9-5-2016

In Vivo And *In Vitro* Studies Of Polyamides That Are Active Antiviral Agents Against HPV16

Edith Csiki-Fejer
University of Missouri-St. Louis, efhv3@umsl.edu

IN VIVO AND *IN VITRO* STUDIES OF POLYAMIDES THAT ARE ACTIVE ANTIVIRAL AGENTS AGAINST HPV16

by

Edith Csiki-Fejer

M.S. in Chemistry / Biochemistry, May, 2014, University of Missouri-St. Louis

M.S. in Chemistry / Physical, June, 1996, University of Transylvania-Brasov

A Thesis

Submitted to The Graduate School of the

University of Missouri-St. Louis

in partial fulfillment of the requirements for the degree

Doctor of Philosophy

In

Chemistry

With an Emphasis in Biochemistry

August 2016

Advisory Committee

James K. Bashkin, Ph.D.

Chairperson

Benjamin J. Bythell, Ph.D.

Keith J. Stine, Ph.D.

Chung F. Wong, Ph.D.

For

Ágota, Mátyás, Anna,

and Bálint

Acknowledgements

It has been a privilege to work under the direction of Prof. James K. Bashkin at UMSL. Under his supervision, I had the freedom and support to be able to pursue my ideas. Working in the Bashkin group over the last five years has allowed me to learn many research techniques and collaborate with a large group of scientists. I would like to express my appreciation for his guidance. I am tremendously thankful for his patience and assistance.

I am also grateful to the supportive faculty at UMSL, a distinguished group of scientists, and to the members of my thesis committee who have always been available for advice or assistance. I will always be grateful to Prof. Cynthia M. Dupureur for her useful insights on footprinting experiments and Prof. Michael E. Hughes and his assistant Erin Arant for their training on next generation sequencing and stimulating scientific discussions.

Among the former members of the Bashkin group, I'm glad to have known Gaofei He, former post-doc, Priyanka Bapat, Faten Tamimi, and Silke Evdokimov, great colleagues who made tough days bearable. I was fortunate to have worked with Fanny Hakami on the bioavailability project and enjoy her uplifting spirit, which helped me through my first years of graduate school. The precursory work to my projects, the synthesis and purification of the polyamides, was done by Dr. Kevin Koeller and Dr. George Harris. I must also thank Dr. E. Vasilieva, J. Niederschulte, Y. Song (members of Dupureur lab), and Dr. Shana Terrill and Astha Ahuja, members of the Nichols Lab, for introducing me to cell culture work. Among current members of the group, thanks to Jose' Scuderi and Carlos Castañeda, who always were helpful lab mates during four years of intensive work. Our discussions helped shape the direction of the work presented herein.

I would also like to acknowledge our collaborators at NanoVir, LLC, Dr. T.G. Edwards and Dr. C. Fisher, for their antiviral potency studies. Special thanks to our collaborators at the Univ. of Missouri, Colombia, the DNA Core Facility members (especially E. Kessler) who ran numerous samples prepared for the CE, as well as N. Bivens, who performed the sequencing on a HiSeq 2500 Illumina instrument. Also, I am eagerly looking forward to the sequencing results' interpretation that is being performed at Informatics Research Core Facility by Christopher Bottoms under the supervision of Dr. Scott Givan.

Many thanks to my family, the most important people in my life. Without their mutual understanding and support this work would not have been completed.

Abstract

Long-term, persistent infection with high-risk strains of human papillomavirus (HPV) is the precursor of most cervical cancers and an increasing number of the head and neck cancers. While HPV vaccines can protect patients under twenty-five years old from possible infections, no HPV antiviral drugs are available for the definitive treatment of preexisting or future viral infections. To treat current or future HPV infections, drugs that selectively block virus-specific processes, but do not damage the host cells, are needed. The compounds developed in our group are unique Imidazole-pyrrole polyamides, analogs of natural products Distamycin A and Netropsin that function by interfering with natural virus-host interactions.

Within the overall program of my group, my work focuses on three specific projects. The first project (Chapter 2) includes bioavailability LC-MS/MS analysis study of one of the leading compounds in plasma and whole blood.

The second project includes the biophysical study of interactions between two polyamides, called TMG Asymmetric Hairpin Polyamides (TMG-AHP), and an essential viral DNA segment. The aim of this study was to determine where and how strongly the compounds bind to the viral DNA. In this work, two methods were used: the quantitative deoxyribonuclease (DNase I) footprinting method and the affinity cleavage assay (AC). The results are presented in Chapters 3 and 4. The remarkable findings resulted in binding location maps that help us to better understand the mechanism of action of AHPs and to address the question: what is the primary reason for neutralizing a virus by our polyamides.

Based on the knowledge gathered from the results of the second project, a complex study (RNA-Seq) concerned with genome expression was performed. The aim of the third project is to find the common features and differences in the polyamides' mechanism of action by quantifying messenger RNA (mRNA), after treatment of HPV infected skin cells with 8 different polyamides.

In parallel with the analysis of human and viral transcriptome, a separate study concerned with the differential expression profile of seven DDR genes that are components of the homologous recombination (HR) pathway have been studied by RT-qPCR method. The results are presented in Chapter 5 and 6. Data gathered from the third project helps us further understand the mechanism of polyamide action.

Table of Contents

Chapter 3

Binding Studies of TMG Asymmetric Hairpin Polyamides (TMG-AHP) NV1078 and NV1087 with Designed, 120 bp DNA Sequences

Chapter 4

Interactions Between two TMG Asymmetric Hairpin Polyamides (TMG-AHP), NV1078 and NV1087, with the HPV16,18 LCR segment

Chapter 5

Differential DDR Gene Expression Profile of W12E Cells After Treatment with Active Anti-HPV Hairpin Polyamides

Chapter 6

The Analysis of Viral and Human Transcriptomes After Treatment of W12E Cells with Polyamides

Chapter 1
Introduction

Human papillomavirus. Polyamides as potential HPV antiviral drugs

1.1 Abstract

Long-term, persistent infection with high-risk strains of human papillomavirus (HPV) is the precursor of most of the cervical carcinomas (1, 2) and an increasing number of oropharyngeal squamous cell carcinomas, which represent 90-95% of head and neck cancers (3). While vaccines can protect the approved population from possible infections (4), no HPV antiviral drugs are available for the definitive treatment of preexisting or future viral infections. The approved population has a complex definition given the existence of a bivalent vaccine (for HPV16 and 18), a quadrivalent vaccine (for HPV6, 11, 16, and 18) and a recent nonavalent vaccine), but is generally stated as "boys and girls from ages 11-13" to "men and women through age 26" (http://www.cdc.gov/std/hpv/stdfact-hpv-vaccine-hcp.htm). To treat current or future HPV infections, drugs that selectively block virus-specific processes, but do not damage the host, are needed.

The compounds developed by our group are unique imidazole-pyrrole polyamides (PA), analogs of the natural products Distamycin A and Netropsin that function by inducing parts of the DNA Damage Response (DDR) (5, 6) and perhaps by interfering with natural virus-host interactions. Note that when this thesis project was started, no information about the DDR had yet been understood in a mechanistic sense and all focus was still on more-traditional, sequence-specific DNA recognition rules for biologically-active polyamides.

Within the overall program of my group, my work focuses on three specific projects. The first project (**Chapter 2**) includes a bioavailability LC-MS analysis study of one of the leading compounds PA25 (NV1042) in plasma and whole blood.

The second project includes the biophysical study of interactions between two polyamides, called TMG Asymmetric Hairpin Polyamides (TMG-AHP), and an essential viral DNA segment. The aim of this study was to determine where and how strongly the compounds bind to the viral DNA. In this work, two methods were used: the quantitative deoxyribonuclease I (DNase I) footprinting method and the affinity cleavage assay (AC). The results are presented in **Chapters 3** and **4**. The remarkable findings resulted in binding-location maps that help to better understand the mechanism of action of AHPs

and to address the question: “What is the primary reason for neutralizing a virus by our polyamides?”

Based on the knowledge gathered from published reports on gene expression with subsets of genes (7-9), the results of parallel studies with more-traditional but long, antiviral polyamides from our group (5, 6), and the second project (**Chapters 3** and **4**), the third project was devised, and a complex study (RNA-Seq) concerned with genome-wide gene expression, at the mRNA level, was performed. The aim of the third project was to find the common features and differences in the polyamides’ mechanism of action by quantifying messenger RNA (mRNA), i.e. determining gene expression, after treatment of HPV-infected skin cells with 8 different polyamides. In parallel with analysis of the human and viral transcriptome, a separate study concerned with the differential expression profile of seven DDR genes that are components of the homologous recombination (HR) pathway were studied by quantitative PCR, or the qPCR method. The results are presented in **Chapters 5** and **6**. Data gathered from the third project helped us further understand the mechanism of polyamide action so we may further develop future antiviral compounds to prevent cervical cancer and other HPV-derived diseases.

Our group’s collaboration with NanoVir led to the discovery of compounds useful for eliminating HPV, unique large polyamides which decrease the amount of HPV DNA in infected cells and tissue and have great potential as possible drugs for the treatment of HPV infections (5, 6, 10-12). Biophysical results helped lead to further antiviral discoveries with new viruses, the polyomaviruses (U.S. Patent Appln. 14/818881, JK Bashkin, TG Edwards, C Fisher, GD Harris Jr, KJ Koeller, Antiviral compounds and methods for treating infections caused by double-stranded DNA viruses, 2015/11/19).

1.2 Human papillomavirus (HPV)

Human papillomaviruses (HPVs) are small dsDNA viruses which belong to the large family Papillomaviridae, a group of viruses that infect the stratified epithelia, keratinocytes, of the skin and mucous membranes. More than 140 types of human papillomaviruses have been fully sequenced and they are grouped by sequence homology into five genera. The high-risk HPV16, HPV18, and HPV31 are part of the alpha genus (13). Human papillomavirus possess a circular genome (HPV16 being 7904 base pairs long) that typically contains 8 genes. Genes and proteins of HPV are named according to timelines for viral gene expression as either early (E1, E2, E5, E6, E7) or late (L1, L2, and E4) (13-15). E1, E2, E4, E5, E6, and E7 are nonstructural proteins, while L1 and L2 are

structural proteins. The nucleotide sequence of L1, one of the late genes encoded by virus DNA, constitutes a basis for HPV classification. The different HPV types are divided into those preferentially infecting the skin versus those infecting mucosal surfaces (16).

Among the mucosal HPVs, at least 15 types are considered high-risk (16), with HPV 16 and 18 accountable for >70% of HPV-induced cancers in the US (17). Cervical cancer is the second largest cause of cancer deaths in women worldwide (2). Approximately 11,000 women are diagnosed with cervical cancer in the USA annually and a third die of this disease (17). The high-risk HPVs are associated with cervical cancer and also with many penile, vulvar and anal carcinomas, and contribute to over 40% of oral cancers (1, 18).

For the discovery that high-risk HPVs cause cervical cancer, Harald zur Hausen was awarded Nobel Prize in 2008, cementing the importance of knowledge related to HPV (https://www.nobelprize.org/nobel_prizes/medicine/laureates/2008/hausen_lecture.pdf). Both prior and subsequent to this, many studies have been published concerned with the HPV-induced carcinogenesis, which is still not completely understood (1, 16, 19). However, it is well-established that malignant transformation (16) is related to the persistent overexpression of HPV E6 and E7, the oncogenic proteins (15, 16, 20).

Only a few HPVs are oncogenic, however, from which we infer that the protein activities of high-risk HPVs are different in a few aspects from the corresponding protein activities of HPVs associated with benign diseases, with limited oncogenic potential (20). Although the relative oncogenicity of E6 vary depending upon the HPV type, there is important overlap in the pathways altered by low- and high-risk HPV E7 (20). Carcinomas occur in the transformation zone of the cervix where the columnar and squamous epithelia meet (20). The intracellular milieu of the basal cells in this region may limit the completion of the productive life cycle (20, 21): therefore, if the productive life cycle cannot be completed (20), then no virions can be formed and the transformation of the cells might be favored.

The integration of the viral genome into the genome of the host cell is common; it is not a necessary pre-requisite for the development of cancerous tumors. HPV 16, 18 and 45 are more frequently found in an integrated state than HPV 31 and 33 (22). When integration of the viral DNA into the host's chromosomal DNA occurs, no virions are produced. HPV integration usually occurs near or within known fragile sites (5) and often disrupts E2 expression (15). E2 autoregulates its own expression and regulates expression of E6 and E7. Disruption of E2 results in increased levels of expression of E6

and E7 which then induce the formation of tumors**.** A second path for malignant transformation could occur in cells that maintain viral episomes (1), and the expression of E5 would increase the activity of E6 and E7, leading to subsequent tumor progression (1).

The E6 and E7 proteins provide functions necessary for the differentiation-dependent life cycle of HPVs in addition to their roles in immortalization. These two HPV proteins also play critical roles in modulating immune evasion. Expression of E6 and E7 oncogenes is regulated on the transcriptional level (23) in the viral long control region (LCR) (23), which is well-conserved in different HPV16 isolates (24).

E7 proteins do not possess intrinsic enzymatic or DNA-binding activities but function by binding to several cellular factors (1) such as the Retinoblastoma tumor suppressor (pRb). pRb which normally prevents cell growth by inhibiting cell cycle progression is destabilized by E7. E7 protein propagate immortalization through interaction with proteins that control cell cycle progression (1). E7 proteins also target members of the interferon pathway (1). Interferons trigger the protective defenses of the immune system in response to the presence of viruses. The efficient binding of E7 to Rb family members and other cell cycle regulator proteins also results in the increase in the levels of the tumor suppressor p53, which limits cellular growth and increases the susceptibility of E7-expressing cells to apoptosis (1, 25).

To diminish increased levels of p53, high-risk **E6 proteins** use several mechanisms to interfere with p53 functions (1), such as the formation of an E6–E6-associated protein (E6AP)–p53 (1); formation of the trimeric complex results in p53 degradation (1). Also, the interaction of E6 with the histone acetyltransferases prevents p53 acetylation, inhibiting the transcription of p53-responsive genes (1). High-risk E6 proteins also activate transcription of telomerase reverse transcriptase (TERT), another essential step in immortalization (24, 25). E6 also interacts directly with components of the IFN response (1, 26).

The expression of E6 and E7 is necessary but not sufficient for malignant progression (1). In the amplification phase of the HPV life cycle (see HPV lifecycle below), E6 and E7 promote the proliferation of undifferentiated and differentiated cells, in the same time allowing the avoidance of apoptosis (1). This phenomenon leads to an accumulation of DNA damage and genomic mutations (1).

E5 protein can augment the function of E6 and E7 and contribute to tumor progression (1). E5 proteins are small, hydrophobic polypeptides approximately 83 amino acids in size that localize primarily to the endoplasmic reticulum (1).

1.2.1 The HPV lifecycle

Three major phases in the HPV lifecycle can be detected (1) the initial amplification post infection (2) the non-productive or maintenance phase and (3) the productive phase, or amplification phase **(Figure 1.1)**.

All HPVs infect the basal cells of stratified squamous epithelia via micro abrasions (20) . Basal cells are undifferentiated skin cells and are the only proliferating (dividing) cells in normal epithelia (1). HPV virions infect the basal layer via heparan sulfate proteoglycans and integrin proteins (27). Integrin proteins are upregulated during wound healing and have been reported to play a role in viral attachment and entry. After being taken up, the viral particles travel to the nuclear membrane by the endosomal pathway (28), where the major and minor capsid proteins (L1 and L2) are left behind and the viral genome in the form of a chromatinized double-stranded DNA circular episome is transported into the host cell nucleus (27). After membrane penetration of mitotic cells and nuclear import, the virus locates to nuclear bodies, or nuclear domains, known as ND10 (28). It was proposed that ND10 bodies are associated with regions of DNA damage in uninfected cells (8). In the basal cells, the viral genome is maintained extrachromosomally (20) as episomes that can be replicated independently from the host chromosomal DNA.

In the initial amplification stage, the viral genome replicates but maintains a low copy number as a consequence of the expression of the E8^E2 transcript which is the repressor of viral transcription and replication (29-31). To produce a persistent infection, the virus has to be able to evade immune responses. HPV is effective at evading immunological detection and produces only a slow and weak immune response which varies among patients (32). Not just in the initial amplification stage but during the whole lifecycle of HPV, the exposure to the host immune system is limited. Infections do not produce viremia, the presence of viruses in the blood, and a limited load of antigenic HPV proteins is produced as a result (4).

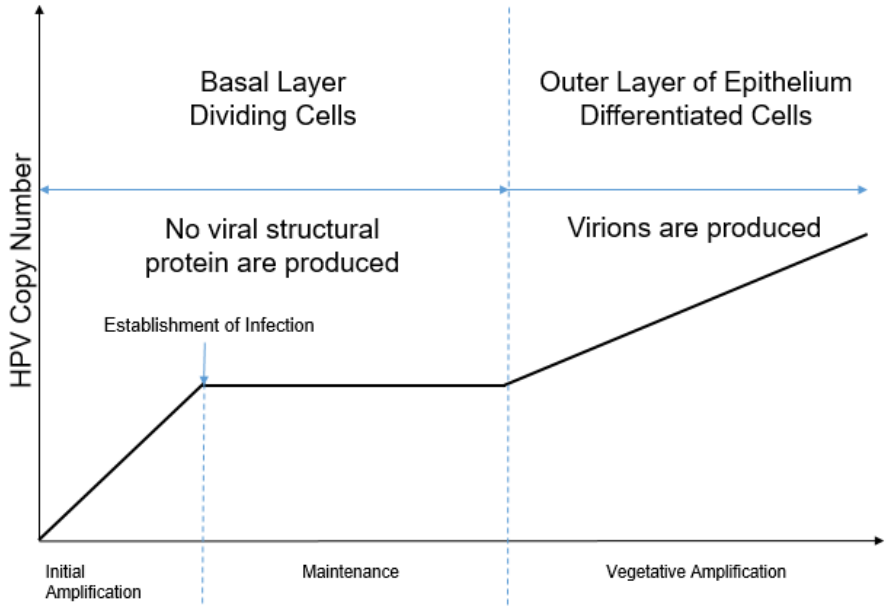


Figure 1.1. Episome copy numbers in different phases of replication. Adapted from (8)

The maintenance phase is the phase in which the episomal DNA is maintained at a few hundred copies of episomes per cell and no viral structural proteins are produced (15, 20). The W12E cervical keratinocyte cells were purified from a clinical isolate and are undifferentiated cells which harbor 100-500 viral episomes per cell (5, 6, 10). These cells, which mimic cells harboring viral episomes in the maintenance phase, were used as the model for the gene expression studies. They were a kind gift of Dr. Paul Lambert of the Medical School at The University of Wisconsin, Madison.

To transition or progress to the maintenance phase, the viral genome has to be tethered to beneficial regions of chromatin and to modulate DDR responses. In the maintenance phase, depending on the HPV type, replication of the episomes occurs during S-phase along with the host genome, or by a random choice mechanism (33). High levels of E1 promote mostly random choice replication in W12 cells (33) and trigger a DDR response imposed by ATM (34) to ensure replication of the virus. In the S-phase, the episome is (or may be, see later discussion) attached to the host chromatin using E2 as an intermediate. E2 binds the episome on its DNA binding domain and also binds the BRD4**.**

The BRD4 tethers the episomes to the host chromatin (at fragile sites as demonstrated in C-33 cells (35)) to ensure replication. BRD4 is important for genome tethering during maintenance and as well as a viral replication center initiator present in the amplification phase (8, 36). This tethering mechanism ensures that the low copy of the

viral genome is retained in the nucleus and is distributed to daughter cells (36). The binding model for HPV18 is shown in **Figure 1.2**.

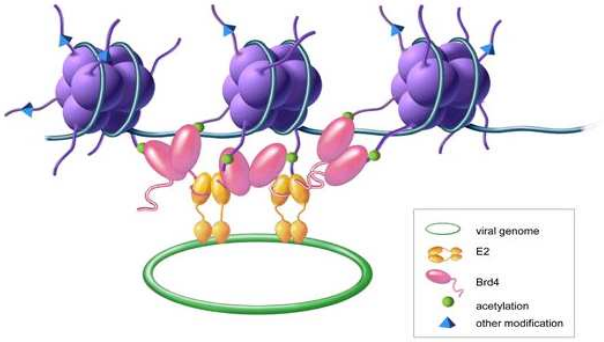


Figure 1.2. Model of E2-mediated Tethering of the Viral Genome to Host Chromatin (36) The DNA-binding domain is of the HPV18 DNA binding domain bound to DNA (shown in light purple) from the PBD file 1JJ4 (36).

Binding of E2 to Brd4 is not sufficient for tethering the viral episome to the mitotic chromosome; E2 also has to bind ChIRI (37, 38). Recently the HPV16-BRD4 complex was analyzed by bimolecular fluorescence complementation experiments (8, 39) and with the help of experiments where E2 is expressed as a GFP-fusion protein (40). Many studies concerned with the role of BRD4 as a key regulator of transcription also found that BDR4 may be associated with DDR elements and also associates with fragile sites (35). BDR4 is an insulator chromatin that can modulate the signal of DNA damage (41).

However, new studies display a model in which the viral episomes (HPV 31) bind to the host chromatin in a way that is independent of BRD4 protein (42). This controversy has not been resolved and is complicated by the fact that the majority of studies on BRD4 were done with chimeric human/bovine papillomavirus systems and/or with E2 highly overexpressed.

Viral replication centers are distinct nuclear foci that contain recruited nuclear replication factors associated with the host DDR; with the help of which viral genomes are replicated. Replication foci can be detected in differentiated cells (8). BRD4 is initially part of the replication process by concentrating viral proteins in specific regions of host chromatin (43). It was shown that DDR factors are present in the replication centers in the case of differentiated keratinocytes (19).

In the amplification stage of the viral life cycle, a rapid increase of viral DNA and virion production also occurs (8, 20). The synthesis of virions, or viral particles, that are able to shed and infect other organisms occurs only after cells have differentiated,

migrating towards the outer layer of the epithelium (15, 20). For virion production, the late genes L1 and L2, which code for capsid proteins, must be expressed (16). The genome amplification in this stage is thought to be a rolling cycle type and may be linked to DDR (9). Recent findings identified SMC1 (cohesin protein) and CTCF (insulator transcription factor) as critical regulators of differentiated high-risk human papillomaviruses (42).

Viral gene expression is controlled through two **viral promoters**. The early promoter (**p97** for HPV 16, **Fig.1.3)** is located in the upstream regulatory region (URR) (8) adjacent to the E6 open reading frame (ORF) and is active in the infection process and during the maintenance phase (44). Promoter p97 directs expression of the E1 and E2 replication proteins (8), which leads to the establishment of viral genomes, as well as the expression of the viral oncoproteins E6 and E7 which regulate cell cycle progression. Transcription stops at the early polyadenylation (poly(A)) site AE and the transcript is processed by using the early splicing signals (45).

As HPV-infected cells differentiate, the **late HPV promoter p670 (Figure 1.3**) located in the middle of the E7 ORF is activated (44), leading to expression of late gene products such as E4, L1, and L2, and increased levels of E1 and E2 (15, 46). In differentiated cells, the transcripts expressed from the late promoter utilize a late poly(A) signal, AL, and late splicing signals (45). The early and late splicing signals compete for the splicing factors, so their usages are exclusive (45).

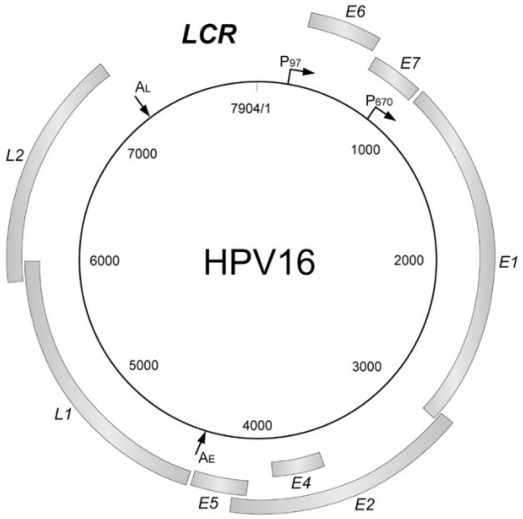


Figure 1.3. The HPV circular dsDNA genome structure. Viral genes are transcribed in clockwise direction. The LCR contains a DNA replication origin and functions as a regulator for DNA replication (45, 46).

1.2.2 The DDR's role in viral episome maintenance (Chapter 5)

In general, infection by DNA viruses induces the activation of many cellular pathways, including those involving p53, stress, apoptosis and the **DNA damage response (DDR)** (7). The DDR can detect foreign DNA and can also sense aberrant DNA structures that form as a result of chromatin changes during nucleic acid transactions such as transcription and replication (47). DDR response is also triggered by changes in chromatin structure (35).

The presence of HPV's circular dsDNA episomes initiates a DDR response. The HPV, a small virus, which contains no polymerase, just two replication proteins (E1 and E2) employs some of the DDR elements for its replication.

In **Figure 1.4** the DDR to single stranded and double-stranded breaks is described (48).

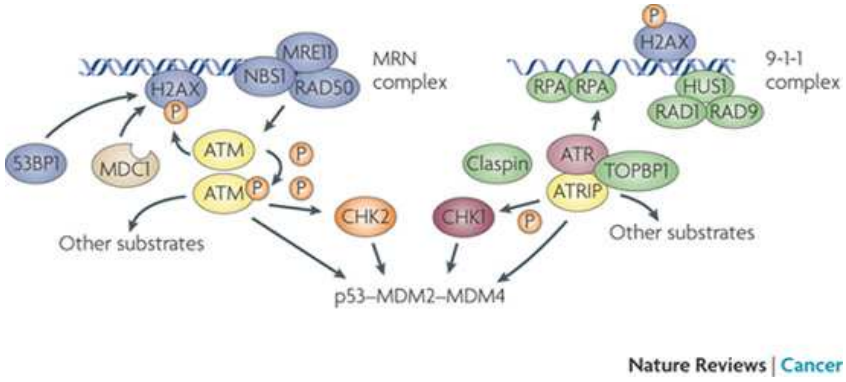


Figure 1.4. DNA damage response to single-stranded and double-stranded breaks (48). ATM: Ataxia-telangiectasia mutated kinase; ATR: Ataxia-telangiectasia and RAD3-related kinase; ATRIP: Ataxia-telangiectasia and RAD3-related kinase-interacting protein; RPA: Replication protein A; TOPBP: Topoisomerase binding protein.

RPA coats single-stranded DNA and is subsequently recognized by an ATR-ATRIP complex (48). The RAD9-RAD1-HUS1 (9-1-1) complex binds to the 5′ overhang near the exposed ssDNA and brings in TOPBP1 which activates ATR in the presence of ATRIP (48). ATR phosphorylates many downstream targets including Chk1, the cell cycle checkpoint kinase, and p53 (48). When Chk1 is phosphorylated, it will arrest the cell cycle (34).

Double-stranded DNA breaks are detected by the Mre11–Rad50–Nbs1 complex, which promotes monomerization and autophosphorylation (activation) of the ATM kinase

(48). Activated ATM will lead to phosphorylation and activation of its downstream effector, Chk2 kinase, and p53, H2AX (γH2AX) and Nbs1, among many more targets (48).

The ATM pathway activated by ds-breaks and the ATR pathway activated by single-strand breaks were thought to function in parallel. However, recent studies reviewed by Hong and Laimins (15) indicate that these two pathways are linked to each other. ATR seems to be subordinated to elements of ATM. The ATR–CHK1 pathway requires activation by ATM. ATR is sensitive to radiation-induced ds-breaks and the MRE11/RAD50/NBS1 (MRN) complex that induces autophosphorylation of ATM is also needed for ATR activation (15).

The present study will involve W12E cells, undifferentiated keratinocytes which maintain the episomal high-risk HPV16 genome. Four viral proteins, E1, E2, E6, and E7, are required for the maintenance of viral DNA in cell cultures (45). The capsid proteins L1 and L2, essential for the infectious particle (virion) formation, are not expressed in sufficiently high levels in the undifferentiated cells that will be used.

1.3 Polyamides as potential HPV antiviral drugs

N-methylpyrrole-N-methylimidazole polyamides (PA) are versatile homologs of the naturally occurring antibiotics Distamycin A and Netropsin, small, crescent shaped polypyrrole peptides that bind to repeating A/T tracts (49) in minor grooves with modest affinity and specificity (49-51). Distamycin has antitumor and antiviral proprieties but also a high toxicity (10, 49, 52).

To overcome the high toxicity but maintain antitumor and antiviral properties, polyamides were designed following the Distamycin model. By introduction of imidazole (Im) building blocks, recognition of G and tracks that include G became possible, leading to reported pairing rules (53, 54) pyrrole (Py) binds to the three nucleotides which present H-bound acceptors in the minor groove A, T, and C. γ-turn, γ-aminobutyric acid binds to A/T or T/A nucleotides. The cationic tail (Ta) aids binding and recognizes A/T or T/A nucleotides; β-alanine (β-spring) binds to A, T or C. Then β-alanine units confer additional flexibility and better follow the DNA curvature (54, 55). The flexibility of the β units can help optimize the interactions of polyamides with the targeted DNA bases, however, the number and position of the substitutions influence the binding affinity of polyamides in the minor groove of DNA, though this level of sophistication was ignored in the original literature but was recently reported (56-58). Polyamides specifically recognize and bind with high affinity to dsDNA in minor groove regions by hydrogen-bonds (49, 59).

The binding of PAs is followed by alterations of the dsDNA shape by widening the minor grove, narrowing the major groove (54, 60), and stiffening the double helix (61).

By binding with high affinity in the minor groove, the small polyamide molecules can fulfill different functions such as the regulation of gene expression, or the molecules may be used as molecular probes. Polyamides are able to displace many proteins from DNA interfering with gene expression at transcription levels (62). Also, PAs can be specifically designed to bind to recognition sites of the transcription factors (49, 63) of dsDNA, silencing aberrant gene expression or infectious diseases in this way. A strategy of suppressing viral DNA transcription was successfully employed for destabilizing the Epstein-Barr virus (64).

Conjugated polyamides can function as site-specific alkylating agents able to covalently modify DNA (49). Alkylated DNA can mimic methylated DNA, interrupting transcription and selectively silencing targeted genes (49). A possible chemotherapeutic choice for cancer treatment is the use of polyamide-alkylator conjugates to silence the oncogenes (49, 65). Bifunctional PAs could also be used to alkylate human telomeres (1000 or more tandem repeats of a short G-rich sequence, TTAGGG, on the 3' end of each chromosome (66)) and abrogate the increased telomerase activity of cancer cells (49, 67). Polyamide-peptide conjugates can function as synthetic transcriptional activators as well (49). Recent work showed the results of coupling a hairpin polyamide to a suberoylanilide hydroxamic acid (SAHA) moiety (68), known to act as a histone deacetylase inhibitor with broad epigenetic activity (49). Histone acetylation will loosen up the nucleosome, allowing the genes to be expressed (68, 69).

Polyamides are also used as molecular probes. When PAs bind to dsDNA match sites, the fluorescent signal produced by the polyamide-fluorescent dye conjugates is enhanced, allowing them to be used in cellular localization studies (70). Although the fluorescent dye used in imaging experiments can affect the cellular uptake of PAs, many studies (59, 71, 72) provide significant results to provide evidence that PA uptake is achievable over a wide range of PA molecular weight (up to 4000 Da) and cell types (73).

PAs accumulate within cells by either passive diffusion, regardless of their molecular size (73, 74), or active diffusion, which is energy-dependent, such as in the case of HeLa cells (74). Effective cellular uptake of PAs was also observed in a range of human primary keratinocytes, including keratinocytes harboring HPV infections (5, 10)**.** Therefore, the process of PA uptake into cells is extremely well studied. Efficient PA uptake was also exhibited in some animal studies (71, 75). Molecular size does not influence the cellular

uptake (polyamides can cross plasma membranes) (73). However, the permeability of the nuclear membrane is sometimes reduced for larger polyamides (73). Many studies show that PA uptake in cells is achievable, but varies with cell line and PA structure (imidazole content, structure of the "tail", hydrophobic qualities, β-alanine (57), and gamma turns (54, 71, 74)). According to cell and animal studies, some polyamides have cytotoxic effects (76-78), while others, such as the case of anti-HPV polyamides that decrease viral DNA levels, exert their effects without toxicity. Sometimes, efficiency is diminished or impaired because polyamides are not able to reach the target destination, the nucleus of the cells. Quantitative measurements have shown that the concentration in the nucleus of polyamide–fluorescein conjugates is lower than the media concentration and differs according to the cell line (75). Accumulation of polyamides in the cytoplasm (79) and sequestration in acidic vesicles was evidenced by confocal laser scanning microscopy and flow cytometry measurements (73). Multidrug resistance (MDR) causes polyamides to be stored in acidic vesicle in some transformed cells; nevertheless, uptake can be improved using inhibitors of MDR (70)

Earlier studies show that much of the DNA in the nucleosome (the elemental repeating unit of all eukaryotic chromatin (80), consisting of histone proteins around which 146 bp of DNA are wrapped (80)) is freely accessible for polyamide binding in the minor groove (54, 81). The sites on nucleosomal DNA facing away from the histone octamer and the sites partially facing the histone octamer are fully accessible for DNA binding (80) (PDB: 1M18, 1M19, 1M1A). Polyamides not bind where sites are completely blocked by interactions with the histone octamer (81). The structure of the histone octamer and its interaction with the DNA is unaffected by PA binding (80). However, the nucleosomal DNA undergoes structural changes at PA binding sites and the adjacent regions, such as the widening of the minor groove and alterations of the DNA twist angles, which are propagated to distant regions (80). The details of the changes upon PA binding are revealed by the crystal structures of the nucleosome (80). The study also suggested that nucleosome-bound DNA is capable of adjusting its structure to allow binding.

In vivo studies suggest that polyamides have an unexplained preference for some nucleosomal DNA binding positions. Research done in Dervan's lab revealed that a polyamide-DNA alkylator, chlorambucil, with high affinity, binds solely to two of the perfect matching sites of the H4C (highly expressed histone protein in cancer cells) gene. No other gene downregulation (a consequence of the polyamide binding) was observed, even if possible binding sites of this polyamide are present many times in the human genome

(77). The binding of the polyamide to the H4C gene in colon cells may be caused by differences in nucleosome placing or nuclear localization of PA (59, 74, 77). Other studies focusing on nuclear chromatin access by a polyamide-chlorambucil conjugate showed that PA binding was restricted to a limited number of genes (82).

In conclusion, even though nucleosomal DNA is partially accessible for high-affinity binding, in vivo studies show that the binding patterns of PA to a certain cell type genome are not completely predictable. Comprehensive binding studies are required to understand the complex, dynamic processes that take place in the cells.

A special class of polyamides and potential drug candidates, long N-methylpyrrole-N-methylimidazole hairpin polyamides, were synthesized in Dr. Bashkin's lab using Boc-chemistry solid phase synthesis procedure (83-85). These polyamides were found to have anti-HPV activity against three different HPV genotypes: HPV16, 18, and 31 (5, 10, 26). Binding in the minor groove of dsDNA (54, 57, 60) causes structural alterations of supercoiled dsDNA, perhaps causing ssDNA exposure or double strand breaks (DSBs) within HPV episomes, but in any case resulting in the activation of DNA damage response (DDR) pathways (6). Activation of DDR genes eliminates (causes enzymatic destruction of) viral episomes without having significant effects upon cell growth or apoptosis (6). These special polyamides have the ability to decrease HPV16 episome levels in undifferentiated W12E cell monolayers and tissue culture in the absence of integration of viral DNA.

In the present study**,** a new class of derivatives of antiviral polyamide NV1042, also called PA25, was used. These compounds have a modified N-terminal capping group, 1,1,3,3 tetramethylguanidine (TMG) or guanidine at the N-terminus. Previous work results (NanoVir, LLC, **Table 1.1 and 1.2**) demonstrated the antiviral activity of the compounds by attesting the decrease of HPV DNA levels in human keratinocytes by more than 90% in cells that maintain HPV16, HPV18, or HPV31 DNA. Moreover, the tetramethylguanidines (TMG) displayed as much as a 5-fold greater efficiency in human keratinocyte cell culture against HPV16, 18 and 31 than the best dIm-terminated members of the library on which they were based.

Table 1.1. Polyamides used in the following studies. The IC50 and IC90 values measured for W12E cells were determined by T. G. Edwards in the Fisher lab, unpublished data**.**

Poyamide (PA)	HPV16 IC50 [μM]	HPV16 IC90 [μM]	HPV18 IC50 [μM]	HPV18 IC90 [μM]
NV1011 (PA11)	NOT ACTIVE	NOT ACTIVE	NOT ACTIVE	NOT ACTIVE
NV1028 (PA1)	0.100	1.113	0.717	>10
NV1042 (PA25)	0.036	0.351	0.056	1.462
NV1078	0.046	0.307	0.030	0.091
NV1087	0.031	0.200	0.024	0.297
NV1111	0.017	0.121	0.018	0.136
NV1113	0.304	>10	0.228	>10
NV1115	0.103	0.378	0.167	0.937

Table 1.2. Sequences of polyamides (PAs). PA1 and PA25 are the original preclinical lead compounds. Abbreviations: Im represents N-methylimidazole, Py is N-methylpyrrole, β is β–alanine, γ is γ-aminobutyric acid, and Ta is $CH_3N(CH_2CH_2CH_2NH_2)_2$ or 3,3'-diamino-N-methyldipropylamine. The polyamides were synthesized by solid phase synthesis as previously described (10, 26).

Polyamide (PA)	PA sequence	N-terminal group
NV1011 (PA11)	ImPPIm-γ-PPPPβDp	Im
NV1028 (PA1)	ImPPβPPP-γ-PPβPPPPβTa	Im
NV1042 (PA25)	ImPPβPPImβPP-γ-PPβPPPβPPPβTa	Im
NV1078	TMG-PβPPImβPP-γ-PPβPPPβPPPβTa	TMG
NV1087	TMG-PPPβPPβPIm-γ-PβPPβPPPβPβTa	TMG
NV1111	Guan-PPPβPPβPIm-γ-PβPPβPPPβPβTa	Guan
NV1113	TMG-ImPPβPPP-γ-PPβPPPPβTa	TMG
NV1115	Guan-ImPPβPPP-γ-PPβPPPPβTa	Guan

Chapter 2

Polyamide Bioavailability

Analytical Studies of Polyamide Content in Plasma and whole Blood

2.1 Introduction

This chapter includes the LC-MS analysis of different blood fractions provided by MPI RESEARCH (2026-006) and whole blood analysis of K_2EDTA-treated rat blood (Fisher Scientific). Prior to this study, animal studies (MPI 2026-003) were conducted by WIL Research Laboratories, LLC, 1407 George Road, Ashland, OH 44805-8946, using an analytical method developed by MPI Research, a subsidiary of WIL Research, and qualified under MPI Research Study Number 2026-001.The project was approved by NIH and animal safety was approved by the Institutional Animal Care and Use Committee (IACUC) established by WIL labs.

Polyamide (NV1042) was administrated to female Sprague Dawley (SD) rats to characterize its pharmacokinetic properties, acute and subacute toxicity, estimate the maximum tolerated dose (MTD), and evaluate the toxicokinetics of the polyamide in the female rats. The MTD is a dose that does not produce mortality, more than a 10% decrement in body weight, or significant clinical signs of toxicity.

Since the absorption following topical application (intra-vaginally), the intended route of clinical administration, was apparently insufficient and special formulations for oral bioavailability were not available, the intravenous route was selected for drug administration in order to determine the NV1042 concentration in blood compartments. This method was thought to be the only alternative to achieve toxicologically-relevant systemic plasma concentrations. (MPI 2026-002, 2026-003).

NV1042 polyamide was administered by a slow (60 $\pm$ 5 s) intravenous injection (IV) at 1, 2, or 5 mg/kg doses. A detailed protocol of the animal studies can be found in MPI 2026-002 research study, cited above. For plasma analysis, blood samples were collected in tubes containing K_3EDTA as an anticoagulant. The blood samples were centrifuged and the plasma was stored at -50 to -90^{0}C until analyzed.

Polyamide plasma concentration levels were quantified by the LC-MS/MS MPI method (details in MPI 2026-0010B-01,2 Analytical method). This method was adapted for our work as described in the paragraph 2.2.6 LC-MS Analytical Method.

Separate analyses for plasma spiked with polyamide as well as plasma of whole blood samples extracted from animals after treatment were performed. The results were different. The polyamide was quantified in plasma samples but not in plasma samples originating from animals after they received treatment. Plasma concentration levels after the intravenous administration of polyamide NV1042 were not detected by the MPI method. (The company claimed that all polyamides were metabolized in the animals.)

Opposing the Contract Research Company's inference, it was hypothesized that D5W (5% dextrose, the vehicle used in animal studies) intravenous (IV) solutions interfered with polyamide analysis and might have caused the polyamide to precipitate during plasma processing. The differences in the bioavailability of the two situations arises from the differences between the polyamide sample preparations. In the case when plasma was spiked with PA, the PA was dissolved in DMSO first and the plasma concentrations were quantifiable. On the other hand, when PA dissolved in D5W was injected into the blood stream and analyzed from the subsequently extracted blood plasma, concentration levels were too low. D5W might have helped absorption, for example into the cells, though the team leaders did not consider this at the time; they were more concerned with inappropriate precipitation by the plasma purification procedure.

The purpose of this work was both exploratory and rather desperate: animals had been sacrificed in the course of carefully designed experiments. Without proof that the compound was present, the FDA would not accept data from the study to establish a maximum tolerated dose. This would mean the inacceptable waste of living animals and the loss of large sums of research dollars and large amounts of test compound prepared at great expense and effort. If the compound could be found in the samples, as Dr. Bashkin believed it must be, then large portions of the study would be salvaged in spite of the loss of pharmacokinetic data.

2.2 Materials and Methods

2.2.1 Polyamide PA25 (NV1042) and Internal Standard NV1057

The polyamide used in the present investigation is one of the leading HPV antiviral compounds PA25 (NV1042) (5, 10, 12, 86). NV1042 is a large hairpin polyamide with the following structure: ImPP-β-PPIm-β-PP-γ-PP-β-PPP-β-PPP-β-Ta, where: Im represents N-methylimidazole, Py is N-methylpyrrole, β is β–alanine, γ is γ-aminobutyric acid, and Ta is $CH_3N(CH_2CH_2CH_2NH_2)_2$ or 3,3'-diamino-N-methyldipropylamine (**Figure 1**). NV1042 was synthesized by solid phase synthesis as previously described (10).

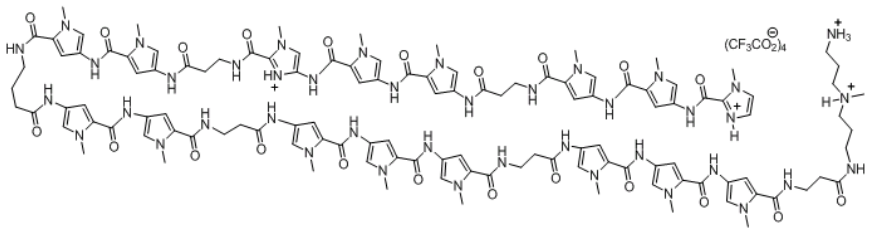


Figure 2.1. Structure of NV1042. The compound, ImPPβPPImβPPγPPβPPPβTa and the 4 counterions (TFA), has the molecular weight 2982.79

Polyamide NV1042, a member of the programmable heterocyclic oligomers N-methylpyrrole–imidazole polyamide class of the hairpin type, has a unique importance, having anti-HPV activity against three different HPV genotypes HPV16, 18, and 31 (Chapter 1, Table 2) (10, 87, 88). The binding of NV1042 by noncovalent bonds in the minor groove of dsDNA was extensively studied (5, 10, 86). As a consequence of the binding, structural alterations of superhelical dsDNA cause ssDNA exposure or double strand breaks (DSBs) within HPV episomes resulting in the activation of DNA damage response (DDR) pathways (6). Activation of DDR genes eliminate and degrade the viral episomes without significantly affecting cell growth or apoptosis (6). This special characteristic made NV1042 an excellent drug candidate to be a subject in the animal studies.

As internal standard polyamide NV1057(KJK5115) was used.

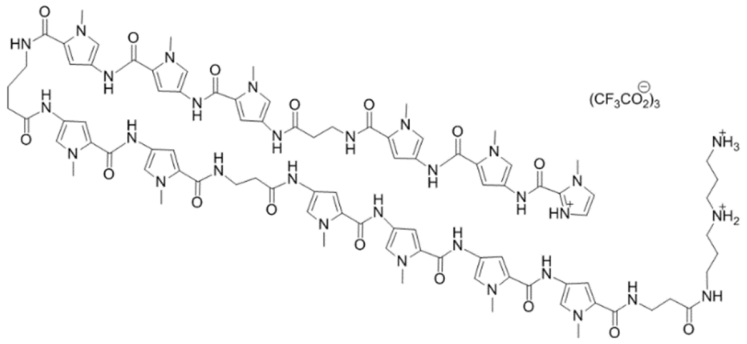


Figure 2.2. Chemical structure of NV1057 internal standard

Table 2.1. Positive electrospray ionization of the NV1057, the most abundant protonated ions, generated at 28 eV collision energy are highlighted (data collected on LC-MS/MS Quadrupole TOF Bruker Maxis, Dr. Bythell's instrument).

Polyamide	Exact mass	[M+H]	[M+2H]/2	[M+3H]/3	[2M+H]	[2M+3H]/3	[3M+4H]/4
NV1057	1884.06	1885.06	943.03	629.02	3769.1	1257.04	1414.045

2.2.2 Plasma and Blood Pellet Provided by MPI Research

Four refrigerated pellets and a total of 8 frozen plasma samples (Analytical method, MPI RESEARCH, LC-MS/MS analysis for NV1042 in SD Rat K_3-EDTA **2026-006**) were received on May 2012 from Christine Wisniewski, MPI Research, Mattawan, MI. The received samples were processed according to the "Analytical method, MPI RESEARCH, LC-MS/MS analysis for NV1042 in SD Rat K_3 -EDTA 2026-006" protocol. These samples were collected on 4/27/2012. The summary of the protocol for processing the samples is provided in the following four paragraphs.

Blood samples (at least 4 mL each) were collected from 3 rats into tubes containing K_3-EDTA anticoagulant. Whole blood samples were transferred (and pooled) into a large conical tube, inverted gently and equally divided into 2 approximately equal aliquots.

From one aliquot (whole blood), four aliquots (1 mL each) were placed into plastic tubes and identified as A, B, C, and D. The whole blood samples were spiked with the polyamide solution to achieve the required concentrations and then spun down to harvest plasma and the pellets. The collected samples were:

4 plasma samples aliquot A (0 mg/mL PA, received no polyamide treatment), aliquot B (1000 ng/mL PA), aliquot C (5000 ng/mL PA), and aliquot D (25000 ng/mL PA) and

4 pellet samples: aliquot A (0 mg/mL), aliquot B (1000 ng/mL), aliquot C (5000 ng/mL), and aliquot D (25000 ng/mL).

The second aliquot was centrifuged at 3000 RPM for 10 minutes at room temperature. The resulting plasma was extracted and separated in four aliquots (1 mL each) into plastic tubes. Plasma blood (not whole blood) was spiked to prepare 4 other plasma aliquots. The aliquots were identified as plasma aliquots AA (0 mg/mL), BB (1000 ng/mL), CC (5000 ng/mL), and DD (25000 ng/mL).

All PA stock solutions were prepared in D5W (5% Dextrose, USP) and added directly to either the blood or plasma samples. The dose formulations were vortexed for 1-2 minutes, then sonicated until complete dissolution.

2.2.3. Whole Blood

Rat whole blood in K2EDTA (50-412-666) was purchased from Fisher Scientific. The blood was aseptically drawn from normal healthy animals; but it was not intended for use in transfusions or other medical applications. The rat blood was stored in the recommended conditions, 2-8 °C, and remained stable for the purposes of these experiments for three weeks.

2.2 4. Plasma Sample Preparation

Plasma samples were prepared for LC-MS analysis by mixing 100 µL of plasma with 25 µL of working internal standard (WIS) solution. The WIS was 10 µg/mL NV1057 and was prepared by diluting a 100 µg/mL NV1057/DMSO solution with WIS diluent solution. WIS diluent being: DMF/HPLC grade Water/FA (50:50;0.1, v/v/v). Formic acid was used to prevent the compound from aggregating in the solvent and from sticking to the container.

After the internal standard was added and gently mixed, 75 µL of precipitation solution was also added. The protein precipitation solution was either acetonitrile with 0.05% acetic acid or methanol (89). The mixture was vortexed, and 200 µL precipitation solution was added again. The mixture was vortexed again for 10 minutes and centrifuged at 4000 rpm. The polyamide was extracted with DMSO from both pellet and supernatant obtained after centrifugation.

2.2.5 Pellet Sample Preparation. Extraction of Genomic DNA from Spiked Whole Blood and Pellet.

The extraction of genomic DNA from blood cells was performed using the DNAzol BD (Life Technologies) procedure at room temperature (90). The lysis of the cells was performed with DNAzol lysing solution, a guanidine-detergent that hydrolyzes RNA and allows a precipitation of DNA from a cell lysate. (1 mL DNAzol + 0.1 mL liquid sample). The DNA precipitation with 100% ethanol or isopropanol (lysate + 0.5 mL ethanol/1 mL of DNAzol) was followed by the precipitated DNA wash (1 mL of 75% ethanol, 2 times). Subsequently, DNA was solubilized with 8 mM NaOH. DNA isolated with a DNAzol regent does not resuspend well in water or Tris buffer.

To separate the polyamides from DNA, DNA digestion was performed with a DNA-Free™–Applied Biosystems (AM1906) kit which contains deoxyribonuclease I (DNase I). DNase I digest single- and double-stranded DNA to oligodeoxyribonucleotides containing a 5'-phosphate. DNase I is suitable for removing DNA from samples containing proteins.

Following the digestion of DNA, protein precipitation was performed by two methods. The first method used was precipitation with acetonitrile + 0.5% acetic acid, the method proposed by MPI. The second method was based on work done by the Dervan group and employed methanol (91). The polyamide was extracted from every fraction, supernatant or pellet, obtained after centrifugations. For polyamide extraction DMSO was used. If the samples contained cell residues that cannot be removed by centrifugation, the

samples were filtered and the filter was washed with DMSO until all polyamide was recovered.

2.2.6 LC-MS Analytical Method

The sample concentrations of polyamide NV1042 were analyzed by LC-MS using an Agilent 1100 system (Agilent, Santa Clara, CA, USA) connected with an Agilent 1956B MSD, or single quadrupole mass spectrometer. HPLC separation was accomplished using a Phenomenex Jupiter Proteo (C12), 4.6 x 50 mm, 4 µm particle size, 90 Å pore size, part number OOB-4396-EO column. The column temperature during the experimental analysis was maintained at 40 °C. The mobile phase A consisted of 100% HPLC grade water with 0.1% TFA. The mobile phase B consisted of 100% HPLC grade acetonitrile. The following gradient program was used: 5% B (0-3 min, 2.0 mL/min), 60% B (4 min, 2.0 mL/min), 60% B (4.25 min, 2.0 mL/min), 5% B (5.5 min, 2.0 mL/min), and 5% B (6.5 min, 2.0 mL/min). The total run time for a sample was 6.5 min. The auto-injector temperature was maintained at RT. The needle wash solution was MeOH. The injection volume used varied between 1 and 100 µL according to the sample concentration. The detection wavelength was set to 300 nm (**Figure 2.3 and 2.5**). Data were processed with Agilent ChemStation software, which was also used to control the HPLC/MS. Positive-mode electrospray ionization of the compound NV1042 produced abundant, protonated, molecular ions $(M+3H^+)^{3+}/3$ =842; $(M+2H^+)^{2+}/2$ = 1264; $(M+H^+)^{+}/1$ =2526.8 $(2M+3H^+)^{3+}/3$ = 1684.5 (**Figure 2.3**). These ions were generated by the parent species at HPLC retention time 2.9 min. Note that in order to see good UV/Vis spectra by diode array detection, high concentrations of compound were injected relative to what was necessary for mass spec detection alone. These high concentrations were the likely cause of the gas-phase aggregates such as $[2M+3H]^{3+}$ that we commonly observed.

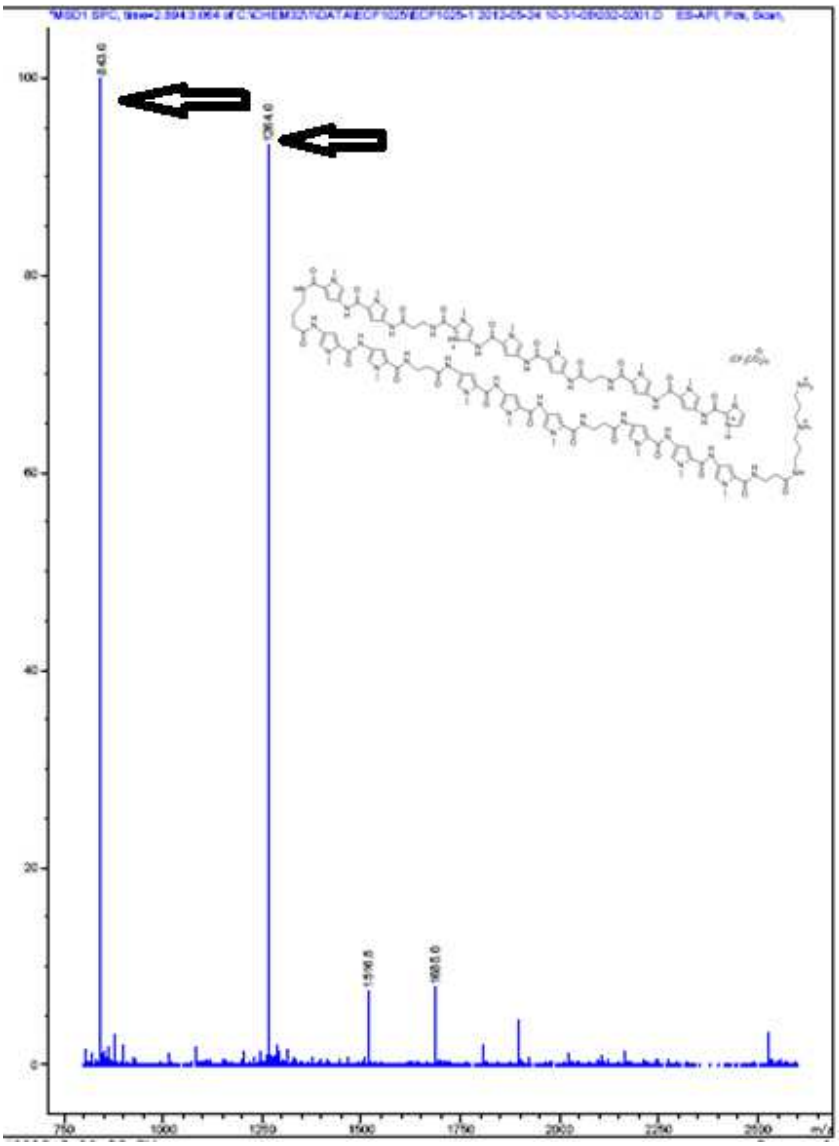


Figure 2.3. Representative mass spectra of NV1042 precursor ions. The positive electrospray ionization of the NV1042 compound produced abundant protonated molecular ions $(M+3H^+)^{3+}/3$ =842; $(M+2H^+)^{2+}/2$ = 1264. 600<m/z<2300 vs ion current

2.3 Results and Discussion

Whole blood fractionation was performed for separation of the blood plasma from the blood pellet. Blood plasma, which represents 55% of total blood volume, is the noncellular liquid part that contains dissolved proteins, glucose, clotting factors, electrolytes, hormones and carbon dioxide (92).

The blood pellet is composed of two fractions. The buffy coat consists of leukocytes (white blood cells) mixed with platelets, and erythrocytes or the red blood cells.

It was proposed that the disappearance of the polyamide molecules in the MPI samples was caused by the molecules being "trapped" in the cells. To recover the polyamide, the pellet originating from the sample treated with the highest polyamide concentration was studied first. Every fraction obtained after centrifugation steps was run on HPLC.

A standard curve used for quantification was prepared by mixing 1 mL DNAzol + 0.2 mL NaOH (8mM) + 6.4 µL 0.1 M HEPES with various concentrations of polyamide (**Table 2.1, Figure 2.4**). HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) is a zwitterionic organic buffering agent and was used to adjust the pH to 8.4.

Table 2.1. HPLC peak areas obtained following the pellet sample preparation procedure. (0.1 mL PA + 1 mL DNAzol + 0.2 mL NaOH (8mM) +6.4 µL HEPES 1M)

Sample	Concentration (ng/mL)	Peak area (mAU*s)	Injection volume(µL)
D1	25000	1020,74988	50
D2	5000	198,15282	50
D3	1000	30,21107	50
D4	50	37	100

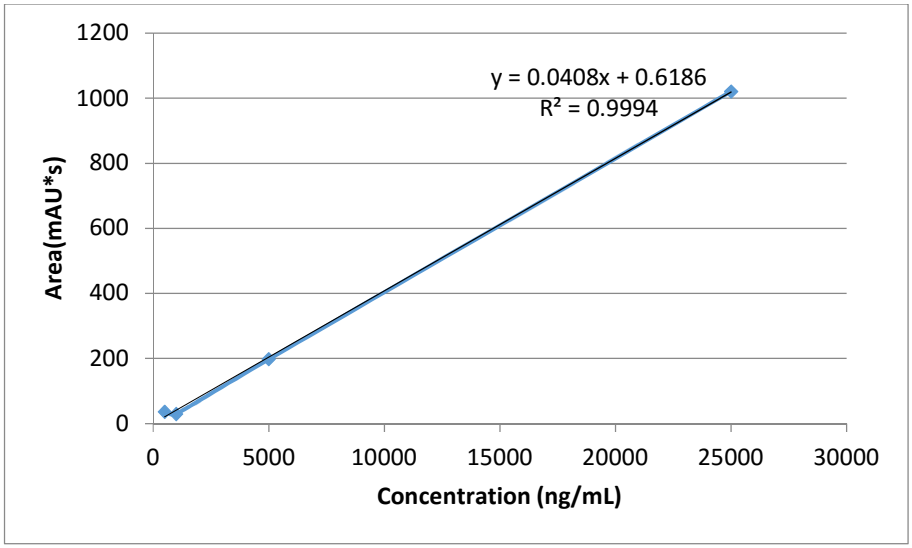


Figure 2.4. The standard curve, as determined by linear regression, displayed good linearity ($R^2 > 0.99$) over the range tested from 50 ng/mL to 25000 ng/mL NV1042.

The quantity of PA recovered from MPI pellet samples (D and DD spiked with 2.5 µg/mL polyamide) was low but detectable. The highest recovery was 33%. However, when fresh whole blood was spiked with 2.5 µg/mL NV1042 a higher quantity was recovered from the blood pellet (cells). When plasma was spiked with the polyamide, the recovery was almost complete. More than 50% of the polyamide precipitated with the protein fraction, and the rest was found in the supernatant of the spiked plasma. The highest protein recovery results were obtained when acetonitrile + 0.5% acetic acid was used for protein precipitation. It was hypothesized that D5W aided absorption of polyamide into the cells. (Note: this hypothesis about D5W aiding blood-cell uptake was later confirmed true for polyamide NV1028 by M.S. student Faten Tamimi working with Dr. G. Davis Harris, Jr. and Dr. Puhong Liao, and by parallel work at another contract lab, Ricerca Biosciences (Concord, Ohio), where Ricerca worked with radiolabeled NV1028 whose synthesis was pioneered in our group by Dr. Puhong Liao.)

Alternatively, it is possible that polyamide NV1042 is taken up and retained by the lung tissue. This phenomenon has been previously described for drugs given intravenously (91, 93, 94). The lung serves as a reservoir for exogenous compounds and as a pathway for elimination. (This alternative was not supported by later radiolabeling studies at Ricerca, at least with NV1028.)

2.4 Supplementary Material

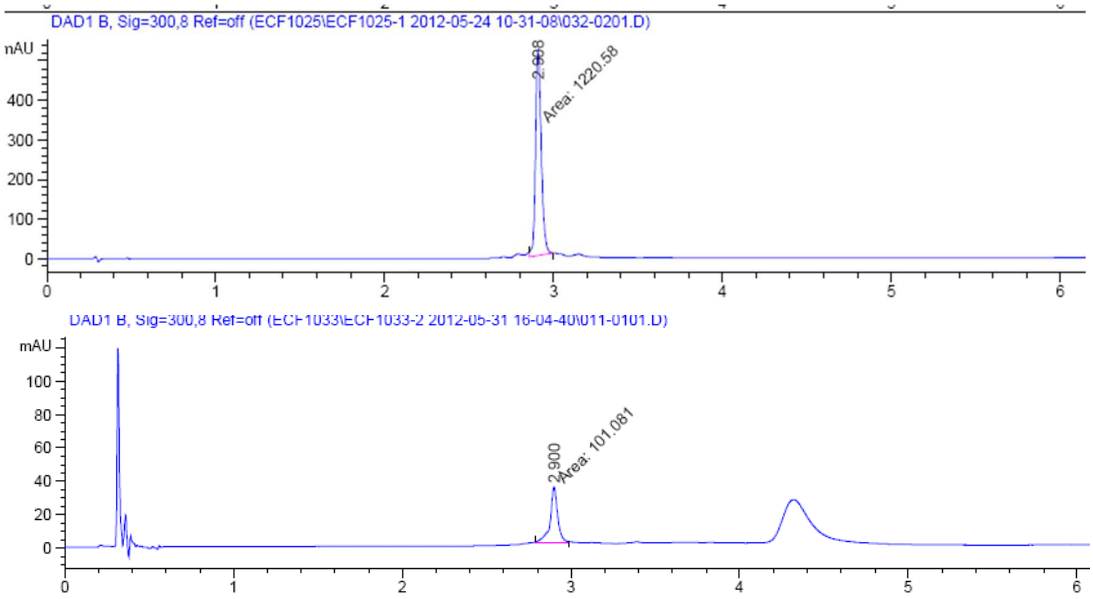


Figure 2.5. Representative chromatograms of NV1042. NV1042 in D5W (ECF1025), NV1042 in plasma fraction (ECF1033). LC-MS Method: 5-60% ACN/0.1% aq. formic acid over 6 min, 2.0 mL/min, 800-2600 MW. Peaks were detected and integrated at 300 nm.

Chapter 3

Binding Studies of TMG Asymmetric Hairpin Polyamides (TMG-AHP) NV1078 and NV1087 with Designed, 120 bp DNA Sequences

3.1 Introduction

The goal of this work was to determine the binding events that take place in vitro between a designed 120 bp DNA and the two large antiviral TMG polyamides, NV1078, and NV1087. TMG polyamides contain a modified N-terminus (1, 1, 3, 3 tetramethylguanidine) (**Table 1.1 and 1.2,** also **Figures 3.4 and 3.5).** The 120 bp DNA fragment was designed to eliminate the possibility of superposed or otherwise confused binding sites.

Superposition of multiple binding sites was often observed when large polyamides bind to dsDNA and made differentiation among binding sites and determination of all associated dissociation constants (Kd) impossible in prior reported work (86). It is impossible to tell which binding site corresponds to the calculated Kd since large polyamides can bind with high affinity even though the binding sequence is not ideal or "correct" according to previously-established binding rules (60).

Binding events were determined using DNase I footprinting (11, 86) and affinity cleavage techniques (95). The binding sites of the polyamides have been analyzed by a capillary electrophoresis (CE) procedure (86, 96). The dissociation constants (Kd's) derived from the isotherms of the binding sites were calculated according to Langmuir and Hill equations (86, 96, 97) based on assumption that the binding is infinitely cooperative for all ligands.

3.2 Materials and Methods

3.2.1. Design of the 120 bp DNA Fragments

The 120 bp DNA sequence was designed to include one "perfect matching site" for PA NV1078 (green) and one perfect matching site for PA NV1087 (red) (**Table 3.1**). Binding events have been analyzed for four 120 bp DNA sequences, following systematic changes in the DNA fragment at TMG binding sites, or locations, to look for sequence preferences (A changed to T, G, and C). The primers (forward and reverse), used in polymerase chain reaction (PCR) mutagenesis by overlap extension are presented in **Table 3.1**. The primers extended the 90 bp DNA fragment to 120 bp mutated DNA. The primers' suitability was analyzed with the Mfold Web Server which predicts secondary

structure of DNA (or RNA) and can be used for examining the folding possibility or the hybridization of the primers.

Table 3.2. Primer pairs that extend the 90 bp DNA to 120 bp DNA. The nucleotide segments that overlap with the fluorescently labeled primers are represented in bold. The perfect matching site for PA NV1087 is highlighted in red, while green nucleotides show the perfect matching site for PA NV1078.

#	Primer 40 bases	Type	TMG binding site
1	5' TAA ATC ACT ATG CGC **CAA CGC CGA TAT ATA TTA AGT GCA T-3'**	Forward	A
2	5' CCT GTG GGT CCT GAA **ACA TTG CGT AAT CAA TTT TAT CGC C-3'**	Reverse	A
3	5' TAA ATC ACT ATG CGC **CAA CGC CGA TGT ATA TTA AGT GCA T-3'**	Forward	G
4	5' CCT GTG GGT CCT GAA **ACA TTG CGT AAT CAA TTC TAT CGC C-3'**	Reverse	G
5	5' TAA ATC ACT ATG CGC **CAA CGC CGA TCT ATA TTA AGT GCA T-3'**	Forward	C
6	5' CCT GTG GGT CCT GAA **ACA TTG CGT AAT CAA TTG TAT CGC C-3'**	Reverse	C
7	5' TAA ATC ACT ATG CGC **CAA CGC CGA TTT ATA TTA AGT GCA T-3'**	Forward	T
8	5' CCT GTG GGT CCT GAA **ACA TTG CGT AAT CAA TTA TAT CGC C-3'**	Reverse	T

A second PCR reaction used the same nucleotide sequence primers labeled with FAM or HEX dyes attached at the 5' ends, to label the 120 bp DNA with CE-compatible fluorescent dye (**Table 3.2**). 6-FAM specifies the fluorescein dye and HEX designates the hexa-chlorofluorescein dye (IDT Coralville, Iowa)

Table 3.2. Fluorescently labeled primers

Fluorescently labeled primers	Denotation
5'/ 5, 6 FAM / TAA ATC ACT ATG CGC **CAA CGC** 3'	SR 1130-F (E V0160F)
5'/HEX / CCT GTG GGT CCT GAA **ACA TTG** 3'	EV 0132-R

Double-stranded oligonucleotides were formed by annealing equimolar amounts of complementary single strands by heating at 90 °C for 5 min and cooling the samples to ambient temperature in 5 mM potassium phosphate and 50 mM NaCl (pH 7.5).

The PCR settings for the 90 bp DNA extension were: 95 °C for 3 min followed by 30 cycles of 95°C for 30 sec. allowing denaturation, 56 °C for 1 min annealing, and 68 °C for 1 min extension temperature, in the presence of Taq Polymerase (NEB). The DNA fragments were separated using a 1.2 % agarose gel, visualized with ethidium bromide, and purified with Qiagen Gel Extraction Kit (Qiagen, Valencia, CA 91355).

The nucleotide order of the 120 bp DNA fragment was verified by Sanger sequencing (USB Affymetrix Thermosequenase) and Maxam-Gilbert sequencing (96, 98).

For the interpretation of capillary electrophoresis (CE) results and to correctly assess the footprinting and affinity cleavage events at the nucleotide level, indexing (size calling) was performed for every 120 bp DNA fragment. Therefore, the hydroxyl radical cleavage products (99) (which include 3'-phosphate and 3'-phosphoglycolate, with 3'-

phosphate being the primary cleavage product) produced in the affinity cleavage reactions were correlated with Maxam-Gilbert sequencing products (96). The fragments generated by DNase I footprinting method (3'-dideoxy ribonucleotide) were associated with the end-termination Sanger products (96).

All primers and DNA oligomers were purchased from Integrated DNA Technologies, Inc.

3.2.2 Sanger and Maxam-Gilbert Sequencing

The DNA sequences used in the study have either a FAM or a HEX dye on the 5' end, therefore the mobility of the fragments in the capillary electrophoresis decrease comparing with the mobility of the same DNA fragment missing the dye. Indexing (size calling) must be performed for every DNA fragment, to assess the footprinting or cleavage events correctly at the nucleotide level (86).

According to the chemistry of the fragments generated in the sequencing cleavage reactions, two different indexing methods were used.

Sanger (USB) sequencing, the end-termination PCR method, correlates the CE nucleotide position to the real position of the nucleotides on DNA fragment generated in the DNase I cleavage reactions. (96).

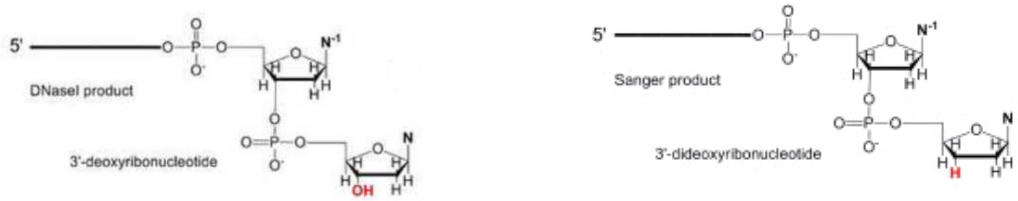


Figure 3.1 Fragments generated in the DNase I reactions and Sanger sequencing (96).

Maxam-Gilbert Sequencing: in order to correlate the hydroxyl radical cleavage products (3'-phosphate or 3'-phosphoglycolate) generated in affinity cleavage reactions with the fragment position in the CE, a second indexing that creates similar cleavage products is necessary (96). The hydroxyl radical generated by the Fenton reaction oxidizes DNA by abstracting hydrogen from the deoxyribose ring (99), causing strand cleavage. In the case of 5'-labeled DNA, the hydroxyl radical cleavage products will be 3'-phosphate or 3'-phosphoglycolate. Fragments with 3'-phosphate are the main species generated in this reaction **(Figure 3.2)**.

In the Maxam-Gilbert method (98), DNA is treated in four different reactions which cleave at G, A+G, C, and C+T., depending on reagents and conditions: formic acid depurinates the purines (A+G), while dimethylsulfate methylates G (and some A).

Pyrimidines (C+T) are hydrolyzed with hydrazine, but addition of NaCl limits this hydrolysis only to C.

Piperidine is then added to cleave the phosphodiester bond where the base has been removed and eliminate both phosphates from the sugar.

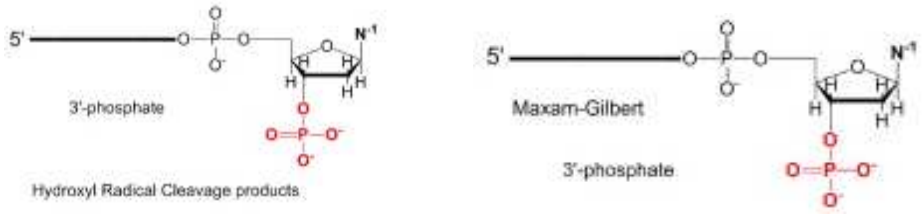


Figure 3.2. Fragments generated in affinity cleavage reactions and Maxam-Gilbert sequencing (96) .

The fragments generated via end termination PCR technique have a 3'-dideoxy ribonucleotide, and moves slower by CE than the corresponding fragment produced by OH radicals, which has the 3'-phosphate end (96).

3.2.3 Quantitative Deoxyribonuclease I (DNase I) Footprinting

DNase I footprinting method by capillary electrophoresis (CE) was employed to determine the binding affinities of TMG polyamides. The four fluorescently-labeled 120 bp DNA segments were used for performing DNase I footprinting titration experiments. Double-stranded DNA (500 pM, 200 pM and 50 pM) was incubated in presence TKMC buffer (with the following composition: 10 mM Tris, 10 mM KCl, 5 mM $MgCl_2$ and 5 mM $CaCl_2$) and 10% final concentration CHAPS (10mM) at 37 °C for 5 hours with polyamides in a concentration range from 0.2 nM to 60 nM according with protocols previously published (57, 86). The samples were digested with the adjusted amount of RQ1 RNase-Free DNase I (Promega, Madison, WI) for 5 min at 37 °C and the reactions were quenched with 200 mM EDTA. The DNase I digested products, obtained following single hit kinetics procedure, were purified with Qiagen PCR purification kit (Qiagen, Valencia, CA) and eluted from column with elution buffer (EB, Tris-Cl, pH 8.5) (57).

The samples were analyzed using an ABI 3730Xl DNA Analyzer using Genescan 600 LIZ as internal lane standard at DNA Core Facility at the University of Missouri. Data were processed using GeneMarker V2.4.0 Software Softgenetics LLC, State Collage, PA. For the interpretation of the electropherograms, the DNase I cleavage products were correlated with Sanger sequencing results of the same fragments.

For the quantitative analysis of the electropherograms KaleidaGraph 4.1 software as fitted to the Hill and Langmuir isotherms was employed.

3.2.4 Affinity Cleavage Assay (AC). Orientation of the TMG-AHP

The four fluorescently-labeled 120 bp DNA segments were used for performing affinity cleavage experiments. In the case of both polyamides, NV1078 and NV1087, chelating agent, EDTA, was covalently linked to the C-terminus (primary amine of the Ta tail) of the polyamides. The PA-EDTA conjugate was incubated with 0.8 equivalence of $Fe(NH_4)_2(SO_4)_2$ to form PA-EDTA-Fe(II) derivatives A volume of 100 µL of 1nM DNA was incubated with PA-EDTA-Fe(II) conjugate at concentrations ranging from 5 nM to 100 nM. The results were registered for 10, 25 or 100 nM. The reaction mixture contained 10 mM Tris, 10mM CHAPS at pH 7.5. The cleavage reaction of the OH radicals formed in a Fenton reaction (see results) was initiated, activated by adding 100 mM DTT reducing agent (86, 98, 100) for a 4-hour incubation time. The reaction was stopped, and the products were purified using Qiagen PCR purification kit.

The samples were analyzed using an ABI 3730Xl DNA Analyzer using Genescan 600 LIZ as internal lane standard at DNA Core Facility at the University of Missouri. Data were processed with GeneMarker V2.4.0 Software Softgenetics LLC, State Collage, PA. For the interpretation of the electropherograms, the hydroxyl radical cleavage products were correlated with Maxam-Gilbert sequencing results of the same fragments (see sequencing).

3.3 Results and Discussion. Binding preferences of NV1078 and NV1087

The sequencing results are included in the tables “Size calling 120 XX_1 top strand” and “Size calling 120 XX_1 bottom strand” where X_1=A, T, C, G, for every strand. The mutation sites that have to be followed were confirmed as follow:

For 120 bp XA (ECF2141): On top strand: peak 81.3, corresponding to real position 88, shows A; On bottom strand: peak 24.4 corresponding to the actual position 88 shows T, peak 88.4 corresponding to position 26 shows T.

For 120 bp XT (ECF3043): On top strand: peak 81, corresponding to real position 88, shows T; On bottom strand: peak 88.4 corresponding to position 26 shows A.

For 120 bp XC (ECF2142): On top strand: peak 81, corresponding to real position 88, shows C; On bottom strand: peak 88.7 corresponding to position 26 shows G.

For 120 bp XG (ECF2159): On top strand: peak 81.2, corresponding to real position 88, shows G; On bottom strand: peak 88.1 corresponding to position 26 shows C.

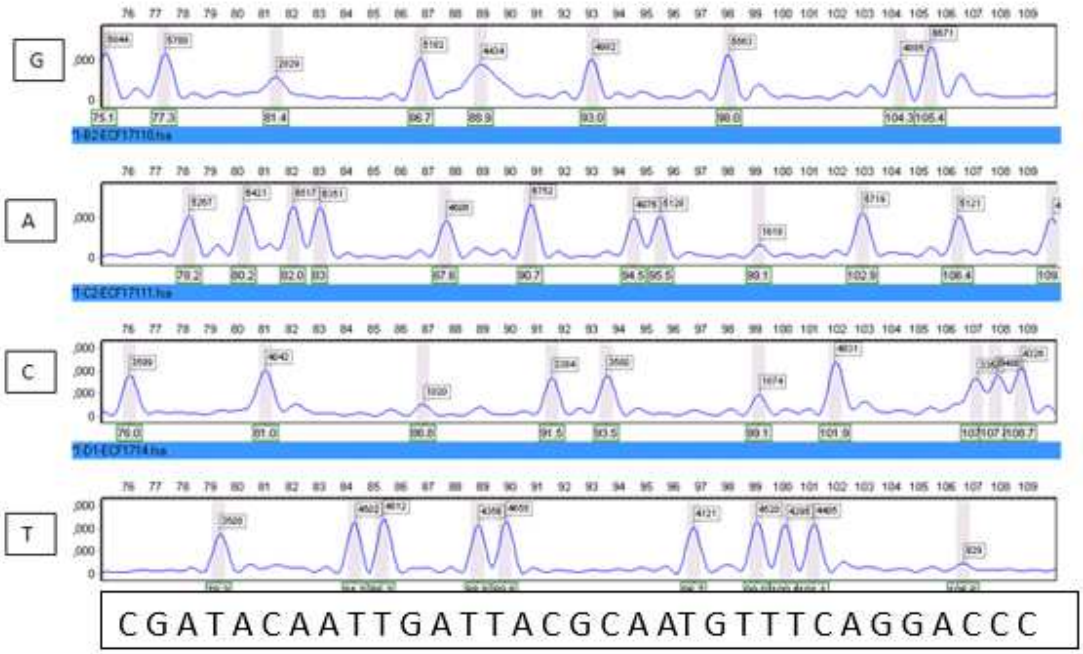


Figure 3.3. Example of Sanger (USB) sequencing electropherogram for 120XC in FAM channel

Quantitative (DNase I) footprinting assay along with affinity cleavage by capillary electrophoresis (CE) was used to assign the TMG binding preferences of NV1078 and NV1087 polyamides. Binding events have been analyzed for designed 120 bp DNA sequence (see methods) following systematic changes in the DNA fragment at TMG binding sites or locations (A changed to T, G, and C) to map and quantify dsDNA-polyamide interactions. Fluorescently-labeled DNA (5'-FAM at the top strand and 5'-HEX for the bottom strand) incubated with varying PA concentrations was used to detect the binding locations (DNase I) and the binding orientation (affinity cleavage) of the polyamides.

DNase I is an enzyme that cleaves the phosphodiester backbone of the DNA double helix (101) relatively nonspecifically (sometimes AT rich sequences being disfavored) (101-103). If no polyamide is bound to the DNA, the digestion will encompass the whole fragment. However, if the polyamide is bound to the DNA, that particular region will not be nicked as much as the polyamide-free control (86). Therefore, the DNase I assay detects DNA-polyamide interactions based on the fact that (reversibly) bound polyamide protects DNA from enzymatic cleavage. The assay uses DNase I enzyme that cleaves the phosphodiester backbone of the DNA double helix in the presence of divalent

cations, introducing single-stranded incisions through hydrolysis of the P–O-C3′ bond (101, 102, 104).

The DNase cleavage pattern of the DNA in the absence of the polyamide is compared to the cleavage pattern of the same DNA sequence in the presence of increasing concentrations of polyamide (from 0.2 nM to 60 nM).

The first round of DNase I footprint experiments explored the binding of PA's to 500 pM of 120 bp DNA followed by experiments using 200 pM substrate (dsDNA) concentration. The electropherograms showed saturation below 5 nM PA. Consequently, the PA's concentration range was lowered from 0, 1, 2, 5, 10, 20, 40, 60 nM to 0, 0.2, 0.3, 0.5, 1, 2, 5, 10 nM. To maintain the proper equilibrium conditions (ligand excess comparing with the substrate concentration; the ligand should be at least five times in excess (97) the DNA concentration was dropped to 50 pM.

The binding measurements were performed after sufficient time (5 hours) to allow the system to reach equilibrium. Therefore, the equilibrium dissociation constant (Kd), the inverse measure of the affinity of the polyamide to dsDNA, and the Hill coefficients were determined in a closed system, under equilibrium conditions. The peak areas (DNase I experiments) or heights (AC experiments) of the resulting electropherograms were taken into consideration. The peak areas in the protected regions were normalized to a neighboring control peak not susceptible to polyamide concentration and plotted as fraction bound vs. TMG-AHP concentration (11, 105).

The electropherogram results were fitted to the Hill (n=1) and Langmuir isotherms based on the assumption that the binding is infinitely cooperative for all ligands. From the fractional saturation θ (Eq. 1), or fraction bound, where L is the free polyamide concentration (97), the apparent dissociation constants, Kds, for the binding sites, and the Hill coefficients (n) were determined, and the results are presented in **Tables 3.2** and **3.3**

$$\theta = ([L]^n/Kd^n) / (1 + ([L]^n/Kd^n)) \qquad \text{(Eq.1)}$$

$$\theta = (Ka\,[PA])^n / 1 + (Ka\,[PA])^n \qquad \text{(Eq.2)}$$

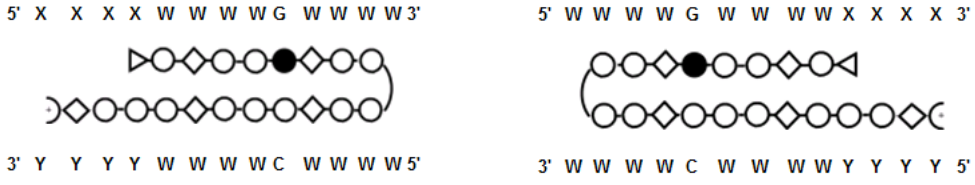
KaleidaGraph 4.1 software outputs the results of the processed data in the form of Eq 2.

The chelating agent, EDTA, was covalently linked to the C-terminus (primary amine of the Ta tail) of the polyamides and the complex was ligated to iron (II). The complex cleaves DNA strand near the polyamide C-terminus binding site following binding of the polyamide in the minor groove. During the reaction, molecular oxygen is reduced to hydrogen peroxide *in situ* (106), followed by Fenton-type chemistry responsible for the production of hydroxyl radicals (107). The hydroxyl radicals generated via Fenton reaction

cleaves the DNA strand near the polyamide C-terminus binding site (up to 2, 3 nucleotides near the Ta-EDTA end) by abstracting hydrogen (H•) from deoxyribose phosphate in the DNA backbone (99) **(Figure 3.3).**

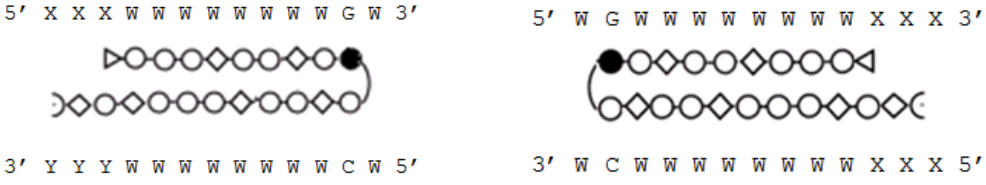
Fenton reaction: $[Fe\ (EDTA)]^{2-} + H_2O_2 \rightarrow [Fe\ (EDTA)]^{-} + OH^{-} + \bullet OH$

The affinity cleavage results give information about the orientation of the polyamides. Hairpin PAs align to adjacent DNA strand in two different ways. N-terminus (TMG) to C terminus (3, 3-diamino-N-methyldiproprylamine: Ta) parallel to the 5'-3' direction of the adjacent DNA strand, as shown in **Figures 3.4A** and **3.5A**, called forward binding, or N to C terminus parallel to the 3'-5' direction, as shown in **Figures 3.4B** and **3.5B**, called reverse binding.



A **B**

Figure 3.4. The structure of hairpin TMG NV1078 polyamide and the code of DNA recognition upon binding to DNA (W: A or T; X, Y: unknown nucleotides). Binding models: filled circles represent imidazole rings; open circles represent pyrrole rings; curves-linking circles represent γ-aminobutyric acid; diamonds represent β-alanine; the triangle represents TMG. (**A) Forward binding**, **(B) reverse binding**.



A **B**

Figure 3.5. The structure of hairpin TMG NV1087 polyamide and the code of DNA recognition upon binding to DNA (W: A or T; X, Y: unknown nucleotides). Binding models: filled circles represent imidazole rings; open circles represent pyrrole rings; curves-linking circles represent γ-aminobutyric acid; diamonds represent β-alanine; the triangle represents TMG. (**A) Forward binding**, **(B) reverse binding**.

When the polyamides align to a sequence suggested by the recognition rules defined by Dervan (54, 60) as in **Figures 3.4 and 3.5**, the binding site is considered a perfect match site. The mismatched sites are the sites where the polyamide aligns to a DNA sequence that is different. For example, a guanine (G) or a cytosine (C) appears in the place of A or T or imidazole aligns to a nucleotide that is different than G.

The binding affinity of PA and PA-EDTA conjugates is similar as demonstrated in the binding study of PA1 (11). The conjugated EDTA group removes one positive charge and adds three negative charges to the polyamide when not bound by Fe(II) (11).

3.3.1 Binding preferences of NV1078

The Map in **Figure 3.6** presents the 120bp DNA sequence and the variety of possibilities of polyamide NV1078 to bind in the minor groove of the ds DNA. The cleavage pattern of the control samples can be compared with the cleavage pattern of the sample treated with 0.5 nM NV1078. The affinity cleavage sites are depicted with the help of the vertical arrows, while and related electropherograms display the DNase I protection sites.

The data represents the interpretation of the following experiments: DNase I Footprinting of 50 pM DNA (ECF3011, ECF3018, ECF3030, ECF3058); Sequencing with USB Kit (ECF2142); Affinity cleavage: 1 nM DNA and 25 nM NV1078 (ECF3037); Maxam-Gilbert Sequencing (ECF2135).

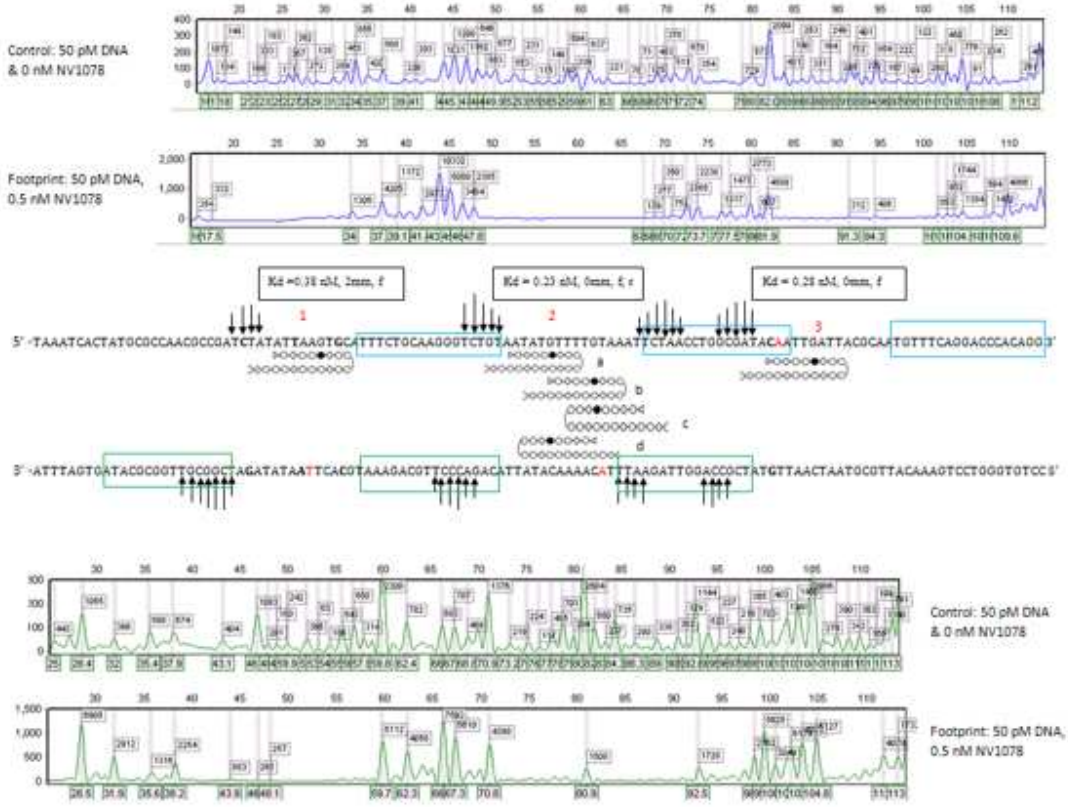


Figure 3.6. Map of 120 XC bp DNA. The first two electropherograms provide information for the top DNA strand, labeled at 5' with FAM dye. The electropherograms in the HEX channel correspond to the mirror image of the bottom nucleotide strand. Nucleotides marked in red are the integration sites, were the dissociation constant values

Kd's were determined. The regions outside the boxes the protected areas were NV1078 binds with high affinity.

The data gathered from electropherograms have been used to calculate the apparent fractional saturation θ (the fraction of polyamide bound to the DNA). The isotherms, fraction bound vs. polyamide concentration, allow the calculation of association/dissociation constants and Hill coefficient.

For the integration site, the fraction bound was calculated as in Eq.3, where I_{site} represents the fluorescence intensity, the peak height in a footprint region, and I_{ref} is the peak height of a reference site that does not change in intensity during titration. I_0 and I_{0ref}, as in Eq.4, are the intensities of the sample peak and reference peak in the control samples (just DNA cut by DNase I or hydroxyl radicals, without polyamide treatment). To obtain the apparent fractional saturation, θ, the results were normalized to one.

$$\Theta = I_{site} / I_{ref}; \quad \text{(Eq.3)}$$

$$\Theta_0 = I_{0site} / I_{0ref} \quad \text{(Eq.4)}$$

$$\Theta_{app} = 1 - \Theta / \Theta_0 \quad \text{(Eq.5)}$$

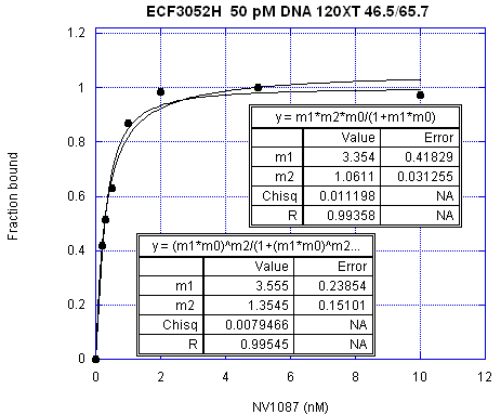


Figure 3.7. Isotherm (KaleidaGraph 4.1 software). Results according with Langmuir equation (upper table) and Hill equation (lower table) of 120 bp DNA, [DNA] = 50 pM, incubated with NV1078 (0 - 10 nM) and digested with 0.35 U DNase I.

The results derived from Hill equation (Eq.2) in this particular case (**Figure 3.7**) are:

Kd = 1/3.5 = 0.74 nM; Hill coefficient, n = 1

From the cumulative DNase I and affinity cleavage data on the 120bp DNA, three binding sites for TMG-AHP NV1078 were identified. The dissociation constants have been calculated from the isotherms obtained in at least three independent experiments. The detailed results are presented in the appendix and in the supplementary material.

(**Appendix 1**: Dissociation/association constants: NV1078 & 120 bp 50 pM DNA; **Figure 3.11:** Isotherm triplicates: NV1078 120 bp, 50 pM DNA).

Table 3.2. Summary of Hill dissociation constants for four binding sites on 120XA, 120XT, 120XC, 120XG DNA of NV1078. The polyamide tail binding site is highlighted in red

Footprint	Binding Sequence	KdHill (nM) A	KdHill (nM) T	KdHill (nM) C	KdHill (nM) G	Binding fashion	TMG
#1	5' ATX-TATATTAAGTGCA-TTT 3'	0.35 ± 0.22	0.35 ± 0.09	0.38 ± 0.18	0.28 ± 0.01	2MM,F	A
#2	5' ATG-TTTTGTAAATTCT-AAC 3'	0.29 ± 0.07	0.23 ± 0.03	0.22 ± 0.08	0.24 ± 0.11	0MM,R	T
#3	5' GCG-ATAXAATTGATTA-CGC 3'	0.42 ± 0.22	0.23 ± 0.09	0.18 ± 0.13	0.67 ± 0.32	0MM,F	X

Table 3.3. Summary of Hill association constants for four binding sites on 120XA, 120XT, 120XC, 120XG DNA of NV1078. The polyamide tail binding site is highlighted in red

Footprint	Binding Sequence	KaHill (nM^{-1}) A	KaHill (nM^{-1}) T	KaHill (nM^{-1}) C	KaHill (nM^{-1}) G	Binding fashion	TMG
#1	5' ATX-TATATTAAGTGCA-TTT 3'	2.86	2.86	2.63	3.57	2MM,F	A
#2	5' ATG-TTTTGTAAATTCT-AAC 3'	3.45	4.35	4.55	4.17	0MM,R	T
#3	5' GCG-ATAXAATTGATTA-CGC 3'	2.38	4.35	5.56	1.49	0MM,F	X

The first binding site, not influenced by mutations (pos. 27-39, 1) in **Figure 3.6**, is a forward binding site where NV1078 tolerates well two mismatched (mm) nucleotides. The high cooperativity at this location shows that an alternative binding mode is also probable at this site, namely a two-mm forward (29-41) site, not supported by AC sites, where the tail binds TATT, TMG at T**.**

Binding site no. 2 (pos. 59-71), **Figure 3.6,** holds the potential for superposition of different binding sites. This site can be associated with a forward 0-mm binding (2a, 5'TGTAATATGTTTT3') or a 0-mm reverse fashion (2c, 5'TTTTGTAAATTCT3'). This binding site is not affected by mutations and shows a nearly constant Kd (Appendix 1). The affinity cleavage patterns suggest that 0-mm reverse binding is preferred over forward binding.

According to the dissociation constant values for the third binding site **Figure 3.6,** 3 (position 85-97), the location with systematic changes in the DNA fragment at TMG binding sites, shows preference for C and T (**Table 3.2**, **Fig. 3.8).** The presence of G is less-preferred than C, T or A at the TMG binding site or location.

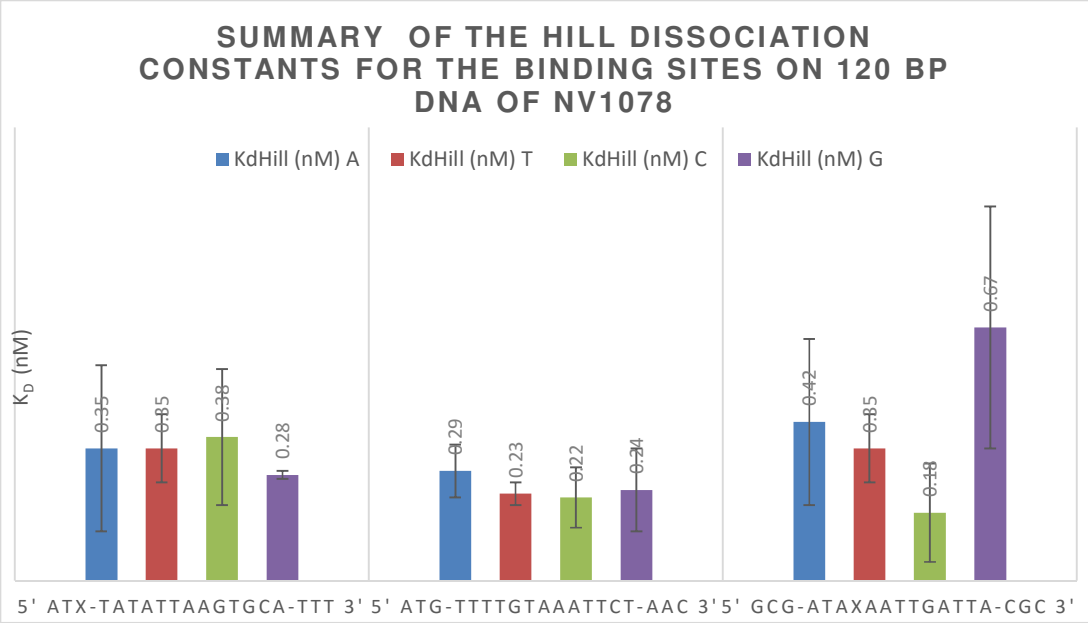


Figure 3.8. Dissociation constants for the binding sites on 120 bp DNA of NV1078, based on Table 3.2

3.3.2 Binding preferences of NV1087

The dissociation constants have been calculated from the isotherms obtained in at least three independent experiments. The detailed results are presented in the appendix. (**Appendix 2**: KaleidaGraph data NV1087 for 120 bp DNA at 50 pM;**)**, **Table 3.4, Figure 3.9**., and **Figure 3.12** (KaleidaGraph Isotherms NV1087 120 bp).

The binding profile of NV1087 on the same 120 bp DNA sequence revealed four binding sites. Binding of NV1087 in site no.1 (position 24-37), the position with systematic changes in the DNA fragment at TMG binding sites, derived from dissociation constant values, shows a preference for C, T, and A (**Table 3.4**, **Fig. 3.5**). Binding site no. 2 (position 55-68) one mm forward binding site is not affected by mutations. Binding site no. 3 (position 66-79) one mismatch reverse binding site is not affected by mutations and is also the most tightly bound site. Binding site no. 4 (position 87-100), 2mm reverse binding site is not affected by mutations.

Table 3.4. Summary of Kd values obtained for NV1087. The polyamide tail binding site is highlighted in red

Footprint	Binding Sequence	KdHill (nM) A	KdHill (nM) T	KdHill (nM) C	KdHill (nM) G	Binding fashion	TMG
#1	5 'CCG-ATXTATATTAAGT-GCA 3'	0.28± 0.10	0.42 ± 0.11	0.28 ± 0.07	0.93 ± 0.29	0MM,F	X
#2	5' CTG-TAATATGTTTTGT-AAA 3'	0.36 ± 0.16	0.45 ± 0.04	0.30 ± 0.09	0.60 ± 0.34	1MM,F	A
#3	5' TTT-TGTAAATTCTAAC-CTG 3'	0.19 ± 0.07	0.24 ± 0.09	0.22 ± 0.07	0.33 ± 0.11	1MM,R	A
#4	5' TTG-CGTAATCAATTYT-ATC 3'	0.34± 0.25	0.51 ± 0.08	0.61± 0.17	0.40 ± 0.12	2MM,R	T

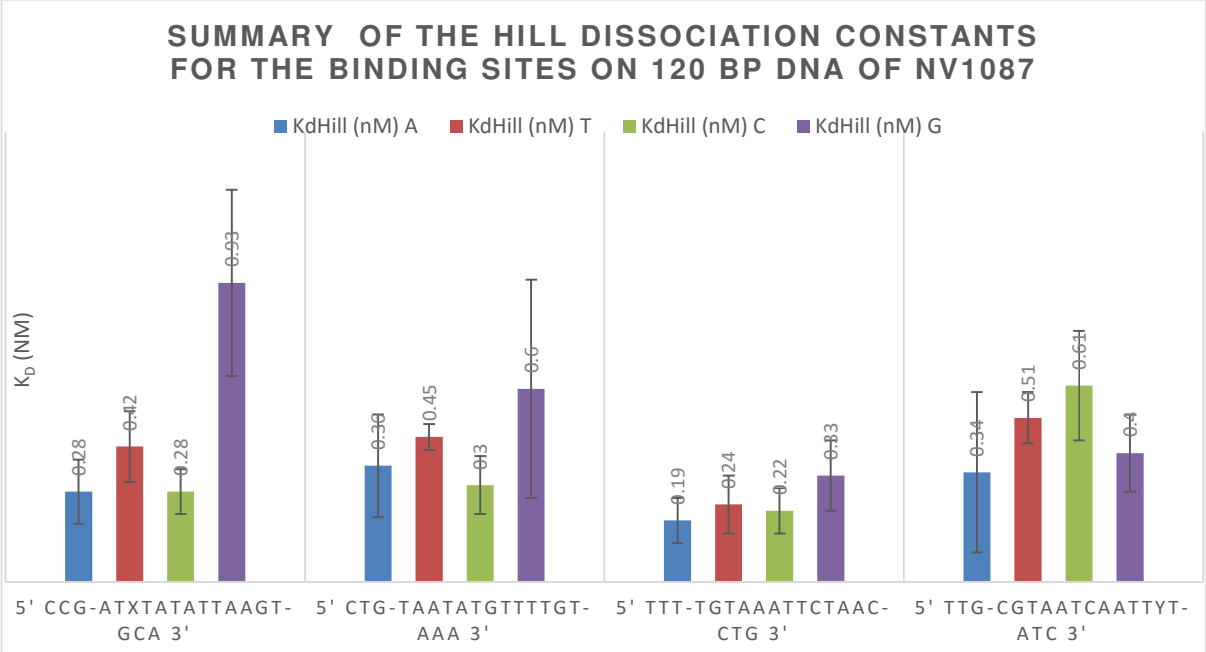


Figure 3.9. Dissociation constants for the binding sites on 120 bp DNA of NV1087 based on Table 3.4.

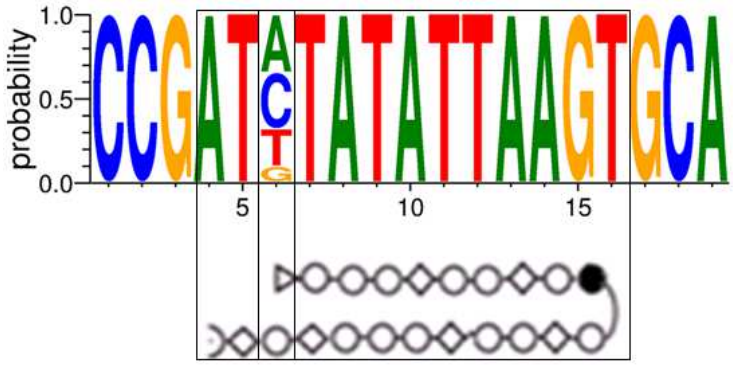


Figure 3.10. Binding of NV1087 in position #1, the position with systematic changes in the DNA fragment at TMG binding or location sites derived from association constant values, the probability plot shows preference for C, T, and A. (WebLogo 3.4)

3.4 Conclusion

The goal of this work was to provide a detailed binding site analysis illustrating how polyamide NV1078 and NV1087 can bind the specified 120 bp DNA sequence. The

second aim was to determine how the DNA sequence where the polyamide tail binds can influence the binding affinity.

In the AC experiments, all significant binding sites appeared at 10 nM PA and became clearer at 25 nM PA. Indexing (size calling) of the electropherograms was achieved with Sanger and Maxam-Gilbert sequencing (11). The 3' shift between the forward and revers strands characteristic for minor groove binders (108) was observed. Although reverse orientation is considered less favorable for small polyamides (109), such patterns are seen for large polyamides in association with the specified DNA fragment. Reverse binding was detected in the present work and also in recent studies of my colleagues (11, 86, 105).

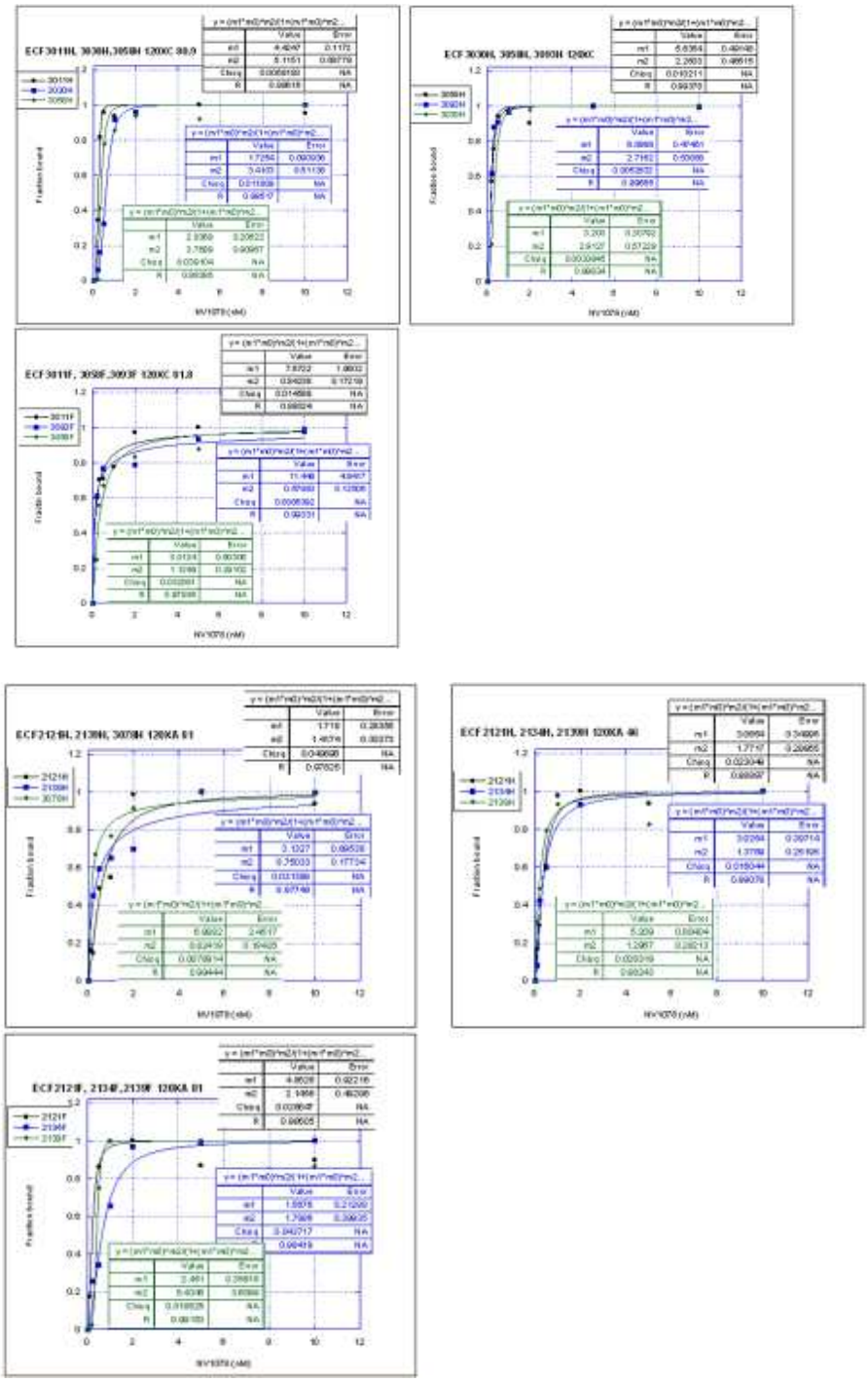
The results for NV1078 show very tight binding for the forward perfect-binding site (K_d = 0.23-0.7 nM), the highest value being registered when TMG binds G, or is located next to G (**Table 3.2** and **Figure 3.8**, binding site #3). However, the reverse perfect match site when TMG binds to T presents the same low value (at site #2). Site #1, a two-mismatch forward-binding site is independent of mutations, having similar Kd values of about 0.3 nM in every strand (TMG at A).

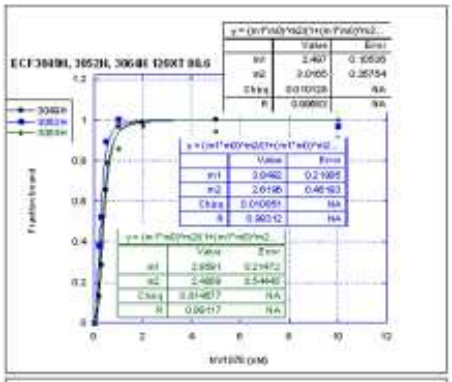
Polyamide NV1087 binds in four different sites (**Table 3.4**). The forward, perfect-binding site presents slightly higher Kd values when TMG binds to G. One-mismatch reverse (TMG at A) has the lowest Kd (0.2 nM), very close to one-mismatch forward (0.4 nM) (TMG at A).

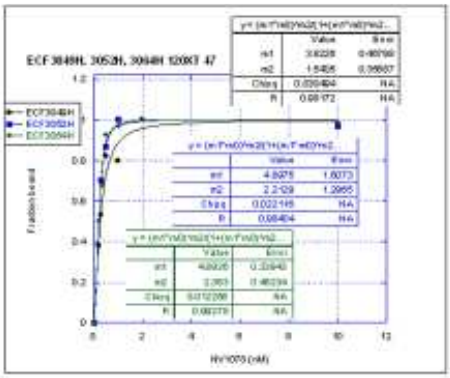
The binding of the polyamides is influenced not just by the nucleotide position at the TMG site but also by the entire nucleotide sequence at the tail of the polyamides. If the polyamides tail bind to A, T, or C, the binding affinity is increased. The presence of G nucleotide is less-preferred than T, C or A at the TMG site for both TMG-AHPs

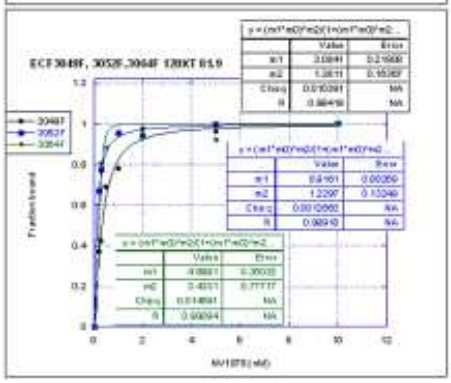
3.5 Supplementary material

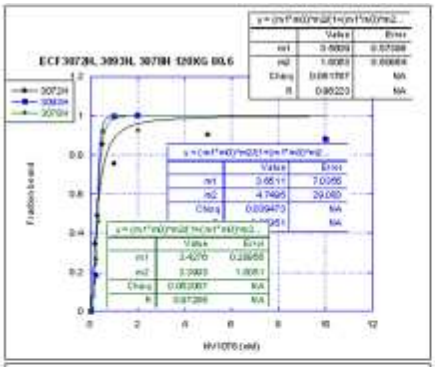
Figure 3.11: Isotherm triplicates of 3 binding sites of NV1078 on 4 DNA sequences. (120 bp DNA, 50 pM DNA)

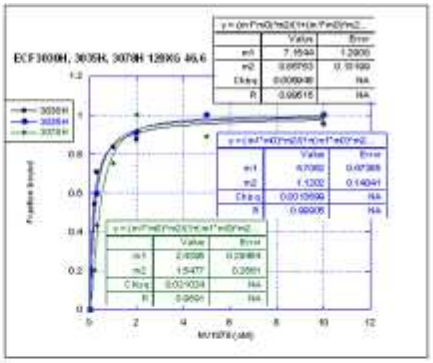












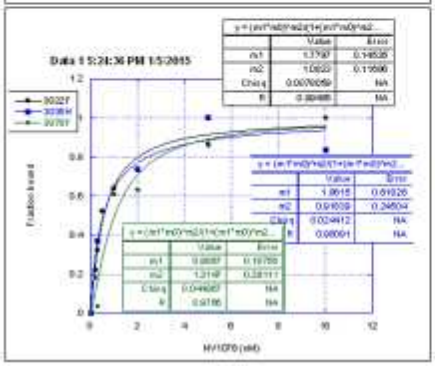
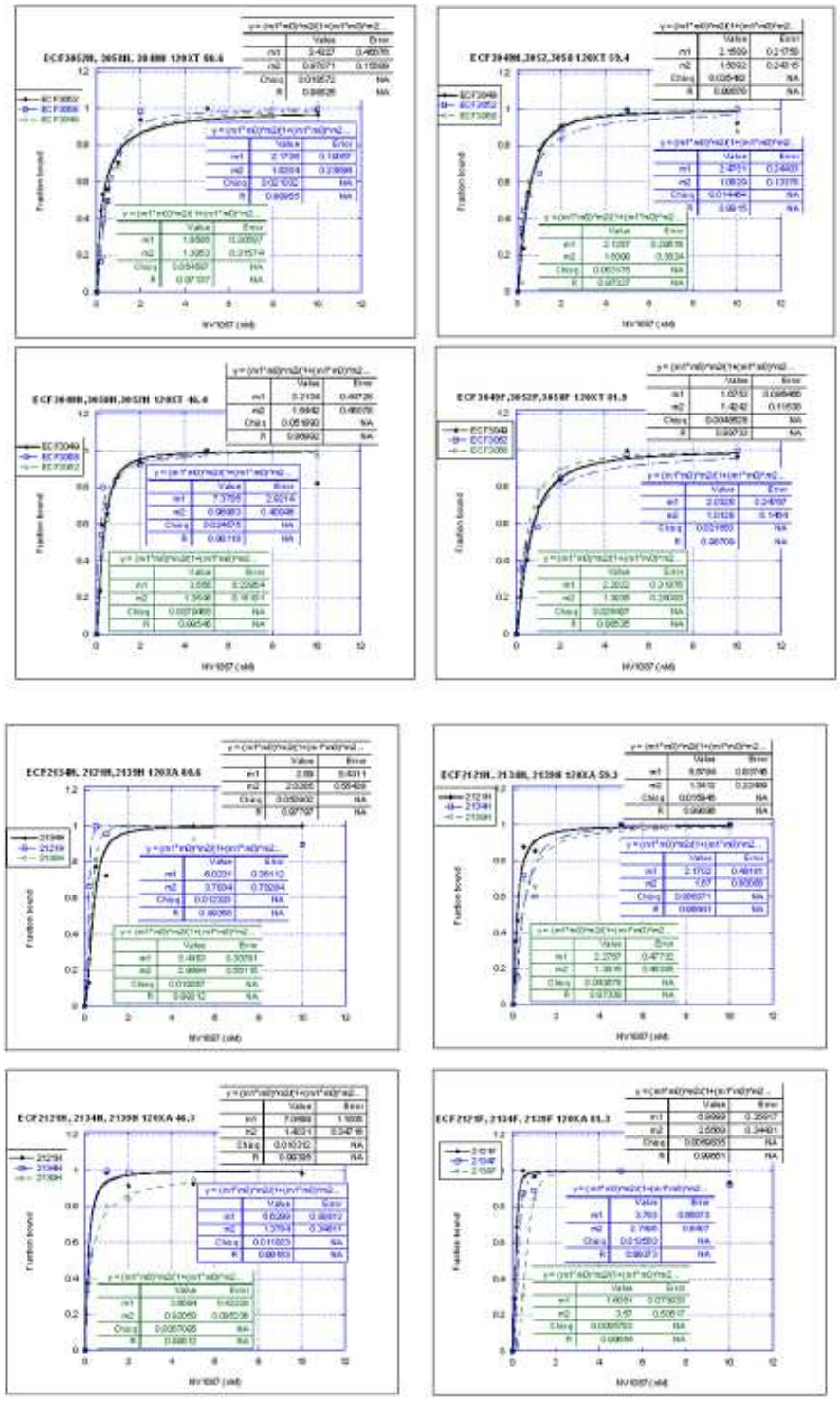
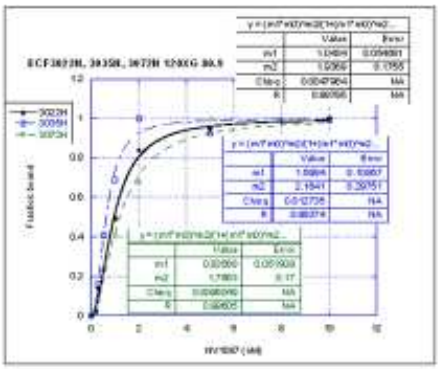
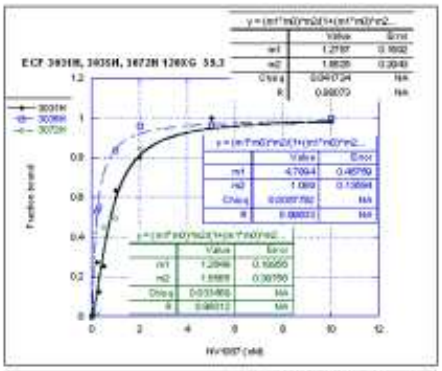
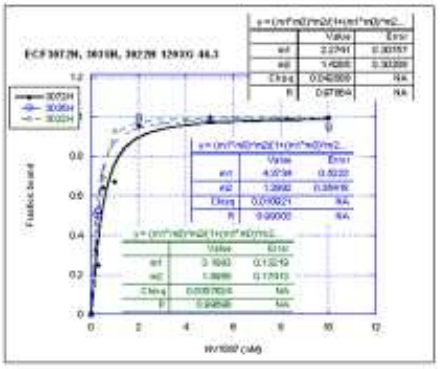


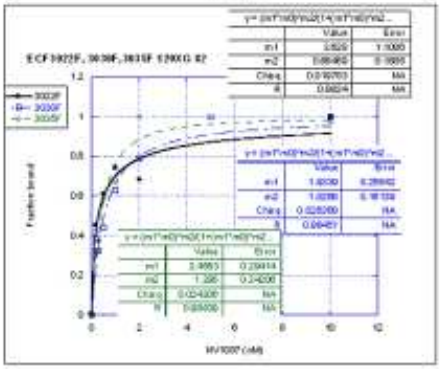
Figure 3. 12: Isotherm triplicates of 4 binding sites of NV1087 on 4 DNA sequences. (120 bp DNA, 50 pM DNA)

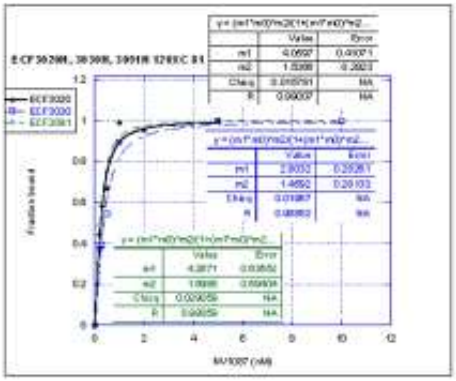


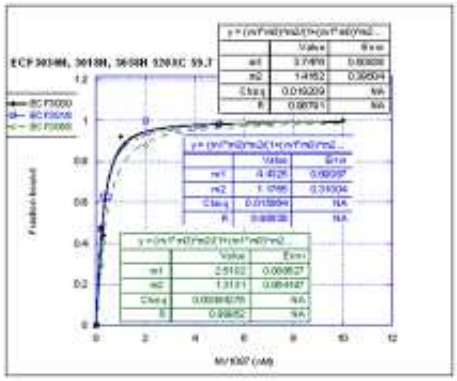


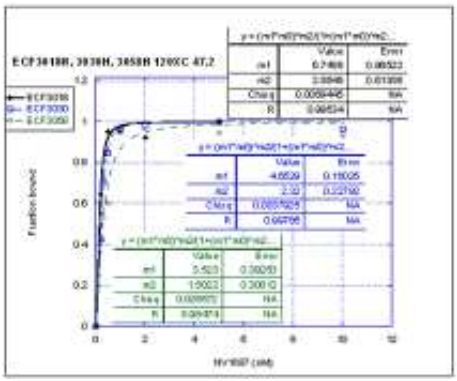


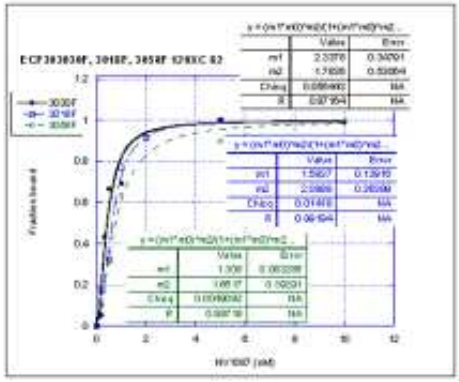












Chapter 4

Interactions Between two TMG Asymmetric Hairpin Polyamides (TMG-AHP), NV1078 and NV1087, with the HPV16,18 LCR segment

4.1Abstract

Long-term, persistent infection with high-risk strains of human papillomavirus (HPV) is the precursor of most of the cervical carcinomas and an increasing number of oropharyngeal squamous cell carcinomas which represent 90-95% of head and neck cancers (3). While the vaccines can protect the targeted population of possible infections (4), no HPV antiviral drugs are available for the definitive treatment of preexisting or future viral infections. To treat current or future HPV infections, drugs that selectively block virus-specific processes, but do not damage the host cells, are needed.

The antiviral compounds employed in the present study, NV1078 and NV1087, represent a subclass of the heterocyclic oligomers known as N-methylpyrrole N-methylimidazole (Py-Im) polyamides (PAs). They have a modified N-terminus capping group, 1,1,3,3 tetramethylguanidine (TMG) and an asymmetric structure (26), and will be referred to as TMG-AHPs for TMG-Asymmetric Hairpin Polyamides. It has been demonstrated (Edwards, T. and Fisher, C.) that the two compounds possess the unique property of unusual antiviral activity against three high-risk HPV strains (HPV16, HPV18, and HPV31) (Table 1). The antiviral compounds eliminate the viral load of infected keratinocytes without integration of the viral DNA episomes in the host chromosome. The TMG-AHPs are analogs of natural products Distamycin A and Netropsin that function, as has been recently determined (5, 6), by interfering with natural virus-host interactions and by stimulating the host cell's DNA Damage Response to destroy viral DNA.

The interactions between the aforementioned polyamides and LCR of viral HPV16 and HPV18 DNA fragments were studied to determine the DNA sequence specificity and binding sites of the polyamides, to help establish the TMG-AHP's mechanism of action. The quantitative deoxyribonuclease I (DNase I) footprinting method was used (11, 97), along with affinity cleavage assays (95), to map and measure the interactions between fluorescently labeled DNA and the active TMG-AHPs. The dsDNA-AHPs binding locations were detected and quantified by capillary electrophoresis (110).

Results in this chapter demonstrate that TMG-AHPs 1078 and 1087 recognize 13 nucleotides, and bind with nanomolar affinity to the expected binding sites, as well as one-

, two-, and even three-mismatch (sites of polyamide binding with a DNA sequence containing one or more bases that do not correspond to the expected polyamide binding sequence) binding sites. Moreover, it was observed that the asymmetric polyamides (APAs) have a conclusive preference for the region adjacent and overlapping with E1 and E2 viral proteins binding sites, where both polyamides bind in a cluster. The data indicate that viral replication for HPV could be disturbed by TMG-AHPs binding at numerous sites with high affinity for the sequence (7880-30) of the HPV16 viral LCR DNA or the preferred binding region of HPV18 LCR (7841-43).

4.2 Introduction

Although human papillomavirus (HPV) vaccines have been developed (4), the occurrence of HPV infections is still extremely prevalent. To treat the already existent or future viral infections, drugs that selectively block virus-specific processes but do not damage the host cells are needed. Persistent infection causes genomic instability and eventually the integration of the viral genome into the host chromosomes (111, 112), and is the main reason for neoplasia and cancer. The elimination of viral DNA during the maintenance stage of a persistent infection is the strategy behind the discovery of the antiviral compounds described here.

The genetic information of Human Papillomaviruses is stored in small, chromatinized circular dsDNA viral genome (episome) that typically contains 8 genes (27). The HPV genome is approximately 8000 base pairs long. Since the viral genome does not encode any polymerases, the only viral enzyme being E1 (an NTPase / helicase), the maintenance of viral load depends on the host, infected keratinocyte cells' replication machinery. Two viral proteins E1 and E2 also have decisive roles in most replication pathways. These viral proteins recruit host proteins, including DNA polymerase alpha and many proteins which are part of the DNA Damage Response (DDR) defense mechanism (8, 9, 15, 48, 113, 114), to initiate and enable the replication process. Also, these viral proteins may serve as bridges between the viral genome and the cellular chromatin (36), ensuring participation of the viral proteins to the replication process. It is important to mention that opposing many studies supporting the importance of E1 and E2 in the viral lifecycle, it was also shown (using various transfected cell lines) that a different replication mechanism, independent of viral trans-factors, E1 and E2 (115), is also possible. HPV DNA is capable of replicating by either random choice or a once-per-S-phase mechanism, depending on cell type and levels of E1 protein (33). In the particular case of W12E cells,

a once-per-S-phase replication mechanism, dependent on E1 and E2, is most frequently observed (116).

Previous studies (10-12, 112, 117) have shown that viral dsDNA can be targeted using small molecules such as pyrrole-imidazole polyamides, homologs of natural products Distamycin A and Netropsin (49). To overcome the high toxicity but maintain antiviral properties of the minor groove binders Distamycin A and Netropsin, hairpin N-methylpyrrole-N-methylimidazole (Py-Im) polyamides (PAs) were designed (54, 60). To exert their effect, the polyamides have to reach and bind to their cellular target, the viral dsDNA in the nuclei. Slight, sometimes minimal structural changes of the compounds may result in distinct cellular or nuclear distribution with dramatic changes of biological activity. Efficient cellular uptake of hairpin PAs was observed in human keratinocytes harboring high-risk HPV episomes (5, 6, 10). Hairpin polyamides specifically recognize and bind with high (nanomolar) affinity to dsDNA in minor groove regions by hydrogen-bonds (Fig.1) (49, 59). The binding of PAs is followed by alterations of the dsDNA shape by widening the minor grove, narrowing the major groove (54, 60), and stiffening the double helix (61).The introduction of β-alanine building blocks confer additional flexibility for the molecules and permit a better mimic of the DNA curvature (55, 57). By binding with high affinity in the minor groove, polyamides displace proteins from DNA, interfering with natural processes (62) and acting as transcriptional activators or repressors.

To decrease or eliminate the HPV viral DNA (episomes) from infected keratinocyte cells, a library of more than 100 polyamides was synthesized. It was found that some of the long hairpin polyamides are active against high-risk HPV strains (HPV16, HPV18, or HPV31) (5, 6, 10-12, 26) and cause a substantial decrease of the HPV DNA episome levels, eradicating the viral load (amount of viral DNA per cell) in undifferentiated keratinocyte monolayers, in the absence of integration of viral DNA (5, 6, 117). A minimum of 10 DNA base-pair coverage is necessary for viral activity (10).

Among the newly synthesized polyamides (work by Kevin J. Koeller, G. Davis Harris. Jr.), two stand out as having the greatest antiviral activity against three high-risk HPV strains: HPV16, HPV18, and HPV31 (26) **(Table 4.1**). The two polyamides employed in the present study, NV1078 and NV1087, have a modified N-terminus capping group, 1,1,3,3 tetramethylguanidine (TMG) and an asymmetric hairpin structure (TMG-AHP). These distinct polyamides have the ability to decrease HPV16, 18, and 31 episome levels in undifferentiated W12E cell monolayers and tissue culture, in the absence of integration

of viral DNA, with greater efficacy than NV1042, the most potent compound published, considered the lead compound (work by Terri G. Edwards, NanoVir LLC).

The interactions between the above-mentioned polyamides and LCR of viral HPV16 and HPV18 DNA fragments were studied to determine the DNA sequence specificity and binding sites of the polyamides, to help establish the TMG-AHP's mechanism of action. The studied DNA sequence includes the core replication origin containing the cis-elements (which do not code for proteins or RNA) imperative in the process of DNA replication: one E1 protein binding sequence, and the three important E2 binding sites (E2BS1, E2BS2, and E2BS3) (118). The E1BS for HPV16 sometimes referred to as the origin-binding domain (OBD), is an 18 base-pair sequence: 5'ATAATACTAAACTACAAT3' (HPV16 episome sequence: 7895-8). The DNA fragment also contains A/T rich regions surrounding the E1BS and E2BSs and transcription factor binding sites.

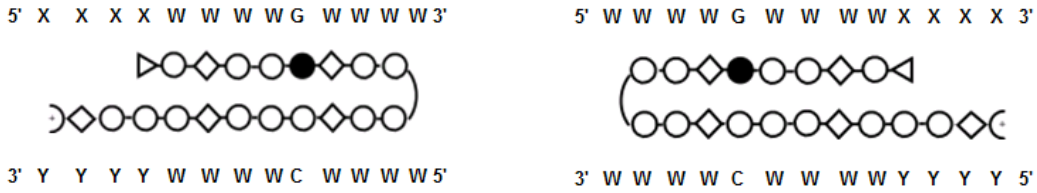
4.2 Methods

The methods used in this work were DNase I footprinting to determine the binding affinities of the TMG polyamides and the affinity cleavage assay (AC) to ascertain the orientation of the TMG-AH polyamides. A detailed description of the methods is provided in **Chapter 3**.

In addition, the HPV16 LCR 7662-122 bp fragment was generated by Polymerase Chain Reaction (PCR). The primers employed in reaction were: (forward) 5'-FAM-TAA ATC ACTATG CGC CAA CGC-3' and (reverse) 5'-HEX-CCT GTG GGT CCT GAA ACA TTG-3'. All oligonucleotides were obtained from IDT (Coralville, Iowa).

4.3 Results

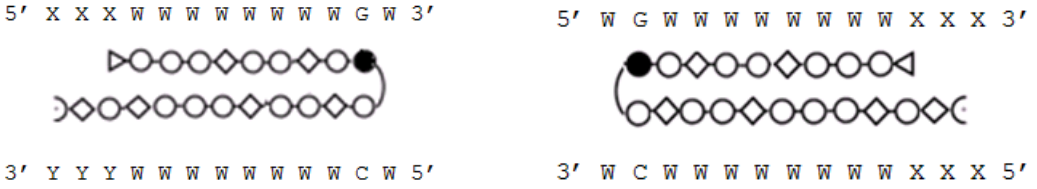
The polyamide NV1078 has the following composition: TMG-PβPPImβPPβγPPβPPPβPPPβTa, where Im represents N-methylimidazole, Py stands for N-methylpyrrole, β–alanine, γ is γ-aminobutyric acid, and Ta represents $CH_3N(CH_2CH_2CH_2NH_2)_2$ or 3,3'-diamino-N-methyldipropylamine. NV1078 is present in solution as a TFA salt. According to the published pairing rules (54, 60) the perfect binding site for NV1078 will be: 5'-XXXWWWWGWWWW-3' where W=A or T, and X stands for the unknown nucleotides for TMG binding. The hairpin structure and the mode of binding to dsDNA it is showed in **Figure 4.1.**



(a) (b)

Figure 4.1. The structure of hairpin TMG NV1078 polyamide and the code of DNA recognition upon binding to DNA (W: A or T; X, Y: unknown nucleotides). Binding models: filled circles represent imidazole rings; open circles represent pyrrole rings; curves-linking circles represent γ-aminobutyric acid; diamonds represent β-alanine; the triangle represents TMG. (**A) Forward binding** corresponds to N to C/5' to 3' orientation, while **reverse binding (B**) corresponds to N to C/3' to 5' direction.

The polyamide NV1087 has the following structure: TMG-PPPβPPβPImγPβPPβPPPβPβTa, where Im represents N-methylimidazole, Py stands for N-methylpyrrole, β is β–alanine, γ is γ-aminobutyric acid, and Ta represents $CH_3N(CH_2CH_2CH_2NH_2)_2$ or 3,3'-diamino-N-methyldipropylamine. According to the pairing rules (54, 60) the perfect forward binding site for NV1087 will be: 5' X-X-X-W-W-W-W-W-W-W-W-G-W 3' where W=A or T, and X stands for the unknown nucleotide for TMG binding **(Figure 4.2).**



(a) (b)

Figure 4.2. The structure of hairpin TMG NV1087 polyamide and the code of DNA recognition upon binding to DNA (W: A or T; X, Y: unknown nucleotides). Binding models: filled circles represent imidazole rings; open circles represent pyrrole rings; curves-linking circles represent γ-aminobutyric acid; diamonds represent β-alanine; the triangle represents TMG. (**A) Forward binding** corresponds to N to C/5' to 3' orientation, while **reverse binding (B)** corresponds to N to C/3' to 5' direction.

4.4.1 NV1078 polyamide binding sites on 365 bp natural HPV16 (7662-122) sequence

The first goal of this study was to provide a detailed binding site analysis, illustrating the binding ability of polyamide NV1078 to the specified DNA sequence. The HPV16 LCR (GeneBank AF125673.1LCR) region contains cis-acting elements that regulate transcription and replication of viral DNA (118).

The studied DNA sequence 7662-122 includes one E1 protein binding sequence, and the three essential E2 binding sites: E2BS1 (35-46), E2BS2 (50-61), and E2BS3 (7858-7869), as shown by the green boxes (**Figure 4.3**). The E1BS for HPV16, referred to as the origin-binding domain (OBD), is an 18 base-pair sequence: 5'ATAATACTAAACTACAAT3' (HPV16 episome sequence: 7895-8) (27). The DNA fragment also contains A/T rich regions surrounding the E1BS and E2BSs and transcription factor binding sites (SP1, TATA motif).

The binding of E2 viral protein to its binding sites and recruitment of E1 viral protein to form a double hexamer is fundamental to viral DNA replication (119). The mapping studies based on data gathered from the DNase I and affinity cleavage experiments indicate that the polyamide NV1078 binds with high affinity to multiple sites adjacent to and overlapping with the E1BS and E2BSs.

The binding information is presented in the ECF3097 Map **(Figure 4.3)** as well as **Tables 4.1 and 4.2.**

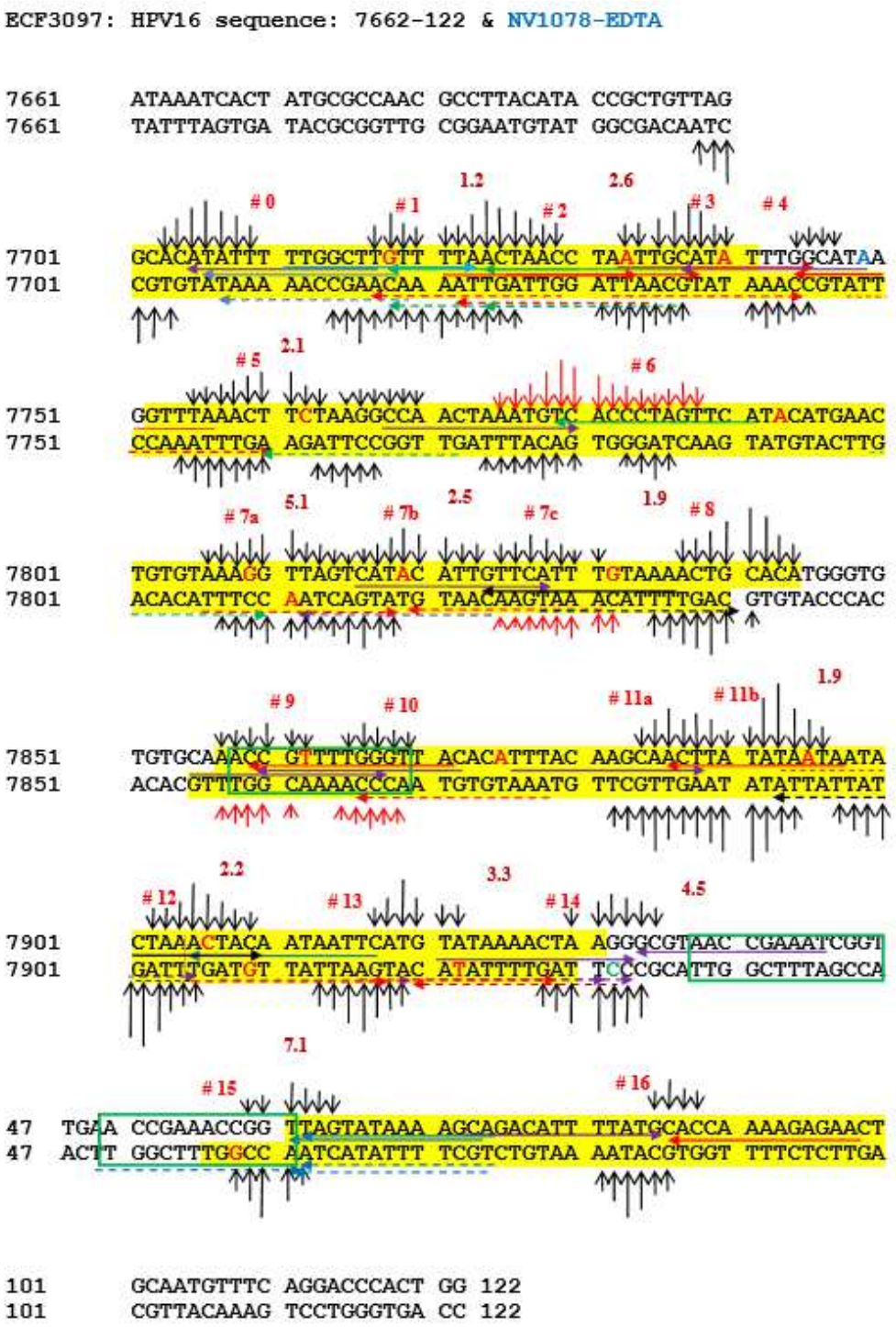


Figure 4.3. Binding sites of NV1078 on 365 bp LCR HPV16 fragment. Horizontal arrows depicting the binding sites are color coded. Black arrow shows perfect binding site; Red arrow - single mismatch (mm) site; Purple arrow - double mm; Blue arrow - triple mm; Green arrow - quadruple mm; Solid arrow - forward fashion of PA binding; Dashes arrow - reverse fashion of PA binding; Green boxes show E2 binding sites; Red letters

represent the integration peaks; Blue or green letters represent the reference peaks. Red numbers show the dissociation constant values Kd [nM] presented in **Table 4.2.** The yellow background highlights the footprinting regions. Red AC sites are present just at high polyamide concentrations (100 nM)**.**

A total of 16 binding sites were identified according to the affinity cleavage patterns. The binding sites and the corresponding nucleotide sequences are presented in the Tables 1 and 2. The affinity cleavage patterns (vertical arrows) indicate the position of the PA tail (**Figures 4.3**) where the hydroxyl radicals are formed (Chapter 3).

The hydroxyl radical cleavage pattern provides information about the PA binding site and the orientation (forward or reverse) as described above. Horizontal arrows show the polyamide binding sites, which may only be occupied by a single polyamide molecule at a time. The length of the arrow corresponds to the DNA sequence to which the polyamide is bound, while the arrow color coding relates to the literature binding rules. The binding of the polyamides in the minor groove is an equilibrium process.

The polyamide associates and dissociates from the minor groove and rapidly rebinds in an alternative way if possible, resulting in a variety of binding patterns. The majority of the sites permit alternative binding modes. Also, overlapping binding sites are often present.

Table 4.1. Binding sites of TMG-AHP, NV1078 on HPV16 (7662-122) fragment. The 13 nucleotide binding sequence of the polyamide has been shown with the associated 5 nucleotide flanking sequences. Blue letters indicate the TMG binding site and the highlighted guanine G indicate imidazole binding site.

AC site	Sequence 5'-3' (XXXXWWWWGWWWW)	Mismatch type	Position	Binding fashion	PA tail 5'-3'
0	GGCAC–ATATTTTTGGCTT–GTTTT	2	7705-7717	forward	ATAT
0	GCACA–TATTTTTGGCTTG–TTTTA	3	7706-7718	forward	TATT
0	TTTAA–ACAAGCCAAAAAT–ATGTA	3	7707-7719	reverse (BS)	AAAT
1	ATTAG–GTTAGTTAAAACA–AGCCA	1	7717-7729	reverse (BS)	AACA
1	GGCTT–GTTTTAACTAACC–TAATT	4	7718-7730	forward	GTTT
1	AATTA–GGTTAGTTAAAAC–AAGCC	4	7718-7730	reverse (BS)	AAAC
2	ATGCA–ATTAGGTTAGTTA–AAACA	1	7722-7734	reverse (BS)	GTTA
2	TTTTA–ACTAACCTAATTG–CATAT	4	7723-7735	forward	ACTA
2	ATATG–CAATTAGGTTAGT–TAAAA	4	7723-7735	reverse (BS)	TAGT
2	TTAGT–TAAAACAAGCCAA–AAATA	3	7711-7723	forward (BS)	TAAA
3	TGCAA–TTAGGTTAGTTAA–AACAA	1	7721-7733	forward (BS)	TTAG
3	AAATA–TGCAATTAGGTTA–GTTAA	1	7726-7738	forward (BS)	TGCA
3	AATTG–CATATTTGGCATA–AGGTT	2	7737-7749	forward	CATA
4	TTATG–CCAAATATGCAAT–TAGGT	1	7733-7745	forward (BS)	CCAA
4	AACCT–AATTGCATATTTG–GCATA	1	7731-7743	reverse	TTTG
4	TATTT–GGCATAAGGTTTA–AACTT	1	7744-7756	forward	GGCA
5	TGGCA–TAAGGTTTAAACT–TCTAA	1	7748-7760	reverse	AACT
5	ATTTA–GTTGGCCTTAGAA–GTTTA	4	7760-7772	reverse (BS)	AGAA
6	AGGGT–GACATTTAGTTGG–CCTTA	2	7768-7780	forward (BS)	GACA
6	AAATG–TCACCCTAGTTCA–TACAT	4	7779-7791	forward	TCAC
7a	ACATG–AACTGTGTAAAGG–TTAGT	4	7798-7810	reverse	AAGG
7a	ATGAA–CAATGTATGACTA–ACCTT	2	7812-7824	reverse (BS)	ACTA
7b	TGTGT–AAAGGTTAGTCAT–ACATT	1	7806-7818	reverse	TCAT
7b	TTTAC–AAATGAACAATGT–ATGAC	1	7819-7831	reverse (BS)	ATGT
7c	ACAAA–TGAACAATGTATG–ACTAA	2	7816-7828	forward (BS)	TGAA
7c	ACATT–GTTCATTTGTAAA–ACTGC	0	7824-7836	forward	GTTC
8	TGTTC–ATTTGTAAAACTG–CACAT	0	7828-7840	reverse	ACTG
9	GCAAA–CCGTTTTGGGTTA–CACAT	2	7859-7871	forward	CCGT
9	CAAAC–CGTTTTGGGTTAC–ACATT	3	7860-7872	forward	CGTT

AC site	Sequence 5'-3' (XXXXWWWWGWWWW)	Mismatch type	Position	Binding fashion	PA tail 5'-3'
10	GTAAC–CCAAAACGGTTTG–CACAC	3	7855-7867	forward (BS)	CCAA
10	CTTGT–AAATGTGTAACCC–AAAAC	1	7866-7878	reverse (BS)	ACCC
11a	ATATA–AGTTGCTTGTAAA–TGTGT	2	7876-7888	forward (BS)	AGTT
11a	AAGCA–ACTTATATAATA–ATACT	1	7886-8898	forward	ACTT
11b	TGTAG–TTTAGTATTATTA–TATAA	0	7893-1	reverse (BS)	ATTA
12	TTATT–GTAGTTTAGTATT–ATTAT	0	7897-5	forward (BS)	GTAG
12	TTATA–TAATAATACTAAA–CTACA	2	7893-1	reverse	TAAA
12	ACTAA–ACTACAATAATTC–ATGT	3	1-13	forward	ACTA
13	TAAAC–TACAATAATTCAT–GTATA	2	3-15	reverse	TCAT
13	TACAT–GAATTATTGTAGT–TTAGT	1	1-13	forward (BS)	GAAT
13	ACGCC–CTTAGTTTTATAC–ATGAA	1	16-28	reverse (BS)	ATAC
14	ATAAT–TCATGTATAAAAC–TAAGG	1	12-24	reverse	AAAC
14	TACGC–CCTTAGTTTTATA–CATGA	2	17-19	forward (BS)	CCTT
14	TCATG–TATAAAACTAAGG–GCGTA	2	17-29	reverse	AAGG
14	ATTCA–TGTATAAAACTAA–GGGCG	2	17-29	reverse	CTAA
14	TAAGG–GCGTAACCGAAAT–CGGTT	2	30-42	forward	GCGT
15	GGTTG–AACCGAAACCGGT–TAGTA	3	49-61	reverse	CGGT
15	ACCGG–TTAGTATAAAAGC–AGACA	3	61-73	forward	TTAG
15	TGTCT–GCTTTTATACTAA–CCGGT	3	61-73	reverse (BS)	CTAA
15	ATGTC–TGCTTTTATACTA–ACCGG	3	62-74	reverse (BS)	ACTA
15	CCGGT–TAGTATAAAAGCA–GACAT	3	62-74	forward	TAGT
16	TGGTG–CATAAAATGTCTG–CTTTT	2	73-85	forward (BS)	CATA
16	TTATG–CACCAAAAGAGAA–CTGCA	1	86-98	forward	CACC

In **Table 4.1**, all binding sites of NV1078 on the HPV16 (7662-122) fragment are shown. The 13 nucleotide binding sequence of the polyamide has been shown with the associated 5 nucleotide flanking sequences. Blue letters in the DNA sequence indicate the binding site of TMG, and the highlighted guanine (G) shows an imidazole binding site. Considering the best fit to the affinity cleavage patterns and the lowest mismatch number the most probable orientations of the NV1078 polyamide (derived from Table 4.1) are presented in **Table 4.2**.

Table 4.2. Considering the best fit to the affinity cleavage patterns and the lowest mismatch number the most probable orientations of the NV1078 polyamide (derived from **Table 4.1**) are presented.

#	Sequence 5'-3' (XXXXWWWWGWWWW)	Mismatch type	Position	Binding fashion	Kd_{Hill} (nM)	Ka (nM^{-1})
1	**ATTAG–GTTAGTTAAAACA–AGCCA**	1	7717-7729	reverse (BS)	1.18	0.8
2	**ATGCA–ATTAGGTTAGTTA–AAACA**	1	7722-7734	reverse (BS)	nd	nd
3	**AAATA–TGCAATTAGGTTA–GTTAA**	1	7726-7738	forward (BS)	2.59	0.4
5	**TGGCA–TAAGGTTTAAACT–TCTAA**	1	7748-7760	reverse	2.1	0.5
7a	**ATGAA–CAATGTATGACTA–ACCTT**	2	7812-7824	reverse (BS)	5.05	0.2
7b	**TGTGT–AAAGGTTAGTCAT–ACATT**	1	7806-7818	reverse	2.5	0.4
8	**TGTTC–ATTTGTAAAACTG–CACAT**	0	7828-7840	reverse	1.98	0.5
11a	**AAGCA–ACTTATATAATA–ATACT**	1	7886-8898	forward	nd	nd
11b	**TGTAG–TTTAGTATTATTA–TATAA**	0	7893-1	reverse (BS)	1.9	0.5
12	**TTATA–TAATAATACTAAA–CTACA**	2	7893-1	reverse	2.2	0.5
13	**TACAT–GAATTATTGTAGT–TTAGT**	1	1-13	forward (BS)	3.3	0.3

The presence of G nucleotide in the tail binding sequence increases the value of Kd (compare site # 1 with sites # 3, # 8, and # 13). The Kd values not always rise with the increasing number of mismatch nucleotides as shown in **Figure 4.4.**

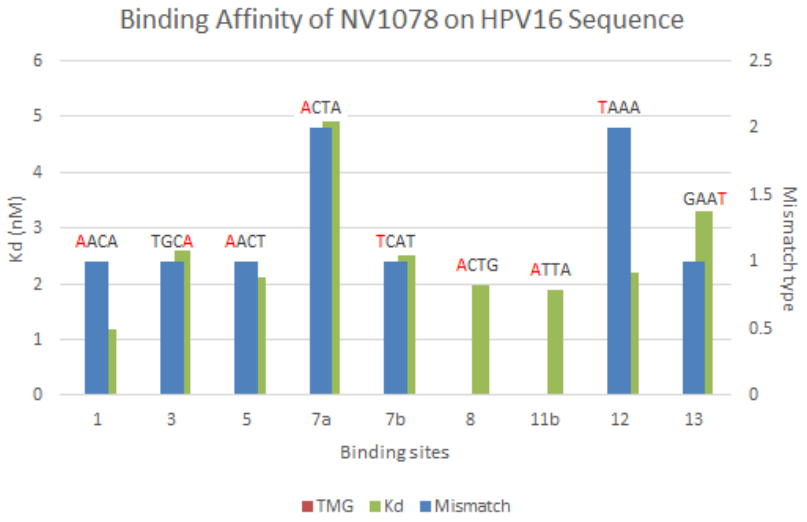


Figure 4.4. Binding affinity of NV1078 on HPV16 LCR sequence

4.4.2 NV1087 polyamide binding sites on 365 bp natural HPV16 (7662-122) sequence.

The map ECF3097NV1087(2) **Figure 4.5**, of DNase I footprinting and affinity cleavage experimental results of NV1087 polyamide binding sites on 365 bp natural HPV16 sequence reveal 13 binding sites. The blue box indicates the E1 protein binding site while E2 binding sites are delineated by green boxes. Multiple cleavage patterns stand out in the region 7881-17.

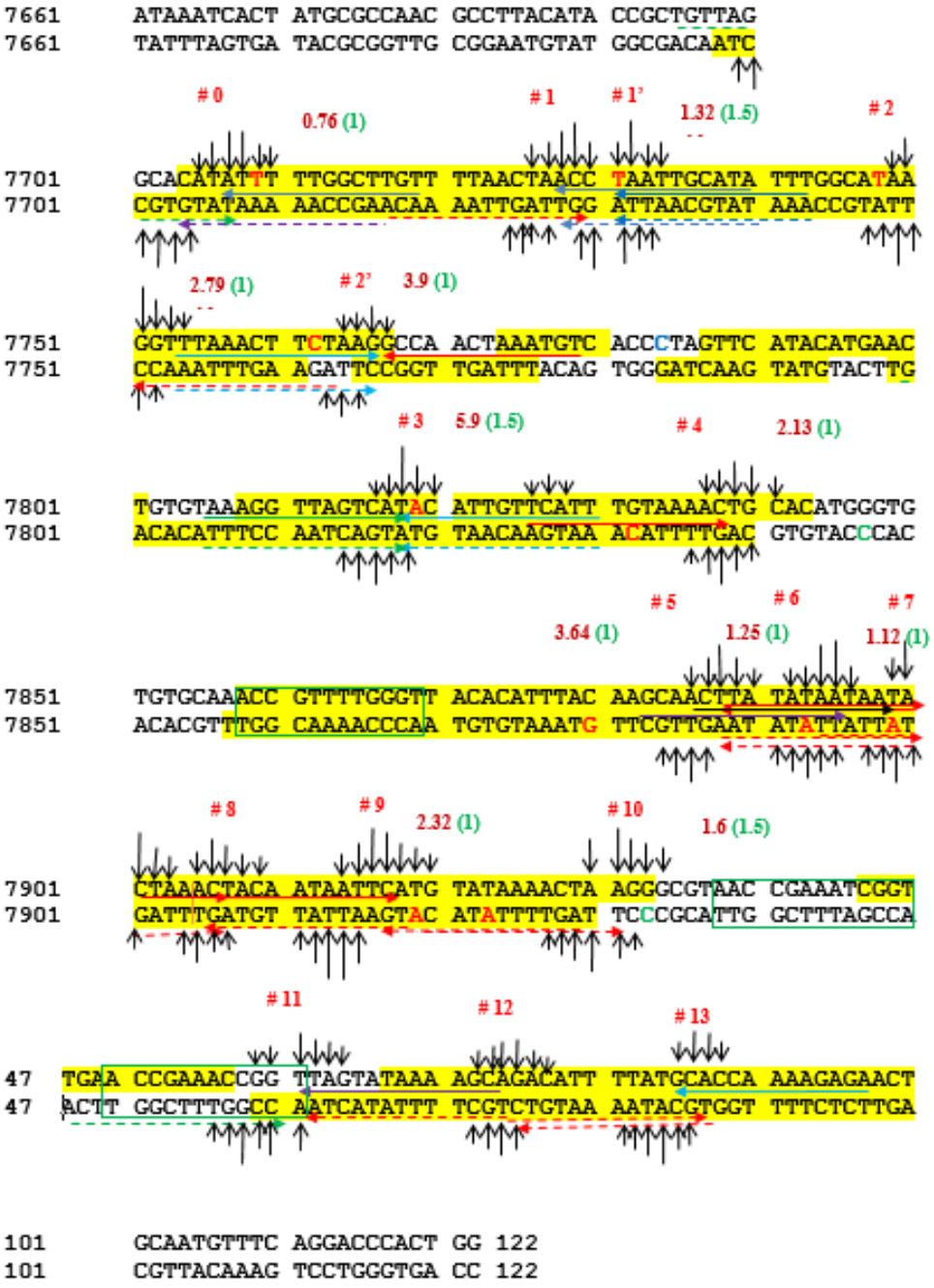


Figure 4.5. Binding sites of NV1087 on 365 bp LCR HPV16 fragment. Horizontal arrows depicting the binding sites are color coded. Black arrow shows perfect binding

site; Red arrow - single mismatch (mm) site; Purple arrow - double mm; Blue arrow - triple mm; Green arrow - quadruple mm; Solid arrow - forward fashion of PA binding; Dashes arrow - reverse fashion of PA binding; Green boxes show E2 binding sites; Red letters represent the integration peaks; Blue or green letters represent the reference peaks. Red numbers show the dissociation constant values Kd [nM] presented in **Table 4.4.** The yellow background highlights the footprinting regions.

Table 4.3. Binding sites of TMG-HAP, NV1087 on HPV16 (7662-122) fragment. The 13 nucleotide binding sequence of the polyamide has been shown with the associated 5 nucleotide flanking sequences. Blue letters indicate the TMG binding site and the highlighted guanine G indicate imidazole binding site

AC site	DNA Sequence (5'-3')	Mismatch type	Position	Binding fashion	PA tail (5'XXX3')
0a	ACCGC–TGTTAGGCACATA–TTTTT	4	7695-7707	reverse	ATA
0b	AAACA–AGCCAAAAATATG–TGCCT	2	7704-7726	reverse (BS)	ATG
0c	CACAT–ATTTTTGGCTTGT–TTTAA	3	7707-7729	forward	ATT
1a	TGGCT–TGTTTTAACTAAC–CTAAT	1	7717-7729	reverse	AAC
1b	AACTA–ACCTAATTGCATA–TTTGG	3	7728-7740	forward	ACC
1c	CCAAA–TATGCAATTAGGT–TAGTT	3	7728-7740	reverse (BS)	GGT
1'a	TAACC–TAATTGCATATTT–GGCAT	3	7731-7743	forward	TAA
1'b	ATGCC–AAATATGCAATTA–GGTTA	3	7731-7743	reverse (BS)	TTA
2	GCCTT–AGTTGTTTAAACC–TTATA	1	7751-7763	reverse (BS)	ACC
2'a	TTGGC–CTTAGAAGTTTAA–ACCTT	3	7754-7766	forward (BS)	CTT
2'b	AAGGT–TTAAACTTCTAAG–GCCAA	3	7754-7766	reverse	AAG
2'c	CTAAG–GCCAACTAAATGT–CACCC	1	7767-7779	forward	GCC
3a	AATGA–ATGACTAACCTTT–ACATT	4	7806-7818	forward (BS)	ATG
3a'	TGTGT–AAAGGTTAGTCAT–ACATT	4	7806-7818	reverse	CAT
3b	TTACA–AATGAACAATGTA–TGACT	3	7818-7830	reverse (BS)	GTA
3b'	AGTCA–TACATTGTTCATT–TGTAA	3	7818-7830	forward	TAC
4	GTGCA–GTTTTACAAATGA–ACAAT	1	7826-7839	forward (BS)	GTT
5a	GCAAC–TTATATAATAATA–CTAAA	1	7888-7890	forward	TTA
5a'	TTTAG–TATTATTATATAA–GTTGC	1	7888-7890	reverse (BS)	TAA
6	TATTA–TTATATAAGTTGC–TTGTA	2	7883-7895	forward (BS)	TTA
7a	TAGTA–TTATTATATAAGT–TGCTT	0	7886-7898	forward (BS)	TTA
7b	GCAAC–TTATATAATAATA–CTAAA	1	7888-7890	reverse	ATA
7b'	TTTAG–TATTATTATATAA–GTTGC	1	7888-7890	forward (BS)	TAT
8a	TATAT–AATAATACTAAAC–TACAA	1	7894-2	reverse	AAC
8a'	TTGTA–GTTTAGTATTATT–TACAA	1	7894-2	forward (BS)	GTT

AC site	DNA Sequence (5'-3')	Mismatch type	Position	Binding fashion	PA tail (5'XXX3')
8b	ATACA–TGAATTATTGTAG–TTTAG	1	2-14	reverse (BS)	TAG
9a	TACAT–GAATTATTGTAGT–TTAGT	1	1-13	forward (BS)	GAA
9b	CCCTT–AGTTTTATACATG–AATTA	1	12-25	reverse (BS)	TGA
10	ATTCA–TGTATAAAACTAA–GGGCG	1	15-27	reverse	TAA
11a	ACCGG–TTAGTATAAAAGC–AGACA	2	61-73	forward	TTA
11b	ATGTC–TGCTTTTATACTA–ACCGG	1	62-74	reverse (BS)	CTA
12a	CGGTT–AGTATAAAAGCAG–ACATT	1	63-75	reverse	CAG
12b	TTTGG–TGCATAAAATGTC–TGCTT	1	75-87	reverse (BS)	GTC
13a	AAAGC–AGACATTTTATGC–ACCAA	1	74-86	reverse	TGC
13b	TTTAT–GCACCAAAAGAGA–ACTGC	3	85-97	forward	GCA

In **Table 4.3**, all binding sites of NV1087 on the HPV16 (7662-122) fragment are presented. Considering the best fit to the affinity cleavage patterns and the lowest mismatch number of the DNA sequence according to the binding rules, the most probable orientations of the polyamide NV1087 are the ones presented in **Table 4.4**.

Table 4.4. The most probable binding sequences of NV1087 on HPV16 (7662-122) 365 bp sequence. Blue letters indicate the TMG binding site and the highlighted guanine G indicate imidazole binding site

AC site	Sequence 5'-3' (XXXXWWWWWWWWGW)	Mismatch type	Position	Binding fashion	KdHill (nM)	KaHill (nM^{-1})
0b	**AAACA-AGCCAAAAATATG-TGCCT**	2	7704-7726	reverse (BS)	0.76	1.3
0c	**CACAT-ATTTTTGGCTTGT-TTTAA**	3	7707-7729	forward	0.76	1.3
1a	**TGGCT-TGTTTTAACTAAC-CTAAT**	1	7717-7729	reverse	1.32	0.8
1'a	**TAACC-TAATTGCATATTT-GGCAT**	3	7731-7743	forward	1.32	0.8
1'a'	**ATGCC-AAATATGCAATTA-GGTTA**	3	7731-7743	reverse (BS)	1.32	0.8
2	**GCCTT-AGTTGTTTAAACC-TTATA**	1	7751-7763	reverse (BS)	2.79	0.4
2'a	**TTGGC-CTTAGAAGTTTAA-ACCTT**	3	7754-7766	forward (BS)	3.9	0.3
2'a'	**AAGGT-TTAAACTTCTAAG-GCCAA**	3	7754-7766	reverse	3.9	0.3
3b	**TTACA-AATGAACAATGTA-TGACT**	3	7818-7830	reverse (BS)	5.9	0.2
3b'	**AGTCA-TACATTGTTCATT-TGTAA**	3	7818-7830	forward	5.9	0.2
4	**GTGCA-GTTTTACAAATGA-ACAAT**	1	7826-7839	forward (BS)	2.13	0.5
5a	**GCAAC-TTATATAATAATA-CTAAA**	1	7888-7890	forward	3.64	0.3
5a'	**TTTAG-TATTATTATATAA-GTTGC**	1	7888-7890	reverse (BS)	3.64	0.3
6	**TATTA-TTATATAAGTTGC-TTGTA**	2	7883-7895	forward (BS)	1.25	0.8
7a	**TAGTA-TTATTATATAAGT-TGCTT**	0	7886-7998	forward (BS)	1.12	0.9
7b	**GCAAC-TTATATAATAATA-CTAAA**	1	7888-7890	reverse	1.12	0.9
7b'	**TTTAG-TATTATTATATAA-GTTGC**	1	7888-7890	forward (BS)	1.12	0.9
8a	**TATAT-AATAATACTAAAC-TACAA**	1	7894-2	reverse	nd	nd
8a'	**TTGTA-GTTTAGTATTATT-TACAA**	1	7894-2	forward (BS)	nd	nd
8b	**ATACA-TGAATTATTGTAG-TTTAG**	1	2-14	reverse (BS)	nd	nd
9a	**TACAT-GAATTATTGTAGT-TTAGT**	1	1-13	forward (BS)	2.32	0.4
9b	**CCCTT-AGTTTTATACATG-AATTA**	1	12-25	reverse (BS)	2.32	0.4
10	**ATTCA-TGTATAAAACTAA-GGGCG**	1	15-27	reverse	1.6	0.6
11	**ACCGG-TTAGTATAAAAGC-AGACA**	2	61-73	forward	nd	nd
11b	**ATGTC-TGCTTTTATACTA-ACCGG**	1	62-74	reverse (BS)	nd	nd
12a	**CGGTT-AGTATAAAAGCAG-ACATT**	1	63-75	reverse	nd	nd
12b	**TTTGG-TGCATAAAATGTC-TGCTT**	1	75-87	reverse (BS)	nd	nd
13a	**AAAGC-AGACATTTTATGC-ACCAA**	1	74-86	reverse	nd	nd
13b	**TTTAT-GCACCAAAAGAGA-ACTGC**	3	85-97	forward	nd	nd

.

The mapping study has shown that the TMG-AHP NV1087 binds with nanomolar affinity to the expected binding sites (0 mismatches), as well as to other binding sites (one,

two, three mismatches), with comparable Kd values in agreement with previous findings for large hairpin polyamides studied by our group.

To determine how the DNA sequences where the polyamide tail binds, can influence the binding affinity the most probable orientations of the polyamide (**Table 4.4**), was taken into consideration. It was found that the presence of G in the tail binding sequence increases the Kd values. As an example, for forward binding 1 mismatch, compare AC site 7b' with 4 and 9a. The Kd value increase from 1.12 nM to 2.13 respective 2.32 at the sites 4 and 9a. The same trend has been observed in the case of reverse binding. For reverse binding 1 mismatch comparing AC site 7b, 1a, and 10 with 9b, or for the 3 mismatch sites comparing 1'a' with 2'a' and 3b, the Kd values have increased (**Figure 4.6**).

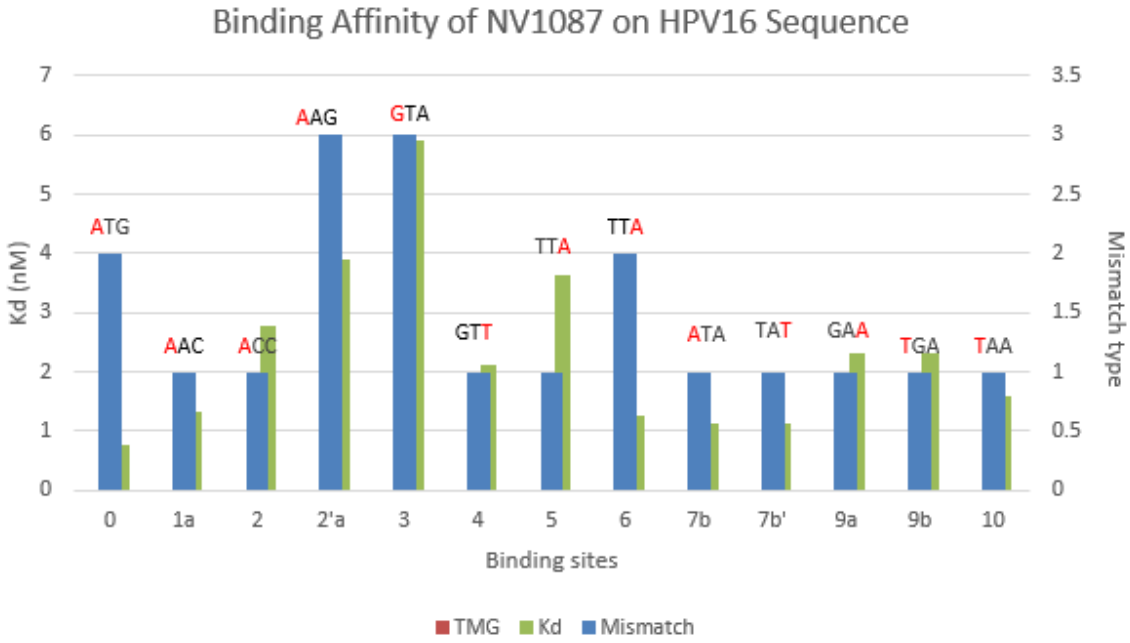


Figure 4.6. Binding affinities of NV1087 on HPV16 LCR sequence

The probabilities of polyamide to bind in forward or reverse orientation had also been analyzed**.** Binding of NV1087 in forward fashion (**Figure 4.7** WebLogo 3.4), have shown a preference for A and T at the TMG binding site. The polyamide can tolerate mismatch nucleotides at almost every position. The middle of the fragment, which contains alternative β/Py is extremely flexible allowing for increased mismatch possibilities.

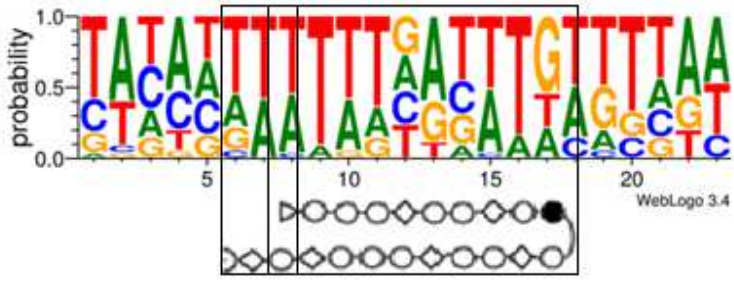


Figure 4.7. DNA binding sequence for forward binding of NV1087 on HPV16 LCR fragment. NV1087 in forward fashion have shown preference for A and T at the TMG binding site. The middle of the fragment, which contains alternative β/Py is extremely flexible allowing for increased mismatch possibilities.

Binding of NV1087 in reverse fashion (**Figure 4.8** WebLogo 3.4**)**, have shown a preference for A and T at the TMG binding site. Although the polyamide can tolerate mismatch nucleotides at almost every position, A and T are preferred through the entire binding pattern, the only exception being the imidazole binding site which favors G.

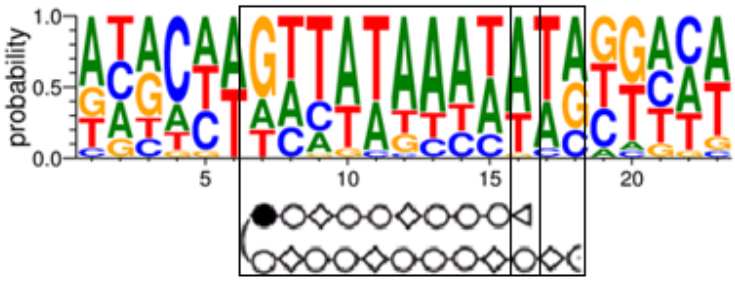


Figure 4.8. DNA binding sequence for reverse binding of NV1087 on HPV16 LCR fragment**.**

NV1087 in reverse fashion have shown preference for A and T at the TMG binding site. Although the polyamide can tolerate mismatch nucleotides at almost every position, A and T are preferred through the entire binding pattern, only exception being the imidazole binding site.

4.4.3 Binding sites of NV1078 on the LCR region of HPV18 (7647-157). Comparison of the binding profiles of NV1078 and NV1042 on the LCR region of HPV18 (7647-157)

To better understand the differences between TMG-AHP and the leading antiviral compound NV1042 (PA 25) (5, 10) synthesized by our group, the binding profiles of NV1078 in the LCR region of HPV18 (7647-157) has been studied. Based on affinity cleavage data, the binding sites for NV1078 (MAP ECF3109LCR HPV18) had been compared with the binding sites for NV1042 (CHC1139).

Figure 4.9 illustrates the affinity cleavage sites of NV1078 (A) and NV1042 (B) on the same DNA sequence. The box shows the binding sites #7, #8, #8' and #9 of NV1078 HPV18 LCR (Map (ECF3109) corresponding to the sites #16, #17, #18, #19 of NV1042 on the same HPV18 LCR sequence. The underlined region is a G/C reach region where the presence of polyamides is weak.

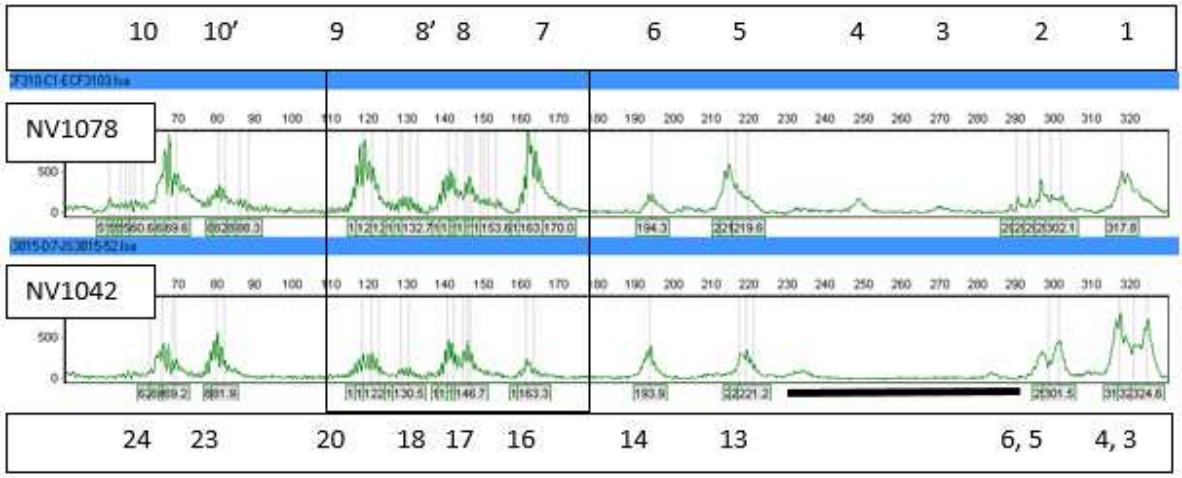


Figure 4.9. Sample electropherogram (bottom strand, Hex dye) illustrating the enhanced affinity cleavage sites of NV1078 (A) comparing with NV1042 (B) on the same HPV18 LCR DNA sequence. The box shows the binding sites #7, #8, #8' and #9 (Map ECF3109) corresponding to the sites #16, #17, #18, #19 (Map NV1042) in the region that includes E1 and E2 proteins binding sites.

As a general overview, an enhancement at four cleavage sites for TMG polyamide can be seen. The region in the black box is an A/T reach region (contains about 14% G). In this area, a highly enhanced cleavage pattern for site #9 compared with #20 and site #7 compared with #16, can be seen. At the position #9 NV1078, most probably binds in a 1mm forward fashion while NV1042 in 3mm forward fashion. For the 7-16 spot the 0 mm

reverse binding has clearly enhanced the cleavage pattern comparing with 2mm, or 3 mm binding possibility of NV1042.

4.5 Discussion

4.5.1 The binding patterns of polyamides to a certain DNA sequence are not completely predictable

Of the presented data and literature, the hypothesis may be derived that the NV1078 polyamide can exhibit the following interactions with dsDNA **(Figure 10**):

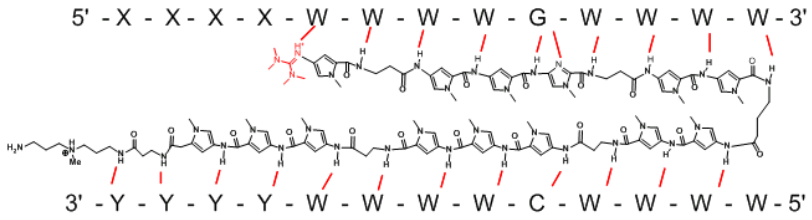


Figure 4.10. Hydrogen bonds formed by NV1078 polyamide

The polyamide NV1078 may form hydrogen bonds with the bases in the minor groove of ds DNA. From a total of 22 intermolecular hydrogen bonds, 19 are formed between the amide hydrogen atoms and the N3 of adenine or O2 of thymine and cytosine (hydrogen bond acceptors) that face the minor groove (denoted in red). Furthermore, one hydrogen bond between N3 of imidazole and a hypothetical C2-NH-H of guanine (red), and two more hydrogen bounds created by the amide hydrogen's of the polyamide tail, according to the bifurcated hydrogen bonds model demonstrated by Dervan, to the N3 of adenine and/or O2 of thymine and cytosine (120). According to this model the tail of the AHPs sits in the middle of the minor groove and forms hydrogen bonds with 2 DNA strands (not marked). An additional third hydrogen bond between TMG and N3 of adenine and/or O2 of thymine is possible.

Positive charges are present at the polyamide C-terminus (Ta) as well as at the N-terminus (TMG). The positive charge at the N-terminus, unique for this class of polyamides, facilitates electrostatic interactions between the tetramethylguanidine and the negatively charged phosphodiester backbone of DNA in addition to hydrogen bond formation with the base pairs in the minor groove (121). It may be postulated that the positive charge at the TMG group has a significant role in the enhanced activity of AHPs. The triad Arg-Pro-Arg at the tail end of an 8 ring polyamide had been shown to compete successfully for the phosphate backbone and displace the GCN4 protein from the major groove (122). In analogy with the guanidinium group of arginine, the TMG may be able to disturb the critical binding of E1 helicase in the major groove.

Also, the polyamides can establish Van der Waals interactions with the groove walls: C2-H of adenine favors van der Waals intermolecular contacts with Py H3 protons. Previous docking studies show that the van der Walls stacking interactions in the minor groove are more favorable for Py/Py than Py/β configurations (123). Therefore, since NV1078 incorporates Py/Py units, favorable van der Waals interactions are expected. The NV1087 incorporate five Py/β units, consequently, this molecule is less prone to take advantage of van der Walls interactions while bound in the minor groove.

The exocyclic amino group of G will impede PA fitting in the minor groove (121) since deep penetration of the minor groove is required. The results revealed that although, the polyamide tolerates mismatch nucleotides well, the presence of the bulky G base is less preferred at the TMG binding sites (**Table 4.2 and 4.4).**

Hydrophobic contacts when polyamides bind in the minor groove were not studied for NV1078 or NV1087 and can be just anticipated. In the case of 8 ring PAs, the hydrophobic contacts are diminished for Py to β substitution (123). If the same rule applies to long polyamides, then for NV1087 (two Py/β pairs) the hydrophobic contacts may be diminished. The NV1078 presents two β/β pairs, therefore stronger hydrophobic contacts are expected. Since the presence of water weakens electrostatic interactions, it is expected that NV1087 have enhanced electrostatic interactions with DNA comparing with NV1078. Positive binding entropies are expected since the release of 4 counterions upon binding is associated with increased binding entropy.

4.5.2 TMG-AHP polyamides display similar Kd values as other related antiviral polyamides but significantly lower IC50 and IC90 values, therefore improved antiviral activity.

During the experimental work, the thermodynamic parameter, equilibrium dissociation constant (Kd), was calculated for a closed system. In these conditions the ligand (i.e. TMG-AHP polyamide), and the receptor (i.e. dsDNA) total concentrations were constant, only the concentration of the free and bound species changes over time until the system reaches equilibrium. The dissociation constant values do not demonstrate a severe affinity decrease of the TMG-AHPs at less ideal (two or three mismatches) **(Table 4.2, 4.4)** sites in agreement with previous findings for large hairpin polyamides studied by our group (11, 86). Moreover, the Kds have shown insignificant variation comparing with the Kd values of polyamides less efficient in eliminating the viral load from the same type of cells (10, 11)

Nonetheless, the IC50 and IC90 values were determined *in vivo* (open system) and showed dramatic improvement in the ability to eliminate the viral DNA from keratinocytes comparing with less effective compounds (**Table 1.1**: IC50, IC90 for different PAs).

These apparently contradictory results can be explained if we consider that in an open system the concentration of the polyamide at the target site can change over time due to metabolic processes or other elimination mechanisms (e.g. drug resistance). In natural systems, the effect of a ligand depends not only on the selectivity of the ligand and its binding properties but also, on the ligand residence time, the amount of time that a ligand resides on a given receptor after binding (124), at the certain sites. A longer residence time may translate into a persistent biological effect (125). The IC50 and IC90 values obtained in open biological systems integrate in a better way the factors that contribute to the overall response of the cells to the TMG-AHP action(s).

4.5.3 Viral DNA is capable of adjusting its parameters (intrinsic bending and minor groove width) to allow binding of TMG-AH polyamides. By interfering with the formation of E1/E2 complex TMG-AHPs block viral DNA synthesis by obstructing the replication process at the start.

The early expressed viral proteins E1 and E2 are essential and sufficient for initiation of HPV DNA replication (126, 127). All other factors required for replication of episomal DNA are obtained from the host cell. E1 protein contains a DNA- binding domain (DBD) and an ATP-ase / helicase domain (128). E1 binds DNA, as a dimer through its DBD, along with E2 protein at the viral origin of replication (126) leading to a final double hexameric helicase, assembled by E2 protein.

The ability of the E1DBD to bind DNA as a dimer is required for the assembly of E1 as a replication- competent double-hexamer (DH) at the ORI (126). Also, the assembly recruits cellular replication factors that will replicate the viral genome. Previous studies showed that polyamides that overlap completely (or partially) with the protein binding sites are effective at displacing the DNA binding protein by impeding essential amino acid-base pair contacts(129). In the cited case, when tandem polyamide binds in the minor-groove the intrinsic bending of the DNA will be altered, therefore E2 DBD won't be able to bind in the major groove of the DNA (129).

From side by side comparison of the AC binding profiles of NV1087 and NV1078 on the same HPV 16 DNA (**Figure 4.10**) it was concluded that the TMG-AHP polyamides have a preferred binding region where AC cleavage patterns stand out (7880-30). Both

polyamides bind in a cluster mode on the ORI of HPV16 that includes the sites of E1 and E2 proteins, sites that are critical in the process of the initiation of DNA replication. The box in **Figure 4.11** shows a cluster of polyamide binding sites.

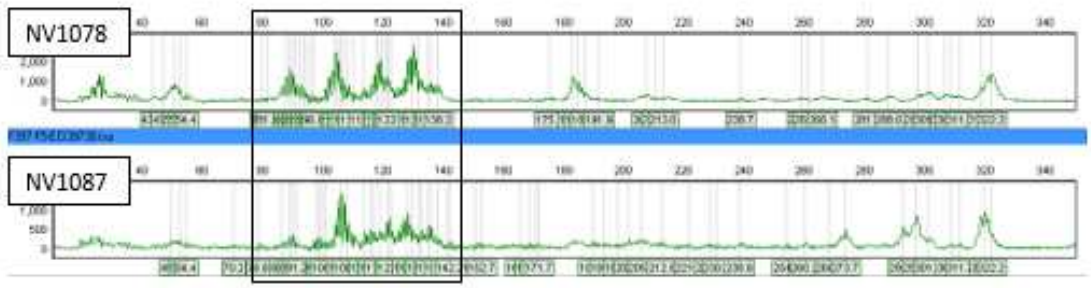


Figure 4.11. Mapping studies performed with NV1078 and NV1087 polyamides on a 365 bp HPV16 DNA sequence indicate that the polyamides bind with high affinity in the region adjacent and overlapping with the E1BS**.The preferred binding sites where AC cleavage patterns stand out are presented in box.**

Since polyamide NV1078 is even more active against HPV18 than HPV16 its binding profile on HPV18 LCR (7647-157) sequence was also studied. Although the polyamide binds in many different sites, it has a noticeable preference for the sequence (7841-43). Moreover, as a result of the comparison of the binding profiles of a TMG-AHP polyamide, NV1078, and the leading antiviral compound PA25 (NV1042), the enhanced affinity cleavage region of NV1078 **Figure 4.9** convinces us that NV1078 presents superior binding proprieties regarding HPV episomes than NV1042.

4.6 Conclusions

Throughout our efforts to discover compounds useful in eliminating HPV, during the maintenance stage of its life cycle, TMG- AHPs were developed and showed potential as possible (antiviral compounds) drugs for the treatment of preexisting and future HPV infections.

The antiviral activity of the compounds was demonstrated by proving the decrease of viral DNA levels in human keratinocytes in cells that maintain HPV16, HPV18, and HPV31 episomal HPV DNA, without integration of viral DNA (T. Edwards, C. Fischer). Two, NV1078 and NV1087, asymmetric TMG polyamides were found to have excellent antiviral activity against all three HPV strains HPV16, HPV18, and HPV31. Similar activity against different HPV strains is a unique property, (only one more compound the NV1087 isomer, NV1111, was found to have similar proprieties) most of our compounds distinguish between different strains.

Evidence from biophysical studies supported the primary step of the underlying mechanism of action. The large polyamides NV1087 and NV1078, which recognize 13 base pairs, tolerate well one, two, and three mismatch binding sites. The high tolerance of mismatched sequences does not affect the biological activity, suggesting that other factors besides sequence specificity, such as the presence of TMG group and possibly the residence time, may be crucial for explaining the unique antiviral proprieties of the TMG-AHPs.

Although, the role of tetramethylguanidine groups in the cellular and nuclear uptake of TMG-AHPs are not completely understood, our findings are in concordance with previous studies that have shown that guanidilated compounds improve the ability of compounds for cellular uptake (130, 131).

Based on the collected data, it may be concluded that the binding of TMG-AHPs in a cluster in the region adjacent and overlapping with the E1, E2 binding sites has a significant contribution to the overall biological activity.

The intensive binding in the minor groove of dsDNA (sequence 7880-30 of HPV16 and sequence 7841-43 of HPV18) of TMG-AHPs structurally alter the dsDNA. The required adjustments of the DNA to allow binding may cause the inability of E1/E2 to assemble, as a replication competent double-hexamer (DH) at the ORI, obstructing the replication process at the start.

Moreover, it is rational to predict that the intensive binding in the region mentioned above may cause ssDNA exposure or double strand breaks within HPV DNA (6) and the

activation of DNA damage response (DDR) pathways (6). Multiple binding sites induce DNA nicks which activates ATR. Our findings complement the notion that the PA binding to negatively supercoiled HPV episomes activate the DNA damage response (DDR) pathways (6). Activation of DDR genes eliminates the viral episomes without having substantial effects on cell growth or apoptosis (6).

The elimination of viral DNA during the maintenance stage it is an effective strategy for the development of antiviral compounds. TMG-AHPs have the potential to be useful for the treatment of preexisting and future HPV infections.

Chapter 5
Differential DDR Gene Expression Profile of W12E Cells After Treatment with Active Anti-HPV Hairpin Polyamides

5.1 Abstract

Human papillomaviruses (HPV) infect the basal layer of the epithelium. DNA damage response (DDR) machinery can recognize viral DNA and replicating viral genetic material as foreign DNA. Low levels of DDR components are expressed in episome bearing cells. Once DDR senses DNA damage, it will recruit an elaborate network of transducers and effectors to repair the damage or initiate apoptosis. In theory viral propagation could be restricted. However, HPV developed ways to take advantage of the DDR machinery and coexist with it. Moreover, HPV uses DDR elements for its replication.

It had been shown that in addition to induction of DDR genes by HPV, anti-HPV polyamide treatment of episome-bearing cells also causes the alteration of some DDR gene expression levels (6, 117). Gene products that act as both enhancers and repressors of anti-HPV activity of polyamide PA25 were identified (6). The goal of the present study is to compare (and find the common features of) the effects of the polyamides NV1078 and NV1042 on genome regulation, using NV1011 as a control like in the previous study (6).

The polyamides' effects on DDR gene expression were analyzed after treatment of W12E cells with the polyamides mentioned above. W12E cells are undifferentiated cells, harboring 100-500 viral episomes per cell. Therefore, they were used as model for gene expression studies of cells mimicking the maintenance phase of the HPV lifecycle. Total RNA was extracted and converted to cDNA. To measure the transcript abundance, seven DDR genes and five reference (housekeeping) genes were analyzed in parallel.

The goal of this project was to have an initial feedback before completing analysis of RNASeq data on polyamide effects. QPCR studies monitored expression of the seven DDR genes (Table 5). These genes are the Transducers (Ataxia-telangiectasia mutated kinase, ATM and Ataxia telangiectasia and Rad3-related protein, ATR (8)) and first line effectors (CHEK1, CHEK2) in the homologous recombination (HR) pathway. Also, RAD1, RAD9A and TP53 were included in the study (**Table 5.5** and Chapter 1).

The gene expression analysis is expected to lead to the identification of genes relevant to the polyamide mechanism of action and provide insight into the complex

regulatory DDR networks. Additional QPCR work will be done to follow up and test the RNASeq results.

5.2 Introduction. DDR response in cells infected with HPV16

The virus HPV has a complex relationship with the infected host cells. In its intricate association with the DDR, the virus has to accomplish two important tasks to replicate in the host cell. First, the virus has to evade the DDR, which senses foreign DNA and DNA-bound viral proteins as a threat to genomic integrity. Moreover, the virus has to use elements of the DDR to complete its own deficient transcriptional machinery.

Suppression of DNA damage response is achieved by viral proteins E6 and E7, which act as direct or indirect antagonists of different DDR components. E6 and E7 interact with DDR components in addition to their other functions. The principal function of E6 and E7 proteins is to maintain the keratinocytes in a proliferative state. This function is achieved in part when E7 inactivates Retinoblastoma protein (pRB) and E6 targets the protein p53 for proteasomal degradation, therefore preventing the cell from undergoing apoptosis. The two oncogenic viral proteins E6 and E7 work cooperatively to avoid downstream signaling, therefore undermining the DDR response in undifferentiated and differentiated host cells (9, 48, 113, 114).

Among numerous functions of viral oncogene E7 (see Introduction, Chapter 1 (1)) is its ability to abrogate the ATM-ATR pathway-induced checkpoints. E7 causes the degradation on claspin, a key regulator of ATR-CHK1 pathway, therefore promoting mitotic entry in the presence of foreign DNA (1).

Similarly to E7, E6 creates interactions with a multitude of cellular proteins. Among those, the relation with p53 is important from a DDR point of view, since activation of the ATM and ATR signaling cascades leads to phosphorylation of the p53 tumor protein (Figure 1). The oncoprotein E6 antagonizes the function of p53, forming an E6-E6 associated protein (E6AP)-p53 and targeting p53 for proteasomal degradation (132). In this way, E6 eliminates the p53-dependent apoptosis (1). Furthermore, E6 interferes with p53 at a genomic level. The interaction of E6 with histone acetyltransferases p300, CREB binding protein, and ADA3 prevents p53 acetylation, and through this, inhibits the transcription of p53 responsive genes (1).

Manipulation of DDR promotes the viral lifecycle. HPV utilizes both the ATM (9, 19) and ATR (5, 133) pathway for replication and vegetative amplification. The ATM pathway components have been shown to be critical regulators of HPV in the amplification phase of differentiated keratinocytes. Nevertheless, the pathway is activated in

undifferentiated cells as well (9). Multiple viral proteins can activate ATM pathway and work independently or cooperatively to promote viral replication as shown in **Table5.1**.

Table 5.1. Viral proteins that activate the DDR pathways to replicate the viral genome by taking advantage of host repair proteins.

Homologus Recombination Pathway, HRP	Viral protein	Effect on cells
ATM	E1 helicase	ds DNA breaks
	E2/ BRD4	Chromatin remodeling; replication centers inducer
	E7	G2/M arrest; CHK1, CHK2 activation; STAT-5 activation
ATR	E1	CHK1 activation

Viral helicase E1 is the major activator of ATM in undifferentiated cells, expressing a complete viral (HPV11, 16, 18, 31) genome (34). Nuclear accumulation of E1, leading to nonspecific DNA unwinding, may result in the induction of undesired DNA breaks in the host genome (133), and consequently, induction of ATM pathways and arrests in the S phase, allowing viral replication. E1 helicase activity is attenuated by complex formation with E2 or by nuclear export (34). Both events, complex formation and nuclear export, are paramount, as they will keep low levels of E1 in undifferentiated cells, desirable to avoid genomic instability generated by ds DNA breaks which could favor integration of the viral genome (34).

The E2 viral protein ameliorates E1 effects but also activates the ATM pathway following an entirely different mechanism. E2 binds to BRD4, which tethers the episomes (HPV 16, 18) to the host chromatin (35) to ensure replication in the initial amplification phase and the maintenance phase. This tethering mechanism ensures that the low copy of the viral genome is retained in the nucleus, and it is partitioned to daughter cells. BRD4 interacts with chromatin and has a significant role in the formation of the viral replication centers(36, 41). Therefore, E2/BRD4 complex is an intermediate component that attracts cellular proteins to viral replication centers to initiate the replication of viral DNA. ATM homologous recombination elements pATM, pH2AX, and pNBS1 are colocalized with BRD4/E2 proteins in the replication centers (43). Replication centers were spotted and detected by confocal imaging in differentiated and undifferentiated (HFK-HPV16E1/E2) cells as well (43).

Viral proteins, in particular oncoproteins E6 and E7, which drive an aberrant proliferation of the cells, provoke a DDR response used by HPV for viral replication. The viral E7 protein has been shown to activate the ATM pathway (1) and its downstream

effector CHK2 in differentiated as well as undifferentiated cells harboring the HPV31 virus (9). Phosphorylation of ATM and its downstream effector CHK2, as well as CHK1, was also observed in cells harboring HPV18 (134).

Activation of the ATM response in cells induces a G2/M arrest that provides an environment favorable to productive viral replication (9). It is important to notice that NBS1 (Nijmegen breakage syndrome 1, also known as nibrin), part of the Mre11–Rad50–Nbs1 MRN sensor complex which promotes monomerization and autophosphorylation of the ATM kinase, is only activated in differentiated cells (1). It was also shown that the formation of viral replication centers in differentiating cells is dependent upon ATM activity, for pATM, pCHK2, pNBS1, and γH2AX are present in the amplification centers (9).

In addition, E7 forms complexes with phosphorylated ATM proteins (113), but not the unphosphorylated form of ATM (9). This interaction may allow ATM to recognize novel substrates (9). In the same paper, it was demonstrated that although the ATM pathway is activated in undifferentiated cells, it is not essential for stable maintenance of viral genome (9). ATM activation is necessary for viral genome amplification in differentiating cells (9).

E7 also activates the cytosolic transcription factor STAT-5 (15). The phosphorylated STAT-5 induces increased levels of PPARγ (a nuclear transcriptional factor) in HPV31 positive cells, which (possibly with the help of additional intermediary proteins) activates the ATM DNA damage pathway (8) leading to genome amplification (15).

Hitherto, it is unclear how ATM factors specifically contribute to replication. It was hypothesized that the primary function of ATM factors is to help resolve concatemers of viral DNA, or that viral replication is mediated by DNA repair polymerases and ligases (15).

Complementing ATM factors, HPV also utilizes the ATR pathway for replication (133) (5, 6). Using U2OS cells transfected with the E1 protein expression constructs, HPV16 and HPV18 were shown to activate the phosphorylation of ATR and CHK1 in undifferentiated and differentiated cells, and members of the ATR pathway localized to HPV replication centers (133).

Inhibition of ATR pathway components reduces the number of HPV episomes that are stably maintained in W12 keratinocyte cell lines, suggesting a role of the ATR pathway in HPV16 replication in undifferentiated cells (5, 6).

Moreover, the previous DDR gene study (6) showed that 11 genes were employed by the virus to maintain the episome levels in cells in natural conditions during the

maintenance phase (6). Among those eleven genes that augment the HPV16 episome levels in W12E cells, three are members of the DDR pathways, ATM, RTEL1 (an HR suppressor), and RUVBL2 (a helicase essential for DBS repair).

The ATM transducer and CHEK2 effector play an important stabilizing role for the HPV episome that does not require kinase activity (5, 6); the ATR transducer and CHEK1 effector stabilize the HPV, lessening the replication fork effects generated by E1 and host replication factors (which is thought to be intrinsically unstable (117).

Large PA's alter the DDR response initiated by HPV episomes

As shown, the HPV episomal DNA by itself mediates a DDR response in cells infected by the virus not just in the amplification phase, but also in the maintenance phase. The extent of DDR involvement is directly proportional to the HPV episome numbers present in cells.

Previous studies have shown that the large PA25 induces a DDR response in cells infected with the HPV virus (5, 6, 117). The significant viral loss registered when infected cells were treated with an active polyamide is a consequence of the degradation of the viral episome, not just the inhibition of viral DNA replication (117). No such response was observed when cells that contained already integrated HPV were treated with the same polyamide (117). Also, when treated with PA 25 the viral episome from infected cells was cleared without integration in the host chromosomes (6, 117). Simultaneously, DDR genes were upregulated or downregulated as a result of PA25 binding in the minor groove of episomal DNA (6, 117).

The activation of DDR pathways and the use of their elements for viral maintenance was evidenced in W12E cells using siRNA-mediated knockdown of the DDR genes. Knockdown or inhibition of the transducer genes ATM and ATR and other proteins in the DDR pathway lead to the decrease of episome numbers. A drastic 40%–50% reduction in the HPV16 copy number in W12 cells was registered (5, 6, 8).

The degradation of the viral episome was supported by the Aphidicolin action on viral episomes. Aphidicolin, a selective α-polymerase inhibitor with antimitotic proprieties, is known to activate a DDR (ATR) response and cause significant episome loss when used for the treatment of cells infected with viral episomes (5). Meanwhile, no significant changes were registered when Aphidicolin was used on infected cells already treated with polyamides. The episomes already degraded by polyamides are uninvolved in the response to Aphidicolin treatment. The goal of the present study is to find the similarities

and differences in the way of action of the PA's presented in **Table 5.2** by the RT-QPCR method.

Table 2. Polyamides used in RT-qPCR studies. The IC50 and IC90 values measured for W12E cells were determined by T. G. Edwards in the Fisher lab, unpublished data.

PA	PA sequence	IC50 [μM]	IC90 [μM]
NV1011 (PA11)	ImPPIm-γ-PPPP-β-Dp	NOT ACTIVE	NOT ACTIVE
NV1042 (PA25)	ImPP-β-PPIm-β-PP-γ-PPβPPPβPPP-β-Ta	0.036	0.351
NV1078 (TMG-AHP)	TMG-PβPPIm-β-PP-γ-PP-β-PPP-β-PPP-β-Ta	0.046	0.307

5.3 Materials and Methods

5.3.1 Cell Culture; W12E Keratinocyte Cells (clone #20850)

The Keratinocyte cell line (W12E clone #20850) containing HPV16 genomes as episomes (10, 135, 136) was maintained in monolayers in E-media (3 parts DMEM with high glucose, and one part F12 media) in the presence of mitomycin C treated 3T3 fibroblast feeder cells.

The 3T3 cell line is a fibroblast cell culture (ATCC NIH/3T3) established from the disaggregated tissue of an embryonic albino Swiss mouse. The 3T3 feeder cells used in this experiment were received from the lab of Paul Lambert on 25 Dec 2009, and were stored in liquid nitrogen (J23T3 GH2012). In culture, 3T3 cells can form confluent monolayers with contact inhibited cell motility. The 3T3 cells could grow indefinitely and do not instigate tumor growth. The 3T3 fibroblasts (Figure 1) are used as feeder cells for keratinocyte cells (W12E) because they secrete growth factors favorable for keratinocytes. Thus, they enhance the growth of human keratinocytes and shorten the cultivation time of said keratinocytes. The 3T3 fibroblasts have been maintained in a monolayer culture to 80% confluence in growth media (DMEMF10 - Dulbecco's Modified Eagle's Medium supplemented with 10% BCS and 1% pen-strep). The cells were grown in a 37℃, 5% CO_2, humidified incubator and passaged to 1:10 every three days based on previously published protocols (137) (Cell culture, 3T3 Protocol: ECF3133). The 3T3 cells were incubated with mitomycin for 2 hours at 37°C in a 5% CO_2 humidifier incubator.

Mitomycin C from Streptomyces caespitosus, which inhibits the proliferation of 3T3 cells, was added to the 3T3 cells before being used as feeders for keratinocytes, to a final concentration of 8 μg/mL (200 μL 50X mitomycin C per 10 mL E-medium (138). Mitomycin C **(Figure 1**) treatment was utilized to stop the division of the 3T3 cells. Mitomycin C reacts

covalently with DNA, crosslinking complementary strands and inhibiting DNA replication (137).

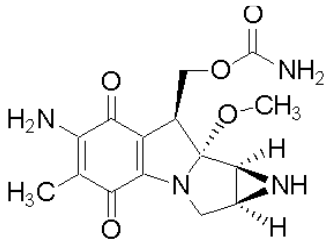


Figure 5.1. Mitomycin C structure

The W12 cell line which is representative of a premalignant HPV16 induced tumor was derived from a low-grade cervical lesion and retains the HPV16 genome as an episome (139). W12E cells contains a frameshift mutation in the E1 gene (140) with the accession number on GenBank as AF125673.The W12E keratinocyte cells (Figure 1.B) harbor extrachromosomal HPV16 genomes with HPV16 copy numbers ranging from 300 to 500 copies per cell (5, 10, 135). In the basal layer, the initial stage of infection, the episome number is considered to have little copy number variation, and it is considered stable.

The W12E clone #20850 cells, maintained at early passage (from P #13 to P #17) have been cultured in monolayers on top of a mitomycin treated 3T3 feeder layer. Clone 20863 W12E cells are known to be differentiated keratinocytes (141)

High passage number was avoided since W12 cells grown in long-term cultures have the ability to differentiate (142). The cells were maintained at 37℃ in a humidifier incubator with 5% CO_2 in a medium consisting of E-media (3 parts DMEM w/ high glucose, pyruvate and 1 part F12 media) supplemented with 0.4 µg/ml Hydrocortisone, 0.1 nM or 10 ng/ml Cholera toxin, 5 µg/ml Insulin, 24 µg/ml Adenine, 10 ng/ml Epidermal Growth Factor, 5 µg/ml Transferrin, 20 pM 3,3,5-Triiodo-thyronine (T3), 1% Penicillin/Streptomycin, and 5% Fetal Bovine Serum (5). The media components were filtered using syringe filters (Millex-GP 0.22 µm) and aliquoted. The prepared E-Complete and E-Incomplete (contains no EGF) cell media were sterilized using Nalgene 0.2 PES filters. The E-Incomplete media was used for culturing keratinocyte and 3T3 feeder cells the first day after passaging. The W12E cells have been passaged at 70% confluence using a split ratio of 1:10 (5) (Epithelial cell culture protocol: ECF3134).

The W12E ($1x10^5$) cells were seeded on the top of $1x10^5$ mitomycin treated J23T3 fibroblast cells in 6 well plates (Sigma-Aldrich, CLS3516-50EA). The working volume of a six-well plate is 1.9-2.9 ml, the growth area is 9.5 cm^2, and the expected cell yield varies

between 9.5 x 10^5 to 10^6 cells. The cells were plated at a density of 1x10^5 cells/ml. **Table 5.3** includes the passage numbers of the cells used for obtaining the biological replicates.

Table 5.3. The W12E and the 3T3 passage numbers used for obtaining 5 series of biological replicates. Also, the RNA extraction dates and the notebook page numbers are included.

	Series 1	Series 2	Series 3	Series 4	Series 5
W12E cells	P13	P14	P15	P16*	P17
J23T3 cells Set I/Set II	P22/P18	P23/P19	P24/P20	P25/P21	P26/P22
Total RNA extraction Set I of PAs	11/7/15 ECF3153	11/10/15 ECF3157	11/13/15 ECF3165	11/16/15 ECF3168	11/19/15 ECF3171
Total RNA extraction Set II of PAs	12/1/15 ECF4007	12/5/15 ECF4011	12/8/15 ECF4015	12/11/16 ECF4021	12/14/15 ECF4022

Set I polyamides refers to polyamides NV1078, NV1087(6102A), NV1042 (KJK6006) (6) NV1028 (GDH1130) NV1111(KA2092) . Set II polyamides refer to NV1011 (KJK1083D 88%), NV1113(KJK6169), and NV1115 (KJK6170)

*According to Garner-Hamrick (135) P16 W12E will have about 1273 HPV copies per cell, while according to Edwards (5) P16 will have 300-500 copies per cell.

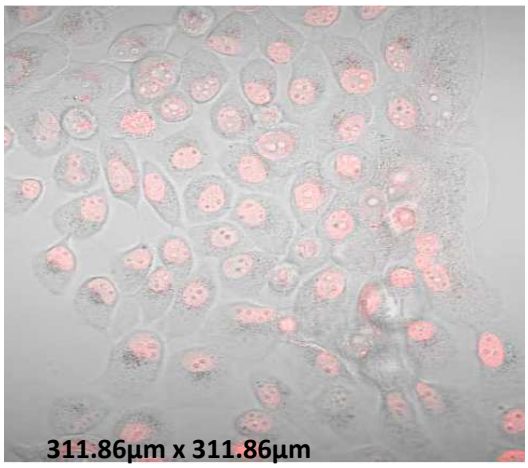


Figure 5.2. Confocal image of live W12E cells incubated with DRAQ5 for 10 min. λ= 639 nm, 40X, **(ECF 1130.6).** The W12E are polygonal in shape with relatively regular dimensions and grow attached to a substrate in discrete patches. The Keratinocyte W12E cells form clusters among 3T3 cells. The image was acquired with a Zeiss LSM 700 laser scanning confocal microscope (LSCM).

The cells were plated in 6 well plates at a seeding density of 1x10^5 J23T3 cells and 1x10^5 W12E cells in 2 ml E-media. The number of cells was determined using a Nexcelom Bioscience Cellometer Auto T4 instrument (cell counting chambers plates SD100).

The W12E cells are identified as biohazardous agents as representatives of Risk Group 2 (RG2). The detailed workflow of the cell culture experiments was submitted for approval to the Institutional Biosafety Committee. The experimental work was done in compliance with standards and regulations according to NIH guidance biosafety level 2 (BL2) (143).

5.3.2 Treatment of the cells with active polyamides

The W12E cells have been treated with 7 active polyamides, an inactive polyamide as a negative control **(Table 5.2),** and 0.1% DMSO vehicle as control. The cells were incubated with polyamides on day two for 24 hours. It was assumed that the selected PA's follow the same pathway during uptake within the cells (see introduction).

The IC50 and IC90 values (Chapter 1) which show the effectiveness of PAs clearing the HPV episomes from infected cells were used as guidance to decide the polyamide concentration. The cells had been treated with 0.1 µM PA for 24 h.

Lyophilized aliquots of PAs were resuspended in 20 µL sterile DMSO (100µM). 2 µL of polyamides were added to 1998 µL of E Complete Media on the second day of culturing the cells. The cells were incubated for another day (24 hours precisely). In this way, the final polyamide concentration in the media was 0.1 µM in 0.1% DMSO. The lyophilization solvent was composed of 50% (water + 0.1% TFA), and 50% (Acetonitrile + 0.1%TFA).

5.3.3 Extraction of total RNA

RNA extraction optimization. Before collecting the RNA used for subsequent experiments, experimental trials were used to decide the proper incubation volume and the protocol steps for harvesting the cells and extracting the RNA. It was found that 300xg centrifugation speed for 5 minutes works well when using either an Eppendorf 5424R, a FA-45-24-11, an Eppendorf mini spin, a F-45-12-11, or a Sorval ST16R centrifuge for cell pellet collection. When using QIAshredder and RNeasy spin columns, 12,500g centrifugation speed was used. It was found that 2 ml of cell suspension containing 2.6 x 10^5 cells yielded an RNA concentration of 0.9 µg/ml. This concentration was found proper for downstream experiments. Therefore, the cells were plated in 6 well plates at seeding density of $1x10^5$ J23T3 cells and $1x10^5$ W12E cells in 2 ml E-media. The RNA concentration was measured using Epoch Microplate BioTek spectrophotometer equipped with Gen 5 software. Based on absorbance values, the concentration of RNA was calculated based on the equation

$$C_{RNA} = 44ng/\mu L \text{ X } A_{260} \text{ X dilution factor}$$

RNeasy procedure. The total RNA (viral and genomic) was extracted following the RNeasy Plus Mini Kit, Qiagen 74134, protocol (ECF4008). Through the RNeasy method, all RNA molecules longer than 200 nucleotides were purified. The process provides enrichment for messenger RNA (mRNA) since most RNAs shorter than 200 nucleotides, such as ribosomal RNA (rRNA, makes up ribosomes along with other proteins), and transfer RNA (carries amino acids to mRNA to form polypeptides) are excluded from the purified pool. The goal of the purification process was to capture the mRNA which carries the genetic information of DNA from the nucleus to the cytoplasm for protein synthesis.

Cells grown in a monolayer seeded at a density of $1x10^5$ J23T3 cells and $1x10^5$ W12E cells in 2 ml E-media were trypsinized and resuspended in medium containing fetal bovine serum to inactivate the trypsin. The cells were collected as a cell pellet before lysis. To improve the RNA yield, complete removal of the cell-culture medium was necessary. The culture media can inhibit lysis and also dilute the lysate, disturbing the binding of RNA to the RNeasy membrane.

Cells were disrupted by adding Buffer RLT, a guanidine-thiocyanate–containing buffer. Guanidine thiocyanate (GATC) is an extremely potent chaotropic agent which quickly denatures macromolecules and inactivates nucleases. GATC disrupts any hydrophilic bonds, therefore lysing the cell membrane, linearizing nucleic acids, and at the same time completely destabilizing RNAses. In the lysis buffer, RNA is perfectly protected. The Qiagen RLT buffer composition is confidential, but usually lysis buffers contain guanidine thiocyanate (5M-10M) which is dissolved in Tris-HCl (0.1 M) + 0.5M EDTA + Tween-20 or Triton X-100.

To ensure total inactivation of RNases, beta-mercaptoethanol (β-ME), 1 µL to the 1 ml buffer, was also added to the RTL buffer. An additional step for homogenization of the lysate with QIAshredder, Qiagen 79654 spin columns was introduced. After homogenization, the lysates were transferred to a gDNA eliminator spin column and centrifuged at high speed (12,500g).

To provide appropriate binding conditions to the column membrane, ethanol was added to the flow-through, and the samples were applied to an RNeasy Mini spin column. The RNA binds to the positively charged columns' (silica, porous glass) membrane in the presence of the high salt concentration. Contaminants were efficiently washed away using Buffer RW1 (contains guanidine salt, which removes carbohydrates) and Buffer RPE (wash buffer containing ethanol).

The total RNA bonded to the column membrane was eluted with 60 µL of RNase-free water. The RNA samples are stored at -80 °C as aliquots of: 20 µL (for further RNASeq experiments), 20 µL (for further QPCR experiments), 17 µL (for DDR-QPCR), and 3 µL (for quality and quantity determination, RIN). The first set of samples was used for TruSeq library preparation (Chapter 6), the second set was stored for RNASeq validation experiments (after bioinformatics will communicate us the whole spectrum of genes that were altered), and the third was used to analyze the polyamide's effects on DDR gene expressions for immediate feedback.

5.3.3.1 The quality of total RNA

Assessing the quality of starting material for downstream experiments is very important since high quality RNA is needed to correctly assign the mRNA expression levels. The quality and quantity of the total RNA was determined using the BioAnalyzer Agilent 2100 instrument in compliance with Agilent RNA 6000 NanoKit protocol (5067-1511).

The RNA integrity of total RNA samples purified from W12E cells bearing HPV16 episomes, was tested using 1 µL of the total RNA. In every well of the chip of interconnected microchannels, 1 µL of total RNA was added to a support of 5 µL of gel-dye. The nucleotide fragments have been separated, as they were driven through electrophoretically and the results were registered as an electrophoretic trace.

The electrophoretic trace (electropherogram) provided the RNA integrity number (RIN), the total RNA concentration of the samples, and the ribosomal ratios. RNA integrity number (RIN) is a software tool that estimates the integrity of the RNA samples and assigns an integrity number to eukaryote tRNA samples when the concentrations of the samples are > 25 ng/µL (Agilent Technologies, Palo Alto, CA, USA). To estimate the RIN, the entire electrophoretic trace is considered, not just the ribosomal bands ratio. The ribosomal ratios are frequently used to determine the integrity of the RNA samples. If 28S/18S > 2 or higher, the 28S/18S ratio is an indication that the purified RNA is intact and hasn't degraded.

The electropherograms obtained after the extraction of the RNA which allow for visual inspection of RNA integrity are provided in the supplementary material. The RIN scores and the concentration of the samples are provided in **Table 5.4.**

5.3.4 Reverse Transcription of RNA

After treatment with polyamides (0.1 µM, 24 h in cell growth media), the W12E total RNA was extracted according to the RNeasy procedure and converted to complementary

DNA (cDNA). cDNA was synthesized from 1 µg of DNA-free total RNA in a 20 µL reaction volume using a Maxima First Strand cDNA Synthesis Kit for Reverse Transcription qualitative PCR (RT-qPCR) (K1641 Thermo Fischer Scientific). The kit employs the Maxima Reverse Transcriptase Enzyme, an advanced enzyme derived by in vitro evolution of M-MuLV RT, and oligo $(dT)_{18}$ and random hexamer primers to prime synthesis of first strand cDNA. The oligo $(dT)_{18}$ is specific for mRNA and initiate reverse transcription at the 3' end of the transcript. Random primers are used only in two-step qRT-PCR reactions. The negative control, reverse transcriptase minus (RT-), was performed to evaluate for genomic DNA contamination of the RNA sample. For every biological replicate, 1µg of total RNA was used for reverse transcription.

Table 5.4. RIN, volumes and concentrations of total RNA samples used for reverse transcription corresponding to 1 µg sample.

PA-Replicate	RIN*	Total RNA (ng/µL)	tRNA µg (17µL)[a]	tRNA(µL)
NV1078-1	10	314	5.3	3.2
NV1087-1	10	644	10.9	1.6
NV1042-1	10	352	6.0	2.8
NV1028-1	10	192	3.3	5.2
NV1111-1	10	330	5.6	3.0
0.1%DMSO-1	10	275	4.7	3.6
NV1078-2	10	369	6.3	2.7
NV1087-2	10	690	11.7	1.4
NV1042-2	10	521	8.9	1.9
NV1028-2	10	472	8.0	2.1
NV1111-2	10	335	5.7	3.0
0.1%DMSO-2	10	542	9.2	1.8
NV1078-3	10	127	2.2	7.9
NV1087-3	10	118	2.0	8.5
NV1042-3	10	99	1.7	10.1
NV1028-3	10	92	1.6	10.9
NV1111-3	10	93	1.6	10.8
0.1%DMSO-3	10	122	2.1	8.2
NV1078-4	10	239	4.1	4.2
NV1087-4	10	307	5.2	3.3
NV1042-4	10	252	4.3	4.0
NV1028-4	10	278	4.7	3.6
NV1111-4	10	277	4.7	3.6
0.1%DMSO-4	10	301	5.1	3.3
NV1078-5	10	330	5.6	3.0
NV1087-5	10	308	5.2	3.2
NV1042-5	10	335	5.7	3.0

PA-Replicate	RIN*	Total RNA (ng/µL)	tRNA µg (17µL)[a]	tRNA(µL)
NV1028-5	10	155	2.6	6.5
NV1111-5	10	394	6.7	2.5
0.1%DMSO-5	10	278	4.7	3.6
NV1011-1	10	134	2.3	7.5
NV1113-1	9.9	126	2.1	7.9
NV1115-1	10	219	3.7	4.6
NV1011-2	10	232	3.9	4.3
NV1113-2	10	110	1.9	9.1
NV1115-2	10	166	2.8	6.0
NV1011-3	10	87	1.5	11.5
NV1113-3	10	82	1.4	12.2
NV1115-3	10	51	0.9	19.6
NV1011-4	10	179	3.0	5.6
NV1113-4	10	112	1.9	8.9
NV1115-4	10	18	0.3	55.6
NV1011-5	10	179	3.0	5.6
NV1113-5	10	170	2.9	5.9
NV1115-5	10	136	2.3	7.4

*Bioanalyzer data from ECF4030

[a]Total RNA aliquot for DDR qPCR experiments

The reaction components were incubated for 10 min at 25 °C followed by 15 min at 50 °C. Reverse transcription of RNA extracted from W12E cells after treatment with polyamides was terminated at 85 °C (5 min) in agreement with the K1641 protocol. cDNA was stored at -20 °C for a week or -80 °C for further experiments.

5.3.5 Real Time - Quantitative Polymerase Chain Reaction (RT-qPCR). Relative quantification method. Normalization against reference genes

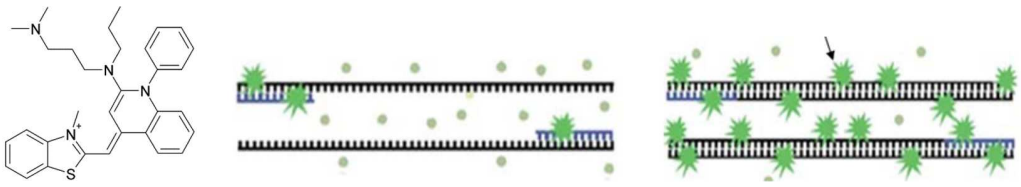
To assess the gene-specific transcription levels, the first strand cDNA was primed with specific primers (DNA Repair II Pathway PCR array, HDRL-II, RealTimePrimer) which hybridized to the reference genes or the chosen target DDR genes. The control genes, further named reference gene primers, were supplied along with the Human DNA Repair Primer Library II (RealTimePrimer)**.**

Real-time quantitative PCR is a fluorescent-based method which permits the detection of small amounts of DNA in a large variety of samples. Single-stranded cDNA was transformed into dsDNA in the presence of Maxima SYBR Green ROX qPCR Master Mix (Fermentas, K0221). The initial amplicon amount was measured with a BioRad CFX96 Real-Time System. The quantitative PCR (q-PCR) reaction was performed using a two-

step cycling protocol, following cycling conditions: 1X 95 °C for 10 min (initial activation of enzyme Maxima Hot) followed by 40 cycles of 95 °C for 10 sec (denaturation), 58 °C for 45 sec (annealing and extension) (6). Hard shell PCR plates (HSP9601 Bio-Rad) were sealed with PCR Microseal “B” film (MSB1001).

The master mix used included Maxima Hot Start Taq DNA Polymerase and dNTPs. Maxima Hot Start Taq DNA Polymerase is a Taq DNA polymerase which is inactive at room temperature. The qPCR buffer contains KCl and $(NH_4)_2SO_4$ to provide high specificity of primer annealing.

The SYBR Green I is an intercalating dye that allows DNA detection and analysis without using sequence-specific probes. SYBR Green I is a fluorescent intercalating dye that emits a fluorescent signal when it binds to the double stranded DNA. The fluorescent signal is proportional to the dsDNA concentration **(Figure 3)**. The excitation maxima of SYBR Green I is 494 nm and emission maxima is 521 nm. SYBR Green I is compatible with the use of BioRad CFX96 Real-Time System.



A B

Figure 5.3. (A) SYBR Green I structure, (B) SYBER Green I upon binding to dsDNA emits fluorescence very brightly, (B) Signal intensities correlate with DNA amplified, thus, the initial sample input amount.

For normalization of the SYBR Green I, ROX is used as a reference dye (**Figure 5.4**). It allows for correction due to pipetting imprecisions and fluorescence variations. ROX does not participate in a PCR reaction. The excitation and emission maxima are 580 nm and 621 nm.

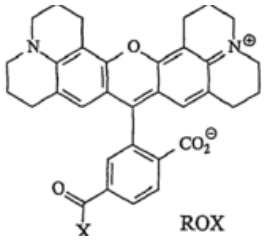


Figure 5.4. ROX chemical formula

After purification of the total RNA from cells, reverse transcription of RNA, real-time PCR reactions were performed for five biological replicates of four cDNA samples.

The samples were: (1) the control sample represented by cells treated with vehicle, 0.1 % DMSO, (2) cells treated with 0.1 µM NV1078 polyamide for 24 hours, (3) cells treated with 0.1 µM NV1042 polyamide for 24 hours, and (4) cells treated with 0.1 µM NV1011 for 24 hours. The transcription level was measured for 7 DDR target genes, ATM, ATR, RAD1, RAD9, CHEK1, CHEK2, and TP53 **(Table 5.5),** and for 5 reference genes: ACTB, PGK1, GAPD, PPIA, and RPL13A **(Table 5.6)**.

The transcription levels of one set of biological replicates were analyzed in the same time in the same 96-well plate (Bio-Rad) to minimize run-to-run differences, and the same threshold setting was used for the determination of the quantification cycle values (Cq). The raw quantification cycle values for the biological replicates are presented in Tables 7, 9, 11, and 13.

Table 5.5. Target DDR genes used in the study

Human Gene	Gene Symbol	Activity/Repair	Citation
Ataxia Telangiectasia Mutated	ATM	DSB Repair	(5, 6, 9) (34)
Ataxia Telangiectasia and Rad3 Related	ATR	SSB Repair	(5, 6, 9)
Rad1 Homolog	RAD1	9-1-1 complex member, exonuclease	(6)
Rad9 Homolog	RAD9A	9-1-1 complex member	(6)
Chk1 Checkpoint Homolog	CHEK1	DDR ATR checkpoint effector	(5, 6)
Chk2 Checkpoint Homolog	CHEK2	DDR ATM checkpoint effector	(5, 6)
Tumor Protein P53	TP53	Tumor suppressor, transcriptional regulator	(6)

The reference genes (Table 5.6) should be stably expressed. The expression level of those genes was measured simultaneously with the target genes to eliminate experimental differences.

Table 5.6. Reference genes level of expression and function

Human Gene	Gene Symbol	Relative expression level	Function
Actin, beta	ACTB 1	+++	Cytoskeletal structural protein
Phosphoglycerate kinase 1	PGK1	+++	Phosphate group transfer in glycolytic pathway
Peptidyl prolyl isomerase A	PPIA	+++	Cis-trans isomerization of proline imidic peptide bonds
Ribosomal protein L13A	RPL13A	+++	Structural component of the large ribosomal subunit
Glyceraldehyde-3-phosphate dehydrogenase	GAPD	+++	Oxidoreductase in glycolysis and gluconeogenesis

The specific primer set sequences for target genes **(Table 5)** as well as reference genes **(Table 6)** are not disclosed by the vendor (RealTimePrimer). A possible primer sequence can be found on the RT Primer Data Base website (RTPimer) which is a collection of submitted primer pairs and assays used in RT-qPCR experiments.

The expression of target genes normalized to 5 reference genes and relative to a calibrator (untreated cells) was calculated based on the Livak (ΔΔCq) method (144) according to Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines (145).

The Livak method requires that the target and reference gene amplify at efficiencies near 100% and have efficiencies within 5% of each other. Three steps are required for calculations (1) normalize Cq (target gene) to Cq (reference gene) for control samples and test samples; (2) normalize ΔCq of the test sample to ΔCq of the calibrator (control); (3) Calculate expression ratio, or fold difference.

1. Normalize Cq of the target gene to Cq of the reference gene

ΔCq(calibrator) = ΔCq(control) = Cq (target, calibrator) - Cq (reference, calibrator)

ΔCq(calibrator) = ΔCq(control) = Cq (target, calibrator) - average Cq (ref., calibrator)

Cells treated with vehicle 0.1% DMSO are the control cells

ΔCq(test) = Cq (target, PA treated) - Cq (reference, PA treated) = ΔCq(PA)

ΔCq(test) = Cq (target, PA treated) – average Cq (reference, PA treated) = ΔCq(PA)

2. Normalize ΔCq of test sample to ΔCq of calibrator

ΔΔCq = ΔCq(calibrator) - ΔCq (test)

3. Calculate the fold difference in expression

Normalized Expression Ratio = RQ=$2^{\Delta\Delta Ct}$

The RQ values obtained in this way represent a change in the expression of the gene of interest between the test and the calibrator conditions normalized for any differences in loading between the reference and test samples.

5.4 Results and Discussion. Effects of polyamides on 7 DDR gene expression

The goal of this project was to have an initial feedback (before the analysis of RNASeq data) on polyamide effects regarding gene expression. RT-QPCR studies monitored expression of the seven DDR genes (**Table 5.5**) at the mRNA level.

Herein, the effects of three polyamides on 7 DDR gene expression compared with the expression of the genes in the control cells is presented. Description of procedure in a nutshell:

W12E cells were seeded in six-well plates (1X10^5 cells per well) over (1X10^5 cells per well) mitomycin treated 3T3 cells and treated with 0.1 μM PA for 24 h. The total RNA was extracted and purified using RNeasy Plus Mini Kit along with QIAshredder homogenizer columns to obtain a homogenized lysate. The quality of the RNA samples was confirmed using an Agilent 2100 Bioanalyzer with an RNA 6000 NanoKit for analysis. The RIN values were 10 for all samples (**Table 5.4**). cDNA was synthesized from 1 μg of DNA-free total RNA in a 20 μL reaction volume using a Maxima First Strand cDNA Synthesis Kit for RT-qPCR. The obtained cDNA samples were diluted 20-fold for RT-PCR reactions. Transcription levels for each specific gene were determined in a 20 μL reaction volume using Maxima SYBR Green/ROX qPCR Master Mix on a CFX96 real-time PCR machine (CFX Manager Version: 3.1.1517.0823, Bio-Rad). Gene expression was analyzed using the relative quantification method. The relative quantification method estimates the level of gene expression against the control. The control samples were W12E cells treated with 0.1% DMSO, used as a vehicle for polyamides. Real-time PCR reactions of four cDNA samples (control, samples treated with NV1078, NV1042, and NV1011) using the same primer sets, seven DDR target **(Table 5.5)** and five control genes **(Table 5.6)** were analyzed together in the same 96-well plate in order to use the same threshold setting for determination of the quantification cycle values (Cq).

Data analysis: Post PCR quality control. Gene expression had been analyzed using the relative quantification method (RQ) named also comparative Cq method which estimates the level of gene expression against the calibrator. In this case, the calibrator sample is W12E cells treated just with 0.1% DMSO vehicle. Five reference genes were used as a loading control to show that the efficiencies of the target DDR genes and control amplification are approximately equal**.**

The typical use of a single gene for normalization leads to errors. Instead, the geometric mean of multiple housekeeping genes (optimal number varies from one to more than ten (146)) was validated as an accurate normalization factor. The expression of reference genes used should not change in the cells under examination or in response to experimental treatment.

In this work, the Vandesompele (147) normalization strategy was used, which employs the calculation of a normalization factor based on three to five reference genes. The expression stability for a particular gene is reflected by the M value calculated by the Bio-Rad CFX Manager 3.1 software and represents the mean standard deviation of the log-transformed expression ratios across samples for the particular gene relative to other reference genes in the gene panel (146).

The most stable reference genes are those with the lowest CV and M value. The recommended stability values for samples are:

(1) Mean CV value ≤ 0.25 and Mean M value ≤ 0.5 for homogeneous sample, and

(2) Mean CV value < 0.5 and Mean M value < 1 are considered acceptable for heterogeneous samples.

The plots of fluorescence vs. cycle number are presented in the supplementary information at the end of the chapter. The raw quantification cycle values for the biological replicates are presented in **Tables 5.7, 5.9, 5.11, 5.13, and 5.15.**

Tables 5.7, 5.9, 5.11, 5.13, and 5.15 present the quantification cycle (Cq) number, previously known as the threshold cycle (Ct). The first seven columns show the values associated with the DDR genes mentioned in the raw above the table. The last five columns display the Cq values related to the reference genes. The reference genes have a high expression level compared with the DDR genes.

Figures 5.5 – 5.9, display the effects of the three different polyamides for the biological replicates according to the normalized relative expression ($RQ=2^{\Delta\Delta Cq}$) values of the seven DDR genes to four reference genes. Although the study employed five reference genes, just four were used in the normalization process. ATCB was not found

to be stably expressed and was eliminated from calculations. Expression stability for a particular gene was assessed based on the M value. The normalization aim was to use reference genes that give a mean **CV value ≤ 0.25** and a **mean M value ≤ 0.5** (see Data Analysis). The bar graphs were obtained through the use of Bio-Rad CFX Manager 3.1 software.

Table 5.7. Cq values of qPCR experiments of C-1 Series, 1th biological replicate. C-1: control 0.1% DMSO-1, 78-1: NV1078-1, 42-1: NV1042-1, 011-1: NV1011-1. ECF4165, 4/11/16

		ATM	ATR	Rad1	Rad9	CHEK1	Chek2	TP53	ACTB	PGK1	GAPD	PPIA	RPL 13A
A	Content	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn
	Sample	C-1	C-1	C-1	C-1	C-1	C-1	C-1	C-1	C-1	C-1	C-1	C-1
	Cq	23.61	21.82	21.67	23.38	20.20	22.88	20.98	16.83	18.98	15.13	15.32	17.49
B	Content	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn
	Sample	78-1	78-1	78-1	78-1	78-1	78-1	78-1	78-1	78-1	78-1	78-1	78-1
	Cq	23.34	21.42	21.40	23.20	19.87	22.39	20.73	15.43	18.59	14.55	14.77	16.10
C	Content	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn
	Sample	42-1	42-1	42-1	42-1	42-1	42-1	42-1	42-1	42-1	42-1	42-1	42-1
	Cq	24.01	21.82	21.72	23.41	20.48	22.92	20.83	15.91	19.06	14.78	15.05	16.45
D	Content	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn
	Sample	011-1	011-1	011-1	011-1	011-1	011-1	011-1	011-1	011-1	011-1	011-1	011-1
	Cq	23.01	20.95	20.71	22.84	19.32	21.65	19.93	15.05	18.07	13.74	14.16	15.30

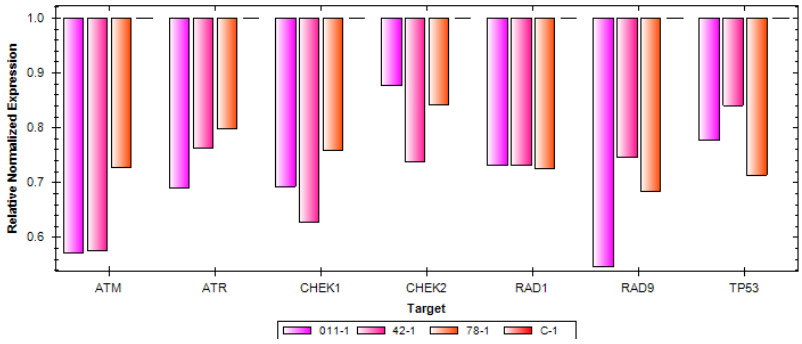


Figure 5.5. Downregulation of DDR genes in cells treated with polyamides. C-1 Series, 1th biological replicate Relative normalized expression (RQ) associated with **Table 7**

Table 5.8. Reference genes stability values associated with qPCR experiments of C-1 Series

C-1	Reference gene	Coeficient variance	M value
1	PGK1	0.1675	0.3140
2	GAPD	0.0441	0.2375
3	PPIA	0.0758	0.2345
4	RPL13A	0.2152	0.4699
	Mean value	0.13	0.31

Table 5.9. Cq values of qPCR experiments of C-2 Series, 2th biological replicate. C-2: control 0.1 % DMSO-2, 78-2: NV1078-2, 42-2: NV1042-2, 011-2: NV1011-2. ECF4157, 4/4/16

		ATM	ATR	Rad1	Rad9	Chk1	Chk2	TP53	ACTB	PGK1	GAPD	PPIA	RPL 13A
A	Content	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn
	Sample	C-2	C-2	C-2	C-2	C-2	C-2	C-2	C-2	C-2	C-2	C-2	C-2
	Cq	23.56	21.91	22.51	24.16	20.87	23.16	21.08	17.86	21.29	16.96	15.98	18.69
B	Content	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn
	Sample	78-2	78-2	78-2	78-2	78-2	78-2	78-2	78-2	78-2	78-2	78-2	78-2
	Cq	23.18	21.29	21.48	23.36	19.77	22.78	20.26	16.44	19.17	15.26	15.32	17.19
C	Content	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn
	Sample	42-2	42-2	42-2	42-2	42-2	42-2	42-2	42-2	42-2	42-2	42-2	42-2
	Cq	22.53	21.08	21.39	23.16	20.14	22.46	20.40	16.77	19.05	15.07	14.95	17.01
D	Content	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn
	Sample	011-2	011-2	011-2	011-2	011-2	011-2	011-2	011-2	011-2	011-2	011-2	011-2
	Cq	22.11	20.65	20.86	22.58	19.58	22.01	19.98	16.26	18.62	14.29	14.61	16.60

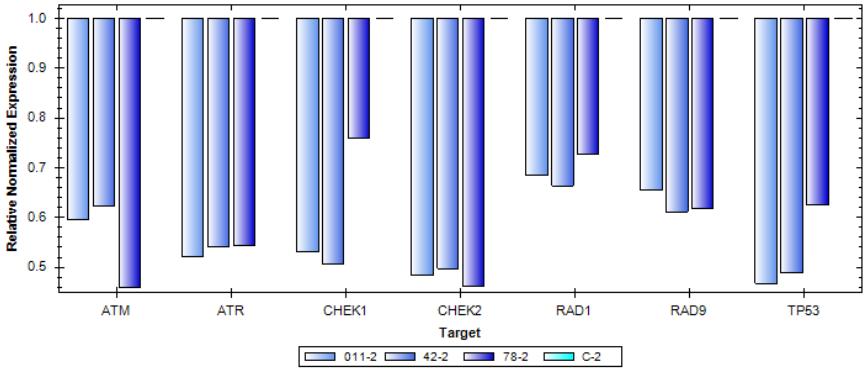


Figure 5.6. Downregulation of DDR genes in cells treated with polyamides. C-2 Series, 2th biological replicate. Relative normalized expression (RQ) associated with Table 9.

Table 5.10. Reference genes stability values associated with qPCR experiments of C-2 Series

C-2	Reference gene	Coeficient variance	M value
1	PGK1	0.1775	0.3973
2	GAPD	0.1367	0.3430
3	PPIA	0.3088	0.5395
4	RPL13A	0.0350	0.3046
	Mean value	0.1645	0.3046

Table 5.11. Cq values of qPCR experiments of C-3 Series, 3th biological replicate. C-3: control 0.1 % DMSO-3, 78-3: NV1078-3, 42-3: NV1042-3, 011-3: NV1011-3. ECF4148, 3/31/16

		ATM	ATR	Rad1	Rad9	CHEK 1	CHEK2	PT53	ACTB	PGK1	GAPD	PPIA	RPL 13A
A	Content	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn
	Sample	C-3	C-3	C-3	C-3	C-3	C-3	C-3	C-3	C-3	C-3	C-3	C-3
	Cq	21.55	20.20	20.20	22.18	18.59	21.22	19.14	17.58	20.38	14.63	14.54	18.28
B	Content	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn
	Sample	78-3	78-3	78-3	78-3	78-3	78-3	78-3	78-3	78-3	78-3	78-3	78-3
	Cq	21.90	19.45	19.93	21.19	18.53	20.55	18.60	13.82	17.59	13.21	13.60	15.78
C	Content	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn
	Sample	42-3	42-3	42-3	42-3	42-3	42-3	42-3	42-3	42-3	42-3	42-3	42-3
	Cq	21.77	19.79	20.30	21.66	18.73	21.17	18.83	14.14	18.05	13.38	13.50	16.30
D	Content	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn
	Sample	011-3	011-3	011-3	011-3	011-3	011-3	011-3	011-3	011-3	011-3	011-3	011-3
	Cq	21.15	19.19	19.48	21.19	18.07	20.08	18.16	13.63	17.35	12.42	13.21	15.51

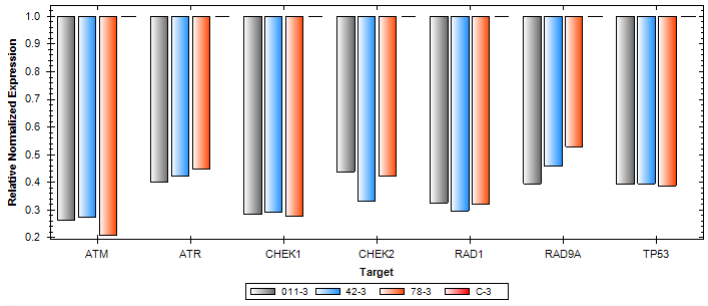


Figure 5.7. Bar chart showing downregulation of DDR genes in cells treated with polyamides. C-3 Series, 3th biological replicate. Relative normalized expression (RQ) associated with Table 11.

Table 5.12. Reference genes stability values associated with qPCR experiments of C-3 Series

C-3	Reference gene	Coeficient variance	M value
1	PGK1	0.2379	0.5294
2	GAPD	0.1598	0.4743
3	PPIA	0.3525	0.6079
4	RPL13A	0.1648	0.4384
	Mean value	0.2288	0.5217

Table 5.13. Cq values of qPCR experiments of C-4 Series, 4th biological replicate. C-4: control 0.1 % DMSO-4, 78-4: NV1078-4, 42-4: NV1042-4. ECF4177, 5/26/16

		ATM	ATR	Rad1	Rad9	CHEK1	Chke2	TP53	ACTB	PGK1	GAPD	PPIA	RPL 13A
A	Content	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn
	Sample	C-4	C-4	C-4	C-4	C-4	C-4	C-4	C-4	C-4	C-4	C-4	C-4
	Cq	23.51	22.12	22.33	24.71	23.10	23.86	21.22	14.91	21.55	16.32	16.25	18.74
B	Content	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn
	Sample	78-4	78-4	78-4	78-4	78-4	78-4	78-4	78-4	78-4	78-4	78-4	78-4
	Cq	24.17	22.23	22.74	25.19	23.34	23.75	21.66	13.44	21.00	16.29	16.66	18.77
C	Content	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn
	Sample	42-4	42-4	42-4	42-4	42-4	42-4	42-4	42-4	42-4	42-4	42-4	42-4
	Cq	23.52	22.04	22.11	24.85	22.98	23.01	21.13	13.73	20.26	16.37	15.69	18.12

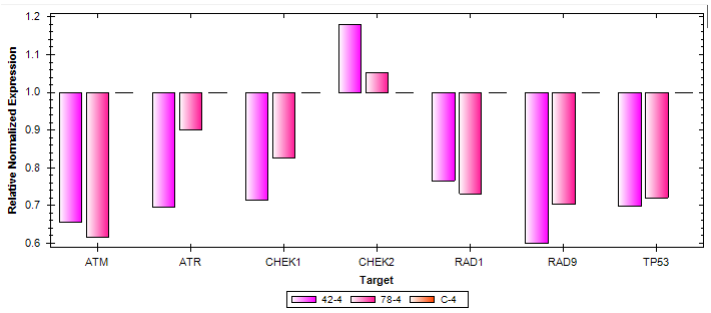


Figure 5.8. Bar chart showing downregulation of DDR genes in cells treated with polyamides. C-3 Series, 3th biological replicate. Relative normalized expression (RQ) associated with Table 13

Table 5.14. Reference genes stability values associated with qPCR experiments of C-4 Series

C-4	Reference gene	Coeficient variance	M value
1	PGK1	0.2317	0.5139
2	GAPD	0.2390	0.5364
3	PPIA	0.1609	0.4096
4	RPL13A	0.0326	0.3243
	Mean value	0.1660	0.4460

Table 5.15. Cq values of qPCR experiments of C-5 Series, 5th biological replicate. C-5: control 0.1 % DMSO-5, 78-5: NV1078-5, 42-5: NV1042-5, 011-5: NV1011-5. ECF4170, 4/22/16

		ATM	ATR	Rad1	Rad9	Chke1	Chke2	TP53	ACTB	PGK1	GAPD	PPIA	RPL 13A
A	Content	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn
	Sample	C-5	C-5	C-5	C-5	C-5	C-5	C-5	C-5	C-5	C-5	C-5	C-5
	Cq	23.92	22.12	22.20	24.42	22.49	23.12	21.08	14.56	19.08	15.45	15.76	17.54
B	Content	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn
	Sample	78-5	78-5	78-5	78-5	78-5	78-5	78-5	78-5	78-5	78-5	78-5	78-5
	Cq	23.95	22.09	22.06	24.65	22.82	23.37	21.39	14.72	19.27	15.58	15.89	17.37
C	Content	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn
	Sample	42-5	42-5	42-5	42-5	42-5	42-5	42-5	42-5	42-5	42-5	42-5	42-5
	Cq	24.02	22.41	22.39	24.91	23.00	23.42	21.73	14.81	19.25	15.94	14.48	17.65
D	Content	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn	Unkn
	Sample	011-5	011-5	011-5	011-5	011-5	011-5	011-5	011-5	011-5	011-5	011-5	011-5
	Cq	24.44	22.58	22.66	25.39	23.38	23.77	21.67	15.12	19.65	15.91	16.26	18.10

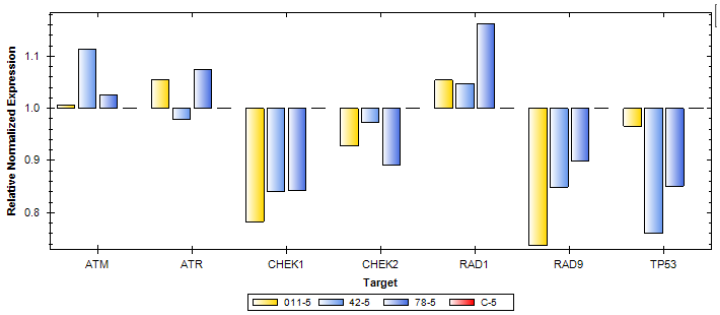


Figure 5.9. Bar-chart showing downregulation of DDR genes in cells treated with polyamides. C-5 Series, 5th biological replicate. Relative normalized expression (RQ) associated with Table 15

Table 5.16. Reference genes stability values associated with qPCR experiments of C-5 Series

	Reference gene	Coeficient variance	M value
1	ACTB	0.0259	0.1165
2	PGK1	0.0573	0.1377
3	GAPD	0.0904	0.1911
4	RPL13A	0.0866	0.1817
	Mean value	0.0650	0.1567

The selected genes' expression was altered in all five biological replicates. Series 1, 2, and 3 presented the same trend in gene regulation. All genes were downregulated. The similarities between the effects induced by the two active polyamides and the effect induced by the NV1011 polyamide were unexpected. It was thought that the small NV1011 polyamide (Table 2) was inactive, and therefore will not have any effect on the cells.

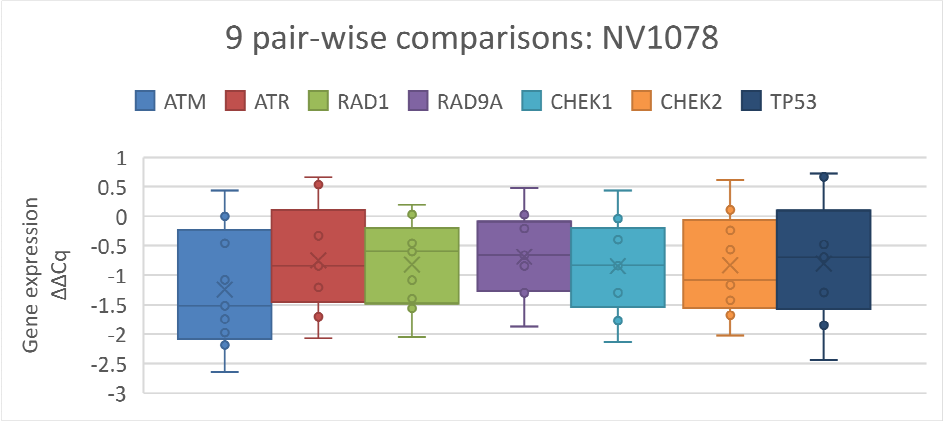
All samples have been treated with 0.1 µM polyamide**. Table 5.17** illustrates the relationship between the IC50 and IC90 values (Chapter 1) and the experimental concentration.

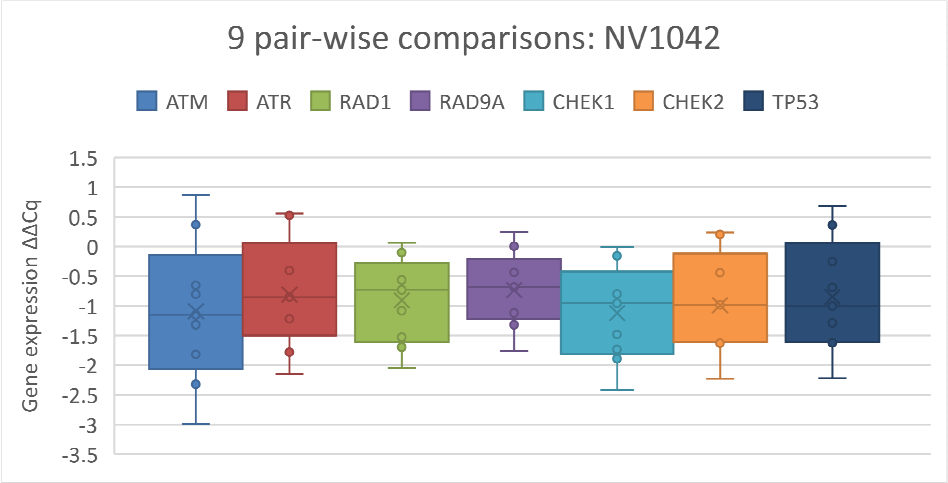
Table 5.17. The IC50 and IC90 values and their relationships to 0.1 µM concentration

Polyamide	IC50 (µM)	IC90 (µM)	0.1µM/IC50	0.1µM/IC90
NV1078	0.046	0.307	2.2	0.3
NV1042	0.036	0.351	2.8	0.3

The polyamide concentration used for the treatment of the cells was 2.8 times higher than the IC50 of the NV1042 polyamide, and 2.2 times greater than the IC50 of NV1078, determined in the same conditions. Consequently, when interpreting the changes in the level of expression for the same gene, this difference has to be taken into consideration.

ΔΔCq was calculated for each target gene against the average of four control genes. For the first three samples, this method produced 9 pair-wise comparisons for each gene (3 DMSO treated ΔCq values x 3 PA treated ΔCq values) (6). The genes are considered affected by polyamide treatment if at least 2 fold ($|\Delta\Delta Cq| \geq 1$) or more change was registered in their expression compared with the same gene expression in samples treated with 0.1% DMSO (control) (6). The value ΔΔCq = 1, accounts for $RQ = 2^{-\Delta\Delta Cq} = 2$, two-fold change, and is therefore considerate to be significant **(Table 5.19 and Figure 5.10**).





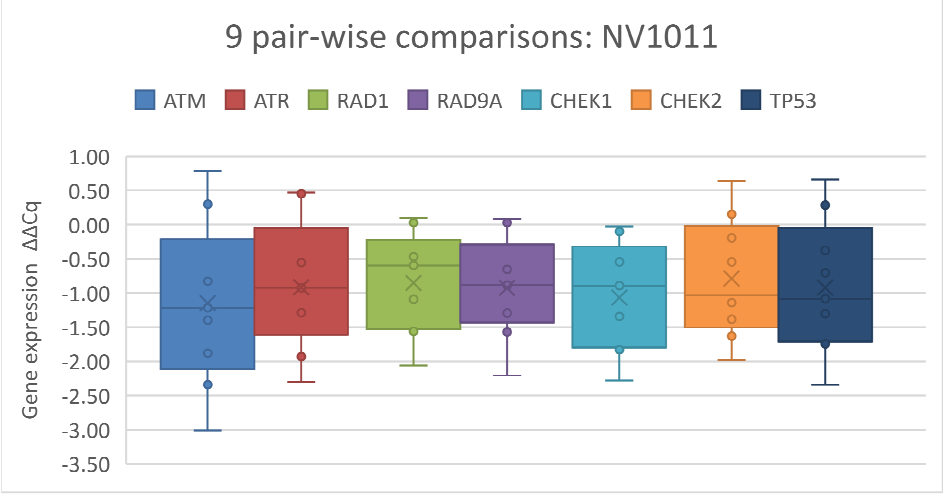


Figure 5.10 Gene expression downregulation induced by NV1078, NV1042 and NV1011. All 9 ΔΔCq values are shown as well as the median line and the average (x) values for every gene.

Taking into consideration that the ATM pathway activated by double-stranded breaks and the ATR pathway activated by single-stranded breaks are linked to each other as demonstrated by recent studies reviewed by Hong and Laimins (15) it makes sense to have the same trend for ATM/CHEK2 and ATR/CHEK1 pairs. ATR seems to be subordinated to elements of ATM. The ATR–CHK1 pathway requires activation by ATM. (15).

In our case, the regulation of ATM was also correlated with the regulation of downstream element P53. In the same way, the regulation of ATR was associated with the expression levels of Rad9, Rad1, and P53. Their expression was also downregulated.

We already know that the large polyamide NV1042 induces a DDR response in the cells infected with the HPV16 virus (5, 6, 117). We also know from the same study that ATM is important for episome maintenance (6) because the siRNA-mediated knockdown of the gene drastically decreased the episome numbers in W12E cells.

It was also demonstrated that the significant viral loss registered when infected cells were treated with the same NV1042 active polyamide is a consequence of the degradation of the viral episome, not just the inhibition of viral DNA replication (117). Therefore, it may be postulated that the downregulation of the DDR elements registered in samples treated with NV1042 and NV1078 polyamides is a direct consequence of the episome degradation. The present results reinforce the principle that the episome degradation was accomplished by the polyamide treatment.

Table 5.18. Comparison of gene expression altered by NV1042 (PA25) in W12E cells, induced in two different conditions. W12E cells were incubated with 10µM NV1042 for 48 hours (6), and this work, W12E cells were incubated with 0.1µM NV1042 for 24 hours (biological replicate #3), Fold Δ: $2^{\Delta\Delta Cq}$

Human Gene	mRNA (Fold Δ)(6)	mRNA (Fold Δ)	Function/Pathway
Ataxia Telangiectasia Mutated, ATM	-3.68	-3.5 ($2^{1.8}$)	HR, DSB Repair, PI-3 kinase
Ataxia Telangiectasia and Rad3 Related, ATR	…	-2.3 ($2^{1.2}$)	HR, SSB Repair
Rad1 Homolog, RAD1	-67.65	-3.3 ($2^{1.7}$)	9-1-1 member, exonuclease
Rad9 Homolog, RAD9A	….	-2.1 ($2^{1.1}$)	9-1-1 complex member
Chk1 Checkpoint Homolog, CHEK1	-4.17	-3.3 ($2^{1.7}$)	DDR ATR checkpoint effector
Chk2 Checkpoint Homolog, CHEK2	-2.62	-3.0 ($2^{1.6}$)	DDR ATM checkpoint effector
Tumor Protein P53, TP53	….	-2.5 ($2^{1.3}$)	Tumor suppressor, transcriptional regulator

Table 5.19. Nine pair-wise (3 DMSO treated ΔCq values x 3 PA treated ΔCq values) comparisons.

GENES		NV1011		NV1042		NV1078	
			ΔΔCt≥1		ΔΔCt≥1		ΔΔCt≥1
ATM	downregulated	7/9	5	7/9	5	7/9	6
	upregulated	2/9	0	2/9	0	2/9	0
ATR	downregulated	7/9	6	7/9	5	7/9	5
	upregulated	2/9	0	2/9	0	2/9	0
RAD1	downregulated	7/9	4	8/9	4	7/9	4
	upregulated	2/9	0	1/9	0	2/9	0
RAD9A	downregulated	7/9	4	7/9	4	7/9	4
	upregulated	2/9	0	1/9	0	1/9	0
CHEK1	downregulated	8/9	5	7/9	5	8/9	4
	upregulated	1/9	0	1/9	0	1/9	0
CHEK2	downregulated	7/9	5	7/9	6	7/9	5
	upregulated	2/9	0	2/9	0	2/9	0
TP53	downregulated	7/9	5	7/9	5	7/9	4
	upregulated	2/9	0	2/9	0	2/9	0

5.5 Annotations

HR: Homologous recombination, exchange of base-paired partners between two homologous DNA molecules (113)

CHEK1,2: checkpoints, places in the cell cycle where DNA sequence and structure fidelity are ascertained before progression to the subsequent phase of the cell cycle (113)

STAT5: Signal transducer and activator of transcription

qPCR: quantitative real-time PCR (145)

RT-qPCR: reverse transcription quantitative real-time PCR (145)

Reference genes: previously known as housekeeping or control genes (145)

Quantification cycle Cq: previously known as the threshold cycle (Ct), the cycle at which fluorescence from amplification exceeds the background fluorescence (145)

5.6 Supplementary Information

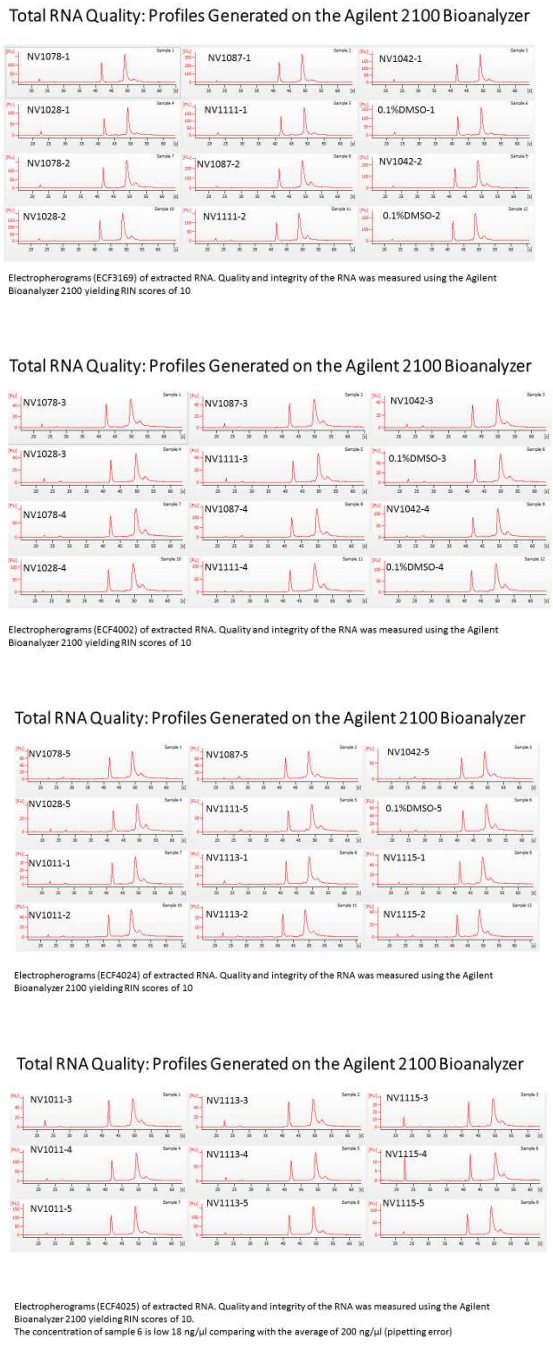


Figure S1. Electropherograms of extracted RNA; quality and integrity of the RNA were measured using the Agilent Bioanalyzer 2100, yielding RIN scores of 10.

Chapter 6

The Analysis of Viral and Human Transcriptomes After Treatment of W12E Cells with Polyamides

6.1 Introduction

Based on the knowledge gathered from the results of the binding studies in **Chapters 3** and **4** and those reported recently (11, 105), coupled with the first gene expression array work on a limited but important set of mRNA levels (6), a complex study concerned with genome-wide expression was designed. With the help of high-throughput sequencing (HiSeq techniques) host and viral mRNA expression will reveal any genes regulated by the polyamides. Thus, the aim of this gene expression project is to find the similarities and differences in different polyamides' mechanism of action related to gene expression by quantifying messenger RNA (mRNA) after treatment of HPV-infected keratinocyte (W12E) cells separately with eight polyamides (**Table 6.1**). The viral transcriptome data obtained after the treatment of cells with polyamides will be compared with the known viral transcriptomes (148, 149) and to each other. By comparing samples treated with different polyamides, different expression levels and different splicing sites may be revealed (150).

The HPV mRNA is polycistonic in both the episomal form and when integrated into human chromosomes (21, 148-150). In undifferentiated cells, the transcription is initiated by the early promoter (p97) and is terminated at the early poly(A) signal, AE, while the mRNA encoding the viral early genes is produced. In differentiated cells, the transcripts are expressed by the late promoter (p670) and utilize a late poly(A) signal, AL, while they produce mRNA for the late genes (45). The transcripts in this situation are terminated at the late poly(A) signal (**Figure 6.1**) (46). By checking the level of E6 and the late L1 and L2 genes, it is possible to assess the replication phase of the virus.

The human transcriptome data obtained after sequencing results are fully-processed will aid in understanding how polyamides help infected cells eliminate the virus from a genomic perspective.

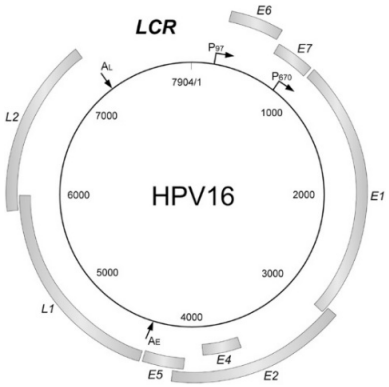


Figure 6.1. The HPV genome circular dsDNA structure (46)

6.2 Materials and Methods

A detailed description of the cell culture (5.3.1), polyamide treatment (5.3.2), and RNA extraction (5.3.3) methods is provided in Chapter 5.

For the RNASeq experiment, an inactive polyamide (NV1011) and 7 active polyamides were used. Among the active polyamides are the two, PA1 and PA42, leading compounds of the HPV viral research. Three TMG compounds NV1078, NV1087 and NV1113, and two guan compounds NV1111 and NV1115 (**Chapter 1, Table 1.2**) were chosen because of their excellent potency in cell cultures harboring HPV 16 or HPV 18.

Table 6.1. Polyamides used for RNASeq studies. The IC50 and IC90 values measured in W12E cells were determined by T. G. Edwards in the Fisher lab (unpublished data).

PA	HPV16 IC50 [µM]	HPV16 IC90 [µM]	HPV18 IC50 [µM]	HPV18 IC90 [µM]
NV1011 (PA11)	NOT ACTIVE	NOT ACTIVE	NOT ACTIVE	NOT ACTIVE
NV1028 (PA1)	0.100	1.113	0.717	>10
NV1042 (PA25)	0.036	0.351	0.056	1.462
NV1078	0.046	0.307	0.030	0.091
NV1087	0.031	0.200	0.024	0.297
NV1111	0.017	0.121	0.018	0.136
NV1113	0.304	>10	0.228	>10
NV1115	0.103	0.378	0.167	0.937

6.2.1 TruSeq Stranded mRNA Library Preparation

All total RNA samples used for library preparation were of excellent quality (RIN=10, see **Chapter 5**) and high concentration (**Table 5.4 and Figure S1 in Chapter**

5), and had fallen into the range recommended by ILLUMINA for sequencing is (0.1 - 4) µg/mL.

The total RNA (tRNA, not meaning transfer RNA in this context) samples stored at -80^{0}C were used for preparing the NextGen sequencing libraries according to TruSeq Stranded mRNA Library LS Prep protocol for 45 samples using 9 adapters (Illumina, RS122-2102 set B Kit). The adapter includes the index, a unique sequence of 6 nucleotides. The index is used to identify the sample. Different input quantities, ranging from 500 to 1000 ng of total RNA, were used to allow the determination of the ideal amount for preparing a complementary ds DNA library with the selected method. The best library quality was obtained using lower tRNA quantities (**Table 6.3**). Therefore, the majority of libraries were prepared from 500 ng total RNA as starting material.

The tRNA samples were subjected to a series of steps to obtain ds cDNA ready for sequencing. The first step was the purification of messenger RNA (mRNA) from total RNA and its subsequent fragmentation. In this way, non-polyA RNA, or ribosomal RNA, was separated from the pool of interest. PolyA mRNA was purified using polyT-oligo-attached magnetic beads provided with the Illumina kit. Two rounds of elution were performed. During the first round of purification, the mRNA was denatured and bonded to the PolyT sequences attached to the beads. During the second purification, mRNA was fragmented into small pieces with a median size of 150 bp, using divalent cations under elevated temperature. The purified and fragmented mRNA was primed with random primers for cDNA synthesis.

The purified and fragmented mRNA was primed with random primers for cDNA synthesis in the presence of reverse transcriptase SuperScript II. This process reverse transcribed the cleaved RNA fragments primed with random hexamers into first strand of cDNA. Actinomycin improved the specificity of the process, allowing RNA-dependent synthesis. For the interpretation of the RNA-seq results only the first cDNA strand, which is the anti-sense strand, is sequenced. During the process of the second strand cDNA synthesis, the template mRNA was removed, and the complementary strand generated the final ds cDNA.

Agencourt PAMPure XP beads, which selectively bind DNA sequences longer than 100 bp, were used to capture the ds cDNA from the second strand reaction mix. To the generated, blunt-ended, ds cDNA, a single “A” was added to the 3’ ends to which a corresponding single “T” at 3’ end of the adapter makes available a complementary overhang for ligating the adapter to the fragment.

Different y-shaped indexing adapters were ligated to the ends of different ds cDNA samples (**Table 6.2**) with the use of T4 DNA ligase. The adapters contain the (1) sequencing primers, (2) the index sequence and (3) oligonucleotides complementary to the sequencing machine's flowcell DNA sequences. Thus, the adapters hybridize with the flowcell oligos.

Table 6.2. Adapters associated with different samples

Sample (1-9)	Adapter	Index Sequence
NV1078	AR001	ATCACG(A)
NV1087	AR003	TTAGGC(A)
NV1042	AR008	ACTTGA(A)
NV1028	AR009	GATCAG(A)
NV1111	AR010	TAGCTT(A)
Control 0.1% DMSO	AR011	GGCTAC(A)
NV1011	AR020	GTGGCC(T)
NV1113	AR022	CGTACG(T)
NV1115	AR025	ACTGAT(A)

To increase the amount of ds cDNA, enrichment with PCR was performed with a Primer Cocktail that anneals just to the ends of the adapters. In this way, just ds DNAs that have adapters ligated to the both ends will be amplified. After a new purification with AMPure XP beads the quantity and the quality of libraries were evaluated.

6.2.2 NextGen Library Quality Control. Validation and Quantification of TruSeq library

The size distribution after fragmentation of ds cDNA and adapter ligation was assessed on a BioAnalyzer Agilent 2100 instrument according to the Agilent DNA 1000 Kit (5067-1504) protocol. The results are presented in **Table 6.2**. In addition, a Qubit 2.0 Fluorometer was employed to determine the concentration of the TruSeq library with the help of Qubit dsDNA BR Assay Kit, catalog no. Q32850. The molar concentration of the samples was estimated by averaging the concentrations of both Qubit and BioAnalyzer measurements. The molar concentration was double-checked with a Qubit 3.0 Fluorometer at DNA Core Facility, University of Missouri, Colombia. 10 nM stock libraries were used to prepare DNA templates for cluster generation.

6.2.3 Sequencing and Alignment

The subsequent work, pooling of the libraries and sequencing on HiSeq 2500 Illumina instrument, was performed by Nathan Bivens and his colleagues at the DNA Core Facility, University of Missouri, Colombia. The sequencing results' interpretation is being

performed at Informatics Research Core Facility (IRCF) by Christopher Bottoms under the supervision of Dr. Scott Givan

The latent Illumina adapter sequences were identified and removed from the RNA-Seq data using bcl2fastq version 2.17.1.14 (run with command-line parameters "--minimum-trimmed-read-length=10 --mask-short-adapter-reads=9"). Subsequently, input RNA-Seq data was trimmed and filtered to remove low quality nucleotide calls and whole reads, respectively, using the Fastx-Toolkit, version 0.0.13. From the remaining RNA-Seq data, foreign or undesirable reads were removed by sequence matching to filter sequences, using Bowtie, version 0.12.9. Filter sequences included the following (National Center for Biotechnology Informatics accessions in parentheses): Phi-X genome sequence (NC_001422.1), the pUC19 vector (M77789.2), human ribosomal RNA sequences, the human mitochondrial genome (from Genomic Reference Consortium assembly h38), an additional human mitochondrial sequence (NCBI accession S64650.1), and Human genomic repeat elements in RepBase, version 20.02. The remaining RNA-Seq data was subjected to genome alignment, transcript assembly and expression quantification using TopHat, version 2.1.0, via the RNA-Seq-Toolkit (https://github.com/sgivan/RNA-Seq-Toolkit, (151) The differential expression analysis was done using Cufflinks and Cuffdiff, version 2.2.1. Computational hardware used included the BioCluster of the Informatics Research Core Facility (University of Missouri) and the Lewis Cluster of Research Computing Support Services (University of Missouri). **Just unique reads-anti-sense expression of the transcripts is being used since the library was prepared using a strand specific preparation kit that only sequenced the first cDNA strand.**

Analysis of the summary file will lead to the discovery of transcripts that are statistically differentially expressed in these data.

6.3 Partial Results

The size distribution of the TruSeq stranded mRNA libraries after adapter ligation shows ideal distribution, 260 bp, as can be seen in the profiles generated by Agilent 2100 Bioanalyzer (**Supplementary Figures**). The molar concentration of the 45 libraries is presented in **Table 6.3.**

Table 6.3. TruSeq Stranded mRNA Libraries characteristics. The table includes the amount of total RNA used for library preparation, and the average size and the concentration of ds DNA, determined with BioAnalyzer Agilent 2100. The molar concentration was calculated with ScienceLauncher software http://www.sciencelauncher.com/mwcalc.

Sample	RNA Amount (ng)	BioA dsDNA (ng/µL)	BioA [DNA] (nM)	dsDNA Average Size (bp)
NV1078-1	628	18.75	88.35	327
NV1028-3	500	24.83	125	306
0.1%DMSO-1	1000	20.27	98.2	317
NV1087-1	500	11.67	57.63	312
NV1042-1	500	20.92	105.69	305
NV1028-1	500	13.46	69.6	298
NV1111-1	500	22.79	118.23	297
NV1042-2	500	14.64	78.33	288
NV1078-2	738	20.27	104.11	300
NV1087-2	690	18.6	97.48	294
NV1028-2	944	24.06	127.4	291
NV1111-2	670	28.23	152	286
0.1%DMSO-2	542	15.41	81.32	292
NV1078-3	500	22.37	115.67	298
NV1087-3	500	18.86	97.85	297
NV1111-3	500	24.62	127.73	297
0.1%DMSO-3	500	17.48	94.84	284
NV1042-3	500	17.67	92.29	295
NV1011-1	500	14.25	73.68	298
NV1113-1	500	15.04	74.76	310
NV1115-1	500	7.04	38.6	281
NV1011-2	500	15.14	75.99	307
NV1113-2	500	15.16	76.09	307
NV1115-2	500	13.52	71.73	284
NV1011-3	500	19.74	108.24	281
NV1113-3	500	20.67	103.74	307
NV1115-3	500	43.68	213.66	315
NV1078-4	500	22.47	116.57	297
NV1087-4	500	23.93	119.33	309
NV1042-4	500	17.16	86.41	306
NV1028-4	500	17.71	91.88	297
NV1111-4	500	29.77	150.89	304
0.1%DMSO-4	500	24.59	127.14	298
NV1011-4	500	19.55	99.42	303
NV1113-4	500	23.3	116.19	309
NV1115-4	500	27.05	137.56	303

Sample	RNA Amount (ng)	BioA dsDNA (ng/µL)	BioA [DNA] (nM)	dsDNA Average Size (bp)
NV1078-5	500	19.89	102.16	302
NV1087-5	500	22.09	113.46	301
NV1042-5	500	18.51	97	294
NV1028-5	500	35.52	186.8	293
NV1111-5	500	31.02	159.32	300
0.1%DMSO-5	500	35.9	180.18	307
NV1011-5	500	17.27	85.29	312
NV1113-5	500	20.67	108.7	293
NV1115-5	500	22.88	111.92	315

The nine samples (a biological replicate series) were multiplexed per lane, and every set of sample was run in three lanes to obtain a statistically relevant amount of reads (50M). The possibility of pooling the maximum of 9 samples together in one lane was verified prior to library preparation with Illumina Experiment Manager (IEM) software.

The sequencing runs were of high quality Q30 score was over 90%, % PF clusters over 90, and the perfect barcode over 99% for every sample (Flowcell Summary and Lane Summary ECF4172-4175).

6.4 Overall conclusions

To treat current or future HPV infections, drugs that selectively block virus-specific processes, but do not damage the host cells, are needed. The compounds developed in Dr. Bashkin's lab are unique imidazole-pyrrole polyamides which have anti-HPV activity.

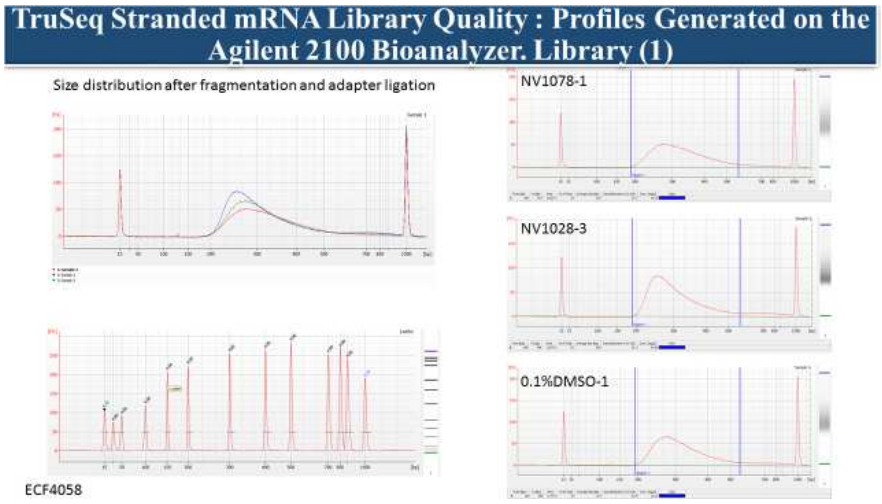
Biophysical studies of two TMG asymmetric hairpin polyamides (NV1078 and NV1087) on HPV LCR segment revealed that although the polyamides bind in many different sites, they have a noticeable preference for the region adjacent and overlapping with the E1, E2 binding sites (**Figure 4.11**). The intensive binding in the minor groove of dsDNA (sequence 7880-30 of HPV16 and sequence 7841-43 of HPV18) of TMG-AHPs structurally alter the dsDNA. The required adjustments of the DNA to allow binding may cause the inability of E1/E2 to assemble, as a replication competent double-hexamer at the ORI, obstructing the replication process at the start. The intensive binding in the region mentioned above may also cause ssDNA exposure or double strand breaks within HPV DNA and the activation of DNA damage response pathways (6). Multiple binding sites induce DNA nicks which activates ATR and ATM.

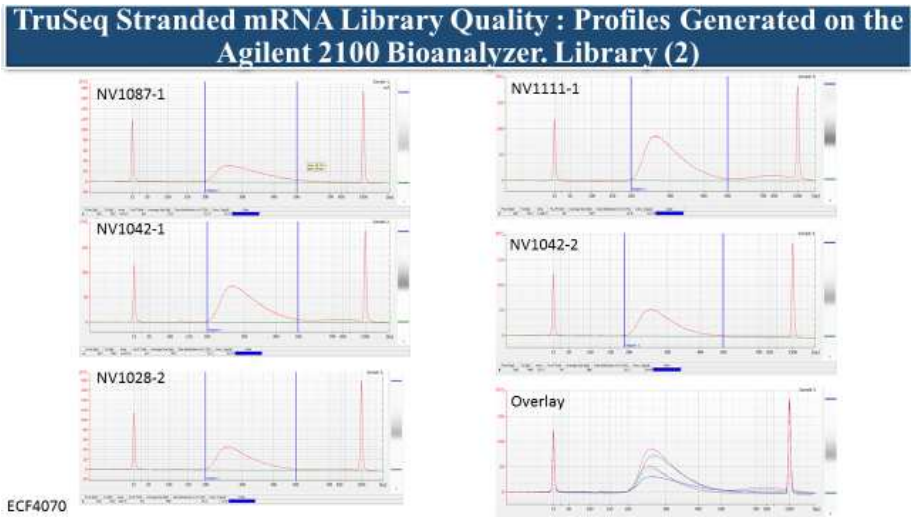
The partial results of gene expression analysis exposed genes that are relevant to the polyamide mechanism of action. After treatment of W12E cells with active polyamides a downregulation of ss and ds signal transducers ATM and ATR was registered. The regulation of ATM was also correlated with the regulation of downstream element TP53. In the same way, the regulation of ATR was associated with the expression levels of Rad9, Rad1, and P53. Their expression was also downregulated. The downregulation of the DDR elements registered in samples treated with NV1042 and NV1078 polyamides is a direct consequence of the episome degradation. The present results reinforce the principle that the episome degradation was accomplished by the polyamide treatment.

Our findings complement the notion that the PA binding to negatively supercoiled HPV episomes activate the DNA damage response (DDR) pathways (6). Activation of DDR genes eliminates the viral episomes without having substantial effects on cell growth or apoptosis (6).

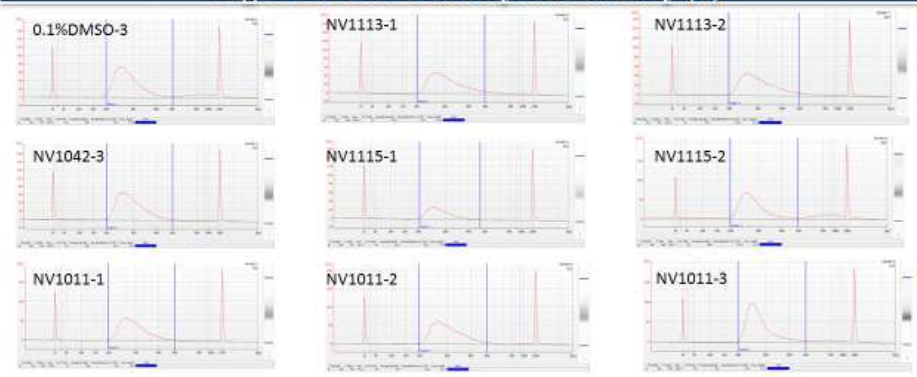
The elimination of viral DNA during the maintenance stage it is an effective strategy for the development of antiviral compounds. The large Im-Py polyamides synthetized in the Bashkin's lab have the potential to be useful for the treatment of preexisting and future HPV infections.

6.5 Supplementary Material.S1



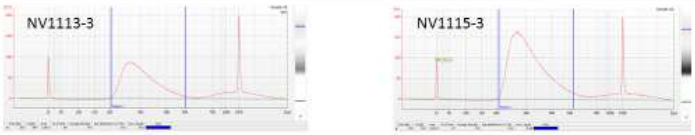


TruSeq Stranded mRNA Library Quality : Profiles Generated on the Agilent 2100 Bioanalyzer. Library (4)



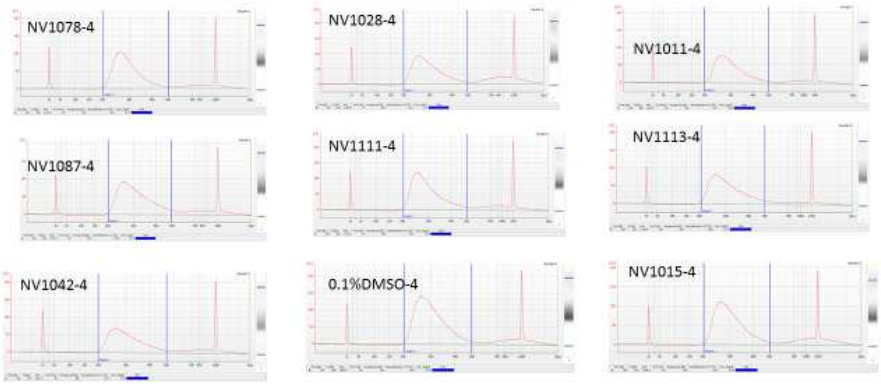
ECF4082

TruSeq Stranded mRNA Library Quality : Profiles Generated on the Agilent 2100 Bioanalyzer. Library (4)



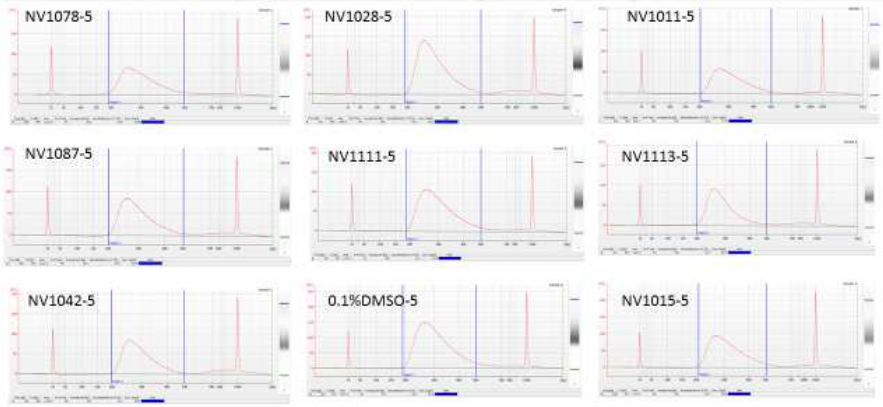
ECF4082/ECF4098

TruSeq Stranded mRNA Library Quality : Profiles Generated on the Agilent 2100 Bioanalyzer. Library (5)



ECF4098

TruSeq Stranded mRNA Library Quality : Profiles Generated on the Agilent 2100 Bioanalyzer. Library (6)



ECF4108

6.6 References

1. Moody CA, Laimins LA. Human papillomavirus oncoproteins: pathways to transformation. Nat Rev Cancer. 2010;10(8):550-60.
2. Hung C-F, Ma B, Monie A, Tsen S-W, Wu TC. Therapeutic human papillomavirus vaccines: current clinical trials and future directions. Expert opinion on biological therapy. 2008;8(4):421-39. doi: 10.1517/14712598.8.4.421. PubMed PMID: PMC3074340.
3. HPV-Associated Cancers Statistics. Available online: http://www.cdc.gov/cancer/hpv/statistics/ (accessed on 8 Sept 2015).
4. Frazer I. Correlating immunity with protection for HPV infection. Int J Infect Dis. 2007;11 Suppl 2:S10-6. Epub 2008/03/08. doi: 10.1016/s1201-9712(07)60016-2. PubMed PMID: 18162240.
5. Edwards TG, Helmus MJ, Koeller K, Bashkin JK, Fisher C. Human Papillomavirus Episome Stability Is Reduced by Aphidicolin and Controlled by DNA Damage Response Pathways. Journal of Virology. 2013;87(7):3979-89. doi: 10.1128/jvi.03473-12.
6. Edwards TG, Vidmar TJ, Koeller K, Bashkin JK, Fisher C. DNA damage repair genes controlling human papillomavirus (HPV) episome levels under conditions of stability and extreme instability. PLoS One. 2013;8(10):e75406. Epub 2013/10/08. doi: 10.1371/journal.pone.0075406. PubMed PMID: 24098381; PMCID: PMC3788802.
7. Everett RD. Interactions between DNA viruses, ND10 and the DNA damage response. Cellular Microbiology. 2006;8(3):365-74. doi: 10.1111/j.1462-5822.2005.00677.x.
8. McKinney CC, Hussmann KL, McBride AA. The Role of the DNA Damage Response throughout the Papillomavirus Life Cycle. Viruses. 2015;7(5):2450-69. doi: 10.3390/v7052450. PubMed PMID: PMC4452914.
9. Moody CA, Laimins LA. Human Papillomaviruses Activate the ATM DNA Damage Pathway for Viral Genome Amplification upon Differentiation. PLoS Pathogens. 2009;5(10):e1000605. doi: 10.1371/journal.ppat.1000605. PubMed PMID: PMC2745661.
10. Edwards TG, Koeller KJ, Slomczynska U, Fok K, Helmus M, Bashkin JK, Fisher C. HPV episome levels are potently decreased by pyrrole-imidazole polyamides. Antiviral Res. 2011;91(2):177-86. Epub 2011/06/15. doi: 10.1016/j.antiviral.2011.05.014. PubMed PMID: 21669229; PMCID: PMC3135741.
11. He G, Vasilieva E, Harris GD, Jr., Koeller KJ, Bashkin JK, Dupureur CM. Binding studies of a large antiviral polyamide to a natural HPV sequence. Biochimie. 2014;102:83-91. Epub 2014/03/04. doi: 10.1016/j.biochi.2014.02.011. PubMed PMID: 24582833; PMCID: PMC4047171.
12. Koeller KJ, Harris GD, Aston K, He G, Castaneda CH, Thornton MA, Edwards TG, Wang S, Nanjunda R, Wilson WD, Fisher C, Bashkin JK. DNA Binding Polyamides and the Importance of DNA Recognition in their use as Gene-Specific and Antiviral Agents. Med Chem (Los Angeles). 2014;4:338-44. Epub 2014/05/20. doi: 10.4172/2161-0444.1000162. PubMed PMID: 24839583; PMCID: PMC4022477.
13. Bernard H-U, Burk RD, Chen Z, van Doorslaer K, Hausen Hz, de Villiers E-M. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. Virology. 2010;401(1):70-9. doi: http://dx.doi.org/10.1016/j.virol.2010.02.002.
14. Hesselberth JR, Chen X, Zhang Z, Sabo PJ, Sandstrom R, Reynolds AP, Thurman RE, Neph S, Kuehn MS, Noble WS, Fields S, Stamatoyannopoulos JA. Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nat Methods. 2009;6(4):283-9. Epub 2009/03/24. doi: 10.1038/nmeth.1313. PubMed PMID: 19305407; PMCID: PMC2668528.

15. Hong S, Laimins LA. Regulation of the life cycle of HPVs by differentiation and the DNA damage response. Future microbiology. 2013;8:1547-57. doi: 10.2217/fmb.13.127. PubMed PMID: PMC3951404.
16. Hughes M, Gao L. Skin cancer viruses: bench to bedside – HPV, HHV8 and Merkel cell carcinoma virus. Drug Discovery Today: Disease Mechanisms. 2013;10(3–4):e91-e4. doi: http://dx.doi.org/10.1016/j.ddmec.2013.05.001.
17. Statistics H-AC. Center for Disease Control and Prevention, HPV-Associated Cancers Statistics. . 2015.
18. The Cancer Genome Atlas N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature. 2015;517(7536):576-82. doi: 10.1038/nature14129 http://www.nature.com/nature/journal/v517/n7536/abs/nature14129.html#supplementary-information.
19. Gillespie KA, Mehta KP, Laimins LA, Moody CA. Human Papillomaviruses Recruit Cellular DNA Repair and Homologous Recombination Factors to Viral Replication Centers. Journal of Virology. 2012;86(17):9520-6. doi: 10.1128/JVI.00247-12. PubMed PMID: PMC3416172.
20. Klingelhutz AJ, Roman A. Cellular transformation by human papillomaviruses: lessons learned by comparing high- and low-risk viruses. Virology. 2012;424(2):77-98. Epub 2012/01/31. doi: 10.1016/j.virol.2011.12.018. PubMed PMID: 22284986; PMCID: PMC3703738.
21. Doorbar J. Molecular biology of human papillomavirus infection and cervical cancer. Clinical Science (London). 2006;110(5):525-41.
22. Vinokurova S, Wentzensen N, Kraus I, Klaes R, Driesch C, Melsheimer P, Kisseljov F, Durst M, Schneider A, von Knebel Doeberitz M. Type-dependent integration frequency of human papillomavirus genomes in cervical lesions. Cancer Res. 2008;68(1):307-13. Epub 2008/01/04. doi: 10.1158/0008-5472.can-07-2754. PubMed PMID: 18172324.
23. Kurvinen K, Yliskoski M, Saarikoski S, Syrjänen K, Syrjänen S. Variants of the long control region of human papillomavirus type 16. European Journal of Cancer. 2000;36(11):1402-10. doi: 10.1016/S0959-8049(00)00121-0.
24. Kämmer C, Tommasino M, Syrjänen S, Delius H, Hebling U, Warthorst U, Pfister H, Zehbe I. Variants of the long control region and the E6 oncogene in European human papillomavirus type 16 isolates: implications for cervical disease. British Journal of Cancer. 2002;86(2):269-73. doi: 10.1038/sj.bjc.6600024. PubMed PMID: PMC2375181.
25. Eichten A, Rud DS, Grace M, Piboonniyom S, Zacny V, Munger K. Molecular pathways executing the "trophic sentinel" response in HPV-16 E7-expressing normal human diploid fibroblasts upon growth factor deprivation, Virology 319, 81-93. Virology. 2004;319:81-93.
26. Bashkin J, Edwards TG, Fisher C, Harris GD, Koeller KJ, inventorsGuanidinyl-substituted polyamides useful for treating human papilloma virus2013.
27. Archambault J, Melendy T. Targeting human papillomavirus genome replication for antiviral drug discovery. Antivir Ther. 2013;18(3):271-83. doi: 10.3851/imp2612. PubMed PMID: 23615820.
28. Day PM, Schelhaas M. Concepts of papillomavirus entry into host cells. Current opinion in virology. 2014;0:24-31. doi: 10.1016/j.coviro.2013.11.002. PubMed PMID: PMC3951680.
29. Straub E, Dreer M, Fertey J, Iftner T, Stubenrauch F. The Viral E8^E2C Repressor Limits Productive Replication of Human Papillomavirus 16. Journal of Virology. 2014;88(2):937-47. doi: 10.1128/JVI.02296-13. PubMed PMID: PMC3911659.
30. Lace MJ, Anson JR, Thomas GS, Turek LP, Haugen TH. The E8(∧)E2 Gene Product of Human Papillomavirus Type 16 Represses Early Transcription and Replication but Is Dispensable for Viral Plasmid Persistence in Keratinocytes. Journal of Virology. 2008;82(21):10841-53. doi: 10.1128/JVI.01481-08. PubMed PMID: PMC2573160.

31. Harden ME, Munger K. HUMAN PAPILLOMAVIRUS MOLECULAR BIOLOGY. Mutation Research/Reviews in Mutation Research. 2016. doi: http://dx.doi.org/10.1016/j.mrrev.2016.07.002.
32. Porras C, Bennett C, Safaeian M, Coseo S, Rodríguez AC, González P, Hutchinson M, Jiménez S, Sherman ME, Wacholder S, Solomon D, van Doorn L-J, Bougelet C, Quint W, Schiffman M, Herrero R, Hildesheim A. Determinants of seropositivity among HPV-16/18 DNA positive young women. BMC Infectious Diseases. 2010;10:238-. doi: 10.1186/1471-2334-10-238. PubMed PMID: PMC2931509.
33. Hoffmann R, Hirt B, Bechtold V, Beard P, Raj K. Different modes of human papillomavirus DNA replication during maintenance. J Virol. 2006;80:4431 - 9. PubMed PMID: doi:10.1128/JVI.80.9.4431-4439.2006.
34. Fradet-Turcotte A, Bergeron-Labrecque F, Moody CA, Lehoux M, Laimins LA, Archambault J. Nuclear Accumulation of the Papillomavirus E1 Helicase Blocks S-Phase Progression and Triggers an ATM-Dependent DNA Damage Response. Journal of Virology. 2011;85(17):8996-9012. doi: 10.1128/JVI.00542-11. PubMed PMID: PMC3165840.
35. Jang MK, Shen K, McBride AA. Papillomavirus Genomes Associate with BRD4 to Replicate at Fragile Sites in the Host Genome. PLoS Pathogens. 2014;10(5):e1004117. doi: 10.1371/journal.ppat.1004117. PubMed PMID: PMC4022725.
36. McBride AA, Jang MK. Current Understanding of the Role of the Brd4 Protein in the Papillomavirus Lifecycle. Viruses. 2013;5(6):1374-94. doi: 10.3390/v5061374. PubMed PMID: PMC3717712.
37. Parish JL, Bean AM, Park RB, Androphy EJ. ChlR1 is required for loading papillomavirus E2 onto mitotic chromosomes and viral genome maintenance. Mol Cell. 2006;24(6):867-76. Epub 2006/12/26. doi: 10.1016/j.molcel.2006.11.005. PubMed PMID: 17189189.
38. Parish JL, Rosa J, Wang X, Lahti JM, Doxsey SJ, Androphy EJ. The DNA helicase ChlR1 is required for sister chromatid cohesion in mammalian cells. J Cell Sci. 2006;119(Pt 23):4857-65. Epub 2006/11/16. doi: 10.1242/jcs.03262. PubMed PMID: 17105772.
39. Helfer CM, Wang R, You J. Analysis of the papillomavirus E2 and bromodomain protein Brd4 interaction using bimolecular fluorescence complementation. PLoS One. 2013;8(10):e77994. Epub 2013/11/10. doi: 10.1371/journal.pone.0077994. PubMed PMID: 24205059; PMCID: PMC3808292.
40. Chang SW, Liu WC, Liao KY, Tsao YP, Hsu PH, Chen SL. Phosphorylation of HPV-16 E2 at serine 243 enables binding to Brd4 and mitotic chromosomes. PLoS One. 2014;9(10):e110882. Epub 2014/10/24. doi: 10.1371/journal.pone.0110882. PubMed PMID: 25340539; PMCID: PMC4207782.
41. Floyd SR, Pacold ME, Huang Q, Clarke SM, Lam FC, Cannell IG, Bryson BD, Rameseder J, Lee MJ, Blake EJ, Fydrych A, Ho R, Greenberger BA, Chen GC, Maffa A, Del Rosario AM, Root DE, Carpenter AE, Hahn WC, Sabatini DM, Chen CC, White FM, Bradner JE, Yaffe MB. The Bromodomain Protein Brd4 Insulates Chromatin from DNA Damage Signaling. Nature. 2013;498(7453):246-50. doi: 10.1038/nature12147. PubMed PMID: PMC3683358.
42. Mehta K, Gunasekharan V, Satsuka A, Laimins LA. Human papillomaviruses activate and recruit SMC1 cohesin proteins for the differentiation-dependent life cycle through association with CTCF insulators. PLoS Pathog. 2015;11(4):e1004763. Epub 2015/04/16. doi: 10.1371/journal.ppat.1004763. PubMed PMID: 25875106; PMCID: PMC4395367.
43. Sakakibara N, Chen D, Jang MK, Kang DW, Luecke HF, Wu S-Y, Chiang C-M, McBride AA. Brd4 Is Displaced from HPV Replication Factories as They Expand and Amplify Viral DNA. PLoS Pathogens. 2013;9(11):e1003777. doi: 10.1371/journal.ppat.1003777. PubMed PMID: PMC3836737.

44. Moody CA, Laimins LA. The Life Cycle of Human Papillomaviruses. In: Damania B, Pipas JM, editors. DNA Tumor Viruses. New York, NY: Springer US; 2009. p. 75-104.
45. Kajitani N, Satsuka A, Kawate A, Sakai H. Productive lifecycle of human papillomaviruses that depends upon squamous epithelial differentiation. Frontiers in Microbiology. 2012;3. doi: 10.3389/fmicb.2012.00152.
46. Johansson C, Schwartz S. Regulation of human papillomavirus gene expression by splicing and polyadenylation. Nat Rev Micro. 2013;11(4):239-51.
47. Luftig MA. Viruses and the DNA Damage Response: Activation and Antagonism. Annu Rev Virol. 2014;1(1):605-25. Epub 2014/11/03. doi: 10.1146/annurev-virology-031413-085548. PubMed PMID: 26958736.
48. Nikitin PA, Luftig MA. At a crossroads: human DNA tumor viruses and the host DNA damage response. Future Virol. 2011;6(7):813-30. Epub 2011/09/20. doi: 10.2217/fvl.11.55. PubMed PMID: 21927617; PMCID: PMC3171950.
49. Blackledge MS, Melander C. Programmable DNA-binding small molecules. Bioorg Med Chem. 2013;21(20):6101-14. Epub 2013/05/15. doi: 10.1016/j.bmc.2013.04.023. PubMed PMID: 23665141; PMCID: PMC3789866.
50. Luck G, Zimmer C, Reinert K-E, Arcamone F. Specific interactions of distamycin A and its analogs with (A-T) rich and (G-C) rich duplex regions of DNA and deoxypolynucleotides. Nucleic Acids Research. 1977;4(8):2655-70. doi: 10.1093/nar/4.8.2655.
51. Zimmer C, Puschendorf B, Grunicke H, Chandra P, Venner H. Influence of Netropsin and Distamycin A on the Secondary Structure and Template Activity of DNA. European Journal of Biochemistry. 1971;21(2):269-78. doi: 10.1111/j.1432-1033.1971.tb01466.x.
52. Satz AL, Bruice TC. Recognition in the Minor Groove of Double-Stranded DNA by Microgonotropens. Accounts of Chemical Research. 2002;35(2):86-95. doi: 10.1021/ar0101032.
53. Dervan P. Design of sequence-specific DNA-binding molecules. Science. 1986;232(4749):464-71. doi: 10.1126/science.2421408.
54. Dervan PB. Molecular recognition of DNA by small molecules. Bioorganic & Medicinal Chemistry. 2001;9(9):2215-35. doi: http://dx.doi.org/10.1016/S0968-0896(01)00262-0.
55. Rucker VC, Melander C, Dervan PB. Influence of beta-Alanine on Hairpin Polyamide Orientation in the DNA Minor Groove. Helvetica Chimica Acta. 2003;86:1839-51.
56. Wang S, Aston K, Koeller KJ, Harris GD, Jr., Rath NP, Bashkin JK, Wilson WD. Modulation of DNA-polyamide interaction by beta-alanine substitutions: a study of positional effects on binding affinity, kinetics and thermodynamics. Org Biomol Chem. 2014;12(38):7523-36. Epub 2014/08/21. doi: 10.1039/c4ob01456a. PubMed PMID: 25141096; PMCID: PMC4339220.
57. Bashkin JK, Aston K, Ramos JP, Koeller KJ, Nanjunda R, He G, Dupureur CM, David Wilson W. Promoter scanning of the human COX-2 gene with 8-ring polyamides: unexpected weakening of polyamide-DNA binding and selectivity by replacing an internal N-Me-pyrrole with beta-alanine. Biochimie. 2013;95(2):271-9. Epub 2012/10/02. doi: 10.1016/j.biochi.2012.09.023. PubMed PMID: 23023196; PMCID: PMC3557744.
58. Wang S, Nanjunda R, Aston K, Bashkin JK, Wilson WD. Correlation of local effects of DNA sequence and position of beta-alanine inserts with polyamide-DNA complex binding affinities and kinetics. Biochemistry. 2012;51(49):9796-806. Epub 2012/11/22. doi: 10.1021/bi301327v. PubMed PMID: 23167504; PMCID: PMC3567211.
59. Edelson BS, Best TP, Olenyuk B, Nickols NG, Doss RM, Foister S, Heckel A, Dervan PB. Influence of structural variation on nuclear localization of DNA-binding polyamide-fluorophore conjugates. Nucleic Acids Research. 2004;32(9):2802-18. doi: 10.1093/nar/gkh609. PubMed PMID: PMC419610.

60. Dervan PB, Edelson BS. Recognition of the DNA minor groove by pyrrole-imidazole polyamides. Current Opinion in Structural Biology. 2003;13(3):284-99. doi: http://dx.doi.org/10.1016/S0959-440X(03)00081-2.
61. Chenoweth DM, Dervan PB. Allosteric modulation of DNA by small molecules. Proceedings of the National Academy of Sciences. 2009;106(32):13175-9. doi: 10.1073/pnas.0906532106.
62. Peter B. Dervan, Adam T. Poulin-Kerstien, Eric J. Fechter, Edelson BS. Regulation of Gene Expression by Synthetic DNA-Binding Ligands. Top Curr Chem. 2005;253:1-31.
63. Olenyuk BZ, Zhang G-J, Klco JM, Nickols NG, Kaelin WG, Dervan PB. Inhibition of vascular endothelial growth factor with a sequence-specific hypoxia response element antagonist. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(48):16768-73. doi: 10.1073/pnas.0407617101.
64. Yasuda A, Noguchi K, Minoshima M, Kashiwazaki G, Kanda T, Katayama K, Mitsuhashi J, Bando T, Sugiyama H, Sugimoto Y. DNA ligand designed to antagonize EBNA1 represses Epstein-Barr virus-induced immortalization. Cancer Sci. 2011;102(12):2221-30. Epub 2011/09/14. doi: 10.1111/j.1349-7006.2011.02098.x. PubMed PMID: 21910783.
65. Chou CJ, Farkas ME, Tsai SM, Alvarez D, Dervan PB, Gottesfeld JM. Small molecules targeting Histone H4 as potential therapeutics for chronic myelogenous leukemia. Molecular cancer therapeutics. 2008;7(4):769-78. doi: 10.1158/1535-7163.MCT-08-0130. PubMed PMID: PMC2376757.
66. Moyzis RK, Buckingham JM, Cram LS, Dani M, Deaven LL, Jones MD, Meyne J, Ratliff RL, Wu JR. A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. Proc Natl Acad Sci U S A. 1988;85(18):6622-6. Epub 1988/09/01. PubMed PMID: 3413114; PMCID: PMC282029.
67. Kashiwazaki G, Bando T, Shinohara K-i, Minoshima M, Nishijima S, Sugiyama H. Cooperative alkylation of double-strand human telomere repeat sequences by PI polyamides with 11-base-pair recognition based on a heterotrimeric design. Bioorganic & Medicinal Chemistry. 2009;17(3):1393-7. doi: http://dx.doi.org/10.1016/j.bmc.2008.12.019.
68. Pandian GN, Taniguchi J, Junetha S, Sato S, Han L, Saha A, AnandhaKumar C, Bando T, Nagase H, Vaijayanthi T, Taylor RD, Sugiyama H. Distinct DNA-based epigenetic switches trigger transcriptional activation of silent genes in human dermal fibroblasts. Scientific Reports. 2014;4:3843. doi: 10.1038/srep03843 http://www.nature.com/articles/srep03843#supplementary-information.
69. Pandian G, Sugiyama H. Strategies To Modulate Heritable Epigenetic Defects in Cellular Machinery: Lessons from Nature. Pharmaceuticals. 2012;6(1):1. PubMed PMID: doi:10.3390/ph6010001.
70. Crowley KS, Phillion DP, Woodard SS, Schweitzer BA, Singh M, Shabany H, Burnette B, Hippenmeyer P, Heitmeier M, Bashkin JK. Controlling the intracellular localization of fluorescent polyamide analogues in cultured cells. Bioorg Med Chem Lett. 2003;13(9):1565-70. Epub 2003/04/18. PubMed PMID: 12699756.
71. Nickols NG, Jacobs CS, Farkas ME, Dervan PB. Improved nuclear localization of DNA-binding polyamides. Nucleic Acids Research. 2007;35(2):363-70. doi: 10.1093/nar/gkl1042.
72. Vaijayanthi T, Bando T, Pandian GN, Sugiyama H. Progress and Prospects of Pyrrole-Imidazole Polyamide–Fluorophore Conjugates as Sequence-Selective DNA Probes. ChemBioChem. 2012;13(15):2170-85. doi: 10.1002/cbic.201200451.
73. Nishijima S, Shinohara K-i, Bando T, Minoshima M, Kashiwazaki G, Sugiyama H. Cell permeability of Py–Im-polyamide-fluorescein conjugates: Influence of molecular size and Py/Im

content. Bioorganic & Medicinal Chemistry. 2010;18(2):978-83. doi: http://dx.doi.org/10.1016/j.bmc.2009.07.018.
74. Best TP, Edelson BS, Nickols NG, Dervan PB. Nuclear localization of pyrrole–imidazole polyamide–fluorescein conjugates in cell culture. Proceedings of the National Academy of Sciences of the United States of America. 2003;100(21):12063-8. doi: 10.1073/pnas.2035074100. PubMed PMID: PMC218713.
75. Hsu CF, Dervan PB. Quantitating the concentration of Py-Im polyamide-fluorescein conjugates in live cells. Bioorganic & medicinal chemistry letters. 2008;18(22):5851-5. doi: 10.1016/j.bmcl.2008.05.063. PubMed PMID: 18524590.
76. Sasaki S, Minoshima M, Shimizu T, Fujimoto J, Shinohara K-i, Bando T, Sugiyama H. The biological impact of sequence-specific DNA alkylation by pyrrole-imidazole polyamides. Nucleic Acids Symposium Series. 2006;50(1):111-2. doi: 10.1093/nass/nrl055.
77. Dickinson LA, Burnett R, Melander C, Edelson BS, Arora PS, Dervan PB, Gottesfeld JM. Arresting cancer proliferation by small-molecule gene regulation. Chem Biol. 2004;11(11):1583-94. Epub 2004/11/24. doi: 10.1016/j.chembiol.2004.09.004. PubMed PMID: 15556009.
78. Martinez TF, Phillips JW, Karanja KK, Polaczek P, Wang CM, Li BC, Campbell JL, Dervan PB. Replication stress by Py-Im polyamides induces a non-canonical ATR-dependent checkpoint response. Nucleic Acids Res. 2014;42(18):11546-59. Epub 2014/09/25. doi: 10.1093/nar/gku866. PubMed PMID: 25249630; PMCID: PMC4191428.
79. Belitsky JM, Leslie SJ, Arora PS, Beerman TA, Dervan PB. Cellular uptake of N-methylpyrrole/N-methylimidazole polyamide-dye conjugates. Bioorganic & Medicinal Chemistry. 2002;10(10):3313-8. doi: http://dx.doi.org/10.1016/S0968-0896(02)00204-3.
80. Suto RK, Edayathumangalam RS, White CL, Melander C, Gottesfeld JM, Dervan PB, Luger K. Crystal structures of nucleosome core particles in complex with minor groove DNA-binding ligands. J Mol Biol. 2003;326(2):371-80. Epub 2003/02/01. PubMed PMID: 12559907.
81. Gottesfeld JM, Melander C, Suto RK, Raviol H, Luger K, Dervan PB. Sequence-specific Recognition of DNA in the Nucleosome by Pyrrole-Imidazole Polyamides. Journal of Molecular Biology. 2001;309(3):615-29. doi: http://dx.doi.org/10.1006/jmbi.2001.4694.
82. Jespersen C, Soragni E, Chou CJ, Arora PS, Dervan PB, Gottesfeld JM. Chromatin structure determines accessibility of a hairpin polyamide-chlorambucil conjugate at histone H4 genes in pancreatic cancer cells. Bioorganic & medicinal chemistry letters. 2012;22(12):4068-71. doi: 10.1016/j.bmcl.2012.04.090. PubMed PMID: PMC3362666.
83. Baird EE, Dervan PB. Solid Phase Synthesis of Polyamides Containing Imidazole and Pyrrole Amino Acids. Journal of the American Chemical Society. 1996;118(26):6141-6. doi: 10.1021/ja960720z.
84. Pauff SM, Fallows AJ, Mackay SP, Su W, Cullis PM, Burley GA. Pyrrole-Imidazole Polyamides: Manual Solid-Phase Synthesis. Curr Protoc Nucleic Acid Chem. 2015;63:8 10 1-41. Epub 2015/12/02. doi: 10.1002/0471142700.nc0810s63. PubMed PMID: 26623975.
85. Carpino LA. 1-Hydroxy-7-azabenzotriazole. An efficient peptide coupling additive. Journal of the American Chemical Society. 1993;115(10):4397-8. doi: 10.1021/ja00063a082.
86. He G, Koeller KJ, Harris GD, Dupureur CM, Bashkin JK. DNA binding properties of a large antiviral polyamide via capillary electrophoresis. 243rd ACS National Meeting & Exposition; March 25-29, 2012; San Diego, CA, United States: American Chemical Society; 2012. p. BIOL-254.
87. Edwards TG, Helmus, Michael J., Koeller, Kevin, Bashkin, James K.,Fisher, Chris. Human Papillomavirus Episome Stability Is Reduced by Aphidicolin and Controlled by DNA Damage Response Pathways. Journal of Virology.87(7):3979-89. doi: 10.1128/jvi.03473-12.

88. Edwards TG, Koeller KJ, Slomczynska U, Fok K, Helmus M, Bashkin JK, Fisher C. HPV episome levels are potently decreased by pyrrole-imidazole polyamides. Antiviral Research.91(2):177-86.
89. Fukasawa A, Nagashima T, Aoyama T, Fukuda N, Matsuda H, Ueno T, Sugiyama H, Nagase H, Matsumoto Y. Optimization and validation of a high-performance liquid chromatographic method with UV detection for the determination of pyrrole-imidazole polyamides in rat plasma. J Chromatogr B Analyt Technol Biomed Life Sci. 2007;859(2):272-5. Epub 2007/10/24. doi: 10.1016/j.jchromb.2007.09.032. PubMed PMID: 17950049.
90. Chomczynski P, Mackey K, Drews R, Wilfinger W. DNAzol: a reagent for the rapid isolation of genomic DNA. Biotechniques. 1997;22(3):550-3. Epub 1997/03/01. PubMed PMID: 9067036.
91. Raskatov JA, Hargrove AE, So AY, Dervan PB. Pharmacokinetics of Py-Im polyamides depend on architecture: cyclic versus linear. J Am Chem Soc. 2012;134(18):7995-9. Epub 2012/04/19. doi: 10.1021/ja302588v. PubMed PMID: 22509786; PMCID: PMC3349014.
92. Chuah BY, Putti T, Salto-Tellez M, Charlton A, Iau P, Buhari SA, Wong CI, Tan SH, Wong AL, Chan CW, Goh BC, Lee SC. Serial changes in the expression of breast cancer-related proteins in response to neoadjuvant chemotherapy. Ann Oncol. 2011;22(8):1748-54. Epub 2011/03/01. doi: 10.1093/annonc/mdq755. PubMed PMID: 21355070.
93. Boer F. Drug handling by the lungs. Br J Anaesth. 2003;91(1):50-60. Epub 2003/06/25. PubMed PMID: 12821565.
94. Synold TW, Xi B, Wu J, Yen Y, Li BC, Yang F, Phillips JW, Nickols NG, Dervan PB. Single-dose pharmacokinetic and toxicity analysis of pyrrole-imidazole polyamides in mice. Cancer Chemother Pharmacol. 2012;70(4):617-25. Epub 2012/08/22. doi: 10.1007/s00280-012-1954-3. PubMed PMID: 22907527; PMCID: PMC3456924.
95. Taylor JS, Schultz PG, Dervan PB. DNA Affinity Cleaving. Sequence Specific Cleavage of DNA by Distamycin-EDTA·Fe(II) and EDTA-Distamycin·Fe(II). Tetrahedron 1984;40:457-65.
96. He G, Vasilieva E, Bashkin JK, Dupureur CM. Mapping small DNA ligand hydroxyl radical footprinting and affinity cleavage products for capillary electrophoresis. Analytical Biochemistry. 2013;439(2):99-101. doi: http://dx.doi.org/10.1016/j.ab.2013.04.011.
97. Cantor CR, Schimmel PR. Biophysical Chemistry Part III: The Behavior of Biological Macromolecules: W. H. Freeman and Co., New York; 1980.
98. Maxam AM, Gilbert W. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. 1980;65(1):499-560. Epub 1980/01/01. PubMed PMID: 6246368.
99. Hertzberg RP, Dervan PB. Cleavage of DNA with methidiumpropyl-EDTA-iron(II): reaction conditions and product analyses. Biochemistry. 1984;23(17):3934-45. doi: 10.1021/bi00312a022.
100. Shultz PG, Taylor JS, Dervan PB. Design synthesis of a sequence-specific DNA cleaving molecule. (Distamycin-EDTA)iron(II). Journal of the American Chemical Society. 1982;104(24):6861-3. doi: 10.1021/ja00388a101.
101. Heddi B, Abi-Ghanem J, Lavigne M, Hartmann B. Sequence-Dependent DNA Flexibility Mediates DNase I Cleavage. Journal of Molecular Biology. 2010;395(1):123-33. doi: http://dx.doi.org/10.1016/j.jmb.2009.10.023.
102. Gueroult M, Picot D, Abi-Ghanem J, Hartmann B, Baaden M. How cations can assist DNase I in DNA binding and hydrolysis. PLoS Comput Biol. 2010;6(11):e1001000. Epub 2010/12/03. doi: 10.1371/journal.pcbi.1001000. PubMed PMID: 21124947; PMCID: PMC2987838.

103. Koohy H, Down TA, Hubbard TJ. Chromatin Accessibility Data Sets Show Bias Due to Sequence Specificity of the DNase I Enzyme. PLoS ONE. 2013;8(7):e69853. doi: 10.1371/journal.pone.0069853. PubMed PMID: PMC3724795.
104. Wilson WD, Tanious FA, Barton HJ, Jones RL, Fox K, Wydra RL, Strekowski L. DNA sequence dependent binding modes of 4',6-diamidino-2-phenylindole (DAPI). Biochemistry. 1990;29(36):8452-61. doi: 10.1021/bi00488a036.
105. Vasilieva E, Niederschulte J, Song Y, Harris GD, Jr., Koeller KJ, Liao P, Bashkin JK, Dupureur CM. Interactions of two large antiviral polyamides with the long control region of HPV16. Biochimie. 2016. Epub 2016/05/08. doi: 10.1016/j.biochi.2016.04.022. PubMed PMID: 27155361.
106. Fenton HJH. LXXIII.-Oxidation of tartaric acid in presence of iron. Journal of the Chemical Society, Transactions. 1894;65(0):899-910. doi: 10.1039/CT8946500899.
107. Pogozelski WK, McNeese TJ, Tullius TD. What species is responsible for strand scission in the reaction of [FeIIEDTA]2− and H2O2 with DNA? . J Am Chem Soc 1995(117):6428–33.
108. Schultz PG, Taylor JS, Dervan PB. Design synthesis of a sequence-specific DNA cleaving molecule. (Distamycin-EDTA)iron(II). Journal of the American Chemical Society. 1982;104(24):6861-3. doi: 10.1021/ja00388a101.
109. Herman DM, Baird EE, Dervan PB. Stereochemical Control of the DNA Binding Affinity, Sequence Specificity, and Orientation Preference of Chiral Hairpin Polyamides in the Minor Groove. Journal of the American Chemical Society. 1998;120(7):1382-91. doi: 10.1021/ja9737228.
110. Sivapragasam S, Pande A, Grove A. A recommended workflow for DNase I footprinting using a capillary electrophoresis genetic analyzer. Analytical Biochemistry. 2015;481(0):1-3. doi: http://dx.doi.org/10.1016/j.ab.2015.04.013.
111. Dalianis T. Human papillomavirus and oropharyngeal cancer, the epidemics, and significance of additional clinical biomarkers for prediction of response to therapy. International Journal of Oncology. 2014;44(6):1799-805. doi: 10.3892/ijo.2014.2355. PubMed PMID: PMC4063535.
112. Edwards TG, Helmus MJ, Koeller K, Bashkin JK, Fisher C. Human papillomavirus episome stability is reduced by aphidicolin and controlled by DNA damage response pathways. J Virol. 2013;87(7):3979-89. Epub 2013/02/01. doi: 10.1128/jvi.03473-12. PubMed PMID: 23365423; PMCID: PMC3624211.
113. Weitzman MD, Lilley CE, Chaurushiya MS. Genomes in conflict: maintaining genome integrity during virus infection. Annu Rev Microbiol. 2010;64:61-81. Epub 2010/08/10. doi: 10.1146/annurev.micro.112408.134016. PubMed PMID: 20690823.
114. Wallace NA, Galloway DA. Manipulation of cellular DNA damage repair machinery facilitates propagation of human papillomaviruses. Semin Cancer Biol. 2014;26:30-42. Epub 2014/01/15. doi: 10.1016/j.semcancer.2013.12.003. PubMed PMID: 24412279; PMCID: PMC4050178.
115. Pittayakhajonwut D, Angeletti P. Viral trans-factor independent replication of human papillomavirus genomes. Virology Journal. 2010;7(1):123. PubMed PMID: doi:10.1186/1743-422X-7-123.
116. Flores ER, Allen-Hoffmann BL, Lee D, Sattler CA, Lambert PF. Establishment of the human papillomavirus type 16 (HPV-16) life cycle in an immortalized human foreskin keratinocyte cell line. Virology. 1999;262(2):344-54. Epub 1999/09/30. doi: 10.1006/viro.1999.9868. PubMed PMID: 10502513.
117. Fisher C. Recent Insights into the Control of Human Papillomavirus (HPV) Genome Stability, Loss, and Degradation. J Clin Med 2015;4(2): 204-30. doi: doi:10.3390/jcm4020204.

118. Stünkel W, Bernard H-U. The Chromatin Structure of the Long Control Region of Human Papillomavirus Type 16 Represses Viral Oncoprotein Expression. Journal of Virology. 1999;73(3):1918-30. PubMed PMID: PMC104433.
119. Schuck S, Stenlund A. Role of Papillomavirus E1 Initiator Dimerization in DNA Replication. Journal of Virology. 2005;79(13):8661-4. doi: 10.1128/JVI.79.13.8661-8664.2005. PubMed PMID: PMC1143733.
120. Dervan PB. Design of Sequence-Specific DNA-Binding Molecules. Science. 1986;232:464-71.
121. Marky LA, Breslauer KJ. Origins of netropsin binding affinity and specificity: correlations of thermodynamic and structural data. Proc Natl Acad Sci U S A. 1987;84(13):4359-63. Epub 1987/07/01. PubMed PMID: 3037518; PMCID: PMC305088.
122. Bremer RE, Baird EE, Dervan PB. Inhibition of major-groove-binding proteins by pyrrole-imidazole polyamides with an Arg-Pro-Arg positive patch. Chem Biol. 1998;5(3):119-33. Epub 1998/04/18. PubMed PMID: 9545429.
123. Bashkin JK, Aston K, Ramos JP, Koeller KJ, Nanjunda R, He G, Dupureur CM, David Wilson W. Promoter scanning of the human COX-2 gene with 8-ring polyamides: Unexpected weakening of polyamide-DNA binding and selectivity by replacing an internal N-Me-pyrrole with Î²-alanine. Biochimie. 2013;95(2):271-9. doi: http://dx.doi.org/10.1016/j.biochi.2012.09.023.
124. Copeland RA, Pompliano DL, Meek TD. Drug-target residence time and its implications for lead optimization. Nat Rev Drug Discov. 2006;5(9):730-9.
125. Tummino PJ, Copeland RA. Residence time of receptor-ligand complexes and its effect on biological function. Biochemistry. 2008;47(20):5481-92. Epub 2008/04/17. doi: 10.1021/bi8002023. PubMed PMID: 18412369.
126. Bergvall M, Melendy T, Archambault J. The E1 proteins. Virology. 2013;445(1–2):35-56. doi: http://dx.doi.org/10.1016/j.virol.2013.07.020.
127. Melendy T, Sedman J, Stenlund A. Cellular factors required for papillomavirus DNA replication. J Virol. 1995;69:7857 - 67.
128. Auster AS, Joshua-Tor L. The DNA-binding domain of human papillomavirus type 18 E1. Crystal structure, dimerization, and DNA binding. J Biol Chem. 2004;279(5):3733-42. Epub 2003/11/01. doi: 10.1074/jbc.M311681200. PubMed PMID: 14593106.
129. Schaal TD, Mallet WG, McMinn DL, Nguyen NV, Sopko MM, John S, Parekh BS. Inhibition of human papilloma virus E2 DNA binding protein by covalently linked polyamides. Nucleic Acids Research. 2003;31(4):1282-91. PubMed PMID: PMC150225.
130. Wexselblatt E, Esko JD, Tor Y. On Guanidinium and Cellular Uptake. The Journal of Organic Chemistry. 2014;79(15):6766-74. doi: 10.1021/jo501101s.
131. Herce HD, Garcia AE, Litt J, Kane RS, Martin P, Enrique N, Rebolledo A, Milesi V. Arginine-rich peptides destabilize the plasma membrane, consistent with a pore formation translocation mechanism of cell-penetrating peptides. Biophys J. 2009;97(7):1917-25. Epub 2009/10/07. doi: 10.1016/j.bpj.2009.05.066. PubMed PMID: 19804722; PMCID: PMC2756373.
132. Andersen AS, Koldjaer Solling AS, Ovesen T, Rusan M. The interplay between HPV and host immunity in head and neck squamous cell carcinoma. Int J Cancer. 2014;134(12):2755-63. Epub 2013/08/06. doi: 10.1002/ijc.28411. PubMed PMID: 23913554.
133. Reinson T, Toots M, Kadaja M, Pipitch R, Allik M, Ustav E, Ustav M. Engagement of the ATR-Dependent DNA Damage Response at the Human Papillomavirus 18 Replication Centers during the Initial Amplification. Journal of Virology. 2013;87(2):951-64. doi: 10.1128/JVI.01943-12. PubMed PMID: PMC3554080.
134. Banerjee NS, Wang HK, Broker TR, Chow LT. Human papillomavirus (HPV) E7 induces prolonged G2 following S phase reentry in differentiated human keratinocytes. J Biol Chem.

2011;286(17):15473-82. Epub 2011/02/16. doi: 10.1074/jbc.M110.197574. PubMed PMID: 21321122; PMCID: PMC3083224.
135. Garner-Hamrick PA, Fisher C. HPV episomal copy number closely correlates with cell size in keratinocyte monolayer cultures. Virology. 2002;301(2):334-41. Epub 2002/10/03. PubMed PMID: 12359435.
136. Jeon S, Allen-Hoffmann BL, Lambert PF. Integration of human papillomavirus type 16 into the human genome correlates with a selective growth advantage of cells. Journal of Virology. 1995;69(5):2989-97. PubMed PMID: PMC188998.
137. Meyers C, Frattini MG, Laimins LA. Tissue Culture Techniques for the Study of Human papillomaviruses in Stratified Epithelia. Cell Biology: A Laboratory Handbook. 1994:491-4.
138. Davy C, Doorbar J. Human Papillomaviruses: Methods and Protocols. Methods in Molecular Medicine, Humana Press. 2005;119.
139. Doorbar J, Parton A, Hartley K, Banks L, Crook T, Stanley M, Crawford L. Detection of novel splicing patterns in a HPV16-containing keratinocyte cell line. Virology. 1990;178(1):254-62. Epub 1990/09/01. PubMed PMID: 2167553.
140. Seedorf K, Krammer G, Durst M, Suhai S, Rowekamp WG. Human papillomavirus type 16 DNA sequence. Virology. 1985;145(1):181-5. Epub 1985/08/01. PubMed PMID: 2990099.
141. McPhillips MG, Veerapraditsin T, Cumming SA, Karali D, Milligan SG, Boner W, Morgan IM, Graham SV. SF2/ASF binds the human papillomavirus type 16 late RNA control element and is regulated during differentiation of virus-infected epithelial cells. J Virol. 2004;78(19):10598-605. Epub 2004/09/16. doi: 10.1128/jvi.78.19.10598-10605.2004. PubMed PMID: 15367627; PMCID: PMC516382.
142. Pett MR, Herdman MT, Palmer RD, Yeo GS, Shivji MK, Stanley MA, Coleman N. Selection of cervical keratinocytes containing integrated HPV16 associates with episome loss and an endogenous antiviral response. Proc Natl Acad Sci U S A. 2006;103(10):3822-7. Epub 2006/03/01. doi: 10.1073/pnas.0600078103. PubMed PMID: 16505361; PMCID: PMC1383496.
143. NIH-GUIDELINES. NIH GUIDELINES FOR RESEARCH INVOLVING RECOMBINANT OR SYNTHETIC NUCLEIC ACID MOLECULES2013.
144. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25(4):402-8. Epub 2002/02/16. doi: 10.1006/meth.2001.1262. PubMed PMID: 11846609.
145. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, Vandesompele J, Wittwer CT. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. Clinical Chemistry. 2009;55(4):611-22. doi: 10.1373/clinchem.2008.112797.
146. Ling D, Salvaterra PM. Robust RT-qPCR Data Normalization: Validation and Selection of Internal Reference Genes during Post-Experimental Data Analysis. PLoS ONE. 2011;6(3):e17762. doi: 10.1371/journal.pone.0017762.
147. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 2002;3(7):RESEARCH0034. Epub 2002/08/20. PubMed PMID: 12184808; PMCID: PMC126239.
148. Klaes R, Woerner SM, Ridder R, Wentzensen N, Duerst M, Schneider A, Lotz B, Melsheimer P, von Knebel Doeberitz M. Detection of high-risk cervical intraepithelial neoplasia and cervical cancer by amplification of transcripts derived from integrated papillomavirus oncogenes. Cancer Res. 1999;59(24):6132-6. Epub 2000/01/08. PubMed PMID: 10626803.
149. Wentzensen N, Ridder R, Klaes R, Vinokurova S, Schaefer U, Doeberitz M. Characterization of viral-cellular fusion transcripts in a large series of HPV16 and 18 positive

anogenital lesions. Oncogene. 2002;21(3):419-26. Epub 2002/02/01. doi: 10.1038/sj.onc.1205104. PubMed PMID: 11821954.
150. Wang X, Meyers C, Wang HK, Chow LT, Zheng ZM. Construction of a full transcription map of human papillomavirus type 18 during productive viral infection. J Virol. 2011;85(16):8080-92. Epub 2011/06/18. doi: 10.1128/jvi.00670-11. PubMed PMID: 21680515; PMCID: PMC3147953.
151. Givan SA, Bottoms CA, Spollen WG. Computational Analysis of RNA-seq. In: Jin H, Gassmann W, editors. RNA Abundance Analysis: Methods and Protocols. New York: Humana Press Springer; 2012. p. 201–19.