTTGTTGATACTACACGCAGTACAAATATGTCATTATGTGCTGCCATATCTACTTCAGAAACTACATATAAAAATACTAACTTTAAGG  6720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  6601  GCATTTGTTGGGGTAACCAACTATTTGTTACTGTTGTTGATACTACACGCAGTACAAATATGTCATTATGTGCTGCCATATCTACTTCAGAAACTACATATAAAAATACTAACTTTAAGG  6720

QUERY  6721  AGTACCTACGACATGGGGAGGAATATGATTTACAGTTTATTTTTCAACTGTGCAAAATAACCTTAACTGCAGACGTTATGACATACATACATTCTATGAATTCCACTATTTTGGAGGACT  6840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  6721  AGTACCTACGACATGGGGAGGAATATGATTTACAGTTTATTTTTCAACTGTGCAAAATAACCTTAACTGCAGACGTTATGACATACATACATTCTATGAATTCCACTATTTTGGAGGACT  6840

QUERY  6841  GGAATTTTGGTCTACAACCTCCCCCAGGAGGCACACTAGAAGATACTTATAGGTTTGTAACATCCCAGGCAATTGCTTGTCAAAAACATACACCTCCAGCACCTAAAGAAGATCCCCTTA  6960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  6841  GGAATTTTGGTCTACAACCTCCCCCAGGAGGCACACTAGAAGATACTTATAGGTTTGTAACATCCCAGGCAATTGCTTGTCAAAAACATACACCTCCAGCACCTAAAGAAGATCCCCTTA  6960

QUERY  6961  AAAAATACACTTTTTGGGAAGTAAATTTAAAGGAAAAGTTTTCTGCAGACCTAGATCAGTTTCCTTTAGGACGCAAATTTTTACTACAAGCAGGATTGAAGGCCAAACCAAAATTTACAT  7080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  6961  AAAAATACACTTTTTGGGAAGTAAATTTAAAGGAAAAGTTTTCTGCAGACCTAGATCAGTTTCCTTTAGGACGCAAATTTTTACTACAAGCAGGATTGAAGGCCAAACCAAAATTTACAT  7080

QUERY  7081  TAGGAAAACGAAAAGCTACACCCACCACCTCATCTACCTCTACAACTGCTAAACGCAAAAAACGTAAGCTGTAAGTATTGTATGTATGTTGAATTAGTGTTGTTTGTTGTTTATATGTTT  7200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  7081  TAGGAAAACGAAAAGCTACACCCACCACCTCATCTACCTCTACAACTGCTAAACGCAAAAAACGTAAGCTGTAAGTATTGTATGTATGTTGAATTAGTGTTGTTTGTTGTTTATATGTTT  7200

QUERY  7201  GTATGTGCTTGTATGTGCTTGTAAATATTAAGTTGTATGTGTGTTTGTATGTATGGTATAATAAACACGTGTGTATGTGTTTTTAAATGCTTGTGTAACTATTGTGTCATGCAACATAAA  7320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  7201  GTATGTGCTTGTATGTGCTTGTAAATATTAAGTTGTATGTGTGTTTGTATGTATGGTATAATAAACACGTGTGTATGTGTTTTTAAATGCTTGTGTAACTATTGTGTCATGCAACATAAA  7320
```

```
QUERY  7321  TAAACTTATTGTTTCAACACCTACTAATTGTGTTGTGGTTATTCATTGTATATAAACTATATTTGCTACATCCTGTTTTTGTTTTATATATACTAAATTTTGTAGCGCCAGCGGCCATTT  7440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  7321  TAAACTTATTGTTTCAACACCTACTAATTGTGTTGTGGTTATTCATTGTATATAAACTATATTTGCTACATCCTGTTTTTGTTTTATATATACTAAATTTTGTAGCGCCAGCGGCCATTT  7440

QUERY  7441  TGTAGCTTCAACCGAATTCGGTTGCATGCTTTTTGGCACAAAATGTGTTTTTTTAAATAGTTCTATGTCAGCAACTATAGTTTAAACTTGTACGTTTCCTGCTTGCCATGCGTGCCAAAT  7560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  7441  TGTAGCTTCAACCGAATTCGGTTGCATGCTTTTTGGCACAAAATGTGTTTTTTTAAATAGTTCTATGTCAGCAACTATAGTTTAAACTTGTACGTTTCCTGCTTGCCATGCGTGCCAAAT  7560

QUERY  7561  CCCTGTTTTCCTGACCTGCACTGCTTGCCAACCATTCCATTGTTTTTTACACTGCACTATGTGCAACTACTGAATCACTATGTACATTGTGTCATATAAAATAAATCACTATGCGCCAAC  7680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  7561  CCCTGTTTTCCTGACCTGCACTGCTTGCCAACCATTCCATTGTTTTTTACACTGCACTATGTGCAACTACTGAATCACTATGTACATTGTGTCATATAAAATAAATCACTATGCGCCAAC  7680

QUERY  7681  GCCTTACATACCGCTGTTAGGCACATATTTTTGGCTTGTTTTAACTAACCTAATTGCATATTTGGCATAAGGTTTAAACTTCTAAGGCCAACTAAATGTCACCCTAGTTCATACATGAAC  7800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  7681  GCCTTACATACCGCTGTTAGGCACATATTTTTGGCTTGTTTTAACTAACCTAATTGCATATTTGGCATAAGGTTTAAACTTCTAAGGCCAACTAAATGTCACCCTAGTTCATACATGAAC  7800

QUERY  7801  TGTGTAAAGGTTAGTCATACATTGTTCATTTGTAAAACTGCACATGGGTGTGTGCAAACCGTTTTGGGTTACACATTTACAAGCAACTTATATAATAATACTAA  7904
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SBJCT  7801  TGTGTAAAGGTTAGTCATACATTGTTCATTTGTAAAACTGCACATGGGTGTGTGCAAACCGTTTTGGGTTACACATTTACAAGCAACTTATATAATAATACTAA  7904
```

C

```
BLASTN results
Sequence ID: lcl|Query_54813   Length: 2685   Number of Matches: 1
Range 1: 1 to 2685

Score                Expect    Identities              Gaps              Strand
4948 bits(2679)      0.0       2684/2686(99 %)         1/2686(0 %)       Plus/Plus

pUC18  1     GCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCT  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SAMPL  1     GCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCT  120

PUC18  121   CACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATGATTACGAATTCGAGCT  240
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
SAMPL  121   CACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGA-CATGATTACGAATTCGAGCT  239

PUC18  241   CGGTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCAC  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SAMPL  240   CGGTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCAC  359

PUC18  361   ATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGT  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SAMPL  360   ATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGT  479

PUC18  481   GCGGTATTTCACACCGCATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCC  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SAMPL  480   GCGGTATTTCACACCGCATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCC  599

PUC18  601   CGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGG  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SAMPL  600   CGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGG  719

PUC18  721   TTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGAC  840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SAMPL  720   TTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGAC  839

PUC18  841   AATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAG  960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SAMPL  840   AATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAG  959

PUC18  961   AAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAA  1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SAMPL  960   AAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAA  1079

PUC18  1081  TGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAG  1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SAMPL  1080  TGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAG  1199

PUC18  1201  TCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGC  1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SAMPL  1200  TCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGC  1319

PUC18  1321  TAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAA  1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SAMPL  1320  TAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGCAA  1439

PUC18  1441  CAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTG  1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SAMPL  1440  CAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTG  1559

PUC18  1561  GCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGG  1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SAMPL  1560  GCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGG  1679

PUC18  1681  CAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTT  1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SAMPL  1680  CAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTT  1799

PUC18  1801  AATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAG  1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SAMPL  1800  AATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAG  1919

PUC18  1921  ATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCA  2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SAMPL  1920  ATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCA  2039

PUC18  2041  GAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCA  2160
             |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SAMPL  2040  GAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCA  2159

PUC18  2161  GTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACA  2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SAMPL  2160  GTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACA  2279

PUC18  2281  CCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTC  2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SAMPL  2280  CCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTC  2399

PUC18  2401  CAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGG  2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SAMPL  2400  CAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGG  2519
```

```
PUC18  2521  CCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCA  2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SAMPL  2520  CCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCA  2639

PUC18  2641  GCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGA  2686
             ||||||||||||||||||||||||||||||||||||||||||||
SAMPL  2640  GCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGA  2685
```

Figure 4.13. Reference Genome for HPV16/pUC18 Plasmid and BLASTN Alignment. (A) Reference genome used to align sequence reads from ·OH-treated plasmid experiments. The reference genome started at nucleotide position 1344 of the pUC18 sequence. Blue font corresponds to pUC18 DNA sequence, red underlined font depicts *Bam*HI cloning sites and black bold font corresponds to HPV16 DNA sequence. BLASTN analyses of **(B)** the HPV16 DNA sequence and **(C)** the pUC18 DNA sequence. Red bold font highlights minor differences between the determined and published pUC18 DNA vector sequence.

4.4.5 Bioinformatics Analysis: Bowtie 2 Alignment and CountNicks.pm Cleavage Intensities

Bowtie 2 alignment statistics are summarized in **Table 4.4.** Unique sequencing reads were aligned to the HPV16/pUC18 reference genome using the *--score-min C,0.*

Table 4.4. Alignment Information of NGS Library Samples.

Sample	Total Reads	Aligned Exactly 1 Time	Aligned > 1 Times
Control-1	41,796,887	33,704,078	0
PA1-1 (2 μM)	54,016,161	41,737,458	0
PA1-1 (20 μM)	50,338,723	39,302,839	0
Control-2	23,221,340	17,147,585	0
PA1-2 (20 μM)	21,465,957	14,040,109	0
Control-3	20,910,159	15,826,078	0
PA1-3 (20 μM)	20,527,338	14,508,498	0

Reads were mapped using Bowtie 2 (--score-min C,0)

The mapped reads in the SAM file were then processed with CountNicks.pm. Because attack of hydroxyl radical results in the chemical destruction of the nucleotide base (refer to **Section 4.2.2**), the CountNicks.pm script assumes that the start of the sequence read occurs just 3' of the cleaved nucleotide. The output file obtained from the CountNicks.pm script consists of normalized cleavage intensities in reads per million (RPM) at each nucleotide position of the plasmid sequence. The average of the normalized cleavage intensities for the control samples (n = 3) and for the **PA1**-treated samples (n = 2) were calculated and used for subsequent analysis. **Figure 4.14** summarizes the normalized, global cleavage intensities (RPM) for HPV16/pUC18 at single-nucleotide resolution. Even at this scale, hotspots of **PA1** binding can be observed in the HPV16/pUC18 sequence.

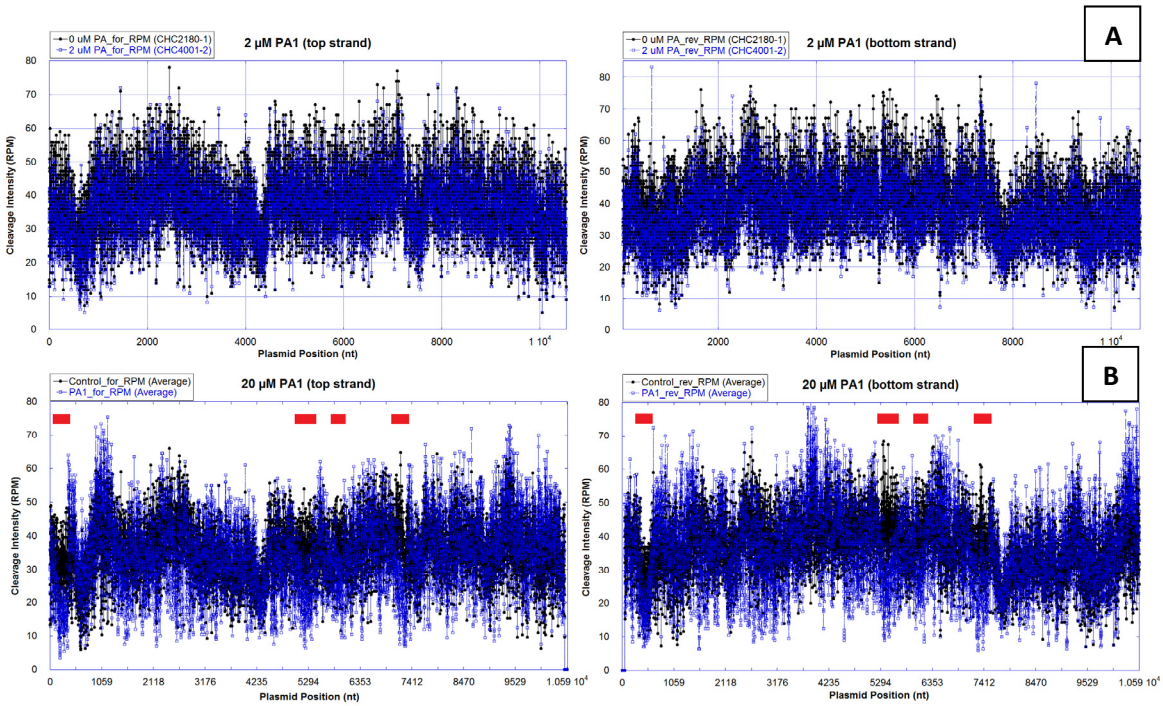


Figure 4.14. Global Cleavage Intensities for HPV16/pUC18. Average cleavage intensities for control (no **PA1**) and **PA1**-treated samples were plotted for the entire HPV16/pUC18 duplex at single-nucleotide resolution. Plots summarize the cleavage intensities observed on the top and bottom strands (*left* and *right panels*, respectively). **(A)** HPV16/pUC18 plasmid subjected to 2 μM **PA1**. **(B)** HPV16/pUC18 plasmid subjected to 20 μM **PA1**. Red bars above cleavage patterns highlighted representative regions of extensive nucleotide (nt) protection from ·OH cleavage by bound **PA1**. Black traces, no **PA1** present. Blue traces, **PA1** present.

The % change in cleavage protection for each nucleotide was then determined by calculating the ratio of the normalized cleavage intensities from the drug-treated sample to the control sample, subtracting this ratio from 1 and multiplying this value by 100 (**Equation 4.4**).

$$\%\ change\ in\ cleavage\ protection = \left(1 - \frac{cleavage\ intensity\ (drug)}{cleavage\ intensity\ (control)}\right) x\ 100 \qquad \textbf{(4.4)}$$

Footprints were identified by plotting the % change in cleavage protection against the nucleotide plasmid position and represented in a histogram form. Positive values correspond to nucleotides that are protected by bound **PA1**.

4.4.6 ·OH-Seq Provides Single-Nucleotide Resolution of PA1 Binding Sites

Our lab previously reported the **PA1** binding sites within the Long Control Region (LCR) of HPV16 (nucleotides 7348-122) by DNase I footprinting and affinity cleavage methodologies.[64] We have focused our attention in this region of the HPV genome because it harbors the promoter and enhancer elements important for gene expression and replication.[75]

PA1 footprints could not be detected at low **PA1** (2 μM) concentration. Under these conditions, the molar ratio of drug added to DNA molecules is approximately 50-fold. As a result, the vast majority of the available sites are not saturated by PA, leading to a low

signal-to-noise ratio. Therefore, the data for hydroxyl radical footprinting at 20 µM **PA1** are provided (500:1 PA:DNA ratio). Results from a representative region from 7351-7450 nt in the HPV16 LCR (**Figure 4.15**) demonstrate **PA1** binding abates cleavage in both the forward and reverse strands by hydroxyl radicals relative to the control.

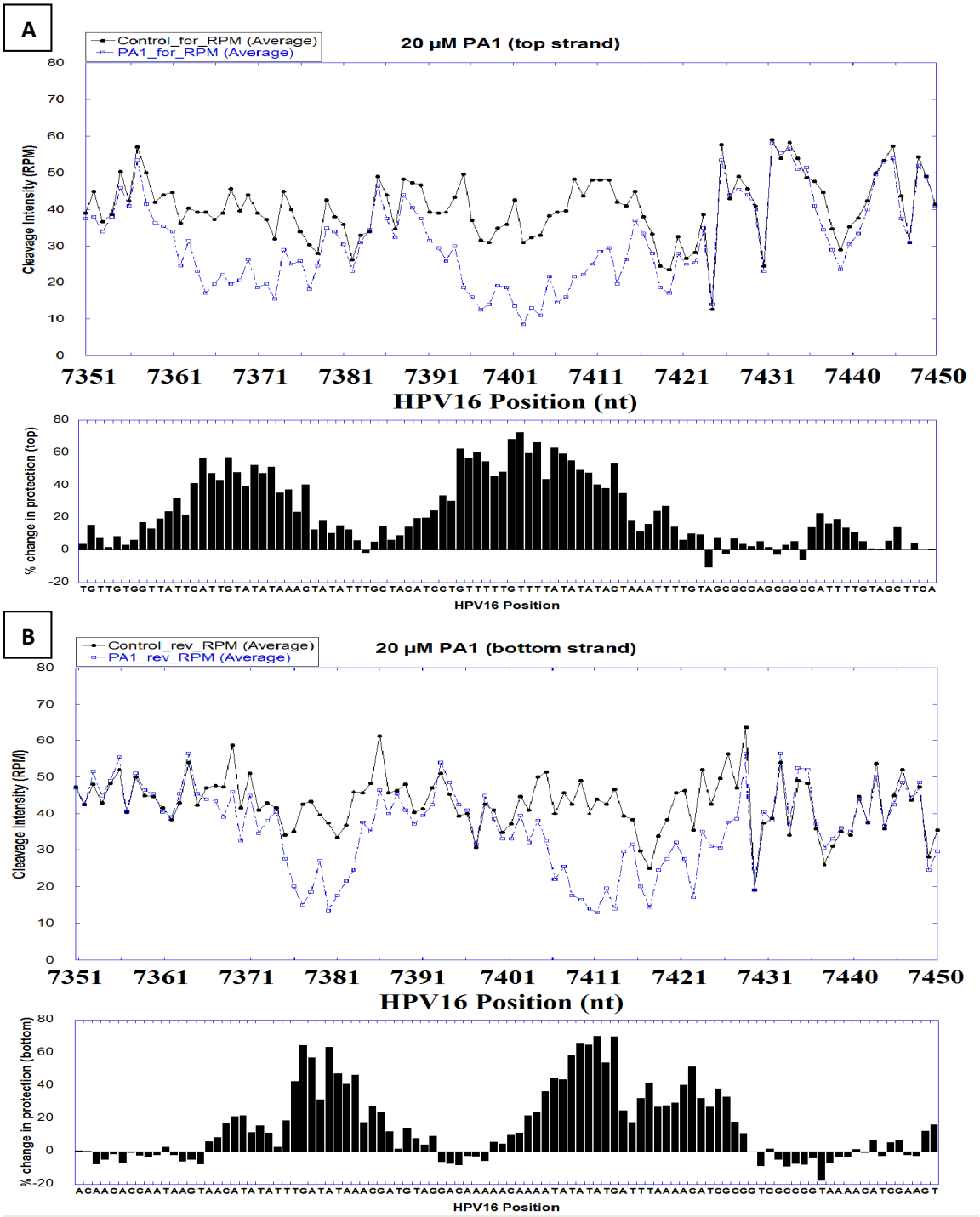


Figure 4.15. Representative Hydroxyl Radical Footprints of 20 µM PA1 in a Region of the HPV16 LCR (7351-7451). The top graphs in each panel correspond to the normalized cleavage intensities per nucleotide for the control (black trace) and the **PA1**-treated samples (blue trace). Below each graph, the % change in cleavage protection per nucleotide is depicted in histogram form. Positive values correspond to nucleotides protected in different degrees by bound **PA1**. Results for **(A)** forward and **(B)** reverse strands are provided.

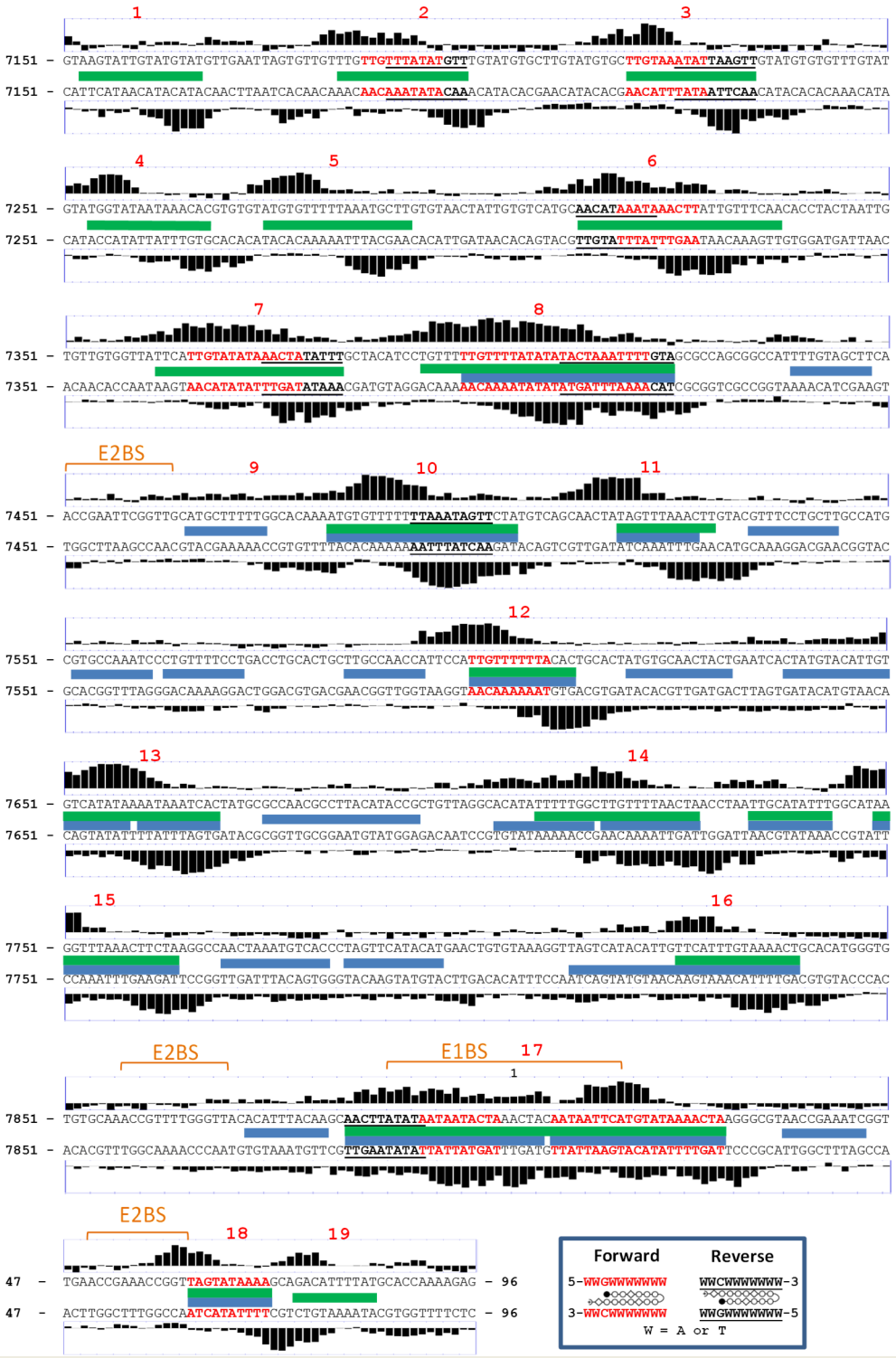


Figure 4.16. Hydroxyl Radical Footprinting of 20 µM PA1 in the HPV16 LCR. HPV16 LCR (7158-83 bp) sequence shown as a DNA duplex with the viral E1- and E2-binding sites indicated in orange. The vertical bars on top and below the sequences correspond to the calculated % change in cleavage protection for the forward/top and reverse/bottom strands, respectively. Each protected region is numbered from 1 to 19 in red font. The blue horizontal lines show the **PA1** binding sites previously determined by DNase I and affinity cleavage experiments from 7390-70 bp. The green horizontal lines depict the **PA1** binding sites consistent with the ·OH-Seq data. *Inset* depicts forward and reverse binding motifs for **PA1**. Sequences highlighted in red correspond to forward cognate sites, while underlined sequences depict reverse cognate sites.

Figure 4.16 summarizes the **PA1** binding sites at 20 μM toward the DNA sequence comprising 7151-96 bp of the HPV16 viral genome. The HPV16 LCR consists of the region from 7158 to 83 bp that harbors the binding sites for the viral E1 helicase and the E2 DNA-binding protein. The latter binds the major groove of the palindromic sequence ACCgN4cGGT (upper case designates required nucleotides, lower case represents preferred nucleotides, and N denotes no nucleotide preference) in close proximity to the viral origin of replication (oriR).[76,77] There, E1 recruits the host replication machinery to initiate viral DNA replication.[53] In this fragment, there are a total of 21 predicted cognate binding sites, with 13 of these sites corresponding to sequences that can accommodate the forward binding motif and 8 sites corresponding to the reverse binding motif. The forward motif refers to alignment of the **PA1** N-terminus to C-terminus vector with the 5'-3' direction of the DNA, whereas a reverse motif describes the **PA1** N→C vector alignment with the 3'-5' direction of the DNA (see **Figure 4.16** ***Inset***). According to the protected regions, there are at least 19 **PA1** binding sites, although most of these regions can accommodate multiple binding events within the ensemble of DNA molecules. The binding sites were assigned to sequences located in the center of the protected regions. These sequences were further identified based on the assumption that **PA1** prefers sites with the least number of mismatches. This may not always be true,[52,62-64] but it is not far off and provides the best starting point for analysis. The observed protected regions determined by hydroxyl radical are in agreement with those previously reported by Vasilieva *et al.* However, the ·OH-Seq results show that under the present experimental conditions where the genome is incorporated into a strained supercoiled plasmid, many of the multiple mismatch sites are not populated by **PA1**: most protected regions containing cognate (forward and reverse motifs) and single-base-pair mismatch sequences. Of course, even a "single-base-pair" mismatch vastly increases the number of binding sites available to **PA1** over published, predictive rules; and "single" is somewhat misleading because many different single-base-pair mismatch sites are observed in these new results.

Protected region **1** can be explained by **PA1** binding to two forward, single-base-pair mismatch sites (5'-AAGTATTGTA and 5'-TTGTATGTAT; mismatch underlined). Protected region **2** can be explained by **PA1** binding to a single-base-pair mismatch site (5'-TTGTTGTTTA) and two cognate sites; one event in the forward (5'-TTGTTTATAT) and the other in the reverse orientation (5'-AACATATAAA). Note that while **PA1** can bind to the latter in a reverse orientation; it may prefer to bind in the forward orientation leading to a **PA1**-DNA single-base-pair mismatch interaction. Similarly, protected region **3** can be explained by **PA1** occupancy of two cognate sites in the forward (5'-TTGTAAATAT) and reverse (5'-AACTTAATAT) binding motifs. Protected region **4** results from **PA1** binding to two single-base-pair mismatch sites (5'-GTGTTTATTA and 5'-TGGTATAATA). Similarly, protected region **5** can be explained by **PA1** binding to three single-base-pair mismatch sites (5'-ATGTGTTTTT, 5'-GTGTTTTTAA and 5'-AAGCATTTAA). Protected region **6** can be explained by **PA1** binding to a single-base-pair mismatch site (5'-TTGAAACAAT) and two cognate sites in the reverse (5'-AACATAAATA) and forward (5'-AAGTTTATTT) orientations.

Protected regions **7**, **8** and **17** exhibit an extended protection pattern as a result of their high A·T content. Protected region **7** can be explained by **PA1** binding to a double-base-pair mismatch site (5'-TTCATTGTAT) and two cognate sites in the forward (5'-TTGTATATAA) and reverse (5'-AACTATATTT) orientations. Protected regions **8** thru

18 have been previously described in the literature.[64] However, most of the protected regions observed in the present study contain cognate (forward and reverse motifs) and a range of single-base-pair mismatch sequences. Furthermore, regions of high G·C content (region flanked by protected regions **16** and **17**, in addition to others not shown) are not bound by **PA1** in this experiment.

Of note, **PA1** populates sequences near, as well as sequences shared by the E2 and E1 protein binding sites. Specifically, an extended protection is observed within the E1 binding site (protected region **17**) and protection near one of the E2 binding sites is also observed (protected region **18**). As a result, **PA1** could disrupt E1- and E2-DNA interactions thus inhibiting their functions within the viral lifecycle. Furthermore, polyamide-binding events impart significant structural perturbations to DNA. Specifically, eight-ring polyamides have been shown to widen the minor groove by up to 4 Å, leading to a narrowing of the major groove by 4 Å and bending of the DNA helix.[78] These perturbations provide an elegant molecular mechanism for the allosteric repression of gene transcription by polyamides since a narrow major groove would fail to properly accommodate the interaction moieties from DNA-binding proteins.[78] This mechanism seems particularly powerful given the high affinity of hairpin PAs to DNA is similar to the affinities observed by many transcription factors to DNA.[79] Similarly, these synthetic DNA-targeting agents can disrupt viral protein-DNA interactions on the viral genome. Yasuda *et al.* designed Py-Im hairpin polyamides targeting the Epstein-Barr Nuclear Antigen 1 binding sites in the origin of plasmid replication (*oriP*) within the viral genome, and demonstrated an efficient inhibition of EBNA1-binding and reduced recruitment of host replication machinery to the *oriP* in chromatin immunoprecipitation (ChIP) assays.[80] Schaal *et al*. showed that tandem hairpin polyamides were able to displace the HPV18 E2, a major groove binding protein, from its canonical E2BS4 (E2 Binding Site 4) DNA binding site. Consistent with the allosteric model, the authors showed that minor-groove binding tandem hairpin polyamides prevented the bending of DNA structure required for E2 binding in the minor groove, and also showed that PAs displace E2 from DNA.[81]

We have previously shown that **PA1** binds to the minor groove of DNA (**Chapter 2**). Because approximately 80 % of hydroxyl radical attacks take place in the minor groove, the protection patterns are expected to be asymmetrical with a 3' shift between opposite strands.[26,27] The unexpected 5'-shift observed and extended protection toward the 5'-end in the **PA1** footprints may be a result of our detection methods. Here, we relied on hydroxyl radicals to cleavage the plasmid DNA. However, depending the site of attack by hydroxyl radicals, different 5'-end groups can be generated. Specifically, abstraction of a hydrogen atom at the C4' site of the deoxyribose ring results in the formation of a 5'-phosphate, while attack at the C5' site leads to the generation of an aldehyde. Because fragments consisting of 5'-phosphates are the only species that can be ligated to NGS adapters in our method, it is possible that binding of **PA1** may result in the increased formation of products containing a 5'-aldehyde.[27,82] Further investigation is required to test this hypothesis, including repeating these experiments using DNase I as the cleavage agent. Similar to our results, Bialonska *et al*. studied the structural details of DAPI- and Hoechst-DNA complexes by a hydroxyl radical footprinting method coupled to capillary electrophoresis. The data also revealed a 5' shift in the protected regions for these minor groove binders that agreed to the predicted solvent accessible surface areas of the

deoxyribose hydrogen atoms.[26] Indexing or other bioinformatics problems could also cause the observed 5'-shift; however, there is no indication of misalignment of our footprinting data.

4.4.7 Hydroxyl Radical as a Chemical Probe Reveals Structural Details

4.4.7.1 *Comparison of Experimentally Determined Hydroxyl Radical Cleavage Patterns and ·OH Radical Cleavage Intensity Database (ORChID) Prediction*

The degree of reactivity of hydroxyl radical toward the different hydrogen atoms correlates with their solvent accessibility within the DNA backbone. As a result, the relative hydroxyl radical pattern reflects the structure and local shape of DNA. Greengaum *et al.* developed ORChID, a tetramer window algorithm, for predicting the hydroxyl radical cleavage patterns of any input sequence based on experimentally determined cleavage data with high accuracy.[21]

To investigate the relative correlation between the experimentally determined hydroxyl radical cleavage patterns and the ORChID predicted cleavage patterns for our case, we calculated the Pearson coefficients for all the pairwise cleavage patterns.[21] Although the determined Pearson correlation coefficients were low ($r_{\text{forward strand}} = 0.27$ and $r_{\text{reverse strand}} = 0.25$), visual comparison of the experimental and predicted cleavage patterns show similar cleavage features (**Figure 4.18**).

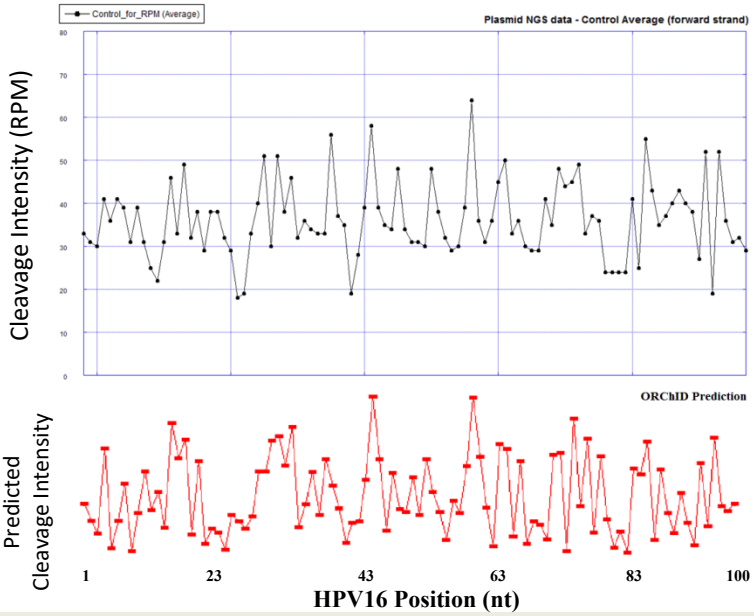


Figure 4.18. Comparison between ·OH-Seq Cleavage Patterns and ORChID Predictions. Experimentally determined hydroxyl radical patterns aligned to ORChID predicted cleavage patterns.

4.4.7.2 *Hydroxyl Radical Cleavage at Adenine Tracts*

It is well-established that A/T-rich regions including poly(dA)·poly(dT) exhibit narrow minor grooves.[83] Indeed, Tullius and colleagues demonstrated that hydroxyl radical cleavage at polyA tracts result in a sequential decrease in the cleavage rate per adenine in the 5' to 3' vector, with the 3'-most adenine displaying the lowest cutting

frequency. Correspondly, the complement strand, harboring the polyT tract, displays a similar decrease in the cleavage rate, except that the lowest cutting frequency is offset one to two bases in the 3' direction. These results are in agreement with the expected narrow minor groove of these adenine tracts.[19,83]

We also observed decreased cleavage intensity in our ·OH-Seq data corresponding to adenine tracts. **Figure 4.19** provides the ·OH cleavage of the region from 904 to 973 bp in the HPV16 genome. The region corresponding to a 9 bp adenine tract (5'-AAAAAAAAA; 941-949 bp) displays decreased cleavage intensity within the A/T-rich sequence. Inspection of the adenine-rich strand reveals an overall decrease of the cleavage rate in the 5' to 3' direction; however the cutting minimum is not seen for the 3'-most adenine. Conversely, the cleavage rate for the thymine-rich strand smoothly decreases in the 3' to 5' direction with lowest cutting near the 5'-most thymine. Comparison of the minima from the adenine-rich and thymine-rich strands reveals an offset by roughly 5 bases in the 3' direction. These results require further investigation.

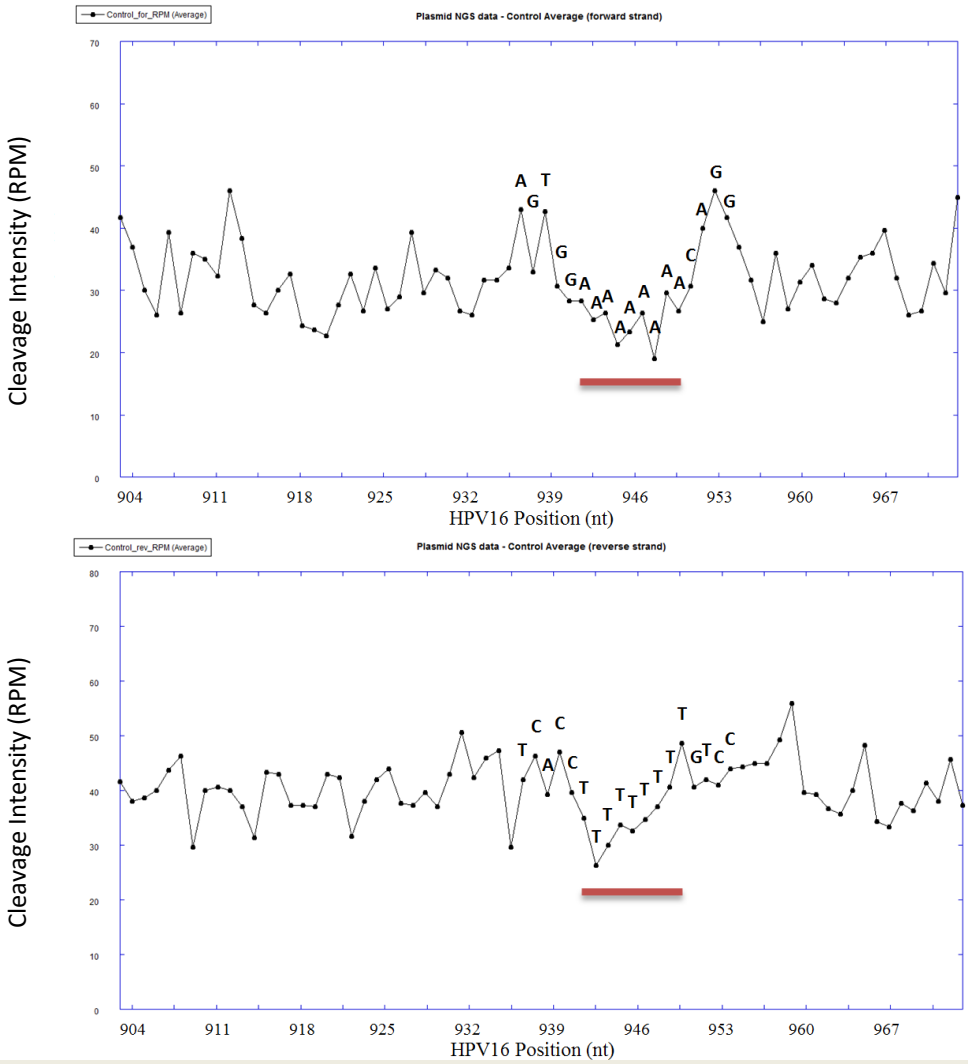


Figure 4.19. Hydroxyl Radical Cleavage of an Adenine Tract within HPV16 (904-973 bp). Hydroxyl radical patterns for the forward and reverse strands of a region in the HPV16 genome harboring a 9 bp adenine/thymine tract for control experiments (n = 3). Sequence is provided above the cleavage pattern. **(A)** Forward strand with adenine tract, **(B)** reverse strand with thymine tract.

4.5 CONCLUSIONS

We have used a hydroxyl radical footprinting method coupled with massively parallel DNA sequencing (·OH-Seq) to map **PA1** binding sites at single-nucleotide resolution in the viral genome of HPV16; the first example of single-nucleotide resolution mapping of PA binding at the genomic level. Because of its high-throughput nature, next-generation sequencing allows for the analysis at a genome-wide scale. Furthermore, this method allows for the detection of fragmented DNA without the requirement of radioactive end-labelling.

Here, we present the **PA1** binding sites determined by hydroxyl radical footprinting in the viral genome of HPV16. The binding sites observed in the HPV16 LCR are in agreement with those previously reported by our group relying on DNase I and affinity cleavage experiments coupled to automated capillary electrophoresis. However, using our adapted ·OH-Seq methodology, a large number of the previously observed binding sites containing multiple mismatches are not populated by **PA1**. Instead, most of the protected regions observed by ·OH-Seq can be described by **PA1** binding to cognate and single-base-pair mismatch sequences.

Although hydroxyl radical footprinting has been successfully applied to examine the structural details of protein-DNA complexes in cells by gamma-radiation, the detection of the footprinting regions typically requires the electrophoretic separation of radiolabeled DNA. Because radiolabeling of DNA in cells is not generally possible, the fragmented DNA must be radioactively end-labelled by primer extension prior to analysis.[24] Therefore, this method is severely limited by the length of DNA that can be interrogated and the loci about which interrogation is carried out. Because the *in vitro* ·OH-Seq experiments yielded promising results in a genome-wide manner, the application of ·OH-Seq methodology to HPV16-infected keratinocyte monolayers may allow for mapping of the **PA1** and other DNA-ligand binding sites within the context of chromatin structures in cells (see **Chapter 5**).

ACKNOWLEDGEMENTS

We thank Professor Paul Lambert (University of Wisconsin-Madison) for providing the HPV16/pUC18 plasmid.

4.6 BIBLIOGRAPHY

(1) Hampshire, A. J.; Rusling, D. A.; Broughton-Head, V. J.; Fox, K. R. *Methods* **2007**, *42*, 128.
(2) Fox, K. R. *Drug-DNA Interaction Protocols* 1997; Vol. 90, p 1.
(3) Galas, D. J.; Schmitz, A. *Nucleic Acids Res* **1978**, *5*, 3157.
(4) Lane, M. J.; Dabrowiak, J. C.; Vournakis, J. N. *Proc Natl Acad Sci U S A* **1983**, *80*, 3260.
(5) Fox, K. R.; Waring, M. J. *Nucleic Acids Res* **1984**, *12*, 9271.
(6) Chaires, J. B.; Fox, K. R.; Herrera, J. E.; Britt, M.; Waring, M. J. *Biochemistry* **1987**, *26*, 8227.
(7) Van Dyke, M. W.; Dervan, P. B. *Nucleic Acids Res* **1983**, *11*, 5555.

(8) Yoshinaga, S. K.; Boulanger, P. A.; Berk, A. J. *Proc Natl Acad Sci U S A* **1987**, *84*, 3585.
(9) Hesselberth, J. R.; Chen, X.; Zhang, Z.; Sabo, P. J.; Sandstrom, R.; Reynolds, A. P.; Thurman, R. E.; Neph, S.; Kuehn, M. S.; Noble, W. S.; Fields, S.; Stamatoyannopoulos, J. A. *Nat Methods* **2009**, *6*, 283.
(10) Van Dyke, M. W.; Sawadogo, M. *Mol Cell Biol* **1990**, *10*, 3415.
(11) Abu-Daya, A.; Brown, P. M.; Fox, K. R. *Nucleic Acids Res* **1995**, *23*, 3385.
(12) Mrksich, M.; Parks, M. E.; Dervan, P. B. *Journal of the American Chemical Society* **1994**, *116*, 7983.
(13) Leblanc, B.; Moss, T. In *DNA-Protein Interactions*; Leblanc, B., Moss, T., Eds.; Humana Press: 2009; Vol. 543, p 37.
(14) Sigman, D. S.; Graham, D. R.; D'Aurora, V.; Stern, A. M. *Journal of Biological Chemistry* **1979**, *254*, 12269.
(15) Sigman, D. S. *Biochemistry* **1990**, *29*, 9097.
(16) Tullius, T. D.; Dombroski, B. A. *Proceedings of the National Academy of Sciences* **1986**, *83*, 5469.
(17) Van Dyke, M. W.; Hertzberg, R. P.; Dervan, P. B. *Proc Natl Acad Sci U S A* **1982**, *79*, 5470.
(18) Xu, L. Ph.D., Boston University, 2009.
(19) Shafer, G. E.; Price, M. A.; Tullius, T. D. *Electrophoresis* **1989**, *10*, 397.
(20) Guo, H.; Tullius, T. D. *Proceedings of the National Academy of Sciences* **2003**, *100*, 3743.
(21) Greenbaum, J. A.; Pang, B.; Tullius, T. D. *Genome Res* **2007**, *17*, 947.
(22) Churchill, M. E. A.; Hayes, J. J.; Tullius, T. D. *Biochemistry* **1990**, *29*, 6043.
(23) Hayes, J. J.; Kam, L.; Tullius, T. D. In *Methods in Enzymology*; Lester Packer, A. N. G., Ed.; Academic Press: 1990; Vol. Volume 186, p 545.
(24) Ottinger, L. M.; Tullius, T. D. *Journal of the American Chemical Society* **2000**, *122*, 5901.
(25) Jain, S. S. T., T. D. *Nature Protocols* **2008**, *3*.
(26) Bialonska, D.; Song, K.; Bolton, P. H. *Mutation research* **2011**, *726*, 47.
(27) Pogozelski, W. K.; Tullius, T. D. *Chemical Reviews* **1998**, *98*, 1089.
(28) Balasubramanian, B.; Pogozelski, W. K.; Tullius, T. D. *Proceedings of the National Academy of Sciences* **1998**, *95*, 9738.
(29) Yindeeyoungyeon, W.; Schell, M. A. *BioTechniques* **2000**, *29*, 1034.
(30) Sivapragasam, S.; Pande, A.; Grove, A. *Anal Biochem* **2015**, *481*, 1.
(31) He, G.; Vasilieva, E.; Bashkin, J. K.; Dupureur, C. M. *Anal Biochem* **2013**, *439*, 99.
(32) Thurman, R. E.; Rynes, E.; Humbert, R.; Vierstra, J.; Maurano, M. T.; Haugen, E.; Sheffield, N. C.; Stergachis, A. B.; Wang, H.; Vernot, B.; Garg, K.; Sandstrom, R.; Bates, D.; Canfield, T. K.; Diegel, M.; Dunn, D.; Ebersol, A. K.; Frum, T.; Giste, E.; Harding, L.; Johnson, A. K.; Johnson, E. M.; Kutyavin, T.; Lajoie, B.; Lee, B.-K.; Lee, K.; London, D.; Lotakis, D.; Neph, S.; Neri, F.; Nguyen, E. D.; Reynolds, A. P.; Roach, V.; Safi, A.; Sanchez, M. E.; Sanyal, A.; Shafer, A.; Simon, J. M.; Song, L.; Vong, S.; Weaver, M.; Zhang, Z.; Zhang, Z.; Lenhard, B.; Tewari, M.; Dorschner, M. O.; Hansen, R. S.; Navas, P. A.; Stamatoyannopoulos, G.; Iyer, V. R.; Lieb, J. D.; Sunyaev, S. R.; Akey, J. M.;

Sabo, P. J.; Kaul, R.; Furey, T. S.; Dekker, J.; Crawford, G. E.; Stamatoyannopoulos, J. A. *Nature* **2012**, *489*, 75.
(33) Nickols, N. G.; Szablowski, J. O.; Hargrove, A. E.; Li, B. C.; Raskatov, J. A.; Dervan, P. B. *Molecular Cancer Therapeutics* **2013**, *12*, 675.
(34) Morozova, O.; Marra, M. A. *Genomics* **2008**, *92*, 255.
(35) Anders, L.; Guenther, M. G.; Qi, J.; Fan, Z. P.; Marineau, J. J.; Rahl, P. B.; Loven, J.; Sigova, A. A.; Smith, W. B.; Lee, T. I.; Bradner, J. E.; Young, R. A. *Nat Biotech* **2014**, *32*, 92.
(36) Erwin, G. S.; Bhimsaria, D.; Eguchi, A.; Ansari, A. Z. *Angewandte Chemie International Edition* **2014**, *53*, 10124.
(37) Mardis, E. R. *Annu Rev Genomics Hum Genet* **2008**, *9*, 387.
(38) Mardis, E. R. *Annu Rev Anal Chem (Palo Alto Calif)* **2013**, *6*, 287.
(39) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc Natl Acad Sci U S A* **1985**, *82*, 1376.
(40) Dervan, P. B.; Edelson, B. S. *Current Opinion in Structural Biology* **2003**, *13*, 284.
(41) Chen, X. R., Boopathy; Rao, Sambhorao T.; Sundaralingam, Muttaiya *Nat Struct Mol Biol* **1994**, *1*, 7.
(42) Marky, L. A.; Breslauer, K. J. *Proc Natl Acad Sci U S A* **1987**, *84*, 4359.
(43) Finlay, A. C.; Hochstein, F. A.; Sobin, B. A.; Murphy, F. X. *Journal of the American Chemical Society* **1951**, *73*, 341.
(44) Herman, D. M.; Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1998**, *120*, 1382.
(45) Pilch, D. S.; Poklar, N.; Gelfand, C. A.; Law, S. M.; Breslauer, K. J.; Baird, E. E.; Dervan, P. B. *Proceedings of the National Academy of Sciences* **1996**, *93*, 8306.
(46) White, S.; Baird, E. E.; Dervan, P. B. *Biochemistry* **1996**, *35*, 12532.
(47) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *Journal of the American Chemical Society* **1996**, *118*, 6160.
(48) Wang, C. C. C. E., U.; Dervan, P. B. *Bioorg Med Chem* **2001**, *9*, 5.
(49) Dervan, P. B. D., R.M.; Marques, M.A. *Current Medicinal Chemistry - Anti-Cancer Agents* **2005**, *5*.
(50) Bremer, R. E.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. *Bioorg Med Chem* **2000**, *8*, 1947.
(51) Edwards, T. G.; Koeller, K. J.; Slomczynska, U.; Fok, K.; Helmus, M.; Bashkin, J. K.; Fisher, C. *Antiviral Res* **2011**, *91*, 177.
(52) Castaneda, C. H.; Scuderi, M. J.; Edwards, T. G.; Harris Jr, G. D.; Dupureur, C. M.; Koeller, K. J.; Fisher, C.; Bashkin, J. K. *MedChemComm* **2016**.
(53) Doorbar, J. *Clin Sci (Lond)* **2006**, *110*, 525.
(54) Bottalico, D.; Chen, Z.; Kocjan, B. J.; Seme, K.; Poljak, M.; Burk, R. D. *J Gen Virol* **2012**, *93*, 1774.
(55) Muñoz, N.; Bosch, F. X.; Castellsagué, X.; Díaz, M.; de Sanjose, S.; Hammouda, D.; Shah, K. V.; Meijer, C. J. L. M. *International Journal of Cancer* **2004**, *111*, 278.
(56) Archambault, J.; Melendy, T. *Antivir Ther* **2013**, *18*, 271.

(57) Walboomers, J. M. M.; Jacobs, M. V.; Manos, M. M.; Bosch, F. X.; Kummer, J. A.; Shah, K. V.; Snijders, P. J. F.; Peto, J.; Meijer, C. J. L. M.; Muñoz, N. *J Pathol* **1999**, *189*, 12.
(58) Organization, W. H. Comprehensive Cervical Cancer Prevention and Control Programme Guidance for Countries. [Online Early Access]. Published Online: 2012.
(59) Faridi, R.; Zahra, A.; Khan, K.; Idrees, M. *Virol J* **2011**, *8*, 269.
(60) Ryser, M. D.; McGoff, K.; Herzog, D. P.; Sivakoff, D. J.; Myers, E. R. *Epidemics* **2015**, *11*, 32.
(61) Clifford, G.; Franceschi, S.; Diaz, M.; Munoz, N.; Villa, L. L. *Vaccine* **2006**, *24 Suppl 3*, S3/26.
(62) He, G.; Vasilieva, E.; Harris Jr, G. D.; Koeller, K. J.; Bashkin, J. K.; Dupureur, C. M. *Biochimie* **2014**, *102*, 83.
(63) Koeller, K. J.; Harris, G. D.; Aston, K.; He, G.; Castaneda, C. H.; Thornton, M. A.; Edwards, T. G.; Wang, S.; Nanjunda, R.; Wilson, W. D.; Fisher, C.; Bashkin, J. K. *Medicinal Chemistry* **2014**, *4*, 338.
(64) Vasilieva, E.; Niederschulte, J.; Song, Y.; Harris Jr, G. D.; Koeller, K. J.; Liao, P.; Bashkin, J. K.; Dupureur, C. M. *Biochimie* **2016**, *127*, 103.
(65) Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1996**, *118*, 6141.
(66) Flores, E. R.; Allen-Hoffmann, B. L.; Lee, D.; Sattler, C. A.; Lambert, P. F. *Virology* **1999**, *262*, 344.
(67) Kibbe, W. A. *Nucleic Acids Res* **2007**, *35*, W43.
(68) DeAngelis, M. M.; Wang, D. G.; Hawkins, T. L. *Nucleic Acids Res* **1995**, *23*, 4742.
(69) Fisher, S.; Barry, A.; Abreu, J.; Minie, B.; Nolan, J.; Delorey, T. M.; Young, G.; Fennell, T. J.; Allen, A.; Ambrogio, L.; Berlin, A. M.; Blumenstiel, B.; Cibulskis, K.; Friedrich, D.; Johnson, R.; Juhn, F.; Reilly, B.; Shammas, R.; Stalker, J.; Sykes, S. M.; Thompson, J.; Walsh, J.; Zimmer, A.; Zwirko, Z.; Gabriel, S.; Nicol, R.; Nusbaum, C. *Genome Biology* **2011**, *12*, R1.
(70) Hawkins, T. L.; O'Connor-Morin, T.; Roy, A.; Santillan, C. *Nucleic Acids Res* **1994**, *22*, 4543.
(71) Lis, J. T.; Schleif, R. *Nucleic Acids Res* **1975**, *2*, 383.
(72) Langmead, B.; Salzberg, S. L. *Nat Methods* **2012**, *9*, 357.
(73) Poptsova, M. S.; Il'icheva, I. A.; Nechipurenko, D. Y.; Panchenko, L. A.; Khodikov, M. V.; Oparina, N. Y.; Polozov, R. V.; Nechipurenko, Y. D.; Grokhovsky, S. L. *Scientific Reports* **2014**, *4*, 4532.
(74) Wheeler, D. L.; Barrett, T.; Benson, D. A.; Bryant, S. H.; Canese, K.; Chetvernin, V.; Church, D. M.; DiCuccio, M.; Edgar, R.; Federhen, S.; Geer, L. Y.; Kapustin, Y.; Khovayko, O.; Landsman, D.; Lipman, D. J.; Madden, T. L.; Maglott, D. R.; Ostell, J.; Miller, V.; Pruitt, K. D.; Schuler, G. D.; Sequeira, E.; Sherry, S. T.; Sirotkin, K.; Souvorov, A.; Starchenko, G.; Tatusov, R. L.; Tatusova, T. A.; Wagner, L.; Yaschenko, E. *Nucleic Acids Res* **2007**, *35*, D5.
(75) Graham, S. V. *Future microbiology* **2010**, *5*, 1493.
(76) Hegde, R. S. *Annu Rev Biophys Biomol Struct* **2002**, *31*, 343.

(77) Lehoux, M.; Fradet-Turcotte, A.; Lussier-Price, M.; Omichinski, J. G.; Archambault, J. *J Virol* **2012**, *86*, 3486.
(78) Chenoweth, D. M.; Dervan, P. B. *Proc Natl Acad Sci U S A* **2009**, *106*, 13175.
(79) Dervan, P. B. *Bioorg Med Chem* **2001**, *9*, 2215.
(80) Yasuda, A.; Noguchi, K.; Minoshima, M.; Kashiwazaki, G.; Kanda, T.; Katayama, K.; Mitsuhashi, J.; Bando, T.; Sugiyama, H.; Sugimoto, Y. *Cancer Sci* **2011**, *102*, 2221.
(81) Schaal, T. D.; Mallet, W. G.; McMinn, D. L.; Nguyen, N. V.; Sopko, M. M.; John, S.; Parekh, B. S. *Nucleic Acids Res* **2003**, *31*, 1282.
(82) Cannistraro, V. J.; Pondugula, S.; Song, Q.; Taylor, J.-S. *Journal of Biological Chemistry* **2015**, *290*, 26597.
(83) Hizver, J.; Rozenberg, H.; Frolow, F.; Rabinovich, D.; Shakked, Z. *Proceedings of the National Academy of Sciences* **2001**, *98*, 8490.

4.7 SUPPLEMENTAL INFORMATION

4.7.1 Ethanol Precipitation of DNA

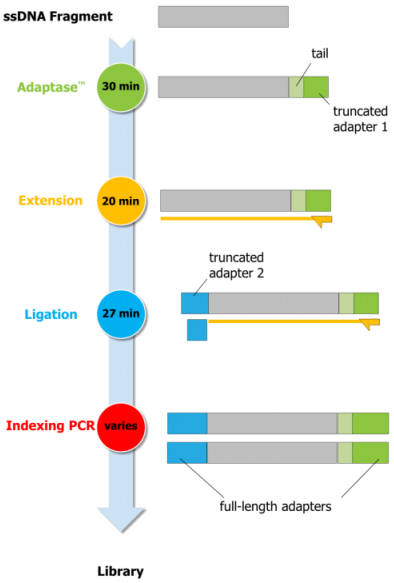
MATERIALS

1. 3 M sodium acetate, pH 5.2 (store at 4 °C)
   Dissolve 2.46 g of sodium acetate (Sigma S7545-250G, MW 82.03 g/mol) in approximately 5 mL of MilliQ H_2O and adjust the pH to 5.2 with glacial acetic acid (~1.5 mL). Add MilliQ H_2O to a final volume of 10 mL. Filter-sterilize using a 0.22 μm filter.
2. Cold (4 °C) 100 % EtOH
3. Cold (4 °C) 70 % EtOH
4. EB buffer (10 mM Tris·Cl, pH 8.5)

PROCEDURE

1. Add 1/10 volume of 3 M sodium acetate, pH 5.2 to DNA sample
2. Mix well by inverting the closed vial/tube ~6 times.
3. Add 3 volumes of cold 100 % EtOH.
4. Mix well by inverting the closed vial/tube ~6 times.
5. Place sample in -20 °C freezer overnight.
6. Spin sample at 17,900*x g* at 4 °C for 30 min.
7. Decant the supernatant being careful not to disturb DNA pellet.
8. Wash DNA pellet with 400 μL of cold 70 % EtOH and centrifuge sample at17,900*x g* at 4 °C for 15 min.
9. Decant the supernatant being careful not to disturb DNA pellet.
10. Air-dry the DNA pellet for ~ 20 min.
11. Resuspend DNA in appropriate EB buffer or Low EDTA TE buffer (usually 50 μL).

4.7.2 Accel-NGS 1S Plus Protocol



- The 1S Plus protocol uses a unique, sequential adaptation process to attach adapters to the ends of single-stranded DNA fragments.
- The Adaptase step is a highly efficient, proprietary reaction that simultaneously performs end repair, tailing of 3′ ends, and ligation of the first truncated adapter to 3′ ends.
- The Extension step is used to facilitate ligation of the second truncated adapter. The synthesized strand does not get sequenced.
- Bead-based SPRI clean-ups are used to remove oligonucleotides and small fragments, and to change enzymatic buffer composition.
- The Ligation reaction is used to add the second truncated adapter to the 5′ ends.
- The PCR step is used to increase yield and add the indexed adapter sequence.

Swift BIOSCIENCES

Cleaning and Preparation of Work Area

To mitigate library and DNA contamination:

1. Physically separate the laboratory areas where pre-PCR and post-PCR steps are executed.
2. Use 10 % bleach to clean work areas.
3. Use wide orifice, barrier pipette tips to reduce DNA fragmentation and cross-contamination.

Prepare a fresh 80 % ethanol solution (5 mL / sample).

DNA Fragmentation: Fragment HPV16/pUC18 plasmid DNA by hydroxyl radical reagent cleavage.

Input DNA Quality Control: Determine the fragmented DNA size distribution using Agilent RNA 6000 Nano kit (ref. CHC2151).

DENATURATION

1. Remove **Adaptase** reagents from -20°C storage and place on ice for at least 10 min in order to allow reagents to reach 4°C.
   - **i.** Buffer G1
   - **ii.** Reagent G2
   - **iii.** Reagent G3
   - **iv.** Enzyme G4
   - **v.** Enzyme G5
   - **vi.** Enzyme G6
2. Once thawed, vortex reagents well (**except enzymes**).
3. Pre-assemble an **Adaptase Reaction Mix** and place on ice.

	Reagent	Volume per Reaction
Pre-assemble	Low EDTA TE	11.5 µL
	Buffer G1	4 µL
	Reagent G2	4 µL
	Reagent G3	2.5 µL

4. Load **1 denature DNA** program on the Eppendorf Mastercycler under carlos/accel 1s ngs.

Incubation Conditions
Pre-heat to 95°C, lid heating ON (105°C) and pause
95°C for 2 min, lid heating ON (105°C)
Immediately place on ice for 2 min then proceed to Adaptase step

5. Transfer each 15 µL of fragmented DNA to a clean 0.2 mL PCR vial.
6. Place each vial in thermocycler and unpause the **1 denarute DNA** program.
7. Immediately place samples on ice for 2 min then proceed to Adaptase step to maximize the amount of ssDNA substrate

ADAPTASE

1. Load **2 adaptase** program on the Eppendorf Mastercycler under carlos/accel 1s ngs.

Adaptase Thermocycler Program
Pre-heat to 37°C, lid heating ON (105°C) and pause
37°C for 15 min, lid heating ON (105°C)
95°C for 2 min, lid heating ON (105°C)
4°C hold

2. Complete the **Adaptase Reaction Mix** by adding Enzymes G4, G5 and G6.

Assembly Order	Reagent	Volume per Reaction
Pre-assemble	Low EDTA TE	11.5 µL
	Buffer G1	4 µL
	Reagent G2	4 µL
	Reagent G3	2.5 µL
Add just before use	Enzyme G4	**1 µL**
	Enzyme G5	**1 µL**
	Enzyme G6	**1 µL**

3. Gently vortex to mix.
4. Transfer 25 µL of the **Adaptase Reaction Mix** to each PCR vial containing a 15 µL denatured DNA sample.

5. Gently vortex the reaction and spin down.
6. Place each vial in thermocycler and unpause the **2 adaptase** program.

EXTENSION

1. Load **3 extension** program on the Eppendorf Mastercycler under carlos/accel 1s ngs.

Extension Thermocycler Program
Pre-heat to 98°C, lid heating ON (105°C) and pause
98°C for 30 s, lid heating ON (105°C)
63°C for 15 s, lid heating ON (105°C)
68°C for 5 min, lid heating ON (105°C)
4°C hold

2. Remove **Extension** reagents from -20°C storage and place on ice for at least 10 min in order to allow reagents to reach 4°C.
   - **i.** Reagent Y1
   - **ii.** Reagent W2
   - **iii.** Buffer W3
   - **iv.** Enzyme W4
3. Once thawed, vortex reagents well (**except enzymes**).
4. On ice, prepare an **Extension Reaction Mix** and place on ice.

Assembly Order	Reagent	Volume per Reaction
Pre-assemble	Low EDTA TE	18.5 µL
	Reagent Y1	2 µL
	Reagent W2	7 µL
	Buffer W3	17.5 µL
Add just before use	Enzyme W4	2 µL

5. Gently vortex to mix.
6. Transfer 47 µL of the **Extension Reaction Mix** to each PCR vial containing a 40 µL Adaptase Reaction.
7. Gently vortex the reaction and spin down.
8. Place each vial in thermocycler and unpause the **3 extension** program.

POST-EXTENSION SPRI

1. Substitute the correct volume for **Sample**, **SPRI**, and **Elution** in the protocol given below to perform fragment size selection by SPRIselect beads.

- For DNA inputs ≥ 1 ng, perform fragment size selection using the following reagent ratios:

Input	Sample	SPRI	Elution
≥ 1 ng, 200 bp	87 µL	104 µL (ratio: 1.2)	20 µL
≥ 1 ng, 350 bp		70 µL (ratio: 0.8)	

- For DNA inputs < 1 ng, perform fragment size selection with two consecutive clean-ups using the following reagent ratios:

Input		Sample	SPRI	Elution
< 1 ng, 200 bp	1st SPRI	87 µL	104 µL (ratio: 1.2)	50 µL
	2nd SPRI	50 µL (1st SPRI eluate)	60 µL (ratio: 1.2)	20 µL
< 1 ng, 350 bp	1st SPRI	87 µL	70 µL (ratio: 0.8)	50 µL
	2nd SPRI	50 µL (1st SPRI eluate)	40 µL (ratio: 0.8)	20 µL

2. Homogenize the SPRIselect beads by gently vortexing the solution.
3. Transfer each **Sample** volume to a clean 1.5 mL Eppendorf microcentrifuge vial and add the appropriate **SPRI** volume to each sample.
4. Homogenize the sample by vortexing or pipetting 10 times. Briefly spin down the samples to collect droplets (no bead-sample droplets on the sides of the tube).
5. Incubate the sample suspensions for 5 min at room temperature.
6. Place the sample suspension on a magnetic stand until the beads settle and a clear solution is observed (~ 2 min).
7. Carefully remove and discard the clear supernatant. Do not disturb the bead pellet, as the DNA is associated with the beads.
8. While on the magnet stand, add 500 µL of the freshly prepared 80 % ethanol solution to the vial without disturbing the bead pellet. Incubate for 30 s, and carefully remove and discard the ethanol solution. Do not disturb the bead pellet, as the DNA is associated with the beads.
9. Repeat step 8.
10. Briefly spin down the sample. Place back on the magnetic stand and remove any residual ethanol from the bottom of the vial.
11. Air-dry the bead pellet.
12. Resuspend the pellet by adding **Elution** volume of Low EDTA TE. Homogenize the suspension by pipetting up and down. Briefly spin down the samples to collect droplets (no bead-sample droplets on the sides of the tube).
13. Incubate suspension for 5 min and place the tube on the magnetic stand.
14. Transfer the entire eluate supernatant to a clean 0.2 mL PCR vial, ensuring not to pipette any magnetic beads. If magnetic beads are present in the eluate, transfer eluate to a new 0.2 mL PCR vial, place on magnetic stand and collect eluate again.

Protocol can be stopped at this point. Store samples at 4°C until ready to continue.

LIGATION

1. Load **5 ligation** program on the Eppendorf Mastercycler under carlos/accel 1s ngs.

Extension Thermocycler Program
Pre-heat to 25°C, lid heating ON (105°C) and pause
25°C for 15 min, lid heating ON (105°C)
4°C hold

2. Remove **Ligation** reagents from -20°C storage and place on ice for at least 10 min in order to allow reagents to reach 4°C.
   - **i.** Buffer B1
   - **ii.** Reagent B2
   - **iii.** Enzyme B3
3. Once thawed, vortex reagents well (**except enzymes**).
4. On ice, prepare a **Ligation Reaction Mix** and place on ice.

Assembly Order	Reagent	Volume per Reaction
Pre-assemble	Low EDTA TE	4 µL
	Buffer B1	4 µL
	Reagent B2	10 µL
Add just before use	Enzyme B3	2 µL

5. Gently vortex to mix.
6. Transfer 20 µL of the **Ligation Reaction Mix** to each PCR vial containing a 20 µL SPRI eluate.
7. Gently vortex the reaction and spin down.
8. Place each vial in thermocycler and unpause the **5 ligation** program.

POST-LIGATION SPRI

1. Substitute the correct volume for **Sample**, **SPRI**, and **Elution** in the protocol given below to perform fragment size selection by SPRIselect beads.

Input	Sample	SPRI	Elution
200 bp	40 µL	40 µL (ratio: 1.0)	20 µL
350 bp		32 µL (ratio: 0.8)	

2. Homogenize the SPRIselect beads by gently vortexing the solution.
3. Transfer each **Sample** volume to a clean 1.5 mL Eppendorf microcentrifuge vial and add the appropriate **SPRI** volume to each sample.
4. Homogenize the sample by vortexing or pipetting 10 times. Briefly spin down the samples to collect droplets (no bead-sample droplets on the sides of the tube).

5. Incubate the sample suspensions for 5 min at room temperature.
6. Place the sample suspension on a magnetic stand until the beads settle and a clear solution is observed (~ 2 min).
7. Carefully remove and discard the clear supernatant. Do not disturb the bead pellet, as the DNA is associated with the beads.
8. While on the magnet stand, add 500 µL of the freshly prepared 80 % ethanol solution to the vial without disturbing the bead pellet. Incubate for 30 s, and carefully remove and discard the ethanol solution. Do not disturb the bead pellet, as the DNA is associated with the beads.
9. Repeat step 8.
10. Briefly spin down the sample. Place back on the magnetic stand and remove any residual ethanol from the bottom of the vial.
11. Air-dry the bead pellet.
12. Resuspend the pellet by adding **Elution** volume of Low EDTA TE. Homogenize the suspension by pipetting up and down. Briefly spin down the samples to collect droplets (no bead-sample droplets on the sides of the tube).
13. Incubate suspension for 5 min and place the tube on the magnetic stand.
14. Transfer the entire eluate supernatant to a clean 0.2 mL PCR vial, ensuring not to pipette any magnetic beads. If magnetic beads are present in the eluate, transfer eluate to a new 0.2 mL PCR vial, place on magnetic stand and collect eluate again.

Protocol can be stopped at this point. Store samples at 4°C until ready to continue.

INDEXING PCR

1. Remove **Indexing PCR** reagents from -20°C storage and place on ice for at least 10 min in order to allow reagents to reach 4°C.
   - **i.** Reagent R1 (index primer) / Index D50X and Index D7XX
   - **ii.** Reagent W2
   - **iii.** Buffer W3
   - **iv.** Enzyme W4
2. Once thawed, vortex reagents well (**except enzymes**).
3. Add indexing reagent(s) directly to 20 µL SPRI eluate.

Reagent	Volume per Sample (SI-IL1SP-12A)	Volume per Sample (DI-IL1SP-48)
Reagent R1 (index primer)	5 µL	-
Index D50X	-	2.5 µL
Index D7XX	-	2.5 µL

4. On ice, prepare a **PCR Reaction Mix** and place on ice.

Assembly Order	Reagent	Volume per Reaction
Pre-assemble	Low EDTA TE	10 µL
	Reagent W2	4 µL
	Buffer W3	10 µL
Add just before use	Enzyme W4	1 µL

5. Gently vortex to mix.
6. Transfer 25 µL of the **PCR Reaction Mix** to each PCR vial containing 20 µL SPRI eluate and 5 µL indexing primers.
7. Mix by pipetting.
8. Place reactions in thermocycler.
9. Load **7 indexing pcr** program on the Eppendorf Mastercycler under carlos/accel 1s ngs.

Indexing PCR Thermocycler Program
98°C for 30 s, lid heating ON (105°C)
PCR Cycles: 98°C for 10 s, lid heating ON (105°C) 60°C for 30 s, lid heating ON (105°C) 68°C for 60 s, lid heating ON (105°C)
4°C hold

Input Quantity	Recommended PCR Cycles
250 ng	3
100 ng	4
10 ng	7
1 ng	10
100 pg	14
10 pg	17

Exact number of PCR cycles must be determined by user

POST-PCR SPRI

1. Substitute the correct volume for **Sample**, **SPRI**, and **Elution** in the protocol given below to perform fragment size selection by SPRIselect beads.

Input	Sample	SPRI	Elution
200 bp 350 bp	50 µL	42.5 µL (ratio: 0.85)	20 µL

2. Homogenize the SPRIselect beads by gently vortexing the solution.
3. Transfer each **Sample** volume to a clean 1.5 mL Eppendorf microcentrifuge vial and add the appropriate **SPRI** volume to each sample.
4. Homogenize the sample by vortexing or pipetting 10 times. Briefly spin down the samples to collect droplets (no bead-sample droplets on the sides of the tube).
5. Incubate the sample suspensions for 5 min at room temperature.
6. Place the sample suspension on a magnetic stand until the beads settle and a clear solution is observed (~ 2 min).
7. Carefully remove and discard the clear supernatant. Do not disturb the bead pellet, as the DNA is associated with the beads.

8. While on the magnet stand, add 500 µL of the freshly prepared 80 % ethanol solution to the vial without disturbing the bead pellet. Incubate for 30 s, and carefully remove and discard the ethanol solution. Do not disturb the bead pellet, as the DNA is associated with the beads.
9. Repeat step 8.
10. Briefly spin down the sample. Place back on the magnetic stand and remove any residual ethanol from the bottom of the vial.
11. Air-dry the bead pellet.
12. Resuspend the pellet by adding **Elution** volume of Low EDTA TE. Homogenize the suspension by pipetting up and down. Briefly spin down the samples to collect droplets (no bead-sample droplets on the sides of the tube).
13. Incubate suspension for 5 min and place the tube on the magnetic stand.
14. Transfer the entire eluate supernatant to a clean 0.2 mL PCR vial, ensuring not to pipette any magnetic beads. If magnetic beads are present in the eluate, transfer eluate to a new 0.2 mL PCR vial, place on magnetic stand and collect eluate again.

 **Store samples at 4°C until ready to quantify and sequence libraries.**

INDEXED ADAPTER SEQUENCES

Single Index "Set A" Adapters	Sequence*	SI-IL1SP-12A
Reagent R1 (I2)	CGATGT(A)	6 µL
Reagent R1 (I4)	TGACCA(A)	6 µL
Reagent R1 (I5)	ACAGTG(A)	6 µL
Reagent R1 (I6)	GCCAAT(A)	6 µL
Reagent R1 (I7)	CAGATC(A)	6 µL
Reagent R1 (I12)	CTTGTA(A)	6 µL
Reagent R1 (I13)	AGTCAA(C)	6 µL
Reagent R1 (I14)	AGTTCC(G)	6 µL
Reagent R1 (I15)	ATGTCA(G)	6 µL
Reagent R1 (I16)	CCGTCC(C)	6 µL
Reagent R1 (I18)	GTCCGC(A)	6 µL
Reagent R1 (I19)	GTGAAA(C)	6 µL

*Base in parentheses is used for phasing calculations, but is not considered part of the index sequence

Dual Index Adapters	Sequence	DI-IL1SP-48
Index D501	TATAGCCT	22 µL
Index D502	ATAGAGGC	22 µL
Index D503	CCTATCCT	22 µL
Index D504	GGCTCTGA	22 µL
Index D505	AGGCGAAG	22 µL
Index D506	TAATCTTA	22 µL
Index D507	CAGGACGT	22 µL
Index D508	GTACTGAC	22 µL
Index D701	ATTACTCG	17 µL
Index D702	TCCGGAGA	17 µL
Index D703	CGCTCATT	17 µL
Index D704	GAGATTCC	17 µL
Index D705	ATTCAGAA	17 µL
Index D706	GAATTCGT	17 µL
Index D707	CTGAAGCT	17 µL
Index D708	TAATGCGC	17 µL
Index D709	CGGCTATG	17 µL
Index D710	TCCGCGAA	17 µL
Index D711	TCTCGCGC	17 µL
Index D712	AGCGATAG	17 µL

4.7.3 Bioinformatics Instructions

1. Install Bowtie 2.
    a. Install Python27 or newer version.
    b. Create a directory for installation (something like C:\Users\jkblab\Desktop\Bowtie).
    c. Download Bowtie 2 (https://sourceforge.net/projects/bowtie-bio/files/bowtie2/2.2.9/bowtie2-2.2.9-mingw-win64.zip/download or newer version).
    d. Unzip it and install in newly created folder.
    e. **Add environmental variables -** Click on Start Menu and right click My Computer. Click on Properties. Click on Advanced System Settings. Under the Advanced tab, click on Environment Variables (this is where we can add Python to the Path environmental variable). Click Edit and add the Variable name: python and Variable value: C:\Python27\ (path to Python installation). Save changes.
2. Create a reference genome using Vector NTI and rename it as a FASTA file (example: HPV16-pUC18.fa). Place it in the Bowtie folder.
3. Align FASTQ result file to the reference genome.
    a. Open the command line (cmd) and navigate to the C:\Users\jkblab\Desktop\Bowtie directory:
    `cd Users\jkblab\Desktop\Bowtie`
    b. Build the reference genome index by entering at the command line:
    `bowtie2-build HPV16-pUC18.fa HPV16-pUC18`
    c. This command will create 6 files with a .bt2 file extension. These will then be used by bowtie2 to map data.
    d. Place the FASTQ file in the C:\Users\jkblab\Desktop\Bowtie directory.
    e. Map the FASTQ file by entering the following command:
    `bowtie2 --score-min C,0 -x HPV16-pUC18 [name of FASTQ file].fastq > [name of desired output SAM file].sam`
    Alternatively, reads with mismatches can be mapped by entering the following command:
    `bowtie2 -x HPV16-pUC18 [name of FASTQ file].fastq > [name of desired output SAM file].sam`
4. Install Strawberry Perl.
    a. Create a directory for installation (something like C:\Users\jkblab\Strawberry).
    b. Download Strawberry Perl (http://strawberryperl.com/download/5.24.0.1/strawberry-perl-5.24.0.1-64bit-portable.zip or newer version).
    c. Unzip it and install in newly created folder.
    d. Run "portableshell.bat" to open Strawberry Perl.
5. Install cpanm.
    a. Download and install cpanm by entering the following command:
    `cpan App::cpanminus`
6. Install CountNicks program written by Christopher Bottoms.
    a. Download Count Nicks (https://github.com/MU-IRCF/Bio-IRCF-CountNicks/blob/master/current_distribution/Bio-IRCF-CountNicks-0.0022.tar.gz or newer version) and copy compressed file in a folder named CountNicks (C:\Users\jkblab\Desktop\CountNicks).
    b. Open portableshell.bat found in the Strawberry Perl folder (C:\Users\jkblab\Strawberry) and navigate to C:\Users\jkblab\Desktop\CountNicks.
    c. Install CountNicks by entering the following command:
    `cpanm Bio-IRCF-CountNicks-0.0022.tar.gz`
7. Count nicks with CountNicks script.
    a. Open portableshell.bat found in the Strawberry Perl folder (C:\Users\jkblab\Strawberry) and navigate to C:\Users\jkblab\Desktop\Bowtie directory.
    b. Run script by entering the following command:
    `Count_nicks [name of sam file].sam`

Chapter 5

Genome-Wide Mapping of Antiviral Hairpin Polyamides using Hydroxyl Radical Footprinting Coupled to Massively Parallel DNA Sequencing (·OH-Seq) and Crosslinking of Small Molecules (COSMIC-Seq)

5.1 ABSTRACT

The rational design of small molecules to selectively target predetermined DNA sequences is a major goal in the development of novel therapeutics. *N*-methylpyrrole/*N*-methylimidazole (Py/Im) hairpin polyamides (PAs) represent a class of synthetic heteroaromatic compounds that can reportedly be rationally designed to target predetermined sequences for the control of gene expression and other biological functions at the DNA level. Here, we have performed COSMIC (crosslinking of small molecules for isolation of chromatin) and hydroxyl radical footprinting coupled with massively parallel DNA sequencing (·OH Seq) to map the binding sites of potent anti-HPV hairpin polyamides in live cells.

5.2 INTRODUCTION

Deciphering both the viral and host genomic sites targeted by our bioactive polyamides (PAs) in live cells will provide important insights into the mechanism of action by which these DNA minor groove binders elicit their functions. These antiviral compounds, belonging to the *N*-methylpyrrole/*N*-methylimidazole (Py/Im) PAs of synthetic molecules, bind avidly to A/T rich sequences in the HPV Long Control Region (LCR) with dissociation constants in the low nanomolar range, leading to an extensive coverage of regions important for viral transcription and replication (see **Chapter 2** and **Chapter 3**).[1-4] They also bind avidly to other A/T-rich regions in the viral genome, as described in **Chapter** 4. To date, we have performed binding studies of our antiviral polyamides across short linear DNA fragments and in a plasmid form under cell-free experiments; however, these experiments disregard the complexity of host and viral genomes found in cells.

In eukaryotic cells, host genomes are packaged in condensed form into nucleosome core particles consisting of 146-bp of dsDNA wound around a histone octamer. These nucleosome core particles fold into higher order structures referred to chromatin, which in turn fold into chromosomes. Similarly, the viral genome is negatively supercoiled, condensed into chromatin-like structures and bound by host and viral proteins in cells.[5-10] Because of the compact nature of genomes in eukaryotic cells, a large number of the potential DNA-binding sequences may be inaccessible to PAs. Although nucleosomal DNA-binding sites that are partially or fully facing away from the histone octamer are accessible to conventional eight-ring polyamides,[11] these sites may not be accessible to our large hairpin polyamides due to their required larger binding site ($\geq$10 bp of DNA). In line with these results, Sugiyama *et al*. reported that alkylating polyamides exhibited limited accessibility toward binding sites found in the central core of well-positioned nucleosomes in isolated nuclei of human cells.[12] Recently, Erwin *et al*. demonstrated that two structurally distinct polyamides (hairpin and linear conformers) were able to populate diverse chromatin architectures including active and repressive states in live H1-hECs.[13] Thus, the relative accessibility of large hairpin PAs to chromatinized genomes requires further investigation.

To investigate the relative chromatin accessibility of our large hairpin, we performed COSMIC (crosslinking of small molecules for isolation of chromatin) and hydroxyl radical footprinting (·OH) coupled with massively parallel DNA sequencing (Seq) to map

the binding sites of anti-HPV hairpin polyamides in H1 human embryonic stem cells (H1-hESC) and HPV16-harboring W12E keratinocytes.

5.2.1 In-Cell Hydroxyl Radical Footprinting

In 1986, Tullius *et al.* introduced hydroxyl radical footprinting as a powerful technique to study the structural interaction details of ligand-DNA complexes under cell-free conditions.[14] The highly-reactive hydroxyl radical moiety is used as a chemical probe to measure the relative accessibility of the cleavage agent toward DNA in the presence of a DNA-ligand. Binding of the ligand to its cognate sequence leads to increased protection of DNA at that particular location, providing a 'footprint'.[15,16] Hydroxyl radicals are an attractive cleavage agent due to their small size, neutral charge, relatively sequence-neutral activity and ease of use.[17,18]

Hydroxyl radicals can be produced in solution by Fenton chemistry[14,19] or ionizing radiation.[20,21] In the former, hydroxyl radicals (·OH) are generated upon the reduction of H_2O_2 by the $[Fe(II)(EDTA)]^{2-}$ complex, while addition of sodium ascorbate reduces Fe(III) to regenerate Fe(II), affording a reaction catalytic in iron (**Equation 5.1**).

$$[Fe(II)(EDTA)]^{2-} + H_2O_2 \xrightarrow[\text{ascorbate}]{} [Fe(III)(EDTA)]^{-} + OH^{-} + \cdot\mathbf{OH} \qquad \textbf{(5.1)}$$

Hydroxyl radicals can also be generated by the gamma radiolysis of water. Of interest, Tullius *et al*. demonstrated using gamma-radiation as the source of hydroxyl radicals that footprinting experiments of the bacteriophage λ repressor/operator complex provided identical cleavage patterns as those generated by Fenton chemistry.[20] Furthermore, the γ-ray technique has also been successfully applied to examine structural details of protein-DNA complexes in cells. In γ-ray methodology, cells are irradiated with γ-rays to generate hydroxyl radicals and the fragmented DNA is then extracted and radioactively end-labelled by primer extension, followed by gel electrophoresis analysis.[21] Recently, Taylor *et al*.[22] devised a novel procedure that permits in-cell hydroxyl radical footprinting using Fenton chemistry (**Equation 5.1**). In this technique, cell membranes are permeabilized by lysolecithin treatment, which allows the Fenton reagents to enter the nucleus, leading to genomic DNA cleavage by hydroxyl radicals.[22]
To map the genome-wide binding sites of bioactive polyamides in HPV16-maintaining W12E cells, we applied hydroxyl radical footprinting by gamma-radiation and Fenton chemistry followed by next-generation sequencing. Footprinting experiments using gamma-radiation were performed in five vehicle-control replicates and six **PA1**-treated biological replicates. Footprinting experiments using Fenton reagents were performed with seven anti-HPV PAs and one inactive PA in five biological replicates. The bioactive compounds were chosen based on their broad range of antiviral activities (IC_{50} = 0.036-0.304 μM) (**Table 5.1**). Chemical structures of these PAs are provided in **Section 5.7.1**.

In these experiments, HPV16-harboring W12E keratinocytes are treated with E media containing 0.1 μM polyamide and a final DMSO concentration of 0.1 %. In other culture plates, cells are simultaneously treated with a vehicle control consisting of E media (complete) with a final DMSO concentration of 0.1 %. After 24 h incubation, the cells are either dissociated with trypsin and subjected to gamma-radiation, or treated with a cell membrane permeabilization agent (*i.e.* lysolecithin) followed by treatment with Fenton

chemistry reagents. After oxidative DNA cleavage, low molecular DNA from W12E cell samples is extracted using a modified Hirt method.[23] The purified DNA is then analyzed by next-generation sequencing as previously described in **Chapter 4** (**Figure 5.1**).

Table 5.1. HPV16 IC_{50} and IC_{90} Values for Polyamides Studied In-Cell Hydroxyl Radical Footprinting Experiments.

Polyamide	DNA Binding Motif	Molecular Weight (g/mol) (TFA salt)	HPV16 IC_{50} [μM]	HPV16 IC_{90} [μM]
PA11	W_2GW_2GW	1564.39	NOT ACTIVE[a]	
PA1	W_2GW_7	2237.12	0.100[a]	1.113[a]
NV1111	$W_{11}GW$	2865.68	0.017[b]	0.121[b]
PA30	W_2GW_7	2464.31	0.304[a]	>10[a]
PA31	W_2GW_7	2408.20	0.103[a]	0.378[a]
NV1078	W_8GW_4	2850.71	0.046[b]	0.307[b]
NV1087	$W_{11}GW$	2921.79	0.031[b]	0.200[b]
PA25	$W_2GW_5GW_4$	2982.19	0.036[a]	0.351[a]

[a] Antiviral activities against HPV16 were previously reported.[3,24,25]

[b] Antiviral activities against HPV16 were determined in W12E monolayers by T.G. Edwards (unpublished data).

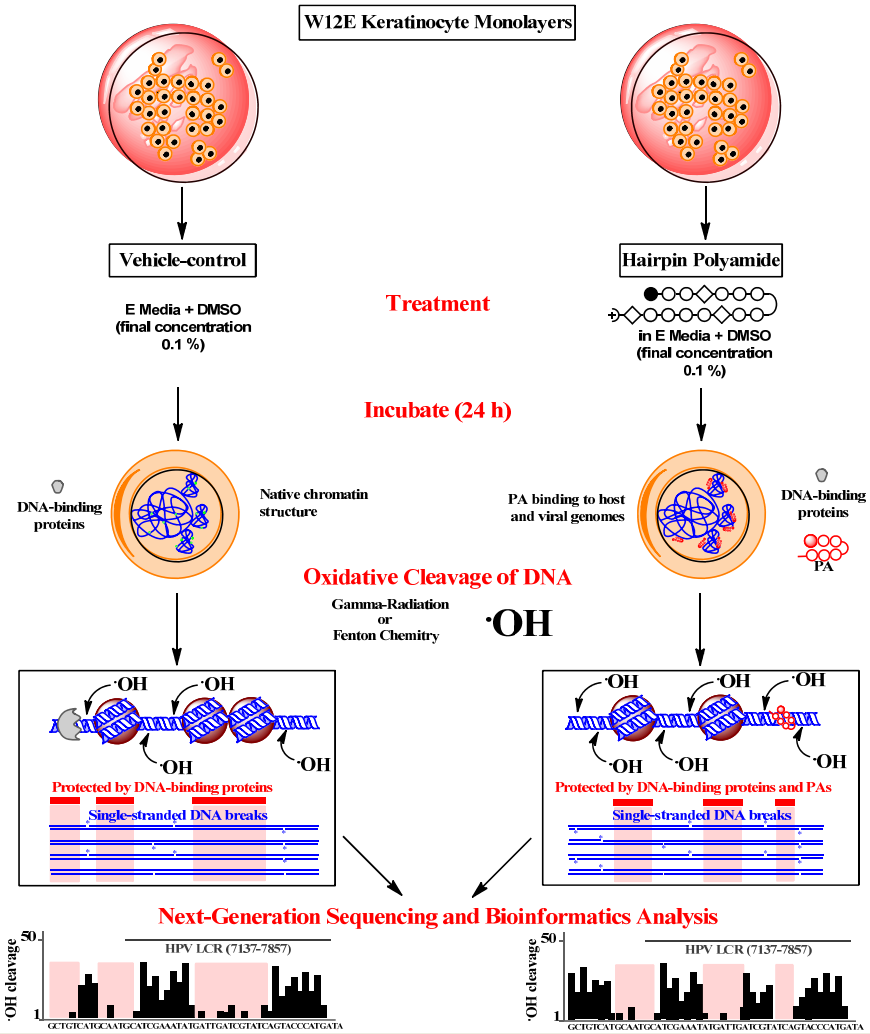


Figure 5.11. Experimental Scheme for ·OH-Seq Experiments. W12E keratinocyte monolayers are separately subjected to vehicle-control or PAs. After 24 h incubation, cells are irradiated with gamma-rays at Missouri Research Reactor or Fenton chemistry reagents. After oxidative DNA cleavage, low molecular DNA is extracted and the fragmented DNA is sequenced by next-generation sequencing. Subsequent bioinformatics analysis is performed to determine the cleavage intensity at single-nt resolution throughout the genome in order to map PA-DNA interactions.

5.2.2 Cognate Site Identification (CSI) Analysis

Cognate Site Identification analysis, developed by Ansari and colleagues, has been successfully applied to examine the DNA-binding preferences of artificial transcription factors and PAs.[26-28] In the SELEX-based approach of CSI, biotinylated DNA-ligands are incubated with a DNA library containing randomized permutations of 20-bp sequences. Affinity capture using streptavidin-coated beads allows for the enrichment of the bound sequences, and subsequent NGS analysis of the enriched sequences provides the specificity profiles of these minor groove DNA-binding ligands. Furthermore, CSI analysis allows for the determination of structural contributions afforded by the sequences flanking the binding motif.[13] Ansari *et al*. have demonstrated that the binding intensities obtained from CSI analysis are directly proportional to the equilibrium association constant values of PA-DNA interactions.[26] As a result, CSI data have been used to predict the probability of binding to specific sequences in a genome-wide manner.[29]

5.2.3 Crosslinking of Small Molecules for Isolation of Chromatin (COSMIC)

In tandem to in-cell hydroxyl radical footprinting experiments, we also performed COSMIC experiments using **PA1** and **PA25** derivatives in H1 human embryonic stem cells at the University of Wisconsin-Madison and in W12E keratinocytes at the University of Missouri-St. Louis in biological duplicates. This method, recently developed by the Ansari group, allows for the mapping of PA interactions with DNA in its native chromatinized form.[29] Indeed, the Ansari group demonstrated that polyamides were able to populate diverse chromatin architectures including actively transcribed and repressed states.[13] Furthermore, the Ansari group reported that polyamides bind similarly to loci presenting multiple clustered binding sites with low- and medium affinities as they bind to loci with few high-affinity binding sites.[13,29]

In COSMIC, the polyamide is conjugated on its C-terminal tail with a bifunctional moiety comprised of psoralen and biotin. The psoralen moiety affords the reversible photo-crosslinking of PA-DNA interactions, whereas biotin allows for the sequence enrichment of photo-crosslinked DNA with streptavidin-coated magnetic beads. Psoralen is a tricyclic furocoumarin which intercalates into DNA base pairs with an affinity of 10^3-10^4 M^{-1} and forms interstrand crosslinks preferentially to thymine bases (5'-TpA sites)[30] upon irradiation with long wavelength UV light (320-380 nm).[29,31] Cells grown as monolayers are treated with the PA-**PB** (psoralen/biotin) conjugate for 24 h. At the end of the treatment period, psoralen-DNA crosslinking is performed at 365 nm under the native chromatin structure in living cells. The cells are lysed and the DNA is sonicated. After sonication, the psoralen-DNA adducts are captured using streptavidin-coated magnetic beads *via* the biotinylated polyamide. After removal of non-specific DNA interactions, the psoralen-DNA crosslinks are reversed by hot alkali treatment and the DNA is purified. The enriched fragments are sequenced and aligned to the genome of interest by next-generation sequencing analysis to uncover genomic regions occupied by PAs (**Figure 5.2**).[13,29]

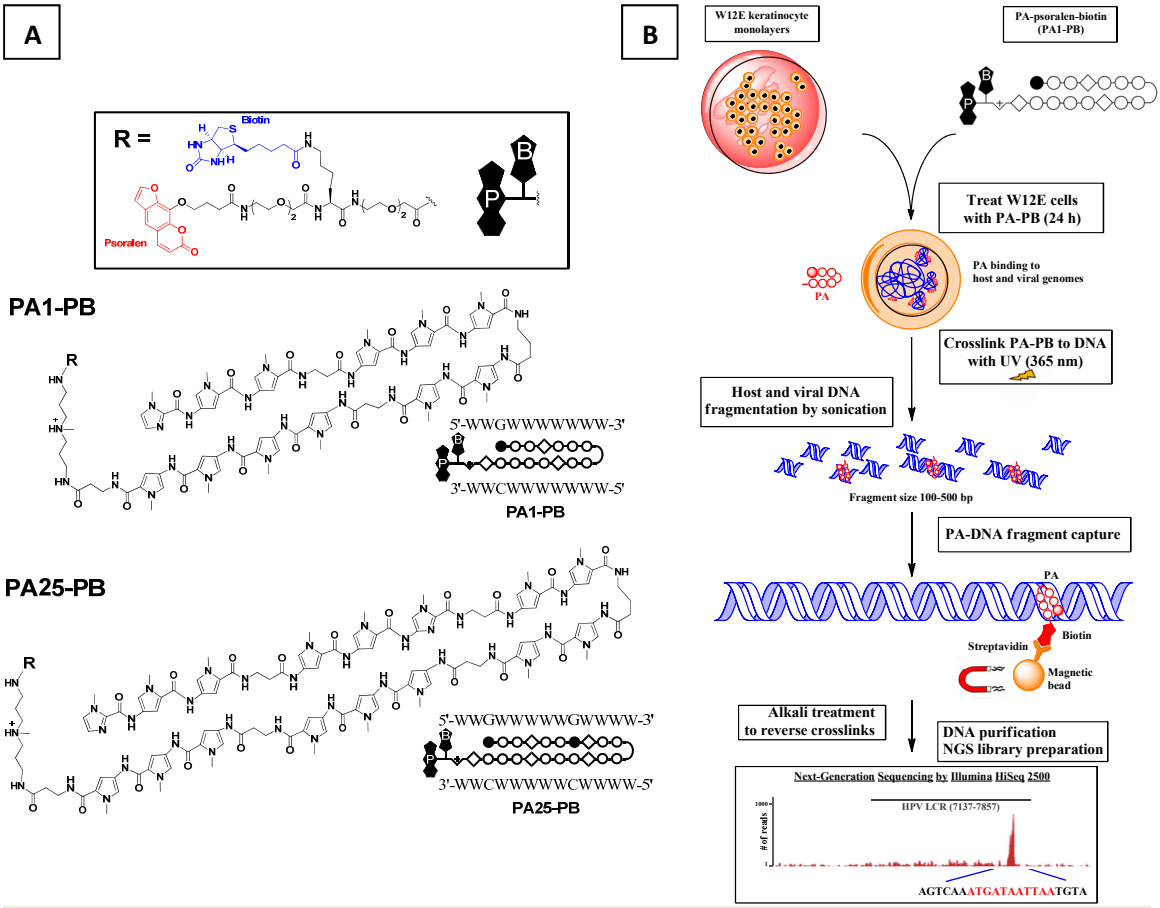


Figure 5.2. Chemical Structures of PA1 and PA25 Functionalized with Psoralen-Biotin (PB) Moiety and Experimental Scheme for COSMIC-Seq. (A) Chemical structures of polyamide-PB conjugates used for COSMIC experiments. According to the PA-DNA recognition rules, **PA1** and **PA25** target the DNA sequences 5'-W_2GW_7-3' and 5'-$W_2GW_5GW_4$-3', respectively (where W can be A or T). Polyamide building blocks are represented by open circles for *N*-methylpyrrole, filled circles for *N*-methylimidazole, diamonds for β-alanine, + for the Ta tail and ⊃ for γ-aminobutyric acid. **(B)** W12E keratinocyte monolayers are treated PA-PB conjugates (**PA1-PB** or **PA25-PB**). The cells are then subjected to UV radiation (365 nm) which leads to the crosslinking of the psoralen moiety to viral and genomic DNA. After the cells are lysed and the DNA is sheared (100-500 bp), streptavidin-coated magnetic beads are used to capture covalent PA-DNA complexes. The captured DNA is then dissociated from beads by alkali treatment to reverse the crosslinks and the released DNA is analyzed by NGS (Illumina HiSeq 2500).

5.3 MATERIALS AND METHODS

5.3.1 Buffers and Reagents

Starting reagents were used without further purification, unless specifically noted. Autoclaved MilliQ H_2O (18.2 MΩ·cm at 25 °C, Millipore Integral 10) was used in the preparation of all reagents. All reagents were filtered using a 0.22 µm MILLEX-GP filter before use. W12E cervical keratinocytes (clonal cell line # 20850) harboring HPV16 episomes and J2 3T3 fibroblasts were a kind gift of Professor Paul Lambert (University of Wisconsin-Madison). Peptide acid psoralen-biotin moiety was synthesized by Dr. Graham Erwin (University of Wisconsin-Madison) as previously described.[29] Biotin-PEG4-NHS ester (Catalog # CP-3026-100MG) was purchased from Conju-Probe. Anhydrous *N*-Methyl-2-pyrrolidinone (NMP) (Catalog # 328634-100ML), trifluoroacetic

acid (Catalog # T6508-2L) and formic acid (Catalog # F0507-100ML) were obtained from Sigma-Aldrich. *N*,*N*-Diisopropylethylamine (DIEA) (Catalog # 400136) was purchased from Applied Biosystems. 1-[Bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-*b*]pyridinium 3-oxid hexafluorophosphate (HATU) (Catalog # AS-60263-100) was obtained from AnaSpec, Inc. HPLC-grade methanol (MeOH) (Catalog # AH230-4) was purchased from Honeywell Burdick & Jackson. LCMS Chromasolv acetonitrile (Catalog # 34967-4L) was obtained from Fluka. Ham's F-12K (Kaighn's) medium (Gibco Catalog # 21127-030), Dulbecco's Modified Eagle's Medium with high glucose, L-glutamine, phenol red, with pyruvate (Gibco Catalog # 11995-081) and without pyruvate (Gibco Catalog # 11965-126), 0.05 % and 0.25 % Trypsin-EDTA with phenol red (Gibco Catalog # 25300-054 and 25200-056) and fetal bovine serum (Catalog # 16140-071) were obtained from ThermoFisher Scientific. Epidermal growth factor (Catalog # PHG0311) was purchased from Invitrogen. Penicillin-streptomycin solution (10,000 U penicillin / 10 mg streptomycin per mL; Catalog # P4333-20ML), bovine serum albumin (Catalog # A4919-1G), 1 M HEPES buffer (Catalog # 83264-100ML-F), hydrocortisone powder (Catalog # H0888-1G), adenine hydrochloride hydrate (Catalog # A9795-1G), insulin from bovine pancreas (Catalog # I6634-50MG), human apo-transferrin (Catalog # T1147-100MG), T3 thyroid hormone, 3,3',5-triiodo-L-thyronine sodium salt (Catalog # T6397-100MG), phenol red (Catalog # P3532-5G), dimethyl sulfoxide (Catalog # D2650) and cholera toxin (Catalog # C9903-.5MG) were purchased from Sigma-Aldrich. 1X phosphate buffer saline, pH 7.4 (Catalog # 10010-031), mytomycin C (Catalog # BP2531-2), sodium chloride (Catalog # BP358-1), potassium chloride (Catalog # BP366-500), sodium bicarbonate (Catalog # BP328-1), sodium phosphate monobasic monohydrate (Catalog # BP330-500), magnesium sulfate heptahydrate (Catalog # M80-500) and ferric nitrate nonahydrate (Catalog # I110-100) were obtained from Fisher Scientific. Sodium hydroxide (Catalog # BP359-500) and hydrogen peroxide (30 % wt/vol; Catalog # H325-500) were purchased from Fisher Scientific. 0.22 μm MILLEX-GP filters were purchased from Merck Millipore Ltd. Sucrose (Catalog # 84097) was obtained from Fluka. 100 % ethanol (Catalog # 2716) was obtained from Decon Laboratories, Inc. Potassium chloride (Catalog # P9333-500G), potassium phosphate dibasic trihydrate (Catalog # P5504-100G), magnesium chloride (Catalog # M8266-100G), calcium chloride (Catalog # C1016-100G), 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (Catalog # H3375-100G), lysolecithin Type I (Catalog # L4129-100MG), thiourea (Catalog # 88810-100G), EDTA disodium salt dehydrate (Catalog # E5134-100G), ammonium iron(II) sulfate hexahydrate (Catalog # 203505-25G), sodium ascorbate (Catalog # 11140-50G) and agarose (Catalog # A9539-100G) were purchased from Sigma-Aldrich. Buffer P1 (Catalog # 19051) and QIAprep 2.0 Spin Columns (Catalog # 27115) were obtained from QIAGEN. Glacial acetic acid (Catalog # A38-212), 1 M Tris-HCl [pH 7.5] (Catalog # 15567-027) and RNase A (Catalog # EN0531; 10 mg/mL) were purchased from ThermoFisher Scientific. Sodium dodecyl sulfate (Catalog # L4509-100G), cesium chloride (Catalog # C4036-50G) and potassium acetate (Catalog # P1190-100G) were purchased from Sigma-Aldrich. TaqMan probes and primers used for real-time quantitative PCR were purchased from Integrated DNA Technologies. CalSun facial tanning sun lamp (Catalog # B001BH0A1A) was obtained from CalSun. Trizma base (Catalog # T1503-1KG), Triton X-100 (Catalog # X100-100ML), deoxycholate acid sodium salt (Catalog # D6750-10G) and urea (Catalog

U5128-100G) were purchased from Sigma-Aldrich. Streptavidin-coated magnetic beads (Dynabeads® MyOne™ Streptavidin C1, Catalog # 65001), benzamidine HCl (Catalog # BP435-25), concentrated HCl (Catalog # A144C-212), proteinase K (Catalog # FEREO0491; 20 mg/mL) and nuclease-free water (Catalog # R0581) were purchased from ThermoFisher Scientific. NP40 (IGEPAL CA-630; Catalog # 198596) was purchased from MP Biomedicals. QIAquick PCR purification kit (Catalog # 28104) was obtained from QIAGEN. Phenylmethylsulfonyl fluoride (PMSF) (Catalog # 786-055) was purchased from G-Biosciences. Pepstatin A (Catalog # J60237-LB0) and potassium hydroxide (Catalog # A16199-0E) were obtained from Alfa Aesar. 1 M Tris-HCl buffer [pH 8.0] (Catalog # 15-568-025) was purchased from Invitrogen. Lithium chloride (Catalog # 199880050) was obtained from Acros Organic. Gel Loading Dye, Blue (6X, Catalog # B7021S) and 1 kb DNA ladder (Catalog # N3232S, 500 µg/mL) were purchased from New England Biolabs. 1X TAE buffer consisted of 40 mM Tris-acetate, 1 M EDTA, pH 8.0. Ethidium bromide was obtained from Spectrum (Catalog # E1031, 1 % solution). Agilent Bioanalyzer RNA 6000 Nanochip kit (Catalog # 5067-1511) was purchased from Agilent Technologies. Qubit dsDNA BR (Broad-Range) Assay kit (Catalog # Q32853), Qubit dsDNA HS (High Sensitivity) Assay kit (Catalog # Q32851) and Qubit assay tubes (Catalog # Q32856) were obtained from Life technologies. Accel-NGS 1S Plus DNA Library Kit (Catalog # DL-IL1SP-12/48) was purchased from Swift Biosciences. Low EDTA TE buffer (Catalog # T0227; 10 mM Tris·Cl, 0.1 mM EDTA, pH 8.0) was purchased from Teknova. SPRIselect reagent kit (Catalog # B23317) was obtained from Beckman Coulter. Agilent DNA 1000 kit (Catalog # 5067-1505) was obtained from Agilent Technologies.

5.3.2 Polyamide Conjugates Synthesis

Hairpin polyamide synthesis was performed by Dr. K. J. Koeller, Dr. G. D. Harris, Jr. and K. Aston. Hairpin polyamides were synthesized by Boc solid-phase methods[32] as previously reported in the literature.[24]

Biotinylated PA conjugates **KA2127-PEG4-Biotin**, **PA1-PEG4-Biotin** and **PA25-PEG4-Biotin** (**Figure 5.3**) were synthesized by reacting a solution of 1 eq. of biotin-PEG4-NHS ester (Conju-Probe, Catalog # CP-3026-100MG) in anhydrous NMP (Sigma-Aldrich, Catalog # 328634-100ML) with a solution of 1.5 eq. of PA (either **KA2127** formate salt, **PA1** TFA salt or **PA25** TFA salt), 50 eq. of DIEA in anhydrous NMP at room temperature for 3-4 h.

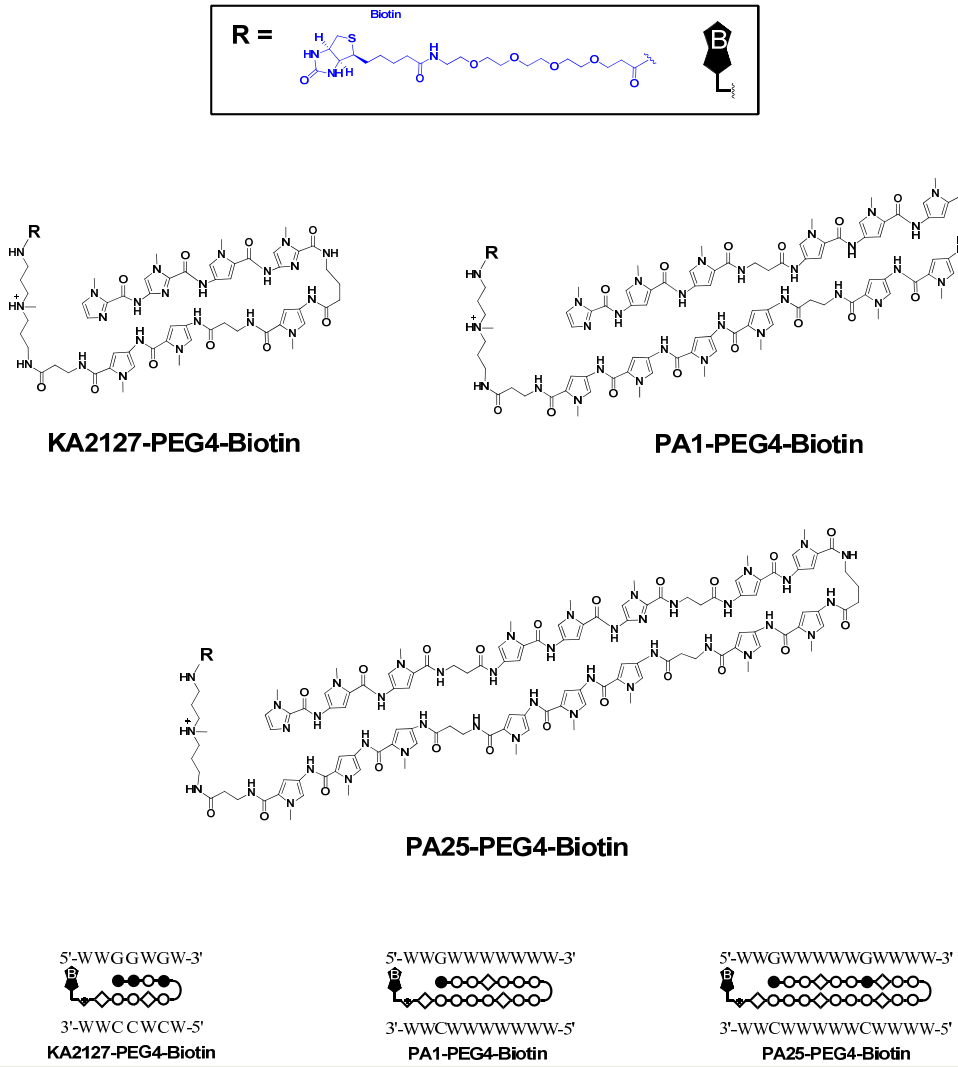


Figure 5.3. Chemical Structures of KA2127, PA1 and PA25 Functionalized with PEG4-Biotin. Chemical structures of polyamide-PEG4-Biotin conjugates used for Cognate Site Identification (CSI) experiments. Biotin conjugation to these polyamides was achieved *via* the primary amine on their TA tails. The PEG spacer minimizes any steric hindrance that may arise from the attachment of the polyamide moiety in the binding of streptavidin by biotin. According to the PA-DNA recognition rules, **KA2127** targets the DNA sequence 5'-W_2GGWGW-3', where W (Weak) can be A or T. Large hairpin polyamides **PA1** and **PA25** target the DNA sequences 5'-W_2GW_7-3' and 5'-$W_2GW_5GW_4$-3', respectively. Polyamide building blocks are represented by open circles for *N*-methylpyrrole, filled circles for *N*-methylimidazole, diamonds for β-alanine, + for the Ta tail and ⊃ for γ-aminobutyric acid.

Covalent conjugation of anti-HPV polyamides **PA1** and **PA25** to a psoralen-biotin moiety (**PB**)[29] afforded **PA1-PB** and **PA25-PB**. Briefly, a solution of 1 eq. of the peptide acid of **PB** in anhydrous NMP was treated with a solution of 3 eq. of PA (either **PA1** or **PA25** as TFA salts), 0.95 eq. HATU and 3 eq. of DIEA in anhydrous NMP at room temperature for at least 2 h. Biotinylated and **PB**-conjugated PAs were then purified by reverse-phase HPLC with a Dionex Ultimate 3000 prep-LC using a Macherey-Nagel Nucleodur C18 column (2 columns in series) (21.0 mm x 50 mm each) and program Gilson M-N short. Mobile phases were 0.2 % formic acid in water (A) and 100 % methanol (B). The gradient consisted of a flowrate of 30 mL/min with a ramp from 25 % to 75 % B over 35 min, followed by a ramp from 75 % to 90 % B over 1 min and held for

1 min at 90 % B. Retention times for biotinylated polyamide conjugates were 17.6, 21.3 and 23.1 min for **KA2127-PEG4-Biotin**, **PA1-PEG4-Biotin** and **PA25-PEG4-Biotin**, respectively. Retention times for polyamide-**PB** conjugates were 21.0 and 24.3 min for **PA1-PB** and **PA25-PB**, respectively (Chromeleon Client Version 6.80 SR10). Fractions were checked with an Agilent Technologies 1260 Infinity analytical HPLC using a Phenomenex Jupiter Proteo (C12) (4.6 x 50 mm, 4 μm, 90 Å) maintained at 40 ˚C. Mobile phases consisted of 0.1 % formic acid in water (A) and 100 % acetonitrile (B). The gradient consisted of 5 % B for 0.75 min followed by a ramp to 60 % B over 3.25 min at 2 mL/min. Retention times for biotinylated polyamide conjugates were 2.80, 3.02 and 3.10 min for **KA2127-PEG4-Biotin**, **PA1-PEG4-Biotin** and **PA25-PEG4-Biotin**, respectively. Retention times for polyamide-**PB** conjugates were 3.18 and 3.25 min for **PA1-PB** and **PA25-PB**, respectively (**Figure 5.4**). Fractions containing pure conjugates were pooled together and lyophilized to yield off-white solids (see **Section 5.7.3** for detailed protocol). Chemical identity was verified with HRMS-ESI (high-resolution mass spectrometry electrospray ionization) by Professor B. Bythell (**Table 5.2**).

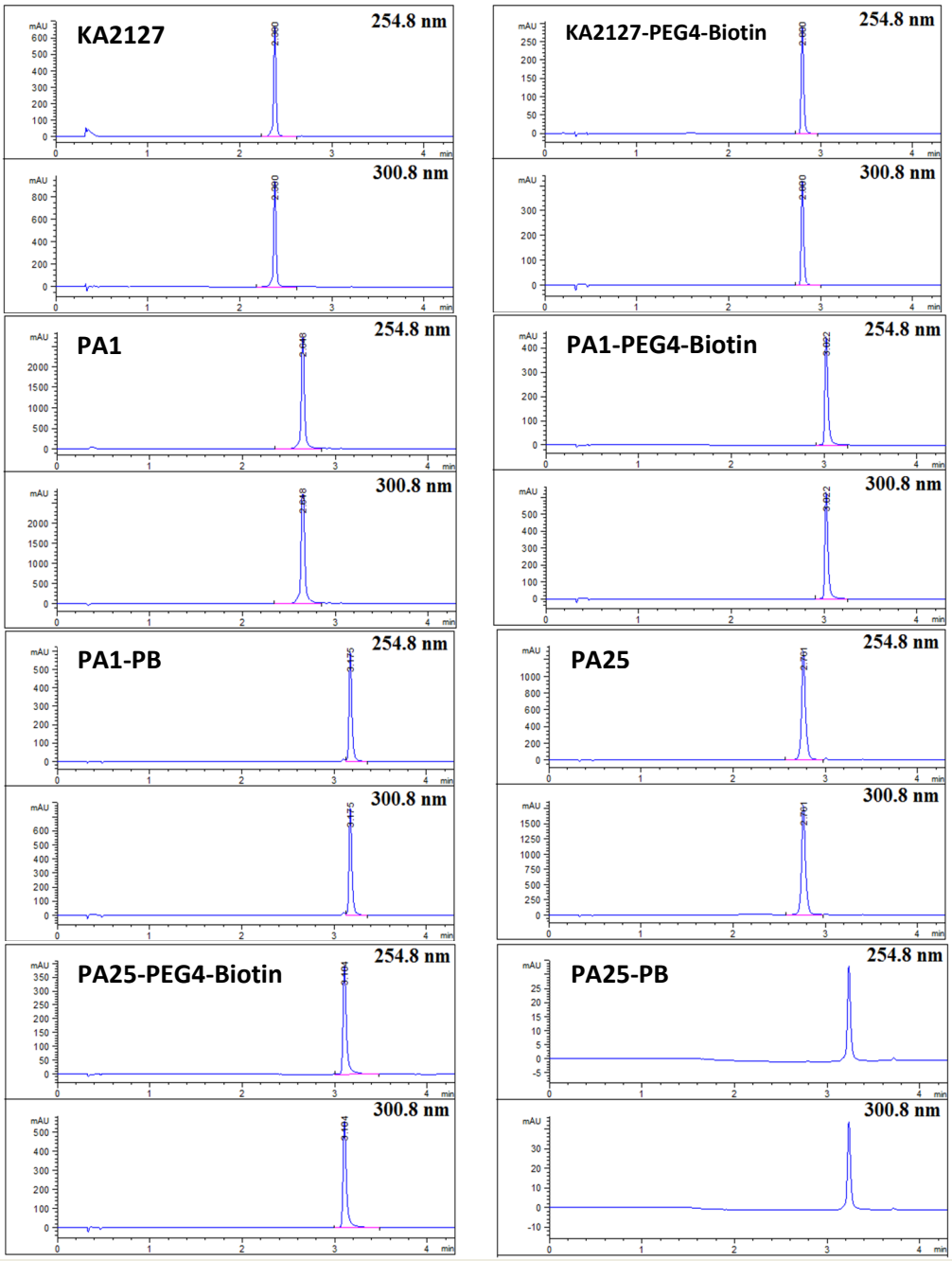


Figure 5.4. Analytical HPLC Traces of Synthesized Polyamide Conjugates with λ Monitored (254.8 and 300.8 nm). Analytical HPLC Traces of polyamide-**PEG4-Biotin** and polyamide-**PB** conjugates used for CSI and COSMIC experiments. HPLC chromatograms were analysed using ChemStation for LC 3D systems Rev. B.04.03[16].

Table 5.2. Characterization of Synthesized Polyamide Conjugates by HRMS-ESI.

Polyamide	Chemical Formula	Exact Mass	Mass Calculated (m/z)	Mass Observed (m/z)
KA2127-B	$C_{77}H_{109}N_{25}O_{17}S$	1687.82	**$[M+2Na]^{2+}$** 866.90	866.93
PA1-B	$C_{112}H_{146}N_{34}O_{23}S$	2367.10	**$[M+2Na]^{2+}$** 1206.54	1206.58
PA1-PB	$C_{134}H_{169}N_{37}O_{30}S$	2808.26	**$[M+H]^{+}$** 2809.27	2809.23
PA25-B	$C_{141}H_{179}N_{45}O_{29}S$	2998.36	**$[M+2Na]^{2+}$** 1522.17	1522.23
PA25-PB	$C_{163}H_{202}N_{48}O_{36}S$	3439.52	**$[M+2Na]^{2+}$** 1742.75	1742.81

5.3.3 Cognate Site Identification (CSI) by High-Throughput SELEX

CSI experiments for **KA2127-PEG4-Biotin**, **PA1-PEG4-Biotin**, **PA25-PEG4-Biotin**, **PA1-PB** and **PA25-PB** were performed at the University of Wisconsin-Madison as previously described.[13] Briefly, polyamide conjugates at three different concentrations (0.01 nM, 1nM and 100 nM) were incubated at room temperature for 1 h with 100 nM of a DNA library containing randomized permutations of 20-bp sequences. The binding buffer consisted of 1X PBS [pH 7.6] with 50 ng/µL Poly(dI-dC) as a competitor to reduce non-specific binding interactions. Polyamide-DNA complexes were affinity purified with streptavidin-coated magnetic beads (Dynabeads® MyOneTM Streptavidin C1, Catalog # 65001). After affinity capture, the beads were washed three times with 100 µL cold binding buffer and resuspended in a PCR master mix. PCR amplification was then performed for 15 cycles, followed by DNA purification with the Econospin Purification Kit (Epoch Life Sciences). The purified, enriched DNA was then used as the input DNA for a subsequent round of enrichment. Illumina sequencing adapters and a unique 6-bp barcode for multiplexing were separately incorporated to each enriched DNA sample and to the starting library by PCR amplification. High-throughput Illumina HiSeq 2500 sequencing on these samples was performed by the University of Wisconsin-Madison DNA Sequencing Facility. Demultiplexing based on the 6-bp barcode was subsequently performed on the sequencing reads to segregate the results into separate data sets. The reads were truncated to 20-bp corresponding to the random region of the library. CSI data analysis was performed as previously described in the literature (**Figure 5.5**).[13]

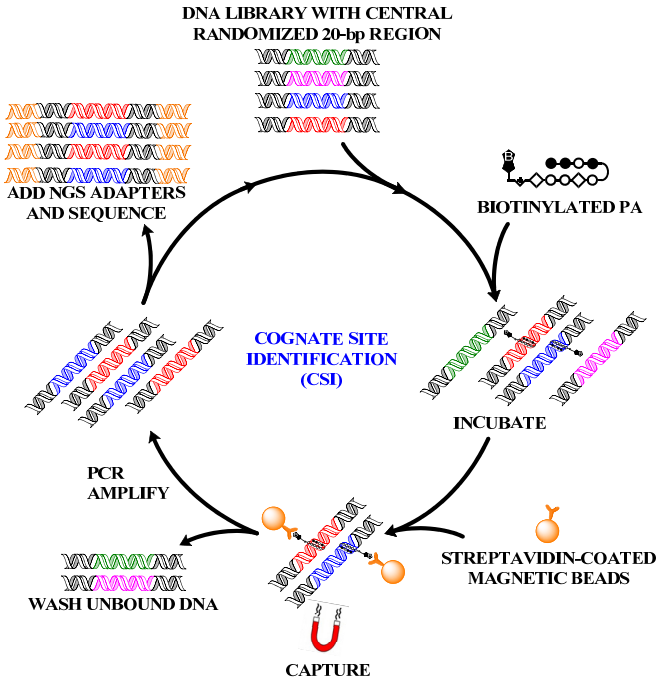


Figure 5.5. Experimental Scheme for Cognate Site Identification (CSI) Analysis. Biotinylated polyamide conjugates are incubated for 1 h at room temperature with a DNA library containing randomized permutations of 20-bp sequences. Non-covalent polyamide-DNA complexes are affinity purified with streptavidin-coated magnetic beads followed by washes to remove unbound DNA. PCR amplification and DNA purification is performed. The purified, enriched DNA is then used as the input DNA for a subsequent round of enrichment. Illumina sequencing adapters and a unique 6-bp barcode for multiplexing are separately incorporated to each enriched DNA sample and to the starting library by PCR amplification. The samples are analyzed by massively parallel DNA sequencing.

5.3.4 Cell Culture

W12E cervical keratinocytes (clonal cell line # 20850) and J2 3T3 fibroblasts were a kind gift of Professor Paul Lambert (University of Wisconsin-Madison). W12E keratinocytes maintaining HPV16 episomes (Genbank Accession No. AF125673) were grown as monolayer cultures in the presence of mitomycin C-treated J2 3T3 fibroblasts as previously described.[24,33-35] Cells were grown at 37 ˚C with a 5 % CO_2 atmosphere (NuAire NU-4750 US AutoFlow CO_2 Water-Jacketed) in E media consisting of three parts Dulbecco's modified Eagle medium (DMEM) and one part Ham's F-12 media. This media was further supplemented with 5 % fetal bovine serum (FBS), 24 µg/mL adenine, 5 µg/mL transferrin, 20 pM 3,3',5-triiodo-L-thyronine (T_3), 0.4 µg/mL hydrocortisone, 10 ng/mL cholera toxin, 5 ng/mL epidermal growth factor (EGF), 5 µg/mL insulin, 100 U/mL penicillin and 100 µg/mL streptomycin. Keratinocytes were passaged at $\leq$ 70 % confluency by trypsin-EDTA treatment (Gibco Catalog # 25300-054) (see **Section 5.7.4.** for detailed cell culture protocols). The number of cells and viability were measured by trypan blue staining (Sigma, Catalog # T8157, 0.4 % solution) using a Cellometer Auto T4 Cell Viability Counter and SD100 cell counting chambers (Nexcelom Bioscience) following the manufacturer's instructions.

Although W12E cells are undifferentiated, differentiation can be induced upon confluence.[24] Thus, passage of W12E keratinocytes was performed at 70 % confluence since differentiation leads to virion packaging, a major safety concern.

5.3.5 Modified Hirt Method: Extraction of HPV16 Episomal DNA

W12E keratinocytes (10^6-10^7 cells) were harvested and lysed in 25 mM Tris-HCl [pH 8.0], 0.6 % SDS, 10 mM EDTA and 50 µg/mL RNase A for 5 min at room temperature. Denatured proteins and chromosomal DNA were then precipitated by the addition of a solution consisting of 3 M cesium chloride, 1 M potassium acetate and 0.67 M acetic acid. The samples were incubated for 15 min at 4 °C and centrifuge. Low molecular DNA from each sample supernatant was then purified using QIAprep 2.0 Spin Columns (QIAGEN) (see **Section 5.7.6** for detailed protocol).

5.3.6 Compound Efficacy Testing: *Taqman®* Real-Time PCR Antiviral Assay

We verified the antiviral activities for **PA1** and **PA25** against HPV16 episomes using W12E keratinocytes maintained in the UMSL cell culture facility. Polyamide antiviral activities were determined by measuring the decrease of episomal HPV16 DNA as a function of polyamide concentration in monolayer W12E keratinocyte cultures. The episomal load was determined by quantitative PCR (Q-PCR) as previously described.[24] Briefly, lyophilized single-use aliquots (0.1 mg) of **PA1** or **PA25** were dissolved in 100 % sterile DMSO (Sigma Life Science, Catalog # D2650) to 10 mM final concentration, and further diluted to 1 mM with autoclaved MilliQ H_2O. Polyamide concentrations were verified by measuring the absorbance at λ_{max} (~305 nm) using a Thermo Scientific Evolution 260 Bio UV-visible spectrophotometer, and molar extinction coefficients of 88,235 $M^{-1}cm^{-1}$ and 147,400 $M^{-1}cm^{-1}$ for **PA1** and **PA25**, respectively.

HPV16-harboring W12E keratinocytes were grown at 37 °C, 5 % CO_2 atmosphere on 10-cm culture dishes in the presence of mitomycin C-treated J2 3T3 fibroblasts in 10 mL

of supplemented E media (complete). After removal of the spent media, cells were separately treated with 10 mL of E media containing PA concentrations ranging from 0.001-3 µM with a final DMSO concentration of 0.1 %. Cells were also treated with a vehicle control consisting of E media (complete) with a final DMSO concentration of 0.1 %. After 24 h incubation at 37 °C and 5 % CO_2 atmosphere, viral DNA was extracted by a modified Hirt method.[23] Episomal load for each sample was quantified by Q-PCR.

Real-time Q-PCR was performed using a CFX96 Real-Time System/C1000 Thermal Cycler (Bio-Rad) as previously described.[24] TaqMan® probe and primers were purchased from Integrated DNA Technologies. The probe consisted of a 5' reporter 6-carboxyfluorescein dye and a 3' quencher 6-carboxytetramethylrhodamine dye. Primers and probes were designed to amplify and quantify an 84-bp region within the HPV16 L1 gene. PCR primers and probe sequences are provided below:

Sense (forward primer): 5′-TGGAGGACTGGAATTTTGGTCTA-3′
Antisense (reverse primer): 5′-CAATTGCCTGGGATGTTACAAA-3′
TaqMan® probe: 5′-56-FAM-TCCCCCAGGAGGCACACTAGAAGATACTT-TAMRA-3′

A standard curve was generated using known concentrations of purified HPV16/pUC18 plasmid DNA (total length = 10,589 bp; see **Chapter 4 Section 4.3.3**). The standard curve covered concentrations of 10^9 to 10^2 episomal molecules per microliter. **Equation 5.2** was used to calculate total episomal molecules per microliter for each standard point.

$$\text{Episomal molecules per microliter} = \frac{(1.82\ x\ 10^{15})(\mu g/\mu L\ std\ point)}{(10{,}589\ bp)(2)} \quad \textbf{(5.2)}$$

Q-PCR reactions contained 2 µL of template DNA, a final concentration of 1X Standard *Taq* Reaction Buffer (NEB, Catalog # B9014S), 200 µM Deoxynucleotide (dNTP) Solution Mix (NEB, Catalog # N0447S, 10 mM), 200 nM each primer and probe and 0.5 U *Taq* DNA Polymerase (NEB, Catalog # M0273L). The Q-PCR reactions were performed with the following thermal profile: 95 °C for 10 min, 39 cycles of 95 °C for 15 s and 60 °C for 70 s, followed by fluorescence reading at the end of each cycle. Q-PCR reagent amounts and parameters used for the PCR reactions are summarized in **Table 5.3**.

Table 5.3. Q-PCR Reagent Amounts and Run Parameters. The amounts given are for a 25 µL reaction.

Program	Reagent	Volume (µL)	PCR Step	Time	
QPCR HPV16.prcl	Autoclaved MilliQ H_2O	18.4	Polymerase Activation at 95 °C	10 min	
	10X Standard *Taq* Buffer	2.5	Denaturation at 95 °C	15 s	39 X
	10 mM dNTP Mix	0.5			
	10 µM Forward Primer	0.5			
	10 µM Reverse Primer	0.5	Annealing/Extension at 60 °C	70 s	
	10 µM TaqMan® Probe	0.5			
	Template DNA	2.0			
	Taq DNA Polymerase	0.1	Plate Read		

Deoxynucleotide (dNTP) Solution Mix (Catalog # N0447S, 10 mM), 10X Standard *Taq* Reaction Buffer (Catalog # B9014S) and *Taq* DNA Polymerase (Catalog # M0273L, 5,000 U/mL) were purchased from New England Biolabs.

Q-PCR results were processed and analyzed by CFX Manager 3.1 (Bio-Rad). The HPV16 episome number was calculated using the standard curve, and the effect of

polyamide on episomal load was plotted as a percent decrease compared to cells treated with the vehicle control. The data were fitted to the Sigmoidal equation included in Kaleidagraph 4.1.1 (Synergy Software) to determine the polyamide antiviral activities.

5.3.7 Hydroxyl Radical (·OH) Footprinting in W12E Keratinocytes

5.3.7.1 PA1 In-Cell Footprinting Using Gamma-Radiation

In-cell PA-DNA interactions were interrogated by hydroxyl radical footprinting experiments relying on gamma-radiation to produce hydroxyl radicals[20,21] in W12E cells treated with vehicle controls or **PA1**. These experiments were performed in five control and six PA-treated biological replicates. Briefly, a lyophilized single-use aliquot (0.1 mg) of **PA1** was dissolved in 100 % sterile DMSO to 10 mM final concentration, and further diluted to 1 mM with autoclaved MilliQ H_2O. Polyamide concentrations were verified by UV/Vis spectrophotometry.

HPV16-harboring W12E keratinocytes were grown at 37 °C, 5 % CO_2 atmosphere on 10-cm culture dishes in the presence of mitomycin C-treated J2 3T3 fibroblasts in 10 mL of supplemented E media (complete). After removal of the spent media, cells were separately treated with 10 mL of E media containing 0.1 μM **PA1** and a final DMSO concentration of 0.1 %. Cells were also treated with a vehicle control consisting of E media (complete) with a final DMSO concentration of 0.1 %. After 24 h incubation at 37 °C and 5 % CO_2 atmosphere, the media was removed and the cells were washed with 10 mL 1X PBS. Versene treatment was then performed to selectively remove 3T3 fibroblasts, followed by dissociation of W12E cells with trypsin. Trypsin was then quenched with an equal volume of E media (complete), and the W12E cells were pelleted by centrifugation. Harvested W12E keratinocytes (approximately 2 x 10^6 cells per sample) in 2 mL of E media (incomplete) and transferred into a 2.0 mL microcentrifuge tube with an internal sealing O-ring (Stockwell Scientific, Catalog # 3306S). These samples were placed in ice. The large volume used for these experiments increases the effective radiation target, thus increasing the rate of hydroxyl radicals generated in solution.[20] W12E samples were subjected to gamma-radiation for 19.5 min (140 kRad) from a ^{60}Co nuclide source (120 Rad/sec) at 37 °C at the Missouri University Research Reactor (MURR) (**Figure 5.6**). Briefly, a sample vial containing either HPV16/pUC18 plasmid or W12E cells was place in an aluminum screw cap container (Elemental Container, Catalog # 0027127; 99.7 % pure aluminum, diameter = 1 inch, height = 5 inches) with a 1” aluminum spacer on the bottom and a 2” aluminum spacer on the top of the sample vial. The aluminum container was then capped (Elemental Container, Catalog # 0027CAP) and tested for leaks. Aluminum spacers served two purposes: 1) they raised the sample to the appropriate height for optimum irradiation, and 2) they weighted down the sample to ensure it would remain submerged during irradiation. Aluminum container was then placed into a submerged DAGIF II aluminum holder with the ^{60}Co gamma-ray source in the middle position.

After gamma-radiation, low molecular DNA from W12E cell samples was extracted using a modified Hirt method[23] and stored at -20 °C.

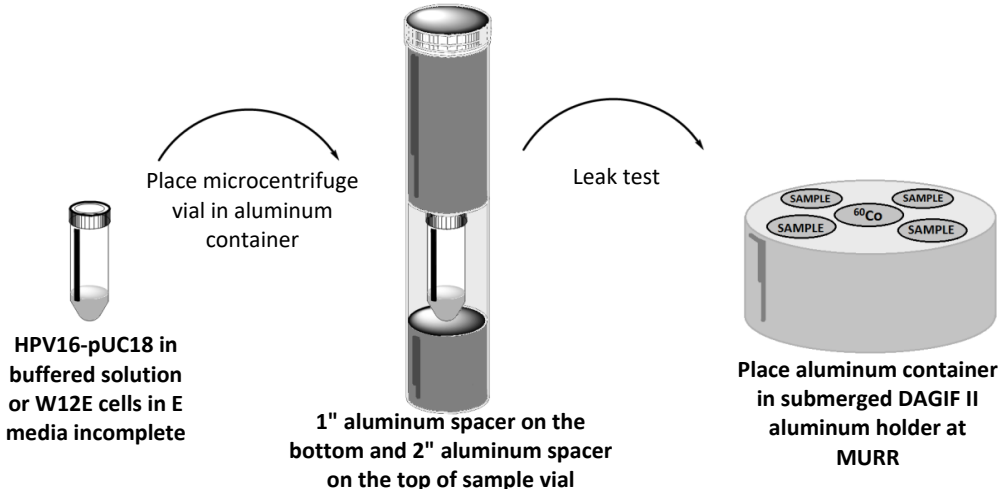


Figure 5.6. DAGIF II Gamma-Radiation Exposure Scheme. Plasmid DNA samples in microcentrifuge vials were subjected to gamma-rays from a ^{60}Co nuclide source (120 Rad/sec) submerged in water at 37 °C. The gamma-radiation exposure setup consisted of placing the microcentrifuge vial in an aluminum container in between a 1" aluminum spacer on the bottom and a 2" aluminum spacer on the top of the sample vial. These containers were leak tested prior to placing them into a submerged DAGIF II aluminum holder containing the ^{60}Co gamma-ray source.

5.3.7.2 In-Cell Hydroxyl Radical (·OH) Footprinting using Fenton Chemistry

In-cell hydroxyl radical footprint was also performed using a method developed by Taylor *et al*. as previously described, but with minor modification.[22] In this method, hydroxyl radicals are generated in cells by the Fenton reaction (**Equation 5.1**), but in contrast to the standard hydroxyl radical procedure used for *in vitro* experiments, a 50-fold and 30-fold higher concentration of sodium ascorbate and hydrogen peroxide is required.[22] In-cell hydroxyl radical footprint experiments were performed in five biological replicates for each treatment regimen. W12E monolayer cells were treated with polyamides exhibiting a broad range of anti-HPV16 activities: one inactive and seven active compounds (**Table 5.1**). Briefly, W12E keratinocytes were grown at 37 °C, 5 % CO_2 atmosphere on 10-cm culture dishes in the presence of mitomycin C-treated J2 3T3 fibroblasts in 10 mL of supplemented E media (complete). Lyophilized single-use aliquots (0.1 mg) of each polyamide were dissolved in 100 % sterile DMSO to 1 mM final concentration, and further diluted to 10 μM with 10 % sterile DMSO in autoclaved MilliQ H_2O. After removal of the spent media, cells were separately treated with 10 mL of E media containing 0.1 μM PA with a final DMSO concentration of 0.1 %. Cells were also treated with a vehicle control consisting of E media (complete) with a final DMSO concentration of 0.1 %. After 24 h incubation at 37 °C and 5 % CO_2 atmosphere, Versene (0.02 % EDTA in 1X PBS) treatment of the plates was performed to selectively remove 3T3 fibroblasts. The attached W12E cells were washed three times with 10 mL PBS, followed by the addition of 10 mL 0.05 % lysolecithin permebilization solution (0.05 % lysolecithin in 150 mM sucrose, 80 mM KCl, 5 mM K_2HPO_4, 5 mM $MgCl_2$, 0.5 mm $CaCl_2$ and 35 mM NaHepes buffer [pH 7.4]) and incubated for 60 s at room temperature. Incubation with the 0.05 % lysolecithin permeabilization solution affords for cell membrane permeabilization allowing the hydroxyl radical reagents to enter the cells. A hydroxyl radical footprinting solution was freshly made by adding 495.5 mg sodium ascorbate (Sigma, Catalog # 11140-50G) to 50 mL of DMEM media (Gibco, Catalog #

11965-084) and mixed by vortexing. Immediately, 1 mL of freshly prepared 50X Fe(II)-EDTA stock (500 mM EDTA and 250 mM $Fe(NH_4)_2(SO_4)_2 \cdot 6H_2O$) plus 1 mL of 30 % H_2O_2 were added to the ascorbate/DMEM media and vortexed for approximately 5 s. After removal of the 0.05 % lysolecithin permeabilization solution, 10 mL of the hydroxyl radical footprinting solution was added onto each W12E culture, followed by gentle swirling of the plates to evenly disperse the solution. The plates were then incubated for 15 minutes at room temperature. At the end of the incubation period, the hydroxyl radical footprinting solution was aspirated and the W12E cells were washed twice with 10 mL stop buffer consisting of 0.76 g of thiourea in 100 mL. The cells were washed with 10 mL PBS, harvested by scrapping and centrifugation (see **Section 5.7.5** for detailed protocol). Low molecular DNA from each sample was then extracted using a modified Hirt method.[23]

5.3.7.3 Assessment of Fragment Size Distribution by Agilent Bioanalyzer RNA 6000 Nanochip Kit

For the assessment of ssDNA fragment size distribution produced by the cleavage reaction, an Agilent Bioanalyzer RNA 6000 Nanochip kit was used following the manufacturer's instructions. Briefly, 4 µL of modified Hirt-extracted DNA was heat denatured at 95 °C for 3 min using a Mastercycler Nexus Thermal Cycler (Eppendorf) and immediately placed on ice until ready to load on the Nanochip. Then, 1 µL of ssDNA was loaded onto a sample well containing 5 µL of Agilent RNA marker solution. The RNA 6000 Nanochip was analyzed in an Agilent 2100 Bioanalyzer using the Eukaryote Total RNA Nano Series II assay conditions. The results were analyzed using Agilent Technologies 2100 Bioanalyzer 2100 Expert Version B.02.08.SI648 (SR3) software.

5.3.7.4 Quantification of dsDNA using Qubit dsDNA BR and HS Assay Kit

DNA concentration of the modified Hirt-extracted DNA from W12E cells and prepared libraries were quantified using Qubit dsDNA BR and HS assay kits with a Qubit 2.0 Fluorometer (Life technologies) following the manufacturer's instructions. Briefly, Qubit dsDNA BR/HS Standards #1 and #2 were prepared by separately diluting 10 µL of either Standard #1 (0 ng/ µL) or Standard #2 (100 ng/µL) in 190 µL of Qubit working solution (1:200 Qubit dsDNA BR/HS Reagent:Qubit dsDNA BR/HS buffer) in separate Qubit assay tubes (Life technologies, Catalog # Q32856). Unknowns were prepared by diluting 2 µL of cleavage reaction products or 4 µL of Illumina NGS library sample to a total volume of 200 µL with Qubit working solution in separate Qubit assay tubes. Standards and unknown solutions were vigorously vortexed and allowed to incubate at room temperature for at least 2 min. After the incubation period, the fluorescence signal of each of these solutions was quantified with a Qubit 2.0 Fluorometer (Life technologies). The fluorescence signal emitted by the proprietary Qubit dsDNA BR/HS Reagent fluorescent dye upon intercalation into dsDNA is directly proportional to the amount of dsDNA in the sample, which allows for the calculation of dsDNA concentration based on a two-point standard calibration curve. The Qubit 2.0 Fluorometer provides a dsDNA concentration in µg/mL of the diluted sample. **Equation 5.3** was used to calculate dsDNA concentration of the sample.

$$Concentration\ of\ unknown = Qubit\ 2.0\ Fluorometer\ value\ (\mu g/mL)\ \text{x}\ \frac{200}{x} \qquad \textbf{(5.3)}$$

$$where\ x\ is\ the\ volume\ of\ sample\ used\ in\ microliters$$

5.3.8 Illumina Next-Generation Sequencing of Hydroxyl Radical Footprinting Experiments

5.3.8.1 Preparation of Illumina Next-Generation Sequencing Libraries

Illumina next-generation sequencing libraries were constructed using an Accel-NGS 1S Plus DNA Library Kit (Swift Biosciences, Catalog # SI-IL1SP-12A/DI-IL1SP-48) following the manufacturer's protocol (refer to **Chapter 4 Section 4.7.2** for detailed protocol). Thermal cycling steps were performed with a Mastercycler Nexus Thermal Cycler. Briefly, a 15 µL (5 ng/µL) of ·OH double-hit fragments in Low EDTA TE buffer was heat denatured in a thermal cycler at 95 °C for 2 min and immediately placed on ice for 2 min before proceeding to the Adaptase step. The Adaptase step is a proprietary reaction that performs fragment end repair, 3' end tailing and 3' ligation of a truncated adapter. 25 µL of the Adaptase Reaction Mix was added to the 15 µL sample aliquot containing the denature ssDNA and cycled with the following thermal profile: 37 °C for 15 min, 95 °C for 2 min and 4 °C hold. Following 3' end repair and truncated adapter ligation, adapter-ligated DNA molecules were amplified by adding 47 µL of the Extension Reaction Mix and cycling with the following thermal profile: 98 °C for 30 s, 63 °C for 15 s, 68 °C for 5 min and 4 °C hold. The newly synthesized strand allows for the ligation of the second truncated adapter to the 5' end of the DNA fragment. This synthesized strand is not sequenced. In order to remove oligonucleotides and small fragments, paramagnetic SPRI-based (Solid Phase Reversible Immobilization) bead purification of the Extension step was performed with a SPRIselect reagent kit (Beckman Coulter, Catalog # B23317). Using a 1.2 ratio of SPRIselect reagent to sample, DNA fragments greater than or equal to 200 bp were size-selected and eluted with 20 µL of Low EDTA TE buffer. Addition of the second truncated adapter were achieved by adding 20 µL of the Ligation Reaction Mix and cycled with the following thermal profile: 25 °C for 15 min and 4 °C hold. Clean-up of the Ligation step was then performed with a 1.0 ratio of SPRIselect reagent:sample to size-select for dsDNA fragments of sizes $\geq$ 200 bp using a 12-tube magnetic separation rack (NEB, Catalog # S1509S), followed by elution from the beads with 20 µL of Low EDTA TE buffer. Addition of the full-length indexed adapters was performed by adding 5 µL of Reagent R1 (index primer, SI-IL1SP-12A) and 25 µL of the PCR Reaction Mix. **Table 5.4** and **Table 5.5** provide the specific indexing adapters used for each sample. The reaction mixture was then amplified by cycling with the following thermal profile: 98 °C for 30 s, 8 cycles of 98 °C for 10 s, 60 °C for 30 s and 68 °C for 60 s, followed by a 4 °C hold. The amplified NGS libraries were then size-selected with a 0.85 ratio of SPRIselect reagent:sample and eluted with 20 µL of Low EDTA TE buffer. This was followed by a right side size selection SPRI of the libraries in order to remove the majority of DNA fragments $\geq$ 800 bp (see **Section 5.8** for detailed protocol). Fragment size distribution was assessed by Agilent Bioanalyzer 2100 analysis using an Agilent DNA 1000 kit (Agilent Technologies, Catalog # 5067-1505).

DNA concentration was determined using a Qubit 2.0 Fluorometer with a Qubit dsDNA BR Assay kit.

Table 5.4. PA1 In-Cell Footprinting Using Gamma-Radiation – Summary of NGS Library Preparation Parameters.

Sample	Adapter (DI-IL1SP-48)	Index Sequence	Adapter (DI-IL1SP-48)	Index Sequence	Number of PCR Cycles
Control-1 (CHC4035-2SS)	D501	TATAGCCT	D701	ATTACTCG	8
Control-2 (CHC4049-3 SS)	D502	ATAGAGGC	D702	TCCGGAGA	8
Control-3 (CHC4054-1 SS)	D506	TAATCTTA	D706	GAATTCGT	8
Control-4 (CHC4054-2 SS)	D507	CAGGACGT	D707	CTGAAGCT	8
Control-5 (CHC4054-3 SS)	D508	GTACTGAC	D708	TAATGCGC	8
PA1-1 (0.1 μM) (CHC4049-4 SS)	D503	CCTATCCT	D703	CGCTCATT	8
PA1-2 (0.1 μM) (CHC4049-5 SS)	D504	GGCTCTGA	D704	GAGATTCC	8
PA1-3 (0.1 μM) (CHC4049-6 SS)	D505	AGGCGAAG	D705	ATTCAGAA	8
PA1-4 (0.1 μM) (CHC4054-4 SS)	D501	TATAGCCT	D709	CGGCTATG	8
PA1-5 (0.1 μM) (CHC4054-5 SS)	D502	ATAGAGGC	D710	TCCGCGAA	8
PA1-6 (0.1 μM) (CHC4054-6 SS)	D503	CCTATCCT	D711	TCTCGCGC	8

Accel-NGS 1S Plus DNA Library Kit was used to prepare NGS libraries. Index sequence refers to a unique DNA sequence identifier in the adapter that allows for multiple libraries to be sequenced simultaneously in a sample lane.

Table 5.5. In-Cell Footprinting Using Fenton Chemistry – Summary of NGS Library Preparation Parameters.

Biological Replicate #	Sample	Adapter DI-IL1SP-48	Sequence	Adapter DI-IL1SP-48	Sequence	Number of PCR Cycles
1	Control-1 (CHC4102-1 SS)	D504	GGCTCTGA	D712	AGCGATAG	8
	PA11-1 (0.1 μM) (CHC4102-2 SS)	D505	AGGCGAAG	D701	ATTACTCG	8
	PA1-1 (0.1 μM) (CHC4102-3 SS)	D506	TAATCTTA	D702	TCCGGAGA	8
	PA25-1 (0.1 μM) CHC4102-4 SS	D507	CAGGACGT	D703	CGCTCATT	8
	NV1078-1 (0.1 μM) CHC4102-5 SS	D508	GTACTGAC	D704	GAGATTCC	8
	NV1087-1 (0.1 μM) CHC4102-6 SS	D501	TATAGCCT	D705	ATTCAGAA	8
	NV1111-1 (0.1 μM) CHC4102-7 SS	D502	ATAGAGGC	D706	GAATTCGT	8
	PA30-1 (0.1 μM) CHC4102-8 SS	D503	CCTATCCT	D707	CTGAAGCT	8
	PA31-1 (0.1 μM) CHC4102-9 SS	D504	GGCTCTGA	D708	TAATGCGC	8

2	Control-2 CHC4114-1 SS	D505	AGGCGAAG	D709	CGGCTATG	8
	PA11-2 (0.1 µM) CHC4114-2 SS	D506	TAATCTTA	D710	TCCGCGAA	8
	PA1-2 (0.1 µM) CHC4114-3 SS	D507	CAGGACGT	D711	TCTCGCGC	8
	PA25-2 (0.1 µM) CHC4114-4 SS	D508	GTACTGAC	D712	AGCGATAG	8
	NV1078-2 (0.1 µM) CHC4114-5 SS	D501	TATAGCCT	D701	ATTACTCG	8
	NV1087-2 (0.1 µM) CHC4114-6 SS	D502	ATAGAGGC	D702	TCCGGAGA	8
	NV1111-2 (0.1 µM) CHC4114-7 SS	D503	CCTATCCT	D703	CGCTCATT	8
	PA30-2 (0.1 µM) CHC4114-8 SS	D504	GGCTCTGA	D704	GAGATTCC	8
	PA31-2 (0.1 µM) CHC4114-9 SS	D505	AGGCGAAG	D705	ATTCAGAA	8
3	Control-3 CHC4117-1 SS	D501	TATAGCCT	D706	GAATTCGT	8
	PA11-3 (0.1 µM) CHC4117-2 SS	D502	ATAGAGGC	D707	CTGAAGCT	8
	PA1-3 (0.1 µM) CHC4117-3 SS	D503	CCTATCCT	D708	TAATGCGC	8
	PA25-3 (0.1 µM) CHC4117-4 SS	D504	GGCTCTGA	D709	CGGCTATG	8
	NV1078-3 (0.1 µM) CHC4117-5 SS	D505	AGGCGAAG	D710	TCCGCGAA	8
	NV1087-3 (0.1 µM) CHC4117-6 SS	D506	TAATCTTA	D711	TCTCGCGC	8
	NV1111-3 (0.1 µM) CHC4117-7 SS	D507	CAGGACGT	D712	AGCGATAG	8
	PA30-3 (0.1 µM) CHC4117-8 SS	D508	GTACTGAC	D701	ATTACTCG	8
	PA31-3 (0.1 µM) CHC4117-9 SS	D501	TATAGCCT	D702	TCCGGAGA	8
4	Control-4 CHC4133-1 SS	D502	ATAGAGGC	D703	CGCTCATT	8
	PA11-4 (0.1 µM) CHC4133-2 SS	D503	CCTATCCT	D704	GAGATTCC	8
	PA1-4 (0.1 µM) CHC4133-3 SS	D504	GGCTCTGA	D705	ATTCAGAA	8
	PA25-4 (0.1 µM) CHC4133-4 SS	D505	AGGCGAAG	D712	AGCGATAG	8
	NV1078-4 (0.1 µM) CHC4133-5 SS	D506	TAATCTTA	D701	ATTACTCG	8
	NV1087-4 (0.1 µM) CHC4133-6 SS	D507	CAGGACGT	D708	TAATGCGC	8
	NV1111-4 (0.1 µM) CHC4133-7 SS	D508	GTACTGAC	D709	CGGCTATG	8
	PA30-4 (0.1 µM) CHC4133-8 SS	D507	CAGGACGT	D710	TCCGCGAA	8
	PA31-4 (0.1 µM) CHC4133-9 SS	D503	CCTATCCT	D711	TCTCGCGC	8

Biological Replicate #	Sample	Adapter DI-IL1SP-12A	Sequence	Number of PCR Cycles
5	Control-5 CHC4141-1 SS	Reagent R1 (I2)	CGATGT(A)	8
	PA11-5 (0.1 µM) CHC4141-2 SS	Reagent R1 (I4)	TGACCA(A)	8
	PA1-5 (0.1 µM) CHC4141-3 SS	Reagent R1 (I5)	ACAGTG(A)	8
	PA25-5 (0.1 µM) CHC4141-4 SS	Reagent R1 (I6)	GCCAAT(A)	8
	NV1078-5 (0.1 µM) CHC4141-5 SS	Reagent R1 (I7)	CAGATC(A)	8
	NV1087-5 (0.1 µM) CHC4141-6 SS	Reagent R1 (I12)	CTTGTA(A)	8
	NV1111-5 (0.1 µM) CHC4141-7 SS	Reagent R1 (I13)	AGTCAA(C)	8
	PA30-5 (0.1 µM) CHC4141-8 SS	Reagent R1 (I14)	AGTTCC(G)	8
	PA31-5 (0.1 µM) CHC4141-9 SS	Reagent R1 (I15)	ATGTCA(G)	8

Accel-NGS 1S Plus DNA Library Kit was used to prepare NGS libraries. Index sequence refers to a unique DNA sequence identifier in the adapter that allows for multiple libraries to be sequenced simultaneously in a sample lane.

5.3.8.2 Quality Control of NGS Libraries

Prepared libraries were quantified by Qubit 2.0 Fluorometer using a Qubit dsDNA BR/HS assay kit and the fragment size distribution was determined by Agilent Bioanalyzer 2100 analysis using an Agilent DNA 1000 kit per manufacturer's directions. Briefly, 1 µL of the library sample was loaded onto a sample well containing 5 µL of Agilent DNA marker solution. The DNA 1000 Nanochip was then analyzed in an Agilent 2100 Bioanalyzer using the DNA 1000 Series II assay conditions. The results were analyzed using Agilent Technologies 2100 Bioanalyzer 2100 Expert Version B.02.08.SI648 (SR3) software. **Equation 5.4** was used to calculate the molar concentration of each NGS library.

$$Molar\ concentration\ of\ NGS\ library = [dsDNA]by\ Qubit\ 2.0\ (\mathrm{ng/\mu L})\ x\ 10^6\ \mu\mathrm{L/L}\ x\ \frac{\mathrm{nmol}}{660\ \mathrm{ng}}\ x\ \frac{1}{N} \quad \textbf{(5.4)}$$

where N is the average fragment size of the library determined by Agilent Qubit 1000 *kit*

The molar concentration of the constructed NGS libraries was also verified by the DNA Core Facility at the University of Missouri using a Qubit 3.0 Fluorometer (Life technologies) prior to high-throughput Illumina HiSeq 2500 or NextSeq 500 sequencing.

5.3.8.3 Illumina Next-Generation Sequencing

High-throughput Illumina sequencing of ·OH-treated W12E samples in the presence or absence of polyamides were performed by the DNA Core Facility at the University of Missouri using single-end, 51 base pair read-length on an Illumina HiSeq 2500. The libraries corresponding to the hydroxyl radical footprinting (Fenton chemistry) experiment in the fifth biological regimen were sequenced using a single-end, 75 base

pair read-length on an Illumina NextSeq 500. Samples were multiplexed per lane since the libraries were constructed using different indexing adapters.

5.3.9 Bioinformatics Analysis of NGS Results from Hydroxyl Radical Footprinting Experiments

5.3.9.1 Sequence Alignment and Analysis of ·OH Cleavage Intensities

Demultiplexing was subsequently performed on the sequencing reads to segregate the sequencing results into separate data sets. Unique sequencing reads were aligned to the HPV16 reference genome (GenBank Accession No. AF125673) and human genome (hg19) using Bowtie 2 (version 2.2.9).[36] The aligned, or mapped reads in the resulting SAM file were then processed with CountNicks.pm to determine the ·OH cleavage intensity at each nucleotide (see **Chapter 4 Section 4.7.3** for bioinformatics instructions).

5.3.10 Crosslinking of Small Molecules for Isolation of Chromatin (COSMIC)

COSMIC experiments using **PA1-PB** and **PA25-B** were performed in H1 human embryonic stem cells (H1-hESC) at the University of Wisconsin-Madison in biological duplicate as previously described.[13] Briefly, approximately 2.5 x 10^7 H1-hESC cells were separately treated with 400 nM of **PA1-PB** or **PA25-PB** in E8 media with a final DMSO concentration of 0.1 %.

COSMIC experiments using **PA1-PB** and **PA25-B** were also performed in W12E keratinocyte monolayers at The University of Missouri-St. Louis in biological duplicate as previously described.[13] Briefly, approximately 2.5 x 10^7 H1-hESC cells were separately treated with 8 mL of E media (complete) containing 100 nM of **PA1-PB** or **PA25-PB** with a final DMSO concentration of 0.1 %. After 24 h incubation at 37 °C and 5 % CO_2 atmosphere, W12E cells were irradiated for 30 min at room temperature with an UV lamp (CalSun, Catalog # B001BH0A1A; 2.4 mW/cm^2) using a glass filter. The latter prevents DNA thymine dimer formation by removing light with wavelengths lower than 300 nm. After crosslinking, Versene (0.02 % EDTA in 1X PBS) treatment of the plates was performed to selectively remove 3T3 fibroblasts. The attached W12E cells were washed with 4 mL PBS, followed by dissociation of W12E cells with trypsin. Trypsin was then quenched with an equal volume of E media (complete), and the W12E cells were pelleted by centrifugation (500*x g* at 4 °C for 5 min). Pelleted cells were resuspended in COSMIC buffer (20 mM Tris-HCl [pH 8.1], 2 mM EDTA, 150 mM NaCl, 1 % Triton-X100, 0.1 % SDS, 1 mM PMSF, 1 mM benzamidine, 1.5 µM pepstatin). Cell lysates were then sonicated in a QSonica Q800R sonicator for 2 h at 4 °C with a cycle of 10 s **on** and 10 s **off** at 50 % power to yield DNA fragments in a range of 100-500 bp. Samples were centrifuged at 12,000*x g* at 4 °C for 10 min and about 10 % of each sample was saved in a new microcentrifuge tube. These samples served as a reference of non-enriched DNA (Input DNA) and they were stored at -80 °C. The remaining of the chromatin samples were treated with streptavidin-coated magnetic beads (120 µL / sample; Dynabeads MyOne Streptavidin C1, Catalog # 65001) to capture PA-DNA crosslinks. Briefly, the chromatin samples were incubated with streptavidin-coated magnetic beads, previously washed with COSMIC buffer, for 12 h at 4 °C on a

shaker/rotator (Barnstead/Thermolyne Labquake 400110). At the end of the incubation period, the samples were placed on a magnetic stand (NEB, Catalog # S1509S) and the immobilized streptavidin beads were washed with 1 mL of COSMIC buffer supplemented with 5 µM of non-functionalized PA (**PA1** or **PA25**) once for 12 h and once for 4 h at 4 °C on a shaker/rotator each. Streptavidin beads were washed once with 1 mL Wash buffer 1 (10 mM Tris-HCl [pH 8.0], 1 mM EDTA, 3 % v/v SDS, supplemented with 5 µM of respective non-functionalized PA) and once with 1 mL Wash buffer 2 (10 mM Tris-HCl [pH 8.0], 1 mM EDTA, 250 mM LiCl, 0.5 % NP40, 1 % sodium deoxycholate, 5 µM of respective non-functionalized PA) for 7 min at room temperature each on a shaker/rotator. Streptavidin beads were then washed twice with Wash buffer 3 (10 mM Tris-HCl [pH 7.5], 1 mM EDTA, 4 M urea, 0.1 % NP40, 5 µM of respective non-functionalized PA) for 7 min at 50 °C on a shaker/rotator. Streptavidin beads were washed twice with 1 mL aliquots of TE buffer (10 mM Tris-HCl buffer [pH 8.0], 1 mM EDTA) and resuspended in 200 µL of TE buffer. These samples of capture DNA are referred to as Affinity-purified (AP) DNA. Psoralen-DNA crosslinks in the Input and AP DNA samples were reversed by the addition of 10X crosslink reversal buffer (100 mM Tris-HCl [pH 8.0], 4 mM EDTA, 1M KOH) to a final concentration of 1X, followed by incubation for 30 min at 90 °C. Upon crosslink reversal by hot alkali treatment, DNA is released from the streptavidin beads onto the supernatant. Sample supernatants were transferred to clean amber microcentrifuge tubes and neutralized with approximately 1 µL concentrated HCl. Samples were then treated with RNase A (Thermo Scientific, Catalog # EN0531)(0.2 µg/µL final concentration) for 1 h at 37 °C, followed by treatment with Proteinase K (Thermo Scientific, Catalog # FEREO0491)(0.2 µg/µL final concentration) for 1 h at 55 °C. Input and AP DNA samples were purified with a QIAquick PCR purification kit (QIAGEN, Catalog # 28104) (see **Section 5.7.7** for detailed protocol). Purified DNA samples were stored -80 °C. Illumina next-generation sequencing libraries for input and AP DNA will be prepared with the TruSeq ChIP Sample Prep Kit (Illumina). High-throughput Illumina HiSeq 2500 sequencing with a read-length of 51 bp will be performed by the University of Wisconsin-Madison DNA Sequencing Facility. Demultiplexing based on the 6-bp barcode will be subsequently performed on the sequencing reads to segregate the results into separate data sets. Analysis of the COSMIC-Seq data will be performed as previously described.[13]

5.3.10.1 Assessment of Fragment Size Distribution by 1.5 % Agarose Gel Electrophoresis

Fragment size distributions of sheared DNA achieved for different sonication times were resolved by agarose gel electrophoresis. Briefly, a 20 µL aliquot of sonicated lysate sample was mixed with 5 µL of 6X electrophoresis loading buffer and 5 µL autoclaved MQH_2O. Samples were loaded onto a 1.5 % agarose gel, and electrophoresis was performed in 1X TAE buffer at a voltage of 110 V for 1.5 h at room temperature. After electrophoresis, the agarose gel was stained with 0.5 µg/mL ethidium bromide for 20 min and washed with autoclaved MilliQ H_2O for 5 min in a staining box. The fragmented DNA and DNA ladders (NEB, Catalog # N3232S and N3236S) were visualized with a FOTO/Convertible UV-light box equipped with an ethidium bromide filter (FOTODYNE

Incorporated). The agarose gel was imaged and analyzed with FOTO/Analyst PC Image version 5.00.

5.4 RESULTS

We relied on Cognate Site Identification (CSI) analysis of a solution-based SELEX (systematic evolution of ligands by exponential enrichment) method coupled to next-generation sequencing[13,28] to determine the binding preferences of **PA1** and **PA25** for the library interrogated. Furthermore, we applied multiple methodologies, including hydroxyl radical footprinting and COSMIC, to map the genome-wide binding sites of seven anti-HPV polyamides and one inactive polyamide in living cells.

5.4.1 Cognate Site Identification (CSI) Analysis

Cognate Site Identification analysis, a SELEX-based approach developed by Ansari and colleagues, was used to examine the binding preferences of three hairpin polyamides (**KA2127**, **PA1** and **PA25**). Toward this goal, we synthesized biotinylated conjugates of these PAs and incubated each with a DNA library containing randomized permutations of 20-bp sequences. Affinity capture using streptavidin-coated beads allows for the enrichment of the bound sequences, and subsequent NGS analysis of the enriched sequences provides the specificity profiles of these minor groove DNA-binding ligands. Furthermore, CSI analysis allows for the determination of structural contributions afforded by the sequences flanking the binding motif.[13] Ansari *et al*. have demonstrated that the binding intensities obtained from CSI analysis are directly proportional to the equilibrium association constant values of PA-DNA interactions.[26] As a result, these data has been used to predict the probability of binding to specific sequences in a genome-wide manner.[29] Because binding cooperativity between nearby sites can pre-organize the minor groove of DNA to convert an incompatible site into a site with the appropriate structural features for PA-binding,[19] a subset of binding sites may not be predicted by this analysis. Next-generation sequencing has been completed for the CSI experiments and bioinformatics analysis is currently in progress.

5.4.2 PA1 and PA25 Antiviral Activity (IC_{50})

Previous cell culture work and antiviral assays were performed by our collaborators at NanoVir. To validate our local cell culture and virology conditions with previous results, we verified the antiviral activities for **PA1** and **PA25** against HPV16 episomes using W12E cells maintained in the UMSL cell culture facility.

The episomal load was determined by quantitative PCR (Q-PCR) as previously described.[24] The standard curve generated from cloned HPV16/pUC18 plasmid DNA was linear over the range used (10^9 to 10^2 HPV16 copies per μL) (**Figure 5.7A**). Based on this standard curve, HPV16 episomal levels in modified Hirt[23] extracts were quantified in monolayer W12E keratinocytes treated with vehicle controls and increasing concentrations of **PA1** or **PA25**. Then, polyamide antiviral activities were determined by measuring the decrease of episomal HPV16 DNA as a function of polyamide concentration (**Figure 5.7B**). The apparent IC_{50} values for **PA1** and **PA25** were 0.13 and

0.010 µM, respectively (**Figure 5.7C**). These values are in agreement with the published IC_{50} values of 0.1 and 0.036 µM for **PA1** and **PA25**.[24]

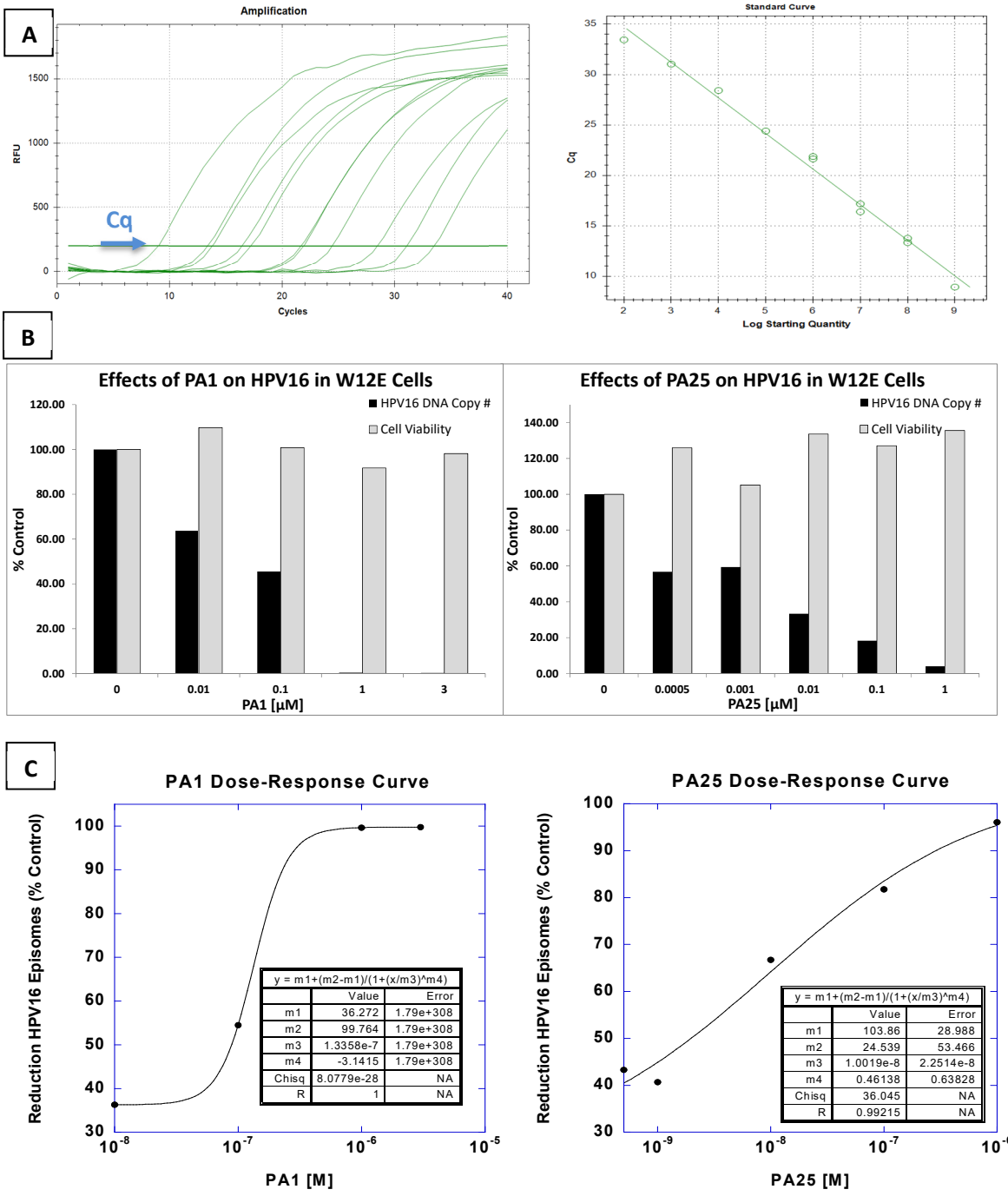


Figure 5.7. Antiviral Activities of PA1 and PA25 Determined by Taqman® Real-Time Q-PCR Assay. (A) HPV16 episomal load was determined using a standard curved generated from known concentrations of cloned HPV16/pUC18 DNA. The representative Q-PCR amplification curve illustrates the linear correlation between the log of HPV16 copies and the quantification cycle (Cq) (slope = -3.529 and y-intercept = 41.795; thus, HPV16 copies/µL = $10^{[(Cq-41.795)/-3.529]}$. **(B)** The effects of **PA1** and **PA25** on HPV16 episomal levels and cell viability of W12E keratinocytes monolayers are plotted as a percent decrease compared to cells treated with the vehicle control. **(C)** Dose-response curves for **PA1** and **PA25** against HPV16 in W12E keratinocytes monolayers. Inset tables provide calculated IC_{50} values (m3).

5.4.3 Hydroxyl Radical Footprinting Coupled to Massively Parallel DNA Sequencing (·OH-Seq) in W12E Keratinocytes

Here we have used hydroxyl radicals generated by either gamma-radiation or Fenton chemistry to produce short single-stranded fragments derived from two ·OH nicks in close proximity within the same DNA strand in W12E cells. Comparison of the relative cleavage intensity, as determined by Illumina next-generation sequencing, between vehicle control-treated cells and polyamide-treated cells should allow us to determine relative ligand-mediated protection within the viral and human genomes. Since oxidative damage to DNA by hydroxyl radicals leads to single-stranded breaks along the DNA molecule, we employed the Accel-NGS 1S Plus DNA Library Kit for library preparation due to its compatibility with denatured, single-stranded, and nicked samples (see **Chapter 4 Section 4.2.2** for hydroxyl radical-DNA chemistry)

5.4.3.1 PA1 In-Cell ·OH Footprinting Using Gamma-Radiation – DNA Concentrations and Fragment Size Distribution of Low Molecular DNA Extracted from W12E Cells Chemistry

In-cell **PA1**-DNA interactions were interrogated by hydroxyl radical footprinting experiments relying on gamma-radiation to produce hydroxyl radicals[20,21] in W12E keratinocytes. In this technique, production of hydroxyl radicals is mediated by gamma radiolysis of water. These experiments were performed in five biological replicates consisting of vehicle control-treated cells, and six biological replicates of cells treated with 0.1 μM **PA1** in E media (final DMSO concentration of 0.1 %) for 24 h at 37 ˚C with a 5 % CO_2 atmosphere. HPV16-harboring W12E cells were subjected to gamma-radiation for 19.5 min from a ^{60}Co nuclide source (120 Rad/sec) at 37 °C at the MURR. The irradiation period was empirically determined. The total radiation load was approximately 140 kRad, which is about twice the amount (75 kRad) previously used by Tullius *et al*. to study protein-DNA complexes in bacterial cells.[21] After gamma-radiation, the low molecular fragments derived from ·OH double-hit kinetics were extracted using a modified Hirt method.[23] DNA sample concentrations were determined by a Qubit dsDNA HS Assay kit (**Table 5.6**) and the fragment size distribution of a representative sample was estimated using the RNA 6000 Nanochip assay (**Figure 5.8**).

Table 5.6. **PA1** In-Cell ·OH Footprinting by Gamma-Radiation – DNA Concentrations of Samples Determined by Qubit dsDNA HS Assay

Sample	Qubit dsDNA Assay (ng/μL)	Sample	Qubit dsDNA Assay (ng/μL)
Control-1 (CHC4035-2)	34.7	**PA1-1 (0.1 μM) (CHC4049-4)**	24.5
Control-2 (CHC4049-3)	31.3	**PA1-2 (0.1 μM) (CHC4049-5)**	23.3
Control-3 (CHC4054-1)	32.5	**PA1-3 (0.1 μM) (CHC4049-6)**	16.4
Control-4 (CHC4054-2)	18.5	**PA1-4 (0.1 μM) (CHC4054-4)**	20.4
Control-5 (CHC4054-3)	24.8	**PA1-5 (0.1 μM) (CHC4054-5)**	19.3
		PA1-6 (0.1 μM) (CHC4054-6)	18.7

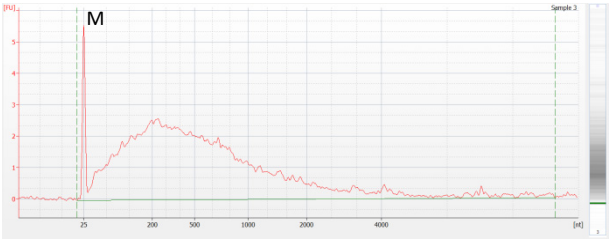


Figure 5.8. ssDNA Fragment Size Distribution of a PA1 In-Cell ·OH Footprinting Sample (Gamma-Radiation). RNA 6000 Nanochip assay analysis of heat denatured ssDNA fragment size distribution for ·OH footprinting sample of **PA1** in W12E cells. Gamma-radiation was performed at MURR for 19.5 min at 37 °C. M: RNA marker.

5.4.3.2 In-Cell Hydroxyl Radical (·OH) Footprinting using Fenton Chemistry – DNA Concentrations and Fragment Size Distribution of Low Molecular DNA Extracted from W12E Cells

To map the binding sites of seven anti-HPV PAs and one inactive PA to the viral and genomic DNA in W12E keratinocytes, we performed in-cell hydroxyl radical footprint experiments as previously described by Taylor *et al*.[22] In this method, highly-reactive hydroxyl radicals are generated by the Fenton reaction between ferrous iron (Fe^{2+}) and hydrogen peroxide (H_2O_2). Addition of sodium ascorbate reduces Fe(III) to regenerate Fe(II), affording a reaction catalytic in iron (**Equation 5.1**). These experiments were performed in five biological replicates of W12E monolayer keratinocytes treated with either one inactive or one of seven active compounds (**Table 5.1**). The cleavage reaction was allowed to proceed for 15 min at room temperature, as longer reaction times of 20 and 25 min led to a significant detachment of W12E from the culture plates. After the cleavage reaction, the low molecular fragments derived from ·OH double-hit kinetics were extracted using a modified Hirt method.[23] DNA sample concentrations were determined by Qubit dsDNA BR or HS Assay kit (**Table 5.7**) and the fragment size distribution was estimated using the RNA 6000 Nanochip assay (**Figure 5.9**).

Table 5.7. In-Cell Footprinting by Fenton Chemistry – DNA Concentrations Determined by Qubit dsDNA HS Assay

Biological Replicate #1		**Biological Replicate #2**		**Biological Replicate #3**	
Sample	**[dsDNA] (ng/µL)**	**Sample**	**[dsDNA] (ng/µL)**	**Sample**	**[dsDNA] (ng/µL)**
Control-1 (CHC4079-1)	5.60	Control-2 (CHC4092-1)	45.6	Control-3 (CHC4097-1)	32.3
PA11-1 (0.1 µM) (CHC4079-2)	13.0	**PA11**-2 (0.1 µM) (CHC4092-2)	68.5	**PA11**-3 (0.1 µM) (CHC4097-2)	25.9
PA1-1 (0.1 µM) (CHC4079-3)	11.0	**PA1**-2 (0.1 µM) (CHC4092-3)	36.1	**PA1**-3 (0.1 µM) (CHC4097-3)	39.6
PA25-1 (0.1 µM) (CHC4079-4)	7.98	**PA25**-2 (0.1 µM) (CHC4092-4)	32.7	**PA25**-3 (0.1 µM) (CHC4097-4)	28.1
NV1078-1 (0.1 µM) (CHC4079-5)	5.68	**NV1078**-2 (0.1 µM) (CHC4092-5)	22.2	**NV1078**-3 (0.1 µM) (CHC4097-5)	23.5
NV1087-1 (0.1 µM) (CHC4079-6)	5.70	**NV1087**-2 (0.1 µM) (CHC4092-6)	27.1	**NV1087**-3 (0.1 µM) (CHC4097-6)	23.6
NV1111-1 (0.1 µM) (CHC4079-7)	7.60	**NV1111**-2 (0.1 µM) (CHC4092-7)	29.1	**NV1111**-3 (0.1 µM) (CHC4097-7)	20.2
PA30-1 (0.1 µM) (CHC4079-8)	9.33	**PA30**-2 (0.1 µM) (CHC4092-8)	35.5	**PA30**-3 (0.1 µM) (CHC4097-8)	17.3
PA31-1 (0.1 µM) (CHC4079-9)	16.6	**PA31**-2 (0.1 µM) (CHC4092-9)	30.1	**PA31**-3 (0.1 µM) (CHC4097-9)	16.5

Biological Replicate #4		Biological Replicate #5	
Sample	**[dsDNA] (ng/µL)**	**Sample**	**[dsDNA] (ng/µL)**
Control-4 (CHC4128-1)	12.4	Control-5 (CHC4136-1)	34.2
PA11-4 (0.1 µM) (CHC4128-2)	16.6	**PA11**-5 (0.1 µM) (CHC4136-2)	42.7
PA1-4 (0.1 µM) (CHC4128-3)	15.4	**PA1**-5 (0.1 µM) (CHC4136-3)	38.2
PA25-4 (0.1 µM) (CHC4128-4)	10.6	**PA25**-5 (0.1 µM) (CHC4136-4)	33.5
NV1078-4 (0.1 µM) (CHC4128-5)	6.4	**NV1078**-5 (0.1 µM) (CHC4136-5)	35.4
NV1087-4 (0.1 µM) (CHC4128-6)	13.3	**NV1087**-5 (0.1 µM) (CHC4136-6)	18.7
NV1111-4 (0.1 µM) (CHC4128-7)	13.3	**NV1111**-5 (0.1 µM) (CHC4136-7)	36.9
PA30-4 (0.1 µM) (CHC4128-8)	8.7	**PA30**-5 (0.1 µM) (CHC4136-8)	13.0
PA31-4 (0.1 µM) (CHC4128-9)	11.0	**PA31**-5 (0.1 µM) (CHC4136-9)	15.5

Accel-NGS 1S Plus DNA Library Kit was used to prepare NGS libraries. Index sequence refers to a unique DNA sequence identifier in the adapter that allows for multiple libraries to be sequenced simultaneously in a single sample lane and be identified and grouped together properly *via* bioinformatics processing of the sequence data. Molar concentrations were calculated according to **Equation 5.4** using the Qubit 2.0 data.

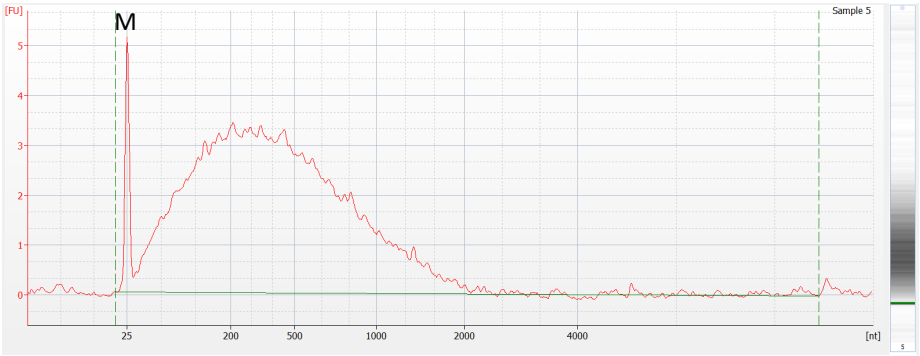


Figure 5.9. Single-Stranded DNA Fragment Size Distribution of a Representative In-Cell ·OH Footprinting Sample (Fenton Chemistry). RNA 6000 Nanochip assay analysis of heat denatured ssDNA fragment size distribution for ·OH footprinting samples in W12E cells. Cleavage reaction was performed for 15 min at room temperature in attached cells using a footprinting solution consisting of 50 mM sodium ascorbate, 10 mM EDTA, 5 mM Fe(II) and 0.6 % H_2O_2. M: RNA marker.

5.4.4 Illumina Next-Generation Sequencing of Hydroxyl Radical Footprinting Experiments

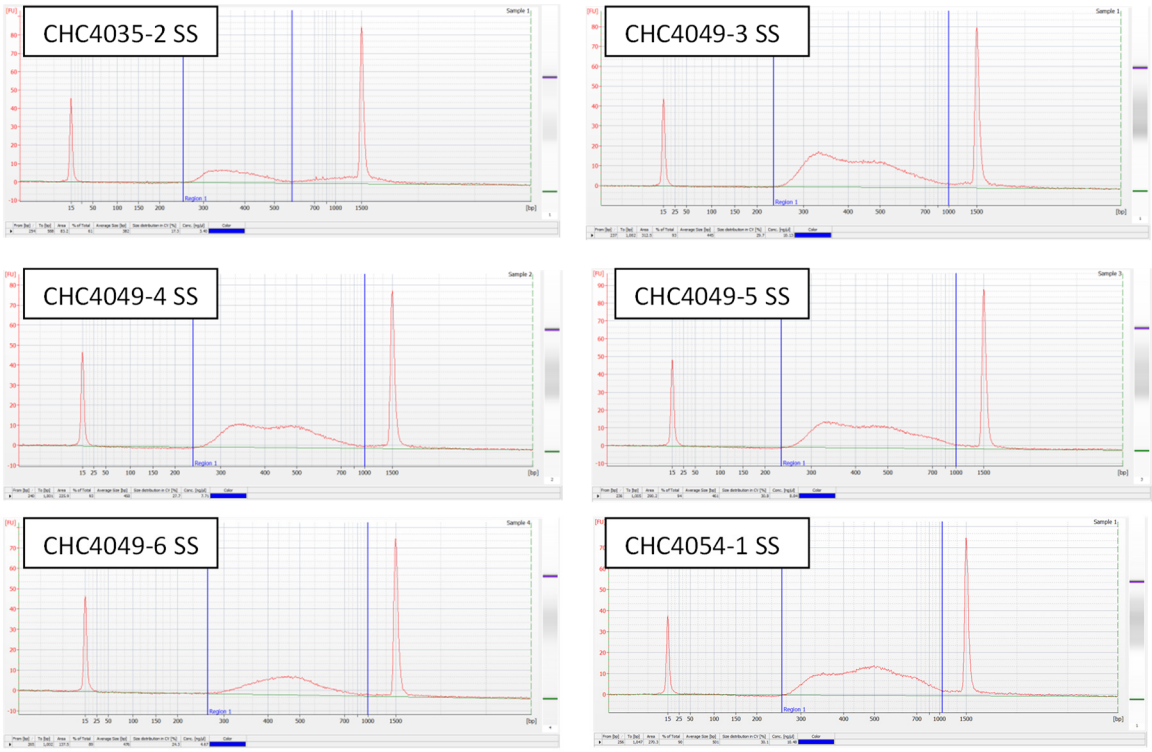
5.4.4.1 Quality Control of NGS Libraries

DNA concentration was measured by the Qubit dsDNA BR/HS assay and fragment size distribution was determined by the Agilent DNA 1000 assay for the constructed libraries. The molar concentrations of the constructed libraries were calculated according to **Equation 5.4** using the DNA concentration measured by Qubit and the average fragment size of the library determined by Bioanalyzer (**Tables 5.8** and **5.9**). Fragment size distribution profiles of NGS libraries are provided in **Figures 5.10** and **5.11**.

Table 5.8. PA1 In-Cell ·OH Footprinting by Gamma-Radiation – Summary of NGS Library Information.

Sample	Adapter (SI-IL1SP-48)	Index Sequence	Qubit 2.0 (ng/µL)	MU Qubit 3.0 (ng/µL)	Bioanalyzer Average Size (bp)	Molar Concentration (nM)
Control-1 (CHC4035-2SS)	D501	TATAGCCT	5.20	6.03	382	20.63
	D701	ATTACTCG				
Control-2 (CHC4049-3 SS)	D502	ATAGAGGC	10.00	11.50	445	34.05
	D702	TCCGGAGA				
Control-3 (CHC4054-1 SS)	D506	TAATCTTA	11.78	13.3	501	35.63
	D706	GAATTCGT				
Control-4 (CHC4054-2 SS)	D507	CAGGACGT	5.29	5.39	493	16.26
	D707	CTGAAGCT				
Control-5 (CHC4054-3 SS)	D508	GTACTGAC	8.14	9.08	404	30.53
	D708	TAATGCGC				
PA1-1 (0.1 µM) (CHC4049-4 SS)	D503	CCTATCCT	7.66	9.79	450	25.79
	D703	CGCTCATT				
PA1-2 (0.1 µM) (CHC4049-5 SS)	D504	GGCTCTGA	8.70	10.20	461	28.59
	D704	GAGATTCC				
PA1-3 (0.1 µM) (CHC4049-6 SS)	D505	AGGCGAAG	3.55	4.99	476	11.30
	D705	ATTCAGAA				
PA1-4 (0.1 µM) (CHC4054-4 SS)	D501	TATAGCCT	11.10	12.2	448	37.54
	D709	CGGCTATG				
PA1-5 (0.1 µM) (CHC4054-5 SS)	D502	ATAGAGGC	8.28	8.92	406	30.90
	D710	TCCGCGAA				
PA1-6 (0.1 µM) (CHC4054-6 SS)	D503	CCTATCCT	8.22	9.06	399	31.21
	D711	TCTCGCGC				

Accel-NGS 1S Plus DNA Library Kit was used to prepare NGS libraries. Index sequence refers to a unique DNA sequence identifier in the adapter that allows for multiple libraries to be sequenced simultaneously in a single sample lane and be identified and grouped together properly *via* bioinformatics processing of the sequence data. Molar concentrations were calculated according to **Equation 5.4** using the Qubit 2.0 data.



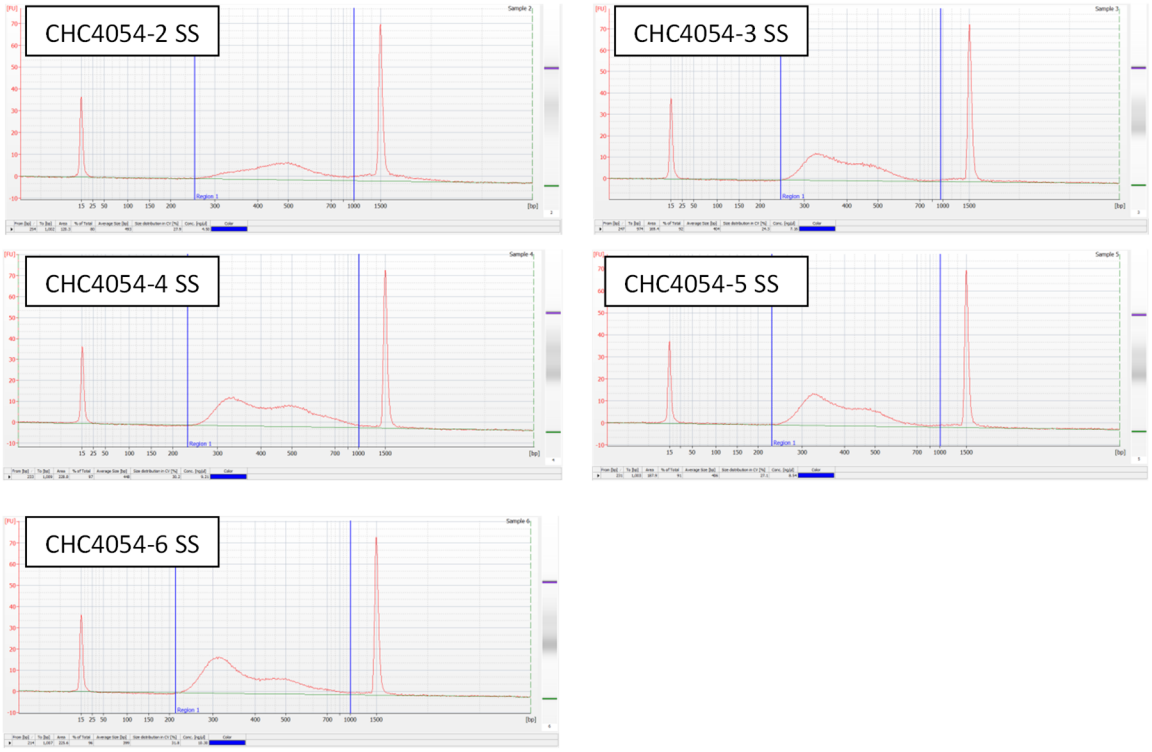


Figure 5.10. Fragment Size Distribution of Illumina NGS Libraries (Gamma-Radiation). Agilent DNA1000 assay analysis of prepared NGS libraries from **PA1** in-cell ·OH footprinting by gamma-radiation. Samples CHC4035-2 SS, CHC4049-3 SS, CHC4054-1 SS, CHC4054-2 SS and CHC4054-3 SS correspond to W12E cells treated with vehicle controls. Samples CHC4049-4 SS, CHC4049-5 SS, CHC4049-6 SS, CHC4054-4 SS, CHC4054-5 SS and CHC4054-6 SS correspond to W12E cells treated with 0.1 µM **PA1** for 24 h.

Table 5.9. In-Cell ·OH Footprinting by Fenton Chemistry – Summary of NGS Library Information.

Biological Replicate #1 Sample	Adapter (SI-IL1SP-48)	Index Sequence	Qubit 2.0 (ng/µL)	MU Qubit 3.0 (ng/µL)	Bioanalyzer Average Size (bp)	Molar Concentration (nM)
Control-1 (CHC4102-1 SS)	D504	GGCTCTGA	12.2	7.29	433	42.69
	D712	AGCGATAG				
PA11-1 (0.1 µM) (CHC4102-2 SS)	D505	AGGCGAAG	19.9	19.9	458	65.83
	D701	ATTACTCG				
PA1-1 (0.1 µM) (CHC4102-3 SS)	D506	TAATCTTA	16.3	16.2	416	59.37
	D702	TCCGGAGA				
PA25-1 (0.1 µM) CHC4102-4 SS	D507	CAGGACGT	16.1	15.6	424	57.53
	D703	CGCTCATT				
NV1078-1 (0.1 µM) CHC4102-5 SS	D508	GTACTGAC	19.5	18.9	457	64.65
	D704	GAGATTCC				
NV1087-1 (0.1 µM) CHC4102-6 SS	D501	TATAGCCT	17.4	18.3	449	58.72
	D705	ATTCAGAA				
NV1111-1 (0.1 µM) CHC4102-7 SS	D502	ATAGAGGC	19.6	17.1	422	70.37
	D706	GAATTCGT				
PA30-1 (0.1 µM) CHC4102-8 SS	D503	CCTATCCT	15.9	16.4	427	56.42
	D707	CTGAAGCT				
PA31-1 (0.1 µM) CHC4102-9 SS	D504	GGCTCTGA	13.3	12.1	422	47.75
	D708	TAATGCGC				

Biological Replicate #2 Sample	Adapter (SI-IL1SP-48)	Index Sequence	Qubit 2.0 (ng/µL)	MU Qubit 3.0 (ng/µL)	Bioanalyzer Average Size (bp)	Molar Concentration (nM)
Control-2 CHC4114-1 SS	D505	AGGCGAAG	13.9	16.9	423	49.79
	D709	CGGCTATG				
PA11-2 (0.1 µM) CHC4114-2 SS	D506	TAATCTTA	16.2	14.0	421	58.30
	D710	TCCGCGAA				
PA1-2 (0.1 µM) CHC4114-3 SS	D507	CAGGACGT	14.3	15.9	424	51.10
	D711	TCTCGCGC				
PA25-2 (0.1 µM) CHC4114-4 SS	D508	GTACTGAC	14.7	15.2	422	52.78
	D712	AGCGATAG				
NV1078-2 (0.1 µM) CHC4114-5 SS	D501	TATAGCCT	15.3	15.4	415	55.86
	D701	ATTACTCG				
NV1087-2 (0.1 µM) CHC4114-6 SS	D502	ATAGAGGC	15.9	18.8	423	56.95
	D702	TCCGGAGA				
NV1111-2 (0.1 µM) CHC4114-7 SS	D503	CCTATCCT	12.1	12.8	421	43.55
	D703	CGCTCATT				
PA30-2 (0.1 µM) CHC4114-8 SS	D504	GGCTCTGA	13.2	15.5	429	46.62
	D704	GAGATTCC				
PA31-2 (0.1 µM) CHC4114-9 SS	D505	AGGCGAAG	13.0	13.7	415	47.46
	D705	ATTCAGAA				
Biological Replicate #3 Sample	**Adapter (SI-IL1SP-48)**	**Index Sequence**	**Qubit 2.0 (ng/µL)**	**MU Qubit 3.0 (ng/µL)**	**Bioanalyzer Average Size (bp)**	**Molar Concentration (nM)**
Control-3 CHC4117-1 SS	D501	TATAGCCT	15.7	16.2	459	51.83
	D706	GAATTCGT				
PA11-3 (0.1 µM) CHC4117-2 SS	D502	ATAGAGGC	17.6	16.6	454	58.74
	D707	CTGAAGCT				
PA1-3 (0.1 µM) CHC4117-3 SS	D503	CCTATCCT	17.3	16.7	491	53.39
	D708	TAATGCGC				
PA25-3 (0.1 µM) CHC4117-4 SS	D504	GGCTCTGA	13.0	13.1	462	42.63
	D709	CGGCTATG				
NV1078-3 (0.1 µM) CHC4117-5 SS	D505	AGGCGAAG	18.4	16.2	441	63.22
	D710	TCCGCGAA				
NV1087-3 (0.1 µM) CHC4117-6 SS	D506	TAATCTTA	15.5	13.2	455	51.62
	D711	TCTCGCGC				
NV1111-3 (0.1 µM) CHC4117-7 SS	D507	CAGGACGT	13.1	11.1	457	43.43
	D712	AGCGATAG				
PA30-3 (0.1 µM) CHC4117-8 SS	D508	GTACTGAC	14.3	10.6	460	47.10
	D701	ATTACTCG				
PA31-3 (0.1 µM) CHC4117-9 SS	D501	TATAGCCT	6.4	5.57	473	20.50
	D702	TCCGGAGA				
Biological Replicate #4 Sample	**Adapter (SI-IL1SP-48)**	**Index Sequence**	**Qubit 2.0 (ng/µL)**	**MU Qubit 3.0 (ng/µL)**	**Bioanalyzer Average Size (bp)**	**Molar Concentration (nM)**
Control-4 CHC4133-1 SS	D502	ATAGAGGC	12.5	11.2	662	28.61
	D703	CGCTCATT				
PA11-4 (0.1 µM) CHC4133-2 SS	D503	CCTATCCT	11.8	11.1	555	32.21
	D704	GAGATTCC				
PA1-4 (0.1 µM) CHC4133-3 SS	D504	GGCTCTGA	3.6	3.19	572	9.54
	D705	ATTCAGAA				
PA25-4 (0.1 µM) CHC4133-4 SS	D505	AGGCGAAG	11.1	9.8	517	32.53
	D712	AGCGATAG				
NV1078-4 (0.1 µM) CHC4133-5 SS	D506	TAATCTTA	12.3	10.8	494	37.73
	D701	ATTACTCG				
NV1087-4 (0.1 µM) CHC4133-6 SS	D507	CAGGACGT	8.1	7.19	514	23.88
	D708	TAATGCGC				
NV1111-4 (0.1 µM) CHC4133-7 SS	D508	GTACTGAC	10.5	9.37	524	30.36
	D709	CGGCTATG				

PA30-4 (0.1 µM) CHC4133-8 SS	D507	CAGGACGT	11.8	10.6	523	34.19
	D710	TCCGCGAA				
PA31-4 (0.1 µM) CHC4133-9 SS	D503	CCTATCCT	9.2	8.28	513	27.17
	D711	TCTCGCGC				
Biological Replicate #5 Sample	**Adapter (DI-IL1SP-12A)**	**Index Sequence**	**Qubit 2.0 (ng/µL)**	**MU Qubit 3.0 (ng/µL)**	**Bioanalyzer Average Size (bp)**	**Molar Concentration (nM)**
Control-5 CHC4141-1 SS	Reagent R1 (I2)	CGATGT(A)	12.5	11.7	474	39.96
PA11-5 (0.1 µM) CHC4141-2 SS	Reagent R1 (I4)	TGACCA(A)	15.7	15.3	498	47.77
PA1-5 (0.1 µM) CHC4141-3 SS	Reagent R1 (I5)	ACAGTG(A)	10.3	9.56	458	34.07
PA25-5 (0.1 µM) CHC4141-4 SS	Reagent R1 (I6)	GCCAAT(A)	14.3	12.8	485	44.67
NV1078-5 (0.1 µM) CHC4141-5 SS	Reagent R1 (I7)	CAGATC(A)	13.4	12.3	468	43.38
NV1087-5 (0.1 µM) CHC4141-6 SS	Reagent R1 (I12)	CTTGTA(A)	11.9	10.4	480	37.56
NV1111-5 (0.1 µM) CHC4141-7 SS	Reagent R1 (I13)	AGTCAA(C)	13.0	12.0	489	40.28
PA30-5 (0.1 µM) CHC4141-8 SS	Reagent R1 (I14)	AGTTCC(G)	12.7	11.3	454	42.38
PA31-5 (0.1 µM) CHC4141-9 SS	Reagent R1 (I15)	ATGTCA(G)	10.2	8.85	459	33.67

Accel-NGS 1S Plus DNA Library Kit was used to prepare NGS libraries. Index sequence refers to a unique DNA sequence identifier in the adapter that allows for multiple libraries to be sequenced simultaneously in a single sample lane and be identified and grouped together properly *via* bioinformatics processing of the sequence data. Molar concentrations were calculated according to **Equation 5.4** using the Qubit 2.0 data.

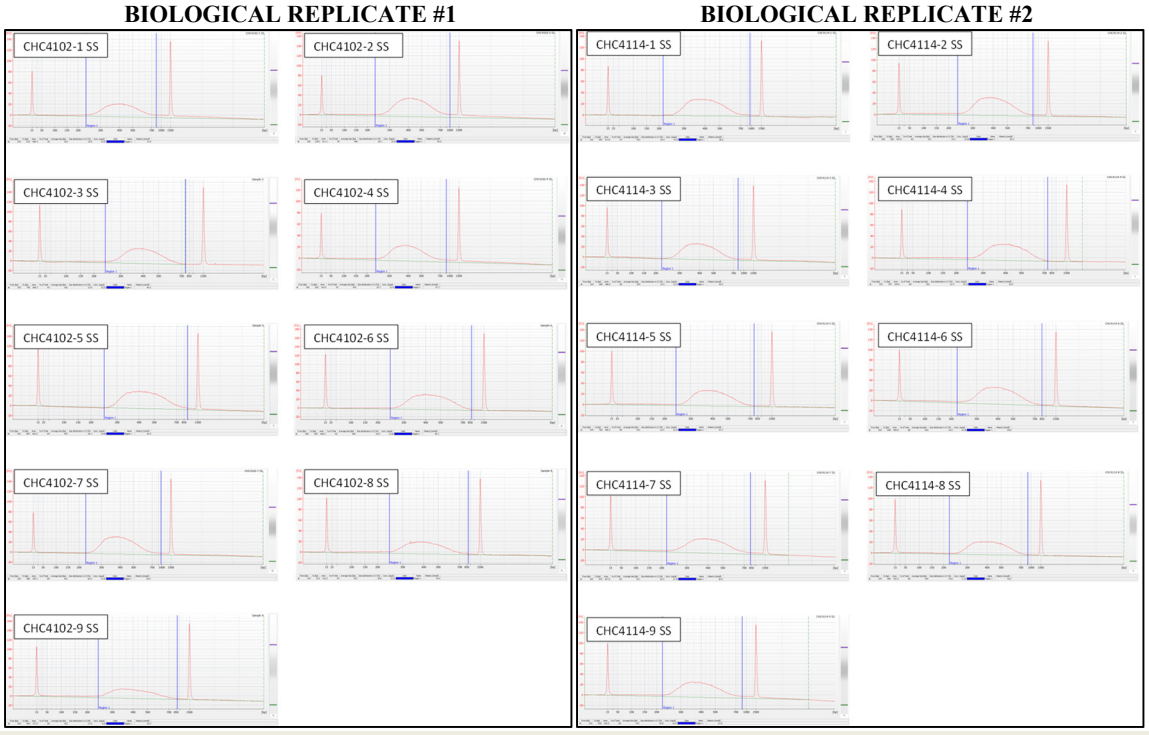


Figure 5.11. *Caption in next page.*

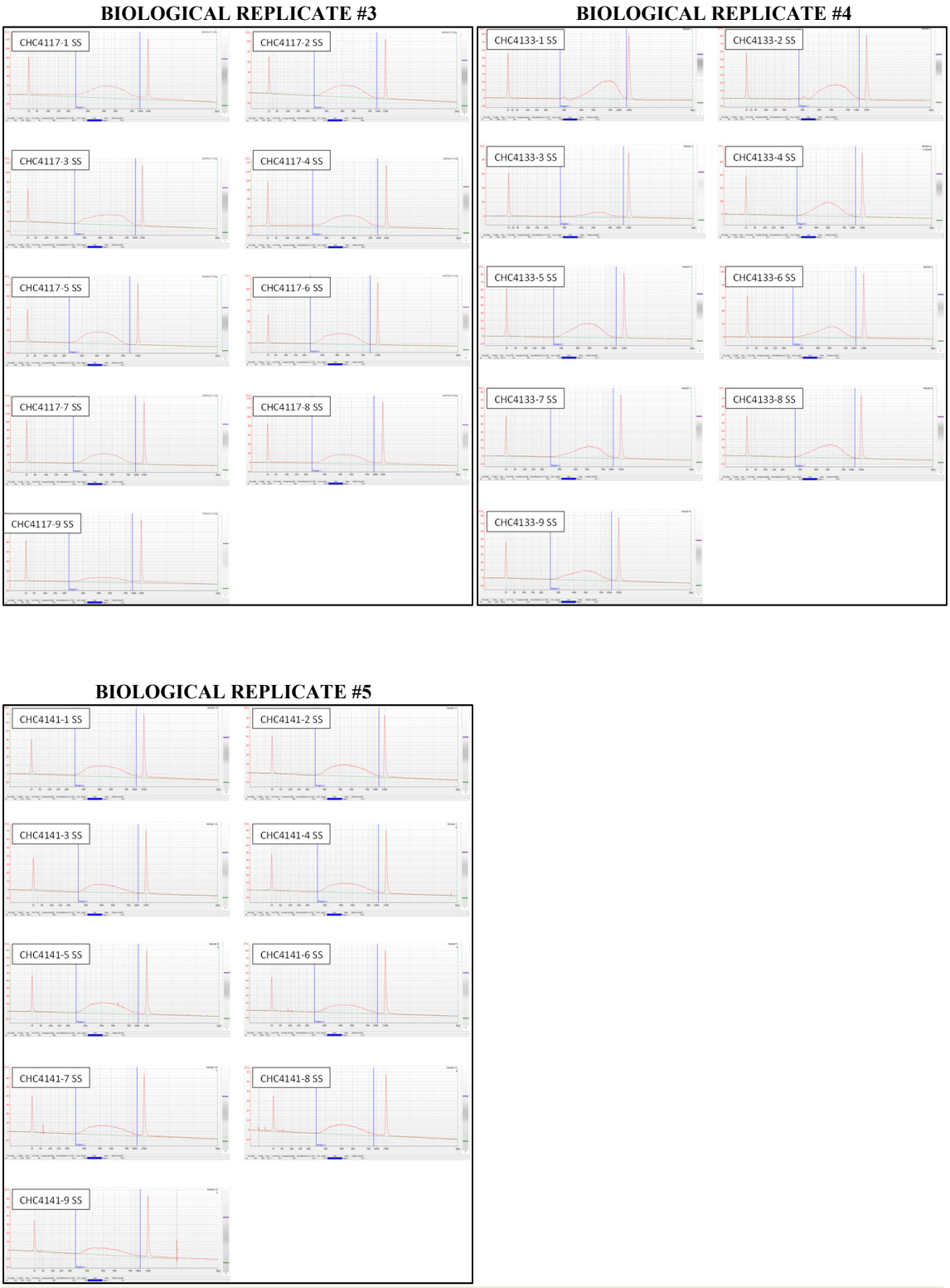


Figure 5.11. Fragment Size Distribution of Illumina NGS Libraries (Fenton Chemistry). Agilent DNA1000 assay analysis of prepared NGS libraries from in-cell ·OH footprinting by Fenton Chemistry (grouped by biological replicates). W12E cells were treated for 24 h with either vehicle controls or 0.1 μM of a respective polyamide.

5.4.4.2 Illumina HiSeq2500 and NextSeq 500 NGS and Bioinformatics Analysis

Illumina NGS libraries of ·OH footprinting samples were sequenced using single-end, 51 base pair read-length on an Illumina HiSeq 2500 or 75 base pair read-length on an Illumina NextSeq 500. Bioinformatics analysis is currently underway following the methods provided in **Chapter 4 Section 4.4.5**.

5.4.5 Crosslinking of Small Molecules for Isolation of Chromatin (COSMIC) with PA1-PB and PA25-PB in H1-hESC and W12E Cells

To map the genome-wide interactions of **PA1** and **PA25** in the context of chromatin structures, we performed COSMIC-Seq experiments in H1 human embryonic stem cells (H1-hESC) and HPV16-bearing W12E keratinocytes. Toward this goal, we synthesized **PA1** and **PA25** derivatives (**PA1-PB** and **PA25-PB**) where the C-termini of these polyamides were covalently conjugated to psoralen and biotin *via* a flexible polyethylene glycol spacer (**Figure 5.2**). Psoralen affords the reversible photo-crosslinking of PA-DNA interactions upon UV irradiation at 365 nm. Crosslinking is performed using a glass filter between the cells and the UV source to remove light with wavelengths lower than 300 nm, significantly decreasing thymine dimer formation. Before performing this COSMIC on W12E keratinocytes, the transmission spectra of various glass panels were determined using a Thermo Scientific Evolution 260 Bio UV-visible spectrophotometer. With the help of Jack Harms from the O'Brien lab, the emission spectrum for the UV source (CalSun, Catalog # B001BH0A1A) was determined using a Hitachi U-3900 Spectrophotometer (**Figure 5.12**). Based on these results, a soda-lime glass panel (3 mm thick) was used to filter UV light below 300 nm, while allowing approximately 87 % of light with 365 nm during the crosslinking step.

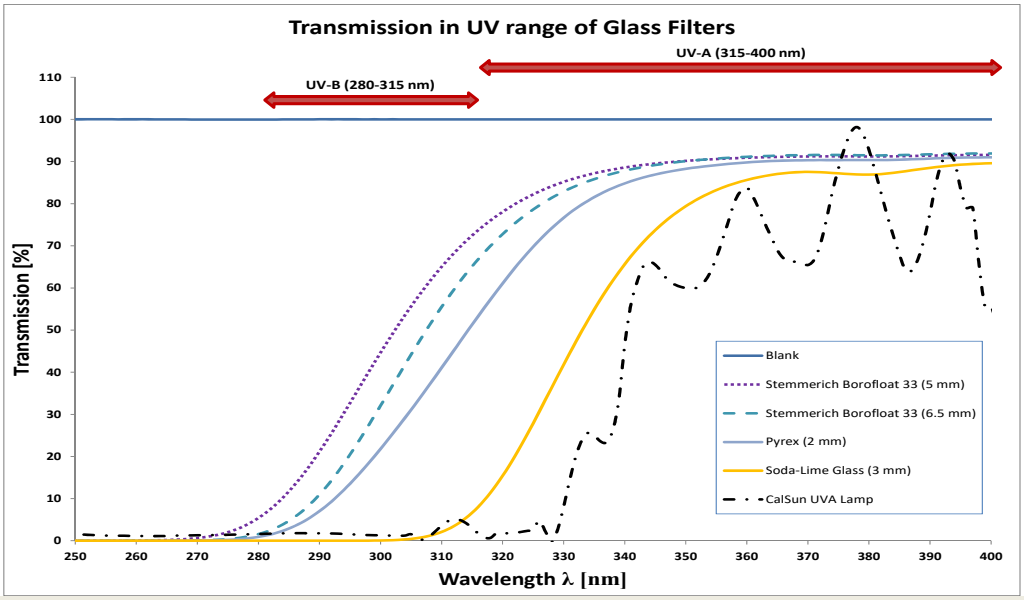


Figure 5.12. Characterization of UV Transmission for Glass Panels and Emission Spectra for UV Source. The transmission spectra of Borofloat (5 and 6.5 mm), Pyrex (2 mm) and soda-lime (3 mm) glass panels (Thermo Scientific Evolution 260 Bio UV-Vis). The emission spectrum of the UV source (power) was collected by directing the emitted light from a CalSun UV lamp into the detector with the spectrometer lamp off (Hitachi U-3900 UV-Vis).

After crosslinking, the cells are lysed and the viral/genomic DNA is sonicated. The sonication time to yield DNA fragments in a range of 100-500 bp was optimized empirically. Using a QSonica Q800R sonicator at 4 °C with a cycle of 10 s **on** and 10 s **off** at 50 % power, the cell lysate solution (2 samples of 600 µL) was sonicated 11 times at 5, 10, 15, 25, 35, 45, 60, 70, 80, 90 and 120 min. A 10 µL aliquot from each solution was sampled at each time interval and diluted with 15 µL of Low EDTA TE buffer. The DNA fragment size distribution for each time interval was determined by electrophoresis with a 1.5 % agarose gel (**Figure 5.13**). The results showed that a sonication time of 2 h afforded DNA fragments in a range of 100-500 bp. Thus, cell lysates for COSMIC experiments were sonicated for 2 h (10 s **on** and 10 s **off**) at 50 % power and 4 °C using QSonica Q800R sonicator.

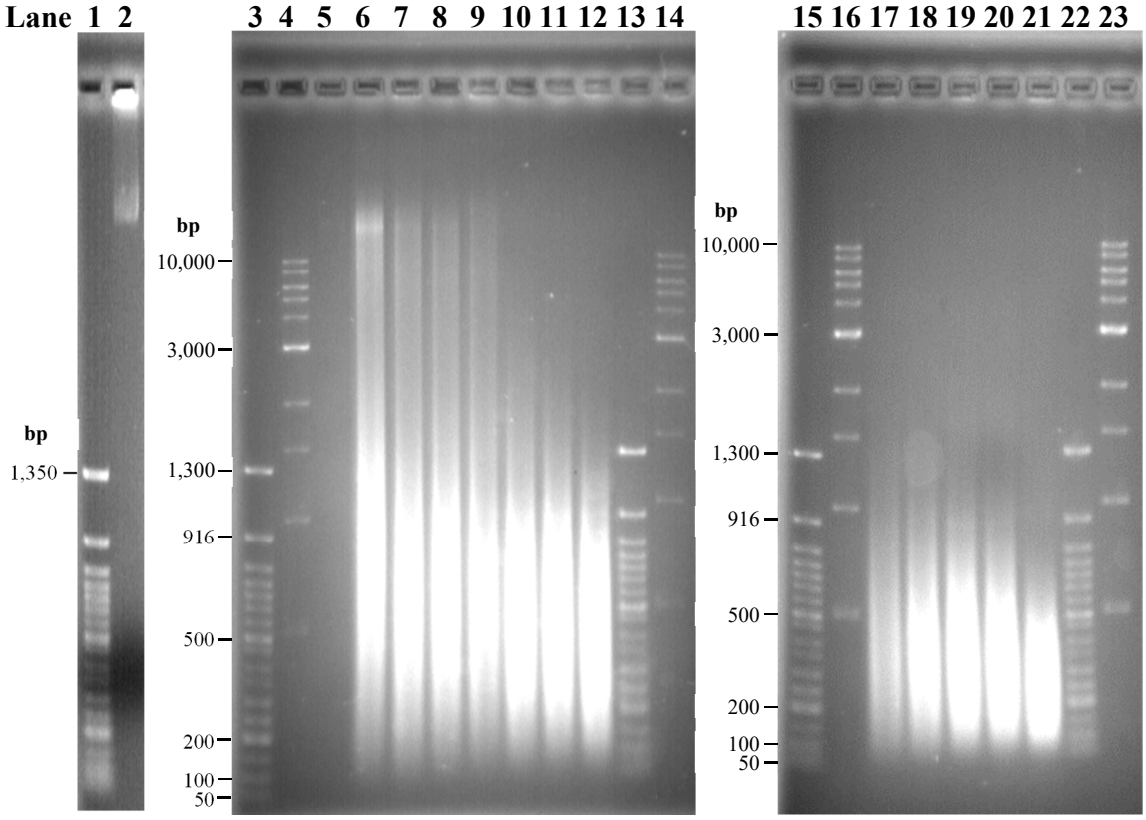


Figure 5.13. Optimization of Sonication Time of W12E Lysates. Agarose gel electrophoresis analysis of DNA fragmentation from W12E lysates by sonication using a QSonica Q800R sonicator at 4 °C with a cycle of 10 s on and 10 s off at 50 % power. Because bromophenol blue tracking dye obscures gel images at ~300 bp, a 6X electrophoresis buffer (25 % v/v glycerol, 1 mM EDTA and 10 mM Tris HCl [pH 7.5]) was used to load the sheared DNA. Lanes 1, 3, 15 and 22 correspond to the 50 bp DNA ladder (NEB, Catalog # N3236S; 1 µg/lane). Lanes 4, 16 and 23 correspond to the 1 kb DNA ladder (NEB, Catalog # N3232S; 0.5 µg/lane). Lane 2 corresponds to initial sample (not subjected to sonication). Samples were sonicated for different times (Lane 6 = 5 min, Lane 7 = 10 min, Lane 8 = 15 min, Lane 9 = 25 min, Lane 10 = 35 min, Lane 11 = 45 min, Lane 12 and 17 = 60 min, Lane 18 = 70 min, Lane 19 = 80 min, Lane 20 = 90 min and Lane 21 = 120 min).

After sonication, the sheared DNA is then subjected to streptavidin-coated magnetic beads in order to capture psoralen-DNA adducts *via* the biotinylated polyamide. After removal of non-specific DNA interactions, the psoralen-DNA crosslinks are reversed by

hot alkali treatment and the purified, enriched sequences are analyzed by next-generation sequencing.[13,29]

COSMIC experiments in W12E keratinocytes are ready for sequencing. On the other hand, next-generation sequencing has been completed for the COSMIC experiments in H1-hESC and bioinformatics analysis is currently in progress.

5.5 CONCLUSIONS

Since the ultimate goal of this project was to decipher polyamide binding across the viral and host genome in the context of chromatin architecture, we have employed hydroxyl radical footprinting (·OH) and COSMIC (crosslinking of small molecules for isolation of chromatin) coupled with massively parallel DNA sequencing (Seq) to assess the binding occupancies of anti-HPV hairpin polyamides across the viral genomes in keratinocytes harboring HPV episomes. These data will allow us to determine PA binding occupancies and whether the putative DNA-binding sites are accessible to PA in cells. Furthermore, PA-DNA interactions and PA-associated perturbations of viral chromatin-like structures will be assessed with the hydroxyl radical footprinting approach.[20,21,37] Consequently, the effects of the antiviral hairpin polyamides in displacing essential viral proteins and transcription factors from the viral genome may be elucidated. These perturbations will be determined by comparing the chromatin landscape of vehicle-treated to PA-treated cells. Furthermore, comparison of the PA occupancies and associated perturbations on chromatin-like structures between HPV-inactive, HPV-weakly active and HPV-potently active polyamides may provide the molecular foundations that underlie the ability of antiviral hairpin PAs to eliminate the viral load.

ACKNOWLEDGEMENTS

We thank Professor Paul Lambert (University of Wisconsin-Madison) for providing W12E cervical keratinocytes (clonal cell line # 20850) and J2 3T3 fibroblasts.

5.6 BIBLIOGRAPHY

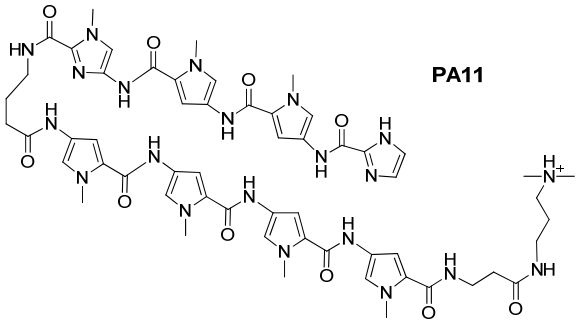
(1) He, G.; Vasilieva, E.; Harris Jr, G. D.; Koeller, K. J.; Bashkin, J. K.; Dupureur, C. M. *Biochimie* **2014**, *102*, 83.
(2) Koeller, K. J.; Harris, G. D.; Aston, K.; He, G.; Castaneda, C. H.; Thornton, M. A.; Edwards, T. G.; Wang, S.; Nanjunda, R.; Wilson, W. D.; Fisher, C.; Bashkin, J. K. *Medicinal Chemistry* **2014**, *4*, 338.
(3) Castaneda, C. H.; Scuderi, M. J.; Edwards, T. G.; Harris Jr, G. D.; Dupureur, C. M.; Koeller, K. J.; Fisher, C.; Bashkin, J. K. *MedChemComm* **2016**.
(4) Vasilieva, E.; Niederschulte, J.; Song, Y.; Harris Jr, G. D.; Koeller, K. J.; Liao, P.; Bashkin, J. K.; Dupureur, C. M. *Biochimie* **2016**, *127*, 103.
(5) Favre, M.; Breitburd, F.; Croissant, O.; Orth, G. *J Virol* **1977**, *21*, 1205.
(6) Lee, D.; Sohn, H.; Kalpana, G. V.; Choe, J. *Nature* **1999**, *399*, 487.
(7) Peña, L. d. M.; Laimins, L. A. *J Virol* **2001**, *75*, 10005.
(8) Stünkel, W.; Bernard, H.-U. *J Virol* **1999**, *73*, 1918.

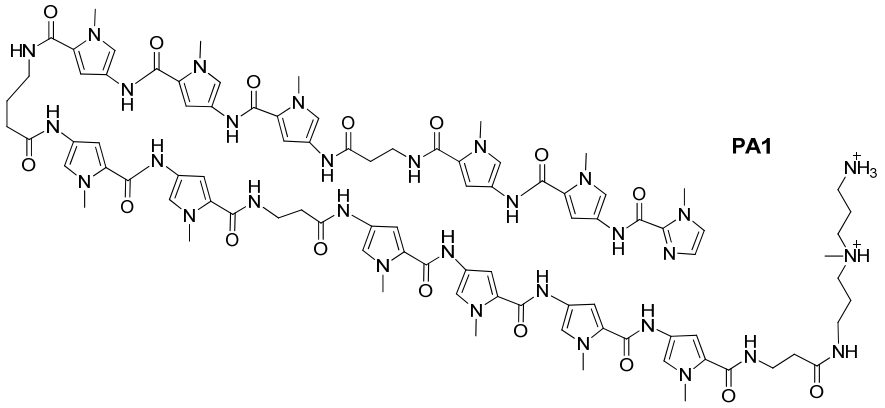
(9) Demeret, C.; Le Moal, M.; Yaniv, M.; Thierry, F. *Nucleic Acids Res* **1995**, *23*, 4777.
(10) Badal, V.; Chuang, L. S. H.; Tan, E. H. H.; Badal, S.; Villa, L. L.; Wheeler, C. M.; Li, B. F. L.; Bernard, H. U. *J Virol* **2003**, *77*, 6227.
(11) Gottesfeld, J. M.; Melander, C.; Suto, R. K.; Raviol, H.; Luger, K.; Dervan, P. B. *J Mol Biol* **2001**, *309*, 615.
(12) Chandran, A.; Syed, J.; Taylor, R. D.; Kashiwazaki, G.; Sato, S.; Hashiya, K.; Bando, T.; Sugiyama, H. *Nucleic Acids Res* **2016**, *44*, 4014.
(13) Erwin, G. S.; Grieshop, M. P.; Bhimsaria, D.; Do, T. J.; Rodríguez-Martínez, J. A.; Mehta, C.; Khanna, K.; Swanson, S. A.; Stewart, R.; Thomson, J. A.; Ramanathan, P.; Ansari, A. Z. *Proceedings of the National Academy of Sciences* **2016**, *113*, E7418.
(14) Tullius, T. D.; Dombroski, B. A. *Proceedings of the National Academy of Sciences* **1986**, *83*, 5469.
(15) Hampshire, A. J.; Rusling, D. A.; Broughton-Head, V. J.; Fox, K. R. *Methods* **2007**, *42*, 128.
(16) Fox, K. R. *Drug-DNA Interaction Protocols* 1997; Vol. 90, p 1.
(17) Pogozelski, W. K.; Tullius, T. D. *Chemical Reviews* **1998**, *98*, 1089.
(18) Balasubramanian, B.; Pogozelski, W. K.; Tullius, T. D. *Proceedings of the National Academy of Sciences* **1998**, *95*, 9738.
(19) Churchill, M. E. A.; Hayes, J. J.; Tullius, T. D. *Biochemistry* **1990**, *29*, 6043.
(20) Hayes, J. J.; Kam, L.; Tullius, T. D. In *Methods in Enzymology*; Lester Packer, A. N. G., Ed.; Academic Press: 1990; Vol. Volume 186, p 545.
(21) Ottinger, L. M.; Tullius, T. D. *Journal of the American Chemical Society* **2000**, *122*, 5901.
(22) Cannistraro, V. J.; Pondugula, S.; Song, Q.; Taylor, J.-S. *Journal of Biological Chemistry* **2015**, *290*, 26597.
(23) Arad, U. *BioTechniques* **1998**, *24*, 760.
(24) Edwards, T. G.; Koeller, K. J.; Slomczynska, U.; Fok, K.; Helmus, M.; Bashkin, J. K.; Fisher, C. *Antiviral Res* **2011**, *91*, 177.
(25) Edwards, T. G.; Vidmar, T. J.; Koeller, K.; Bashkin, J. K.; Fisher, C. *PLoS One* **2013**, *8*, e75406.
(26) Carlson, C. D.; Warren, C. L.; Hauschild, K. E.; Ozers, M. S.; Qadir, N.; Bhimsaria, D.; Lee, Y.; Cerrina, F.; Ansari, A. Z. *Proc Natl Acad Sci U S A* **2010**, *107*, 4544.
(27) Ozers, M. S.; Warren, C. L.; Ansari, A. Z. *Methods in molecular biology (Clifton, N.J.)* **2009**, *544*, 637.
(28) Eguchi, A.; Wleklinski, M. J.; Spurgat, M. C.; Heiderscheit, E. A.; Kropornicka, A. S.; Vu, C. K.; Bhimsaria, D.; Swanson, S. A.; Stewart, R.; Ramanathan, P.; Kamp, T. J.; Slukvin, I.; Thomson, J. A.; Dutton, J. R.; Ansari, A. Z. *Proceedings of the National Academy of Sciences* **2016**, *113*, E8257.
(29) Erwin, G. S.; Bhimsaria, D.; Eguchi, A.; Ansari, A. Z. *Angewandte Chemie International Edition* **2014**, *53*, 10124.
(30) Sastry, S. S.; Ross, B. M.; P'arraga, A. *Journal of Biological Chemistry* **1997**, *272*, 3715.
(31) Hyde, J. E.; Hearst, J. E. *Biochemistry* **1978**, *17*, 1251.

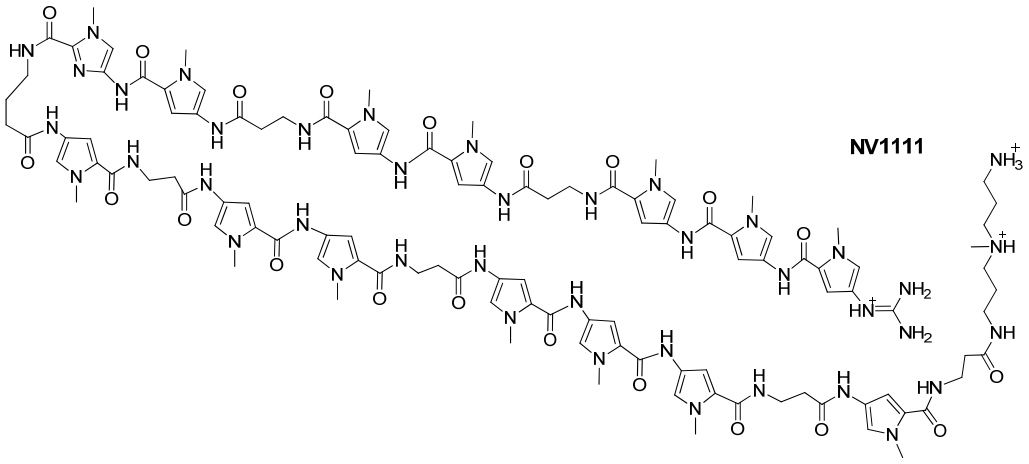
(32) Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1996**, *118*, 6141.
(33) Edwards, T. G.; Helmus, M. J.; Koeller, K.; Bashkin, J. K.; Fisher, C. *J Virol* **2013**, *87*, 3979.
(34) Garner-Hamrick, P. A.; Fisher, C. *Virology* **2002**, *301*, 334.
(35) Meyer, C.; Frattini, M. G.; Laimins, L. A. *Cell Biology: A Laboratory Handbook (Celis, J.E. ed.)* **1998**, 491.
(36) Langmead, B.; Salzberg, S. L. *Nat Methods* **2012**, *9*, 357.
(37) Xu, L. Ph.D., Boston University, 2009.

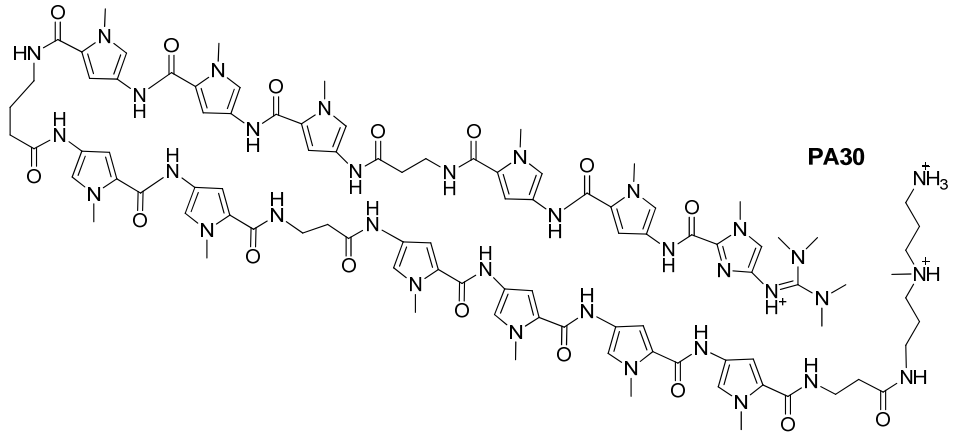
5.7 SUPPLEMENTAL INFORMATION

5.7.1 Chemical Structures of PAs Used for In-Cell Hydroxyl Radical Experiments (Fenton Chemistry)

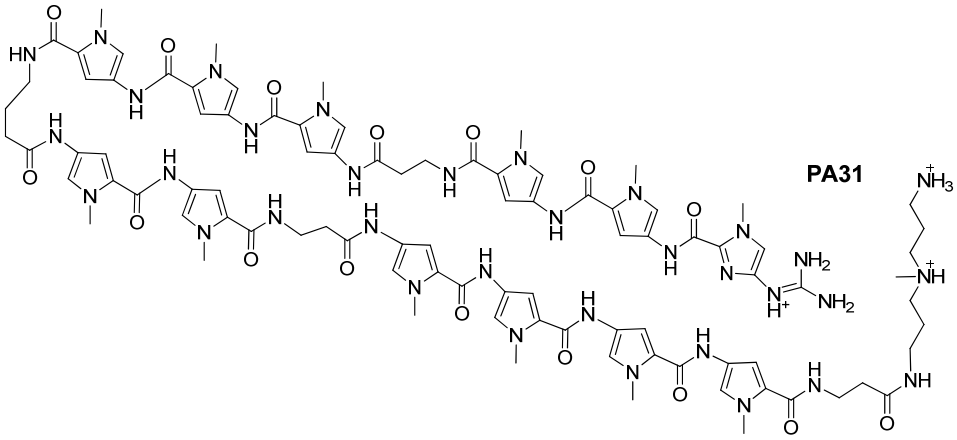




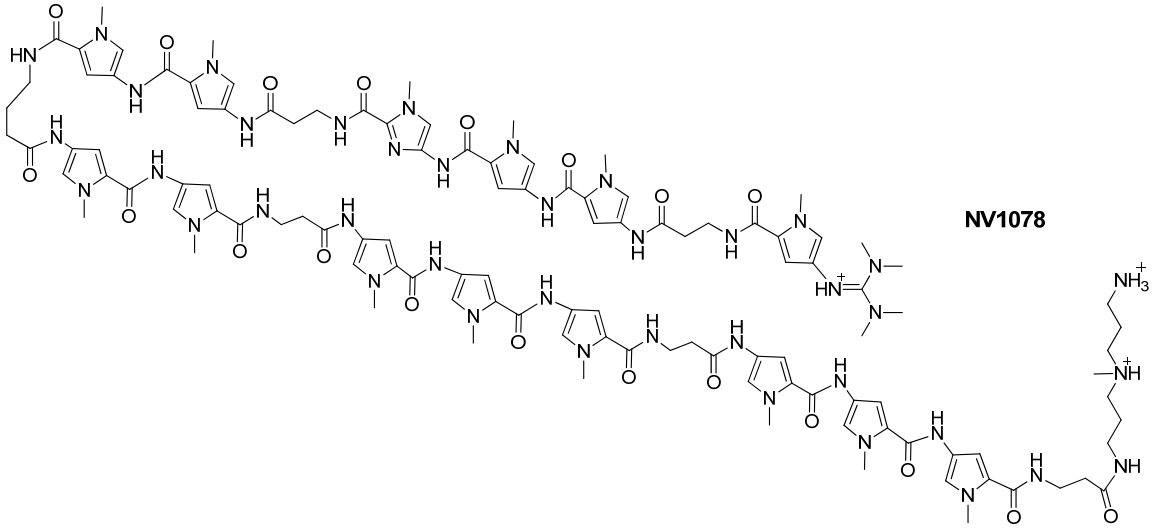




PA30



PA31



NV1078

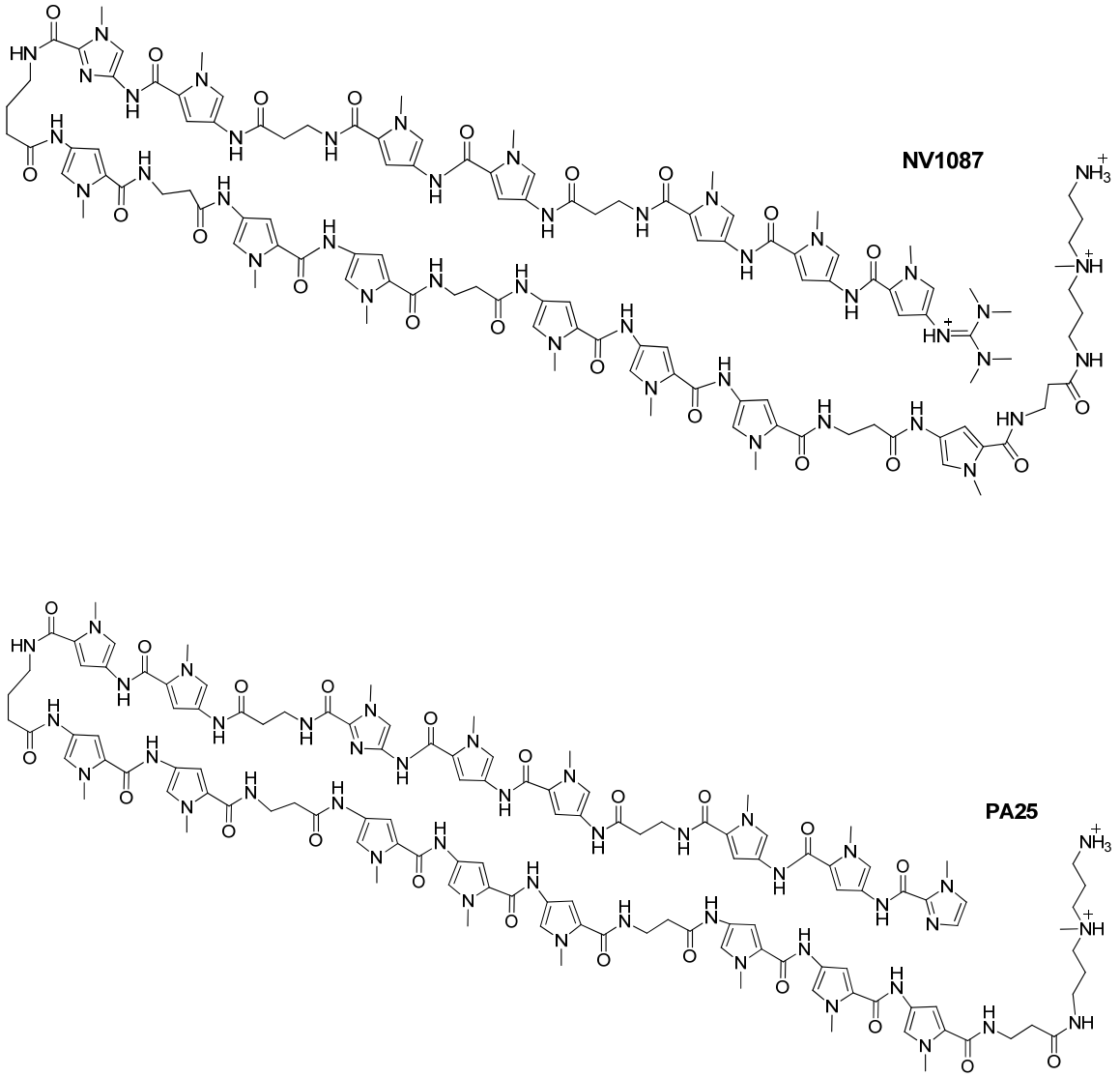


Figure SI5.1. Chemical Structures of PAs Used for In-Cell Hydroxyl Radical Experiments (Fenton Chemistry).

5.7.2 Synthesis of PA1-Alexa Fluor 488 Conjugate

Anti-HPV **PA1** was functionalized with Alexa Fluor 488 (**Figure SI5.2**) for subsequent cell uptake studies by confocal microscopy and kinetic studies conducted by Dr. Dupureur's group. Covalent conjugation of the fluorophore was accomplished by reacting of Alexa Fluor 488 carboxylic acid, succinimidyl ester to the primary amine on the Ta tail of **PA1** in the presence of DIEA in anhydrous NMP.

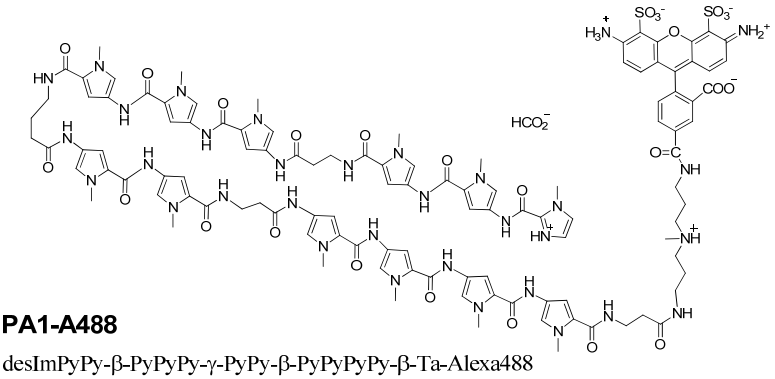


Figure SI5.2. Chemical Structure of PA1 Functionalized with Alexa Fluor 488 as the Formate Salt.

Reagents

Reagents	Molecular Weight (g/mol)	Equivalents	Quantity
PA1 (3 TFA) (GDH1130, 97 %)	2237.12	1	**2.32 mg** 1.04×10^{-6} mol
Alexa Fluor 488 carboxylic acid, succinimidyl ester	643.41	1.5	**1.00 mg** 1.56×10^{-6} mol
***N,N*-Diisopropylethylamine (DIEA)**	129.24 (0.742 g/mL)	50	**9.03 µL** 5.19×10^{-5} mol
***N*-Methyl-2-pyrrolidon (NMP)**	99.13	Solvent	**1000 µL**

Procedure

1. Care must be taken through the process to keep light exposure to a minimum.
2. Dissolve 2.32 mg of **PA1 (3 TFA)** in 1000 µL NMP in 1-dram vial. To this solution, add 9.03 µL of **DIEA**. Cover reaction vessel with aluminum foil to protect from light.
3. Dissolve 1 mg aliquot of **Alexa Fluor 488 carboxylic acid, succinimidyl ester** (ThermoFisher Scientific A-20000) in 150 µL.
4. Add dye/NMP solution to the **PA1**/NMP solution.
5. Allow reaction to proceed at room temperature for at least 4 h. Check for the relative formation of the polyamide-dye conjugate by analytical HPLC.
    a. Aliquot 10 µL of reaction mixture and dilute to 100 µL with DMSO.
    b. Filter through 0.45 µm syringe filter and inject 10 µL into reverse-phase HPLC (C12).
6. Filter the reaction mixture through 0.45 µm syringe filter and transfer to appropriate amber Dionex vial.
7. Rinse reaction vessel with 0.5 mL of DMSO and filter through 0.45 µm syringe filter. Add to solution in Step 7.
8. Add 0.5 mL of MilliQ H_2O (0.2 % formic acid) to the Dionex vial containing the reaction mixture.

9. Purify on Dionex Ultimate 3000 prep-LC using a Macherey-Nagel Nucleodur C18 column (2 columns in series) (21.0 mm x 50 mm each) and program Gilson M-N short. Mobile phase A = H_2O (0.2 % formic acid); mobile phase B = methanol.

$$25\ \%\ B \xrightarrow{35\ min} 75\ \%\ B \xrightarrow{1\ min} 90\ \%\ B \xrightarrow{1\ min} 90\ \%\ B \xrightarrow{0.1\ min} 25\ \%\ B$$

10. Product eluted at 21.24 min.
11. Check fractions for pure polyamide-dye conjugate by analytical HPLC.
    a. Aliquot 100 µL of a given fraction.
    b. Inject into reverse-phase HPLC (C12).
12. Concentrate fractions with rotovap and transfer (~5 mL) to 20 mL vial.
13. Rinse round bottom flask with 1:1 CH_3CN / H_2O (0.2 % formic acid) and add to 20 mL vial.
14. Freeze solution on dry ice and lyophilize overnight.

5.7.3 Synthesis of Biotinylated and Psoralen-Biotin PA Conjugates

KA2127, PA1 and **PA25** were functionalized with a PEG4-biotin affinity handle to produce **KA2127-PEG4-Biotin**, **PA1-PEG4-Biotin** and **PA25-PEG4-Biotin** (**Figure 5.3**) for subsequent Cognate Site Identification (CSI) experiments. Covalent conjugation of the affinity handle was accomplished by reacting biotin-PEG4-NHS ester to the primary amine on the Ta tail of each polyamide in the presence of DIEA in anhydrous NMP.

Reagents

Reagents	Molecular Weight (g/mol)	Equivalents	Quantity
Biotin-PEG4-NHS ester	588.67	1	**1.00 mg** 1.70×10^{-6} mol
KA2127 (5 FA) (KA2127E) OR **PA1 (3 TFA)** (GDH1130, 97 %) OR **PA25 (4 TFA)** (KJK6006, 95 %)	1445.46 / 2237.12 / 2982.79	1.5	**3.68 mg** / **5.70 mg** / **7.60 mg** 2.55×10^{-6} mol
***N,N*-Diisopropylethylamine (DIEA)**	129.24 (0.742 g/mL)	50	**14.8 µL** 8.49×10^{-5} mol
***N*-Methyl-2-pyrrolidon (NMP)**	99.13	Solvent	**1000 µL**

Procedure

1. Dissolve either 3.68 mg **KA2127 (5FA)**, 5.70 mg **PA1 (3 TFA)** or 7.60 mg **PA25 (4 TFA)** in 1000 µL anhydrous NMP in 1-dram vial. To this solution, add 14.8 µL of **DIEA**.
2. Stir solution for 2 minutes with magnetic stir bar. Remove 10 µL aliquot and dilute to 100 µL DMSO, filter through 0.2 µm syringe filter and inject in analytical HPLC.
3. Dissolve 1 mg of **Biotin-PEG4-NHS ester** (Conju-Probe CP-3026-100MG) in 150 µL anhydrous NMP.
4. Add Biotin-PEG4/NMP solution to the polyamide/NMP solution.
5. Allow reaction to proceed at room temperature for 3-4 h. Check for the relative formation of the polyamide-PEG4-Biotin conjugate by analytical HPLC.

   a. Aliquot 10 µL of reaction mixture and dilute to 100 µL with DMSO.
   b. Filter through 0.45 µm syringe filter and inject 10 µL into reverse-phase HPLC (C12).
6. Filter the reaction mixture through 0.45 µm syringe filter and transfer to appropriate amber Dionex vial.
7. Rinse reaction vessel with 0.5 mL of DMSO and filter through 0.45 µm syringe filter. Add to solution in Step 6.
8. Add 0.5 mL of MilliQ H_2O (0.2 % formic acid) to the Dionex vial containing the reaction mixture.
9. Purify on Dionex Ultimate 3000 prep-LC using a Macherey-Nagel Nucleodur C18 column (2 columns in series) (21.0 mm x 50 mm each) and program Gilson M-N short. Mobile phase A = H_2O (0.2 % formic acid) and Mobile phase B = methanol.

$$25\%\ B \xrightarrow{35\ min} 75\%\ B \xrightarrow{1\ min} 90\%\ B \xrightarrow{1\ min} 90\%\ B \xrightarrow{0.1\ min} 25\%\ B$$

10. Biotinylated polyamide conjugates eluted at 17.6, 21.3 and 23.1 min for **KA2127-PEG4-Biotin**, **PA1-PEG4-Biotin** and **PA25-PEG4-Biotin**, respectively.
11. Check fractions for pure polyamide-PEG4-biotin conjugate by analytical HPLC.
   a. Aliquot 100 µL of a given fraction.
   b. Inject into reverse-phase HPLC (C12).
12. Concentrate fractions with rotovap and transfer (~5 mL) to 20 mL vial.
13. Rinse round bottom flask with equal volume of 1:1 CH_3CN / H_2O (0.2 % formic acid) and add to 20 mL vial.
14. Freeze solution on dry ice and lyophilize overnight.

PA1 and **PA25** were functionalized with a psoralen-biotin (photo-crosslinker/affinity handle) to produce **PA1-PB** and **PA25-PB** (**Figure 5.2**) for subsequent CSI (Cognate Site Identification) and COSMIC (crosslinking of small molecules for isolation of chromatin) experiments. Covalent conjugation of the psoralen-biotin moiety was accomplished by reacting the peptide acid of **PB** with HATU to generate the active ester, followed by amide bond formation between the primary amine on the Ta tail of each polyamide and the activated **PB** ester in the presence of DIEA in anhydrous NMP.

Reagents

Reagents	Molecular Weight (g/mol)	Equivalents	Quantity
PA1 (3 TFA) (GDH1130, 97 %)	2237.12	3	**14.39 mg** 6.43×10^{-6} mol
OR			
PA25 (4 TFA) (KJK6006, 95 %)	2982.79		**19.18 mg** 6.43×10^{-6} mol
Peptide acid psoralen-biotin (PB)	933.03	1	**2.00 mg** 2.14×10^{-6} mol
***N*-Methyl-2-pyrrolidon (NMP)**	99.13	Solvent	**150 µL**
1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-*b*]pyridinium 3-oxid hexafluorophosphate (HATU)	380.23	0.95	**0.77 mg** 2.04×10^{-6} mol
***N,N*-Diisopropylethylamine (DIEA)**	129.24 (0.742 g/mL)	3	**1.12 µL** 6.43×10^{-6} mol

Procedure

1. Care must be taken through the process to keep light exposure to a minimum.
2. Dissolve 2.00 mg of the **PB** moiety and 0.77 mg of **HATU** in 150 µL of anhydrous **NMP**. To this solution, add 1.12 µL of **DIEA**. Allow reaction to proceed at room temperature for 5 min. Check for the relative formation of the active ester by analytical HPLC.
    a. Aliquot 1 µL of reaction mixture and dilute to 100 µL with DMSO.
    b. Filter through 0.45 µm syringe filter and inject 10 µL into reverse-phase HPLC (C12).
3. Dissolve 14.39 mg of **PA1 (3 TFA)** or 19.18 mg **PA25 (4 TFA)** in 300 µL of anhydrous **NMP**. Add this solution to the **PB** active ester and allow the reaction to proceed for 2 h at room temperature. Check for the relative formation of the polyamide-**PB** conjugate by analytical HPLC.
    a. Aliquot 1 µL of reaction mixture and dilute to 100 µL with DMSO.
    b. Filter through 0.45 µm syringe filter and inject 10 µL into reverse-phase HPLC (C12).
4. Filter the reaction mixture through 0.45 µm syringe filter and transfer to appropriate amber Dionex vial.
5. Rinse reaction vessel with 0.5 mL of DMSO and filter through 0.45 µm syringe filter. Add to solution in Step 4.
6. Add 0.5 mL of MilliQ H_2O (0.2 % formic acid) to the Dionex vial containing the reaction mixture.
7. Purify on Dionex Ultimate 3000 prep-LC using a Macherey-Nagel Nucleodur C18 column (2 columns in series) (21.0 mm x 50 mm each) and program Gilson M-N short. Mobile phase A = H_2O (0.2 % formic acid); mobile phase B = methanol.

   25% B $\xrightarrow{35\ min}$ 75% B $\xrightarrow{1\ min}$ 90% B $\xrightarrow{1\ min}$ 90% B $\xrightarrow{0.1\ min}$ 25% B

8. Polyamide-**PB** conjugates eluted at 21.0 and 24.3 min for **PA1-PB** and **PA25-PB**, respectively.
9. Check fractions for pure polyamide-PB conjugate by analytical HPLC.
    a. Aliquot 100 µL of a given fraction.
    b. Filter through 0.45 µm syringe filter and inject into reverse-phase HPLC (C12).
10. Concentrate fractions with rotovap and transfer (~5 mL) to 20 mL vial.
11. Rinse round bottom flask with equal volume of 1:1 CH_3CN / H_2O (0.2 % formic acid) and add to 20 mL vial.
12. Freeze solution on dry ice and lyophilize overnight.

5.7.4 Maintenance of Cell Culture

5.7.4.1 Reagents for Cell Culture Growth

Adapted from Tissue Culture Techniques for the Study of Human Papillomaviruses in Stratified Epithelia by Meyers, *et al.*

Materials

1. Autoclaved MilliQ H_2O (18.2 MΩ·cm at 25 °C)
2. Dulbecco's Modified Eagle's Medium with high glucose, L-glutamine, phenol red, **no pyruvate** (Gibco, Catalog # 11965-126)
3. Dulbecco's Modified Eagle's Medium with high glucose, L-glutamine, phenol red, sodium pyruvate (Gibco, Catalog # 11995-081)
4. Ham's F-12 media with L-gluramine, phenol red (Gibco, Catalog # 21127-022)
5. 1X Phosphate Buffer Saline (Fisher, Catalog # 10010-031)
6. 0.05 % Trypsin-EDTA with phenol red (Gibco, Catalog # 25300-054)
7. 0.25 % Trypsin-EDTA with phenol red (Gibco, Catalog # 25200-056)
8. Penicillin-streptomycin solution (10,000 units penicillin and 10 mg streptomycin per mL) (Sigma, Catalog # P4333-20ML)
9. Epidermal growth factor powder (Invitrogen, Catalog # PHG0311)
10. Bovine serum albumin powder (Sigma, Catalog # A4919-1G)
11. 1 M HEPES buffer (Sigma, Catalog # 83264-100ML-F)
12. Hydrocortisone powder (Sigma, Catalog # H0888-1G)
13. Adenine hydrochloride hydrate (Sigma, Catalog # A9795-1G)
14. Insulin from bovine pancreas (Sigma, Catalog # I6634-50MG)
15. Apo-Transferrin human (Sigma, Catalog # T1147-100MG)
16. T3 thyroid hormone, 3,3',5-triiodo-L-thyronine sodium salt (Sigma, Catalog # T6397-100MG)
17. Fetal Bovine Serum (Gibco, Catalog # 16140-063)
18. Mytomycin C powder (Fisher BioReagents, Catalog # BP2531-2)
19. Cholera toxin powder (Sigma, Catalog # C9903-.5MG)
20. Dimethyl sulfoxide (Sigma Life Science, Catalog # D2650)
21. Sodium Chloride (Fisher BioReagents, Catalog # BP358-1)
22. Potassium Chloride (Fisher BioReagents, Catalog # BP366-500)
23. Sodium Bicarbonate (Fisher BioReagents, Catalog # BP328-1)
24. Sodium Phosphate Monobasic Monohydrate (Fisher BioReagents, Catalog # BP330-500)
25. Magnesium Sulfate Heptahydrate (Fisher Chemical, Catalog # M80-500)
26. Ferric Nitrate Nonahydrate (Fisher Chemical, Catalog # I110-100)
27. Phenol Red (Sigma Aldrich, Catalog # P3532-5G)

Solutions

1. **Earle's Salts**
   - **a.** For 500 mL solution weigh and dissolve:
     - i. NaCl 32 g
     - ii. KCl 2 g
     - iii. $NaHCO_3$ 18.5 g
     - iv. $NaH_2PO_4 \cdot H_2O$ 0.625 g
     - v. $MgSO_4 \cdot 7H_2O$ 1 g
     - vi. $Fe(NO_3)_3 \cdot 9H_2O$ 0.0005 g (make 1 mg/mL in autoclaved H_2O)
     - vii. Phenol Red 0.025 g
   - **b.** Dilute to 500 mL with autoclaved MilliQ H_2O.
   - **c.** Filter-sterilize and store tightly capped at room temperature.
2. **HEPES Buffered Earle's Salts (HBES) –** Transfer 25 mL of 1M HEPES buffer into 100 mL of Earle's salts and dilute to 1 L of final volume with autoclaved MilliQ H_2O. Filter-sterelize.
3. **100X Adenine –** Dissolve 0.169 g of adenine hydrochloride hydrate in 50 mL of MilliQ H_2O. Filter-sterilize with a 0.22 μm filter and store at -20 °C in 5 mL aliquots.
4. **100X Transferrin –** Dissolve 50 mg of transferrin in 100 mL sterile PBS. Filter-sterilize with a 0.22 μm filter and store at -20 °C in 5 mL aliquots.
5. **100X T_3 thyroid hormone, 3,3',5-triiodo-L-thyronine sodium salt –** Dissolve 13.6 mg of T_3 in 100 mL of 0.02 M NaOH (2 x 10^{-4} M T_3). Transfer 0.1 mL of 2 x 10^{-4} M T_3 and dissolve in 9.9 mL of sterile PBS (2 x 10^{-6} M T_3). Transfer 0.5 mL of 2 x 10^{-6} M T_3 solution and dilute with 49.5 mL of sterile PBS (2 x 10^{-8} M T_3). Filter-sterilize with a 0.22 μm filter and store at -20 °C in 5 mL aliquots.
6. **100X Hydrocortisone –** Dissolve 25 mg of hydrocortisone in 5 mL 100 % EtOH to generate a 5 mg/mL solution. Transfer 0.4 mL of 5 mg/mL hydrocortisone and dilute with 49.6 mL of HBES buffer. Filter-sterilize with a 0.22 μm filter and store at -20 °C in 5 mL aliquots.
7. **100X Cholera toxin –** Dissolve 0.5 mg of cholera toxin (MW 85 kDa) in 0.6 mL of MilliQ H_2O to generate a 1 x 10^{-5} M stock. Transfer 50 μL of 1 x 10^{-5} M hydrocortisone stock and dilute it to 50 mL of HBES buffer. Filter-sterilize with a 0.22 μm filter and store at 4 °C in 5 mL aliquots.
8. **100X Epidermal Growth Factor (EGF) –** Dissolve the 100 μg EGF vial in 10 mL of MilliQ H_2O. Add 100 mg Bovine Serum Albumin dissolved in 90 mL of HBES buffer. Filter-sterilize with a 0.22 μm filter and store at -20°C in 5 mL aliquots.
9. **50X Mitomycin C –** (Wear gloves! DNA cross-linker) Dilute 2 mg of mitomycin C in 5 mL of HBES. Filter-sterilize with a 0.22 μm filter and store at -20 °C in 5 mL aliquots (Shelf-life: 2 months).
10. **100X Insulin –** Prepare individually and do not freeze. Dissolve 5 mg of insulin in 10 mL of 0.005 M HCl. Filter-sterilize with a 0.22 μm filter by prewetting filter with fetal bovine serum.

Cell culture media

1. **3T3 Fibroblasts Media**
   - **a.** DMEM with high glucose, L-glutamine, phenol red, **no pyruvate** **445 mL**
   - **b.** Fetal Bovine Serum **50 mL**
   - **c.** Penicillin-streptomycin solution (10,000 U Pen and 10 mg Strep per mL) **5 mL**

2. **Keratinocyte Growth Media or E Media (Incomplete)**
   Use a Nalgene 500 mL carboy (Catalog # 1068672)
   - **a.** DMEM with high glucose, L-glutamine, phenol red, **sodium pyruvate** **330 mL**
   - **b.** Ham's F-12 media with L-gluramine, phenol red **110 mL**
   - **c.** Fetal Bovine Serum **25 mL**
   - **d.** 100X Hydrocortisone **5 mL**
   - **e.** 100X Insulin **5 mL**
   - **f.** 100X Cholera toxin **5 mL**
   - **g.** 100X Adenine **5 mL**
   - **h.** 100X Transferrin **5 mL**
   - **i.** Penicillin-streptomycin solution (10,000 U Pen and 10 mg Strep per mL) **5 mL**
   - **j.** 100X T_3 thyroid hormone **5 mL**

3. **E Media (Complete)**
   - **a.** Keratinocyte Growth Medium or E Medium (Incomplete) **495 mL**
   - **b.** 100X Epidermal Growth Factor (EGF) **5 mL**

4. **Freezing Media for 3T3 Fibroblasts (5 % DMSO)**
   - **a.** 3T3 Fibroblasts Media **47.5 mL**
   - **b.** Dimethyl sulfoxide **2.5 mL**

5. **Freezing Media for Keratinocytes (10 % DMSO)**
   - **a.** Keratinocyte Growth Medium or E Medium (Incomplete) **45 mL**
   - **b.** Dimethyl sulfoxide **5 mL**

5.7.4.2 Thawing of 3T3 Fibroblasts (Feeder Cells) from N_2 Chamber

Materials

1. 3T3 fibroblasts received from the laboratory of Paul Lambert and stored in liquid N_2 chamber.
2. 3T3 Fibroblasts Media (DMEM with high glucose, L-glutamine, phenol red, **no pyruvate** supplemented with 10 % fetal bovine serum and 1 % penicillin-streptomycin).

Procedure

1. Place 3T3 Fibroblasts Media in 37 $^{\circ}$C water bath (American Scientific Products S/P Water bath).
2. Obtain 3T3 fibroblast aliquot from the liquid N_2 chamber.
3. Quickly thaw the container by hand in approximately 2 min. Immerse closed contained in 70 % ethanol and wipe with Kimwipes (make sure to wipe neck of container well).

4. Transfer cell suspension to a 15 mL disposable centrifuge tube and add 3T3 Fibroblasts Media to a final volume of 10 mL.
5. Mix suspension well by gently pipetting up and down.
6. Immediately, pipette a 20 µL aliquot from this suspension and add to cellometer slide. Count cells in Cellometer Auto T4 Cell Viability Counter using 3T3 setting.
7. Pellet 3T3 fibroblasts using an Eppendorf Centrifuge 5702 at 500 rcf (relative centrifugal force) for 10 min.
8. Aspirate the media being careful not to disturb the cell pellet.
9. Add the appropriate volume of 3T3 Fibroblasts Media to obtain 4.00 x 10^5 cells/mL (given in step 6). Resuspend the 3T3 cell pellet by gently pipetting up and down.
10. Transfer 1 mL of cell suspension into a 100 mm x 20 mm culture dish (Sigma SIAL0167) and add 9 mL of 3T3 Fibroblasts Media.
11. Place cell culture dish in an incubator at 37 $^{\circ}$C with a 5 % CO_2 atmosphere (NuAire NU-4750 US AutoFlow CO_2 Water-Jacketed).
12. Change 3T3 Fibroblasts Media every other day. Passage 3T3 fibroblasts at 80-90 % confluence. Do not allow cells to reach 100 % confluence.

5.7.4.3 3T3 Fibroblasts (Feeder Cells)

Materials

1. 3T3 fibroblasts in cell culture dish/flask.
2. 3T3 Fibroblasts Media (DMEM with high glucose, L-glutamine, phenol red, **no pyruvate** supplemented with 10 % fetal bovine serum and 1 % penicillin-streptomycin).
3. 1X Phosphate Buffer Saline.
4. 0.05 % trypsin / 0.53 mM EDTA solution.

Procedure

1. Place 3T3 Fibroblasts Media in 37 $^{\circ}$C water bath (American Scientific Products S/P Water bath).
2. Aspirate the media from 3T3 fibroblasts culture and wash cells with 10 mL of 1X PBS. Repeat once.
3. Add 2 mL of 0.05 % trypsin/EDTA and incubate for approximately 5 min. in an incubator at 37 $^{\circ}$C with a 5 % CO_2 atmosphere (NuAire NU-4750 US AutoFlow CO_2 Water-Jacketed) (longer incubation times might be necessary, check dish/flask under microscope).
4. After most of the 3T3 fibroblasts have been detached from the surface, add 2 mL of 3T3 Fibroblasts Media to the plate transfer cell suspension to a 15 mL disposable centrifuge tube. Rinse plate with 6 mL of 3T3 Fibroblasts Media and transfer to the same 15 mL disposable centrifuge tube containing the bulk of the cells.
5. Mix suspension well by gently pipetting up and down.
6. Pellet 3T3 fibroblasts using an Eppendorf Centrifuge 5702 at 500 rcf (relative centrifugal force) for 5 min.
7. Aspirate the media being careful not to disturb the cell pellet.

8. Add 10 mL of 3T3 Fibroblasts Media and resuspend the 3T3 cell pellet by gently pipetting up and down.
9. Transfer 1 mL of cell suspension into a 100 mm x 20 mm culture dish (Sigma SIAL0167) and add 9 mL of 3T3 Fibroblasts Media. This corresponds to a 1:10 split ratio.
10. Place cell culture dish in an incubator at 37 °C with a 5 % CO_2 atmosphere (NuAire NU-4750 US AutoFlow CO_2 Water-Jacketed).
11. Change 3T3 Fibroblasts Media every other day. Passage 3T3 fibroblasts at 80-90 % confluence. Do not allow cells to reach 100 % confluence.

5.7.4.4 Mitomycin C Treatment of 3T3 Fibroblasts (Feeder Cells)

Materials

1. 3T3 fibroblasts in cell culture dish/flask at 80-90 % confluency.
2. 3T3 Fibroblasts Media (DMEM with high glucose, L-glutamine, phenol red, **no pyruvate** supplemented with 10 % fetal bovine serum and 1 % penicillin-streptomycin).
3. E Media (incomplete).
4. 1X Phosphate Buffer Saline.
5. 50X Mitomycin C (Make sure to wear gloves; inhibitor of cell division).
6. 0.05 % trypsin / 0.53 mM EDTA solution.
7. Cellometer Auto T4 Cell Viability Counter slide.

Procedure

1. Place 3T3 Fibroblasts Media and E Media (incomplete) & thaw 50X Mitomycin C in 37 °C water bath (American Scientific Products S/P Water bath).
2. Aspirate the media from 3T3 fibroblasts culture and wash cells with 10 mL of 1X PBS. Repeat once.
3. Add 200 µL of 50X Mitomycin C solution to 10 mL of 3T3 Fibroblasts Media on 3T3 fibroblasts. Swirl gently and incubate for 2-4 hr at 37°C with a 5 % CO_2 atmosphere (NuAire NU-4750 US AutoFlow CO_2 Water-Jacketed).
4. Aspirate the media from 3T3 fibroblasts culture and wash cells with 10 mL of 1X PBS. Repeat once.
5. Add 3 mL of 0.05 % trypsin/EDTA and incubate for approximately 5 min in an incubator at 37 °C with a 5 % CO_2 atmosphere (NuAire NU-4750 US AutoFlow CO_2 Water-Jacketed) (longer incubation times might be necessary, check dish/flask under microscope).
6. After most of the mitomycin C-treated 3T3 fibroblasts have been detached from the surface, add E Media (incomplete) to a final volume of 10 mL and transfer cell suspension to a 15 mL disposable centrifuge tube.
7. Mix suspension well by gently pipetting up and down.
8. Immediately, pipette a 20 µL aliquot from this suspension and add to cellometer slide. Count cells in Cellometer Auto T4 Cell Viability Counter using 3T3 setting.
9. Pellet 3T3 fibroblasts using an Eppendorf Centrifuge 5702 at 500 rcf (relative centrifugal force) for 5 min.

10. Aspirate the media being careful not to disturb the cell pellet.
11. Add the appropriate volume of **E Media (incomplete)** or 3T3 Fibroblasts Media to obtain 4.00 x 10^5 cells/mL (given in step 8). If fibroblasts are to be used within 24 h, resuspend in E Media (incomplete), otherwise resuspend in 3T3 Fibroblasts Media. Use fibroblasts within 48 h after Mitomycin C treatment. Resuspend the 3T3 cell pellet by gently pipetting up and down.
12. Transfer 1 mL of cell suspension into a 100 mm x 20 mm culture dish (Sigma SIAL0167) and add 9 mL of 3T3 Fibroblasts Media (1:10). **Need at least two plates: one for W12E maintenance and one for experiment.**
13. Place cell culture dish in an incubator at 37 °C with a 5 % CO_2 atmosphere (NuAire NU-4750 US AutoFlow CO_2 Water-Jacketed).

5.7.4.5 Thawing of W12E Keratinocytes from N_2 Chamber

Materials

1. W12E keratinocytes received from the laboratory of Paul Lambert and stored in liquid N_2 chamber.
2. E Media (incomplete).
3. Mitomycin C-treated 3T3 fibroblasts in E Media (incomplete).
4. E Media (complete).
5. Cellometer Auto T4 Cell Viability Counter slide.

Procedure

1. Obtain W12E keratinocytes aliquot from the liquid N_2 chamber.
2. Quickly thaw the container by hand in approximately 2 min. Immerse closed contained in 70 % ethanol and wipe with Kimwipes (make sure to wipe neck of container well).
3. Transfer cell suspension to a 15 mL disposable centrifuge tube and add **cold** E Media (incomplete) to a final volume of 10 mL.
4. Mix suspension well by gently pipetting up and down.
5. Immediately, pipette a 20 µL aliquot from this suspension and add to cellometer slide. Count cells in Cellometer Auto T4 Cell Viability Counter using small cell setting.
6. Pellet W12E keratinocytes using an Eppendorf Centrifuge 5702 at 800 rcf (relative centrifugal force) for 5 min.
7. Aspirate the media being careful not to disturb the cell pellet.
8. Add the appropriate volume of E Media (incomplete) to obtain 4.00 x 10^5 cells/mL (given in step 5). Resuspend the W12E cell pellet by gently pipetting up and down.
9. Check under the microscope the dish with the Mitomycin C-treated 3T3 fibroblasts in E Media to ensure cells are attached to the surface. Aspirate the media from 3T3 fibroblasts culture.
10. Add 9 mL of E Media (incomplete) to the 100 mm x 20 mm culture dish (Sigma SIAL0167) containing the mitomycin C-treated 3T3 fibroblasts and transfer 1 mL of the W12E keratinocyte suspension on top of the 3T3 fibroblasts.

11. Incubate cells for 24 h in incubator at 37 °C with a 5 % CO_2 atmosphere (NuAire NU-4750 US AutoFlow CO_2 Water-Jacketed).
12. After the 24 h incubation period, replace the media with 10 mL of E Media (complete).
13. Change E Media (complete) every other day. Passage W12E keratinocytes at 70 % confluence. Do not allow cells to reach 100 % confluence.

5.7.4.6 W12E Keratinocytes

Materials

1. W12E keratinocytes (70 % confluent) in cell culture dish/flask on top of mitomycin C-treated 3T3 fibroblasts.
2. E Media (incomplete).
3. 1X Phosphate Buffer Saline.
4. 0.02 % EDTA / PBS.
5. 0.05 % trypsin / 0.53 mM EDTA solution.
6. Cellometer Auto T4 Cell Viability Counter slide.
7. E Media (complete).

Procedure

1. Place E Media (incomplete) in 37 °C water bath (American Scientific Products S/P Water bath).
2. Aspirate the media from sub-confluent culture.
3. Add 3 mL of 0.02 % EDTA / PBS and incubate for 1 min. After incubation, use a wide-bore tip to spray the plate vigorously with the 0.02 % EDTA / PBS solution. Aspirate solution and wash cells with 10 mL of 1X PBS.
4. Add 3 mL of 0.05 % trypsin/EDTA and incubate for approximately 5 min in an incubator at 37 °C with a 5 % CO_2 atmosphere (NuAire NU-4750 US AutoFlow CO_2 Water-Jacketed) (longer incubation times might be necessary, check dish/flask under microscope).
5. After most of the W12E keratinocytes have been detached, add E Media (incomplete) to a final volume of 10 mL and transfer cell suspension to a 15 mL disposable centrifuge tube.
6. Mix suspension well by gently pipetting up and down.
7. Immediately, pipette a 20 µL aliquot from this suspension and add to cellometer slide. Count cells in Cellometer Auto T4 Cell Viability Counter using HeLa setting.
8. Pellet W12E cells using an Eppendorf Centrifuge 5702 at 500 rcf (relative centrifugal force) for 5 min.
9. Aspirate the media being careful not to disturb the cell pellet.
10. Add the appropriate volume of E Media (incomplete) to obtain 2.00 x 10^5 cells/mL (given in step 7). Resuspend the W12E cell pellet by gently pipetting up and down.
11. Transfer 1 mL of cell suspension into a 100 mm x 20 mm culture dish (Sigma SIAL0167) containing Mitomycin C-treated 3T3 cells and add 9 mL of E Media (incomplete). Incubate cells for 24 h in incubator at 37 °C with a 5 % CO_2 atmosphere (NuAire NU-4750 US AutoFlow CO_2 Water-Jacketed).

12. After the 24 h incubation period, replace media with 10 mL of E Media (complete).
13. Change E Media (complete) every other day. Passage W12E keratinocytes at 70 % confluence. Do not allow cells to reach 100 % confluence.

5.7.4.7 Cryopreservation of 3T3 Fibroblasts (Feeder Cells)

Materials

1. 3T3 fibroblasts in cell culture dish/flask at 80-90 % confluency.
2. 3T3 Fibroblasts Media (DMEM with high glucose, L-glutamine, phenol red, **no pyruvate** supplemented with 10 % fetal bovine serum and 1 % penicillin-streptomycin).
3. Freezing Media for 3T3 Fibroblasts (5 % DMSO).

Procedure

1. Trypsin-treated cells are resuspended in fresh Freezing Media for 3T3 Fibroblasts at approximately 1 x 10^6 cells / mL.
2. Transfer approximately 1 x 10^6 cells into a sterile storage vial that has been wiped with 70 % ethanol.
3. Place vial into an isopropanol holder and transfer them to a -80 $^{\circ}$C chamber to allow gradual cooling of the cells for 24 h.
4. After the 24 h period, place vial into a liquid N_2 chamber.

5.7.4.8 Cryopreservation of W12E Keratinocytes

Materials

1. W12E keratinocytes (70 % confluent) in cell culture dish/flask on top of mitomycin C-treated 3T3 fibroblasts.
2. Freezing Media for Keratinocytes (10 % DMSO).

Procedure

1. Trypsin-treated cells are resuspended in fresh Freezing Media for Keratinocytes at approximately 1 x 10^6 cells / mL.
2. Transfer approximately 1 x 10^6 cells into a sterile storage vial that has been wiped with 70 % ethanol.
3. Place vial into an isopropanol holder and transfer them to a -80 $^{\circ}$C chamber to allow gradual cooling of the cells for 24 h.
4. After the 24 h period, place vial into a liquid N_2 chamber.

5.7.5 Fenton Chemistry: ·OH Radical Cleavage of Viral and Host DNA in W12E (HPV16-harboring) Keratinocytes

MATERIALS

1. **W12E keratinocytes** (70 % confluent) in cell culture dish/flask with mitomycin C-treated 3T3 fibroblasts.
2. **1 M NaOH** (Store at 4 °C)
   Dissolve 2.00 g of NaOH pellets (Fisher Scientific, Catalog # BP359-500) in 50 mL of autoclaved MilliQ H_2O.
3. **1X Phosphate Buffer Saline** (Fisher Scientific, Catalog # 10010-031)
4. **500 mM Sucrose Stock** (Store at 4 °C)
   Dissolve 34.23 g of sucrose (Fluka, Catalog # 84097; MW 342.30 g/mol) in 200 mL of autoclaved MilliQ H_2O. Filter-sterilize using a 0.22 μm MILLEX-GP filter.
5. **1,600 mM KCl Stock**
   Dissolve 5.96 g of potassium chloride (Sigma-Aldrich, Catalog # P9333-500G; MW 74.55 g/mol) in 50 mL of autoclaved MilliQ H_2O. Filter-sterilize using a 0.22 μm MILLEX-GP filter.
6. **100 mM K_2HPO_4 Stock**
   Dissolve 1.14 g of potassium phosphate dibasic trihydrate (Sigma-Aldrich, Catalog # P5504-100G; MW 228.22 g/mol) in 50 mL of autoclaved MilliQ H_2O. Filter-sterilize using a 0.22 μm MILLEX-GP filter.
7. **100 mM $MgCl_2$ Stock**
   Dissolve 476 mg of magnesium chloride (Sigma-Aldrich, Catalog # M8266-100G; MW 95.21 g/mol) in 50 mL of autoclaved MilliQ H_2O. Filter-sterilize using a 0.22 μm MILLEX-GP filter.
8. **10 mM $CaCl_2$ Stock**
   Dissolve 55.5 mg of calcium chloride (Sigma-Aldrich, Catalog # C1016-100G; MW 110.98 g/mol) in 50 mL of autoclaved MilliQ H_2O. Filter-sterilize using a 0.22 μm MILLEX-GP filter.
9. **100 mM NaHepes buffer, pH 7.4**
   Dissolve 11.92 g of 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (Sigma-Aldrich, Catalog # H3375-100G; MW 238.30 g/mol) in 400 mL of autoclaved MilliQ H_2O. Adjust pH to 7.4 using 1M NaOH (~25 mL). Adjust total volume to 500 mL with autoclaved MilliQ H_2O. Filter-sterilize using a 0.2 μm aPES membrane Nalgene filter unit.
10. **Incomplete Permeabilization Solution** (Store at 4 °C)
    Combine 180 mL of 500 mM sucrose stock, 30 mL of 1,600 mM KCl stock, 30 mL of 100 mM K_2HPO_4 stock, 30 mL of 100 mM $MgCl_2$, 30 mL of 10 mM $CaCl_2$, 210 mL of 100 mM NaHepes buffer (pH 7.4) and 90 mL of autoclaved MilliQ H_2O. Filter-sterilize using a 0.2 μm aPES membrane Nalgene filter unit.
11. **0.05 % Lysolecithin Permeabilization Solution** (prepare immediately before use)
    Dissolve 0.5 mg of lysolecithin Type I (Sigma-Aldrich, Catalog # L4129-100MG) per 1 mL of Incomplete Permeabilization Solution. Filter-sterilize using a 0.22 μm MILLEX-GP filter.
12. **Stop Buffer** (Store at 4 °C)
    Dissolve 7.6 g of thiourea (Sigma-Aldrich, Catalog # 88810-100G) in 1000 mL of 1X PBS buffer. Filter-sterilize using a 0.2 μm aPES membrane Nalgene filter unit.
13. **500 mM EDTA disodium Stock** (Store at 4 °C)
    Dissolve 7.44 g of ethylenediaminetetraacetic acid disodium salt dehydrate, EDTA disodium salt dehydrate (Sigma-Aldrich, Catalog # E5134-100G; MW 372.24

g/mol) in 20 mL of autoclaved MilliQ H_2O. Adjust pH to 7.4-8.0 using 1 M NaOH. Adjust total volume to 40 mL with autoclaved MilliQ H_2O. Filter-sterilize using a 0.22 μm MILLEX-GP filter.

14. **50X Fe(II)-EDTA Stock** (prepare immediately before use)
    Dissolve 196 mg of ammonium iron(II) sulfate hexahydrate, $Fe(NH_4)_2(SO_4)_2 \cdot 6\ H_2O$ (Sigma-Aldrich, Catalog # 203505-25G; MW 392.14 g/mol) in 2 mL of 500 mM EDTA disodium stock.
15. **DMEM (high glucose, L-glutamine, phenol red, no pyruvate)** (Gibco 11965-084)
16. **30 % wt/vol H_2O_2 Solution** (Fisher Scientific, Catalog # H325-500)

PROCEDURE

1. Aliquot 100 mL of **Incomplete Permeabilization Solution** and allow to equilibrate to room temperature.
2. Aliquot 100 mL **DMEM Media** and allow to equilibrate to room temperature. Bring 30 % H_2O_2 to cell culture lab.
3. Aspirate the media from sub-confluent culture.
4. Add 5 mL of 0.02 % EDTA / PBS, wait 60 s (shake dish to aid in the Versene treatment). After incubation, use a wide-bore tip to spray the plate vigorously with the 0.02 % EDTA / PBS solution. Aspirate solution and wash cells three times with 10 mL of **1X PBS**.
5. During the Versene treatment incubation period, prepare **0.05 % Lysolecithin Permeabilization Solution** by dissolving lysolecithin (0.5 mg/mL; 50 mg for 100 mL) in **Incomplete Permeabilization solution**. Also, prepare **50X Fe(II)-EDTA Stock** and weigh sodium ascorbate. Set aside for Step 9.
6. Add 10 mL of **0.05 % Lysolecithin Permeabilization Solution** to each cell culture dish and incubate for 60 seconds at room temperature.
7. Aspirate the permeabilization solution.
8. Add 991 mg of (+)-Sodium L-ascorbate (Sigma-Aldrich, Catalog # 11140-50G; MW 198.11 g/mol) to 100 mL of **DMEM Media** and mix well. Add 2 mL of **50X Fe(II)-EDTA Stock** followed by 2 mL of **30 % H_2O_2**. Immediately mix for 5 seconds by vortexing.
9. Add 10 mL of the solution prepare in Step 9 onto each cell culture plate. Gently swirl the plates and incubate at room temperature for 15 minutes. Check each plate under microscope for cell detachment.
10. Aspirate the footprinting solution from the dishes and wash cells twice with 10 mL of **Stop Buffer**.
11. Wash cells with **1X PBS**.
12. Add 3 mL of 0.05 % trypsin/EDTA and incubate for approximately 5 min in an incubator at 37 °C with a 5 % CO_2 atmosphere (NuAire NU-4750 US AutoFlow CO_2 Water-Jacketed) (longer incubation times might be necessary, check dish/flask under microscope).
13. After most of the W12E keratinocytes have been detached, add **DMEM Media** to a final volume of 5 mL and transfer cell suspension to a 15 mL disposable centrifuge tube.
14. Pellet W12E cells using an Eppendorf Centrifuge 5702 at 500 rcf (relative centrifugal force) for 5 min.
15. Aspirate the media being careful not to disturb the cell pellet.
16. Place tubes in ice and perform modified Hirt DNA extraction as soon as possible.

Reference: Cannistraro, V.J.; Pondugula, S.; Song, Q.; Taylor, J.-S. Rapid Deamination of Cyclobutane Pyrimidine Dimer Photoproducts at TCG Sites in a Translationally and Rotationally Positioned Nucleosome *in Vivo*. *Journal of Biological Chemistry* **209**, 26597-26609 (2015)

5.7.6 Modified Hirt Extraction of HPV Episomes from W12E (HPV16-harboring) Keratinocytes

Materials

1. W12E keratinocytes (10^6-10^7 cells).
2. MilliQ H_2O (18.2 MΩ·cm at 25°C).
3. **Cell wash buffer** – Ca^{2+} / Mg^{2+}-free Phosphate-buffered saline (PBS).
4. **Resuspension buffer (Buffer P1)** – 50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 100 µg/mL RNase A.
5. **Lysis solution** – 1.2 % sodium dodecyl sulfate (SDS); 10 mM EDTA.
   Dissolve 1.2 g SDS (MW = 288.37 g/mol) in 80 mL MilliQ H_2O. Add 2 mL of 0.5 M EDTA. Add MilliQ H_2O to a total volume of 100 mL.
6. **Precipitation solution** – 3M cesium chloride, 1 M potassium acetate, 0.67 M acetic acid.
   Dissolve 25.25 g CsCl (MW = 168.36 g/mol), 4.91 g potassium acetate (98.14 g/mol) and 1.93 mL glacial acetic acid (17.4 M) in 40 mL MilliQ H_2O. Add MilliQ H_2O to a total volume of 50 mL.
7. Ice bath.
8. Centrifuge (14,000x *g*).
9. QIAprep Miniprep Kit with 2.0 Spin Columns (up to 20 µg episomal DNA).
10. **Column wash buffer** – 80 mM potassium acetate, 10 mM Tris-HCl, pH 7.5, 40 µM EDTA, 60 % ethanol.
    Dissolve 3.93 g potassium acetate (98.14 g/mol), 5 mL 1M Tris (pH 7.5), 40 µL 0.5 M EDTA in 100 mL MilliQ H_2O. Add 300 mL 100 % ethanol and MilliQ H_2O to a total volume 500 mL.
11. **Elution buffer** – EB Buffer (10 mM Tris-HCl, pH 8.5) or Low EDTA TE.

Procedure

1. Centrifuge cells (10^6-10^7) at 500*x g*, 10 min and wash pellet with cold **cell wash buffer**.
2. Resuspend cells in 250 µL of **resuspension buffer (Buffer P1)**.
3. Lyse cells with 250 µL of **lysis solution.**
4. Gently mix the solution by inverting the microcentrifuge tube (**do not vortex**). Allow solution to incubate for 5 min at room temperature.
5. Add 350 µL of **precipitation solution** to precipitate denatured proteins and chromosomal DNA.
6. Immediately mix solution by inverting the tube (**do not vortex**). Place tube into an ice bath for 15 min.
7. Centrifuge at 16,000x *g* at 4 °C for 30 min.
8. Carefully transfer the supernatant to a QIAprep Spin Column.
9. Centrifuge for 60 s at 14,000x *g* and discard the flow-through.
10. Wash the column by adding 750 µL of **column wash buffer**.
11. Centrifuge for 60 s at 14,000*x g* and discard the flow-through.
12. Centrifuge the column an additional minute at 14,000*x g* to remove the residual **column wash buffer**.
13. Place the inner column into a clean 1.5 mL Eppendorf microcentrifuge tube and elute the DNA by adding 50 µL of **elution buffer** onto the center of the column.
14. Allow the column to stand for 60 s and centrifuge at 14,000*x g* for 60 s.

Reference: Arad, U. Modified Hirt Procedure for Rapid Purification of Extrachromosomal DNA from Mammalian Cells. *BioTechniques* **24**, 760-762 (1998).

5.7.7 Crosslinking of Small Molecules to Isolate Chromatin (COSMIC)

Materials

1. **W12E keratinocytes (~2.5 x 10^7 cells)** (70 % confluent) in cell culture dish/flask with mitomycin C-treated 3T3 fibroblasts.
2. **NuAire NU-4750 US AutoFlow CO_2 Water-Jacketed**
3. **Polyamide-psoralen/biotin (PB)**
4. **E Media (complete) and E Media (incomplete)**
5. **1X Phosphate Buffer Saline (PBS)** (HyClone, Catalog # SH30256.02)
6. **CalSun Facial Tanning Sun Lamp / Face Tanner 110V** (CalSun, Catalog # B001BH0A1A)
7. **Glass filter** (23 cm x 40 cm x 3 mm)
8. **0.05 % trypsin / 0.53 mM EDTA solution** (ThermoFisher, Catalog # 25300-062)
9. **0.02 % EDTA / PBS** (Store at room temperature)
   Combine 1.325 mL of 100 mM EDTA disodium solution with 248.675 mL autoclaved MilliQ H_2O. Filter-sterilize using a 0.22 μm MILLEX-GP filter.
10. **Thermo Scientific Sorvall ST16R centrifuge**
11. **Benchtop shaker/rotator** (Barnstead/Thermolyne Labquake shaker/rotator 400110)
12. **Streptavidin-coated magnetic beads** (Dynabeads® MyOne™ Streptavidin C1, Catalog # 65001)
13. **QIAquick PCR purification kit** (QIAGEN, Catalog # 28104)
14. **1 M Tris-HCl (pH 8.1)** (Store at 4 °C)
    Dissolve 121.14 g of Tris (Trizma base; Sigma-Aldrich, Catalog # T1503-1KG; MW = 121.14 g/ mol) in 800 mL of autoclave MilliQ H_2O. Adjust pH to 8.1 using concentrated HCl. Adjust total volume to 1000 mL with autoclaved MilliQ H_2O. Filter-sterilize using a 0.2 μm aPES membrane Nalgene filter unit.
15. **5 M NaCl** (Store at 4 °C)
    Dissolve 29.22 g NaCl (Fisher Bioreagents, Catalog # BP358-1; MW = 58.44 g/mol) in 90 mL autoclaved MilliQ H_2O. Add autoclaved MilliQ H_2O to a total volume of 100 mL.
16. **500 mM EDTA disodium** (Store at 4 °C)
    Dissolve 7.44 g of ethylenediaminetetraacetic acid disodium salt dehydrate, EDTA disodium salt dehydrate (Sigma-Aldrich, Catalog # E5134-100G; MW 372.24 g/mol) in 20 mL of autoclaved MilliQ H_2O. Adjust pH to 8.0 using 1M NaOH. Adjust total volume to 40 mL with autoclaved MilliQ H_2O. Filter-sterilize using a 0.22 μm MILLEX-GP filter.
17. **10 % Sodium dodecyl sulfate (SDS)** (Store at room temperature)
    Dissolve 5 g SDS powder (Sigma-Aldrich, Catalog # L4509-100G; MW = 288.37 g/mol) in 40 mL autoclaved MilliQ H_2O. Add autoclaved MilliQ H_2O to a total volume of 50 mL.
18. **10 % Triton X-100** (Store at room temperature)
    Dissolve 5 mL Triton X-100 (Sigma-Aldrich, Catalog # X100-100ML) in 40 mL autoclaved MilliQ H_2O. Add autoclaved MilliQ H_2O to a total volume of 50 mL.
19. **COSMIC buffer** (20 mM Tris-HCl [pH 8.1], 2 mM EDTA, 150 mM NaCl, 1 % Triton-X100, 0.1 % SDS) (Store at 4 °C)
    Combine 1 mL of 1 M Tris-HCl (pH 8.1), 0.2 mL of 500 mM EDTA, 1.5 mL of 5 M NaCl, 5 mL of 10 % Triton X-100 (Sigma-Aldrich, Catalog # X100-100ML), 0.5 mL of 10 % SDS and 41.8 mL of autoclaved MilliQ H_2O. Filter-sterilize using a 0.22 μm MILLEX-GP filter.

20. **100 mM Phenylmethylsulfonyl fluoride (PMSF)** (Store at -20 °C)
    Dissolve 17.5 mg of PMSF powder (G-Biosciences, Catalog # 786-055; MW = 174.19 g/mol) in 1 mL of anhydrous ethanol 200 proof (Decon Labs, Inc. Catalog # DSP-MD.43)
21. **100 mM Benzamidine HCl** (prepare fresh each time)
    Dissolve 15.7 mg of benzamidine HCl (Fisher BioReagents, Catalog # BP435-25; MW = 156.61 g/mol) in 1 mL of autoclaved MilliQ H_2O.
22. **150 µM Pepstatin A** (Store at -20 °C)
    Dissolve 0.1543 mg of pepstatin A (Alfa Aesar, Catalog # J60237-LB0; MW = 685.90 g/mol) in 1.5 mL of sterile DMSO. Hydrolysis leads to a more yellow solution.
23. **Wash Buffer 1** (10 mM Tris-HCl [pH 8.0], 1 mM EDTA, 3 % v/v SDS) (Store at 4 °C)
    Add 100 µL of 1 M Tris-HCl buffer pH 8.0 (Invitrogen UltraPure, Catalog # 15-568-025), 20 µL of 500 mM EDTA disodium, 3 mL of 10 % SDS and 6.88 µL of autoclaved MilliQ H_2O. Filter-sterilize using a 0.22 µm MILLEX-GP filter.
24. **Wash Buffer 2** (10 mM Tris-HCl [pH 8.0], 1 mM EDTA, 250 mM LiCl, 0.5 % NP40, 1 % sodium deoxycholate) (Store at 4 °C)
    Add 100 µL of 1 M Tris-HCl buffer pH 8.0 (Invitrogen UltraPure, Catalog # 15-568-025), 20 µL of 500 mM EDTA disodium, 105.95 mg of LiCl powder (Acros Organics, Catalog # 199880050; MW = 42.38 g/mol), 50 µL of NP40 (MP Biomedicals IGEPAL CA-630, Catalog # 198596) and 100 mg of deoxycholate acid sodium salt (Sigma-Aldrich, Catalog # D6750-10G, MW = 414.55 g/mol). Add autoclaved MilliQ H_2O to a total volume of 10 mL and filter-sterilize using a 0.22 µm MILLEX-GP filter.
25. **Wash Buffer 3** (10 mM Tris-HCl [pH 7.5], 1 mM EDTA, 4 M urea, 0.1 % NP40) (prepare fresh each time)
    Add 100 µL of 1 M Tris-HCl buffer pH 7.5 (Invitrogen, Catalog # 15567-027), 20 µL of 500 mM EDTA disodium, 2.4024 g of urea crystals (Sigma-Aldrich U5128-100G, MW = 60.06 g/mol) and 10 µL of NP40 (MP Biomedicals IGEPAL CA-630, Catalog # 198596). Add autoclaved MilliQ H_2O to a total volume of 10 mL and filter-sterilize using a 0.22 µm MILLEX-GP filter.
26. **Tris EDTA (TE) Buffer** (10 mM Tris-HCl buffer [pH 8.0], 1 mM EDTA) (Store at room temperature)
    Add 500 µL of 1 M Tris-HCl buffer pH 8.0 (Invitrogen UltraPure 15-568-025) and 100 µL of 500 mM EDTA disodium. Add autoclaved MilliQ H_2O to a total volume of 50 mL and filter-sterilize using a 0.22 µm MILLEX-GP filter.
27. **10X Crosslink Reversal Buffer** (100 mM Tris-HCl [pH 8.0], 4 mM EDTA, 1M KOH) (Store at RT)
    Add 1.5 mL of 1 M Tris-HCl buffer pH 8.0 (Invitrogen UltraPure 15-568-025), 120 µL of 500 mM EDTA disodium and 0.84 g of KOH powder (Asfa Aesar A16199-0E, MW = 56.11 g/mol). Add autoclaved MilliQ H_2O to a total volume of 15 mL and filter-sterilize using a 0.22 µm MILLEX-GP filter.
28. **6N HCl** (Store at room temperature)
    Dilute 5 mL of concentrated HCl (Fisher Scientific A144C-212) with 5 mL of autoclaved MilliQ H_2O. filter-sterilize using a 0.22 µm MILLEX-GP filter.
29. **pH indicator paper** (Whatman Type CF 2613991)
30. **RNase A** (Thermo Scientific EN0531, 10 mg/mL) (Store at -20 °C)
31. **Proteinase K** (Thermo Scientific FEREO0491, 20 mg/mL) (Store at -20 °C)
32. **Water, nuclease-free** (ThermoFisher Scientific R0581)

Procedure

I. Crosslinking in Live Cells

1. Cell culture media and PBS are always added to the side of culture dishes to prevent disrupting the attached cells.
2. Start with ~2.5 x 10^7 **W12E keratinocytes** (this corresponds to approximately ten 10-cm culture dishes).
3. Grow **W12E keratinocytes** in **E Media (complete)** on 10-cm culture dishes in an incubator at 37 °C with a 5 % CO_2 atmosphere (NuAire NU-4750 US AutoFlow CO_2 Water-Jacketed).
4. From this step forward, care must be taken to keep light exposure to a minimum in order to avoid premature photo-crosslinking.
5. Prepare at least 80 µL of 100 µM polyamide-psoralen/biotin (PB) conjugate in sterile DMSO (each 10-cm culture dish requires 8 µL). Determine PA-PB concentration using an UV-Vis spectrophotometer.
6. Aspirate the spent **E Media (complete)** from the sub-confluent culture and add 8 mL of fresh **E Media (complete)**.
7. Add 8 µL of 100 µM polyamide-PB conjugate (100 nM final concentration in culture media) directly to the culture media of each 10-cm culture dish. Gently swirl the dish to evenly disperse the polyamide-PB conjugate into the media.
8. Incubate **W12E keratinocytes** for 24 hours in an incubator at 37 °C with a 5 % CO_2 atmosphere (NuAire NU-4750 US AutoFlow CO_2 Water-Jacketed). Ensure to protect dishes from light.
9. Aspirate media and wash cells with 4 mL of **1X PBS**.
10. Aspirate **PBS** and add 3 mL of **E Media (incomplete)**.
11. Dim the lights and remove the lid of five dishes. Set the dishes containing **W12E keratinocytes** on a flat surface and place a **Glass filter** (23 cm x 40 cm x 3 mm) over the five culture dishes to filter out UV-light with $\lambda < 300$ nm. Set the UV source (CalSun Facial Tanning Sun Lamp / Face Tanner 110V) on top of G**lass filter** and crosslink cells for 30 minutes with 365 nm UV irradiation (2.4 mW/cm^2). Repeat for remaining dishes. Proceed with protocol while crosslinking is performed for the second set of dishes.
12. Aspirate media and wash cells with 4 mL of **1X PBS**.
13. Add 3 mL of **0.02 % EDTA / PBS** and incubate for 1 min. After incubation, use a wide-bore tip to spray the plate vigorously with the **0.02 % EDTA / PBS** solution. Aspirate solution and wash cells with 4 mL of **1X PBS**.
14. Aspirate **1X PBS** and add 3 mL of **0.05 % trypsin / 0.53 mM EDTA solution** per 10-cm dish. Incubate for 5 min in an incubator at 37 °C with a 5 % CO_2 atmosphere (NuAire NU-4750 US AutoFlow CO_2 Water-Jacketed).
15. Quench trypsin by adding 3 mL of **E Media (incomplete)** per dish and transfer cell suspension to one 50 mL disposable conical tube. Place **W12E keratinocytes** on ice from this step forward.
16. Pellet **W12E keratinocytes** using a **Thermo Scientific Sorvall ST16R centrifuge** at 500 rcf (relative centrifugal force), 4 °C for 5 minutes. Aspirate supernatant.
17. Wash pellet with 3 mL cold **1X PBS** and centrifuge at 500 rcf, 4 °C for 5 minutes. Aspirate supernatant.

Protocol can be stopped at this point. Flash-freeze W12E keratinocytes in liquid nitrogen and store samples at -80 °C until ready to continue.

II. Isolation of Chromatin

1. Add 1.2 mL **COSMIC buffer** and resuspend cells by pipetting up and down several times.
2. To this suspension, add 13.3 µL each of **fresh** 100 mM PMSF, 100 mM benzamidine and 150 µM pepstatin protease inhibitors to a final concentration of 1 mM PMSF, 1 mM benzamidine and 1.5 µM pepstatin.
3. Split cell lysate solution into two amber microcentrifuge tubes.
4. Sonicate cell lysate solutions with sonicator (QSonica Q800R) for 2 hours (10 seconds on, 10 seconds off, 50 % power) at 4 °C to fragment host and viral genomes to a range of 100-500 bp. Ensure level of water in reservoir is parallel to the level of chromatin solution.
5. Place samples in ice.
6. Centrifuge the samples at 12,000*x g*, 4 °C for 10 minutes.
7. Carefully transfer supernatant containing soluble chromatin into a new amber microcentrifuge tube. Discard the pellet.
8. Save 10 % of sample in a new microcentrifuge tube and label it as Input DNA. Store at -80 °C.
9. Place the remaining chromatin sample in ice for the next steps of protocol.

Protocol can be stopped at this point. Store samples at -80 °C until ready to continue.

III. Capture of Ligand-DNA Crosslinks

1. Resuspend the **streptavidin-coated magnetic beads** (Dynabeads® MyOne™ Streptavidin C1 65001) by vortexing for >30 seconds.
2. Transfer 120 µL of streptavidin-coated magnetic beads in an amber microcentrifuge tube.
3. Add 1 mL of **COSMIC buffer** and mix for 5 minutes at room temperature on a shaker/rotator (Barnstead/Thermolyne Labquake shaker/rotator 400110).
4. Place the sample of streptavidin-coated magnetic beads on a magnetic stand until the beads settle and a clear solution is observed (~ 2 minutes).
5. Carefully remove and discard the clear supernatant.
6. Immediately, add chromatin sample (~1 mL) to streptavidin-coated magnetic beads (120 µL) and resuspend the beads by pipetting up and down several times.
7. Incubate the sample suspension for at least 4 hours (or overnight) at 4 °C on a shaker/rotator.

IV. Isolation of Affinity-purified DNA

1. Prepare Wash Buffers 1-3. Wash Buffers 1 and 2 can be stored at 4 °C for several months. Wash Buffer 3 must be prepared fresh each day.
2. Add non-functionalized polyamide to Wash Buffers 1-3 to a final concentration 5 µM.

   Add 4.75 µL of 10 mM non-functionalized polyamide in sterile DMSO to 9.50 mL of Wash Buffer 1.
   Add 2.13 µL of 10 mM non-functionalized polyamide in sterile DMSO to 4.25 mL of Wash Buffer 2.
   Add 2.25 µL of 10 mM non-functionalized polyamide in sterile DMSO to 4.50 mL of Wash Buffer 1.

3. At the end of the incubation period, place the sample suspension on a magnetic stand until the beads settle and a clear solution is observed (~ 2 minutes). Carefully remove and

discard the clear supernatant. Do not disturb the bead pellet, as the DNA—PA-PB is associated with the beads.

4. Add 1 mL of **COSMIC buffer** and mix for 12 hours at 4 °C on a shaker/rotator. Repeat wash and incubate for 4 hours.
5. Add 1 mL of **Wash Buffer 1** to the sample and resuspend the streptavidin-coated magnetic beads. Incubate the sample suspension for 7 minutes at room temperature on a shaker/rotator. Place the sample suspension on a magnetic stand until the beads settle and a clear solution is observed (~ 2 minutes). Carefully remove and discard the clear supernatant.
6. Add 1 mL of **Wash Buffer 2** to the sample and resuspend the streptavidin-coated magnetic beads. Incubate the sample suspension for 7 minutes at room temperature on a shaker/rotator. Place the sample suspension on a magnetic stand until the beads settle and a clear solution is observed (~ 2 minutes). Carefully remove and discard the clear supernatant.
7. Add 1 mL of **Wash Buffer 3** to the sample and resuspend the streptavidin-coated magnetic beads. Incubate the sample suspension for 7 minutes at 50 °C on a shaker/rotator. Place the sample suspension on a magnetic stand until the beads settle and a clear solution is observed (~ 2 minutes). Carefully remove and discard the clear supernatant.
8. Repeat step 7.
9. Add 1 mL of **TE Buffer** to the sample and resuspend the streptavidin-coated magnetic beads. Place the sample suspension on a magnetic stand until the beads settle and a clear solution is observed (~ 2 minutes). Carefully remove and discard the clear supernatant.
10. Repeat step 9.
11. Resuspend the streptavidin-coated magnetic beads in 200 µL of **TE Buffer**. This sample of capture DNA is referred to as Affinity-purified DNA.
12. Reverse the psoralen-DNA crosslinks by adding **10X Crosslink Reversal Buffer** to Input DNA and Affinity-purified DNA aliquots to a final concentration of 1X. Incubate samples for 30 minutes at 90 °C.
13. For the Affinity-purified DNA, place the sample suspension on a magnetic stand until the beads settle and a clear solution is observed (~ 2 minutes). Transfer the entire eluate supernatant to a clean amber microcentrifuge tube, ensuring not to pipette any magnetic beads.
14. Neutralize the Input and Affinity-purified DNA samples to pH 7 by adding ~1 µL **6 N HCl** per 100 µL sample. Confirm samples have been neutralized by adding a drop (~0.5 µL) onto pH indicator paper.
15. Add **RNase A** (Thermo Scientific EN0531, 10 mg/mL) to the Input DNA and Affinity-purified DNA (0.2 µg/µL final RNase A concentration). Incubate for 1 hour at 37 °C.
16. Add **Proteinase K** (Thermo Scientific FEREO0491, 20 mg/mL) to the Input DNA and Affinity-purified DNA (0.2 µg/µL final Proteinase K concentration). Incubate for 1 hour at 55 °C.
17. Purify each DNA sample with a **QIAquick PCR purification kit** (QIAGEN 28104). Elute DNA in 58 µL of **Water, nuclease-free** (ThermoFisher Scientific R0581). Store Input and Affinity-purified DNA samples at -20 or -80 °C.
18. Analyze samples by Next-generation sequencing.

Reference: Erwin, G.S.; Grieshop, M.P.; Bhimsaria, D.; Eguchi, A.; Rodríguez-Martinez, J.A.; Ansari, A.Z. Genome-wide Mapping of Drug-DNA Interactions in Cells with COSMIC (Crosslinking of Small Molecules to Isolate Chromatin). *Journal of Visualized Experiments* **107**, e53510 (2016)

5.7.8 Right Side Size Selection with SPRI of NGS Libraries

Purpose: To remove DNA fragments ≥ 800 bp in prepared Illumina NGS libraries.

1. Homogenize the SPRIselect beads by gently vortexing the solution.
2. Transfer each **Sample** volume to a clean 1.5 mL Eppendorf microcentrifuge vial and dilute to a final volume of 50 µL with Low TE buffer. Add the appropriate **SPRI** volume to each sample.

   Sample volume * 0.6 (ratio) = **SPRI** volume
3. Homogenize the sample by vortexing or pipetting 10 times. Briefly spin down the samples to collect droplets (no bead-sample droplets on the sides of the tube).
4. Incubate the sample suspensions for 1 min at room temperature.
5. Place the sample suspension on a magnetic stand until the beads settle and a clear solution is observed (~ 2 minutes).
6. Carefully remove and **transfer** the clear supernatant to a clean vial. Do not disturb the bead pellet, as the large DNA fragments are associated with the beads.
7. Add the appropriate **SPRI** volume to each transferred supernatant sample.

   Sample (initial) volume * (1.8x – 0.6) = **SPRI** volume
8. Homogenize the sample by vortexing or pipetting 10 times. Briefly spin down the samples to collect droplets (no bead-sample droplets on the sides of the tube).
9. Incubate the sample suspensions for 1 min at room temperature.
10. Place the sample suspension on a magnetic stand until the beads settle and a clear solution is observed (~ 2 minutes).
11. Carefully remove and discard the clear supernatant. Do not disturb the bead pellet, as the DNA is associated with the beads.
12. While on the magnet stand, add 180 µL of the freshly prepared 85 % ethanol solution to the vial without disturbing the bead pellet. Incubate for 30 s, and carefully remove and discard the ethanol solution. Do not disturb the bead pellet, as the DNA is associated with the beads.
13. Briefly spin down the sample. Place back on the magnetic stand and remove any residual ethanol from the bottom of the vial.
14. Air-dry the bead pellet.
15. Resuspend the pellet by adding **Elution** volume (20 µL) of Low EDTA TE. Homogenize the suspension by pipetting up and down. Briefly spin down the samples to collect droplets (no bead-sample droplets on the sides of the tube).
16. Incubate suspension for 1 minute and place the tube on the magnetic stand.
17. Transfer the entire eluate supernatant to a clean 0.2 mL PCR vial, ensuring not to pipette any magnetic beads. If magnetic beads are present in the eluate, transfer eluate to a new 0.2 mL PCR vial, place on magnetic stand and collect eluate again