4-18-2017

DNA-Binding and Cellular Uptake Properties of Pyrrole-Imidazole Polyamides

Maria J. Scuderi
University of Missouri-St. Louis, scuderim@umsl.edu

DNA-Binding and Cellular Uptake Properties of Pyrrole-Imidazole Polyamides

Maria José Scuderi

Master of Science in Chemistry, University of Missouri – St. Louis, 2014

Master in Chemistry and Pharmaceutical Technologies, Università degli Studi di Catania, 2011

A Dissertation Submitted to The Graduate School of the
University of Missouri-St. Louis
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Chemistry
May, 2017

Advisory Committee

James K. Bashkin, Ph.D. (Chair)

Benjamin J. Bythell, Ph.D.

Keith J. Stine, PhD.

Chung F. Wong, Ph.D.

Abstract

Pyrrole-imidazole polyamides (PAs) represent a class of synthetic ligands that bind to the minor groove of DNA through hydrogen bonds and can be programmed to recognize predetermined DNA sequences. Potential applications of these compounds include gene expression modulation, antimicrobial activity and cell imaging.

Studies conducted in our lab in collaboration with NanoVir have shown that large PAs exhibit antiviral activity against Human Papillomavirus (HPV), which is the most common sexually transmitted infectious agent in the US and the primary factor for the development of cervical cancer. PAs act against HPV by targeting the viral DNA and promoting its selective elimination from cells with a mechanism involving the DNA Damage Response (DDR) pathway, although the details underlining this mechanism are still under investigation. In a library of more than 100 PAs, we observed that even minor changes in polyamide chemical structure can cause dramatic differences in antiviral activity.

The work presented in this thesis focuses on exploring some of the factors that may play a role in determining the different effectiveness of compounds that are structurally similar. The first part of this report describes the work intended to assess the interactions occurring between pyrrole-imidazole polyamides and HPV-DNA fragments. The DNA fragments used in these studies correspond to the Long Control Region of HPV, a regulatory region of primary importance for viral replication and transcription. The preferred binding sites as well as the strength of the binding were investigated

through DNase I footprinting and affinity cleavage coupled with capillary electrophoresis. This work provided information on a molecular level on the binding mode of large PAs with a viral DNA sequence, showing interesting differences in comparison to the shorter PAs usually described in the literature and showing how minor structural substitutions can change the binding mode of these compounds when interacting with DNA. The second part of this report focuses on the investigation of cellular uptake of pyrrole-imidazole polyamides in human cells. HPLC and LC-MS/MS methods were developed to assess cell-associated polyamide. The methods were employed to explore the permeability of polyamides differing in their N-terminal groups or polyamides paired with different counterions.

Acknowledgements

First of all I wish to express my gratitude to my research advisor Dr. James K. Bashkin. He has been an invaluable scientific guidance through these five years of research and has allowed me to work with the right level of independence. I will never stop to be amazed on how extremely knowledgeable he is in just about everything. Not many people are dedicated and passionate about their work as he is.

I also wish to thank all the past and present group members that I have had the pleasure to work with: Dr. George Davis Harris J. and Dr. Kevin J. Koeller who guided me during my experience as a visiting researcher in the synthetic lab, Silke Evdokimov, Priyanka Bapat, Faten Tamimi, Debbie Kirchner, Rachel Schafer, Christopher D. Bognar, Magnus Creed, Ali A. Sabi and Brandon M. Wood. I thank Dr. Gaofei He and Dr. Csiki-Fejér who mentored me during my first years of PhD and Carlos H. Castañeda not only for the insightful research related discussions but especially for being always ready to help.

I am grateful to all the people who helped me achieving and complete my research projects, including the members of my committee Dr. Stine, Dr. Wong and Dr. Bythell for their time, interest, and helpful comments. In particular, I thank Dr. Bythell for being an active part of the uptake project giving me access to his lab facilities and for the help provided by his student Shanshan Guan. I also thank Dr. Dupureur for her guidance in the biophysical projects.

I gratefully acknowledge the funding sources NIH, CAS and the UMSL Chemistry and Biochemistry Department that made my Ph.D. work possible. I also would

like to thank the Chemistry and Biochemistry department faculty members: I could not ask for a friendlier environment. Among the Professors I must thank Dr. Alexei Demchenko and Dr. Cristina De Meo for encouraging me to pursue a PhD degree and for being always supportive through this experience.

Last, but by no means least, I would truly like to thank my husband, my family and friends for being always understanding and for their unconditioned support.

Table of Contents

Page

List of Figures and Tables

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 1

INTRODUCTION

1.1 Human Papillomavirus Infection

1.1.1 Introduction

Human Papillomavirus (HPV) is a small tumor DNA virus known to infect the basal layer of stratified squamous epithelia of the skin, anogenital and upper respiratory tracts.[1,2]

To date, more than 120 different HPV genotypes[8] have been identified and can be classified as low- and high-risk according to their association with cancer.[9] Low-risk (LR) HPV types cause predominantly skin and genital warts or benign lesions, while high-risk (HR) strains are the causative agents of squamous cell carcinoma (SCC) of the anogenital tract and the back of the oral cavity. Nearly all cervical cancers and about 60-70% of oral and throat carcinoma, are caused by HPV,[10] with HPV16 and HPV18 types alone accounting for about 70% of cervical cancers.[11] Overall, at least fifteen different high-risk human papillomavirus types have been identified to date.[12]

HPV infection is very common in young adults and in fact about 50-80% of sexually active women and men will encounter an HPV infection at least once in their lifetime.[13] Up to 50 % of the women infected with HPV will also acquire an infection with a different HPV type.[14]

Although the infection is very common, in most of the cases the immune system is able to

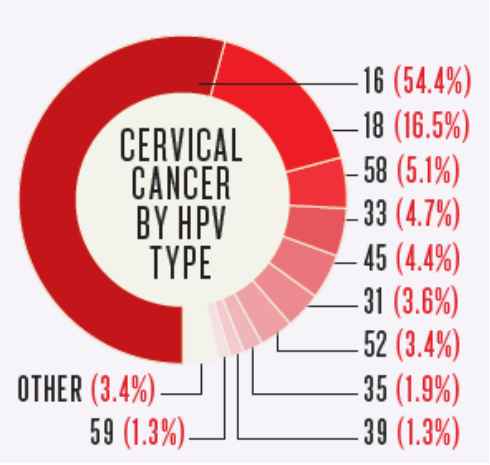


Figure 1.1[5] WHO/ICO information center on HPV and cervical cancer (Permission number 4093280619270)

recognize the viral proteins and generate an antibody response against the virus that clears the infection in 6-12 months.[13] However, failure to develop an appropriate response results in a persistent active infection, increasing the probability of viral DNA integration into the host genome and the progression to precancerous and cancerous lesions.[1,15]

1.1.2 HPV Infectious Cycle

The infection is initiated when the virus enters the epithelial basal cells through microabrasions. During productive infection, the viral genome is found in the nucleus in an extrachromosomal state (episome) and relies on the host machinery for its replication and protein synthesis. In the first stage of the infection the viral DNA is replicated at low copy number (50 to 100 episomes per cells) and, although the expression of viral proteins E6 and E7 is minimal, it is sufficent to enhance proliferation of infected cells.[1,16]

Since viral DNA replication is dependent on keratinocyte differentiation, massive viral replication and protein expression occurs once the virus reaches the suprabasal layers after cell division.[17] Viral assembly and viral particle release occur only in the upper level of the epithelium.[18]

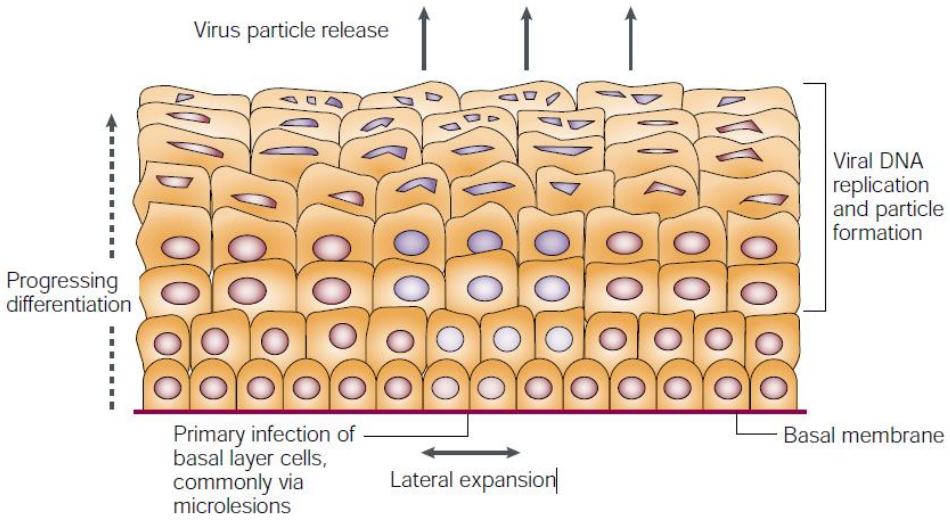


Figure 1.2 HPV life cycle[1] (Permission number 4093261369087)

As mentioned in section 1.1, the infection is in general successfully cleared by the immune system. Nevertherless, in a small percentage of people the infection is not spontaneously resolved and persists for several years. A persistent infection increases the probability of viral DNA integration into the host genome, a key event for the progression to low-grade lesions and eventually to malignant carcinoma.[2] The main consequence of DNA integration is the loss of a regulatory negative feedback loop that leads to the uncontrolled expression of the viral oncogenes E6 and E7. Cells with integrated HPV exhibit higher proliferation and a propensity to genomic instability.[19,20]

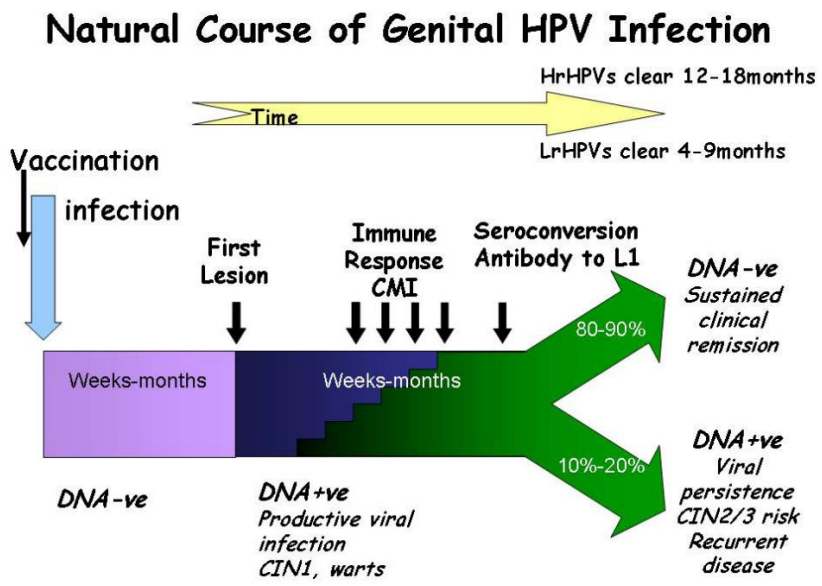


Figure 1.3 Mechanism of HPV infection in the female genital tract[13]

1.1.3 HPV Genome Organization and Proteins Encoded

The genome of HPV is a circular double-stranded DNA of ~8000 bp and includes eight open reading frames (ORFs).[21] E1, E2, E4, E5, E6, E7 are defined as "early genes" and are encoded during early stages of the infection, while the "late genes" L1 and L2 are expressed at late times, although these historical designations are not completely accurate. The genome includes also a regulatory region defined as the Long Control Region (LCR) containing transcription factors binding sites, promoter elements and the origin of replication, and thus important for viral replication and transcription.[22]

Proteins L1 and L2 constitute the viral capsid and are expressed only in the upper layer of the epithelium during virus assembly. These proteins also represent the major antigenic components of the virus.[23]

E1 and E2 proteins play a role in viral replication, though E1-E2 independent pathways are known.[16,24] E1 is a DNA helicase and the only enzyme encoded by the virus, while E2 involved in viral transcription and replication and may play a role in tethering the viral DNA to chromatin during cell division.[25] One of the functions of E2 is to maintain low expression of the oncogenes E6 and E7 through a negative feedback loop early in the HPV life cycle. Integration of the viral DNA to the host genome occurs at the E2 gene, therefore abrogating the repression of the oncogenic proteins E6 and E7. The overexpression of E6 and E7 contributes to the cellular transformation that ultimately results in cancer.[20]

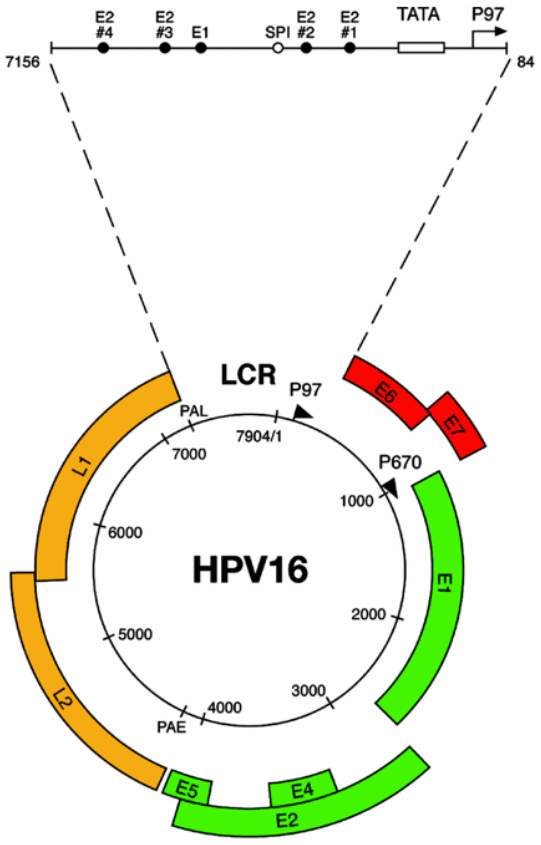


Figure 1.4 HPV genome organization [4]

The main oncoproteins encoded by HPV are E6 and E7.[26] The prominent functions associated to E6 and E7 are the degradation of p53 and retinoblastoma (Rb).[20] The p53 degradation is mediated by E6 and occurs through an ubiquitin-mediated process, making the cells more susceptible to DNA damage. The retinoblastoma (Rb) protein binds to the E2F transcription factor family known to induce the expression of

proteins that increase DNA replication and cell cycle progression. The E7 mediated inhibition of Rb, contributes therefore to tumor progression.[20,27]

Although the p53 and pRb deregulation contribute to cellular transformation, the interaction with other targets involved in cell-cycle, apoptosis, and immune-modulation are also of primary importance for malignant progression. Together, E6 and E7 work in synergy, inducing a multitude of cellular effects that lead to cellular immortalization and are required to develop and maintain a malignant phenotype.

While the functions of E6 and E7 have been widely studied, the roles of E4 and E5 are not fully characterized. E4 seems to play a role in facilitating virus particle synthesis, release and transmission[28] while E5 is involved in viral immune evasion and may enhance the activities of E6 and E7 before viral DNA integration.[29,30]

1.1.4 Prevention and Therapies

Three prophylactic HPV vaccines have currently market approval. Cervarix® is a bivalent vaccine covering infections mediated by the HPV types 16 and 18,[31,32] while Gardasil® is a quadrivalent vaccine including HPV types 6, 11, 16 and 18.[33] Gardasil®9 is the newest vaccine introduced in the market in 2014 and offers protection against HPV 6, 11, 16, 18, 31, 33, 45, 52, 58[34] and will soon be the only vaccine available in the US.

All three vaccines are DNA-free subunit vaccines consisting only of the L1 protein, the main protein constituting the viral capsid. L1 proteins spontaneously assemble into virus-like particles (VLP) which are highly immunogenic.[7] The antibody

response induced by the vaccine is vastly higher than the response produced by a natural infection, remaining high for several years.[35]

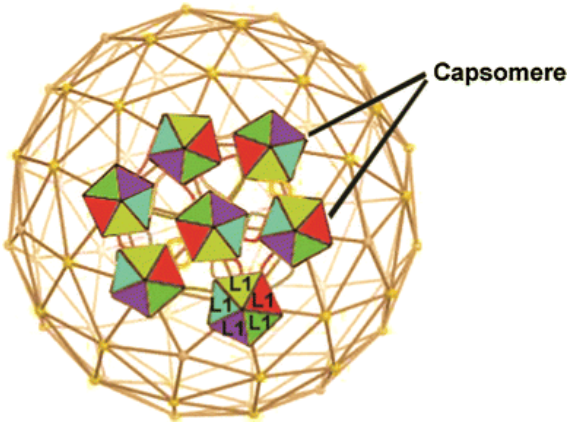


Figure 1.5 Virus-like particle (VLP). The HPV virion and the VLP are equal in size and shape. The VLP lacks a genome. The L1-VLP lacks the L2 protein in its capsomere[7] (Permission number: 4093270358633)

Although the vaccines offer protection against several high-risk HPV types, they offer no coverage against the remaining cancer causing HPV strains and are not effective for patients that have already been infected. Furthermore, while the vaccines have been proved to be largely safe and efficacious, population coverage is still lower than expected.

As for treatments, standard therapies for genital HPV infections are limited to surgical excision accompanied to chemoradiation in advanced stages of cancer.[36] Other treatments include mitotic inhibitors such as Podophyllotoxin or immune modulators such as Imiquimod,[37] both approved as topical drugs for non-mucosal tissues. Imiquimod activates the innate immune response inducing the production of several cytokines.

Compared to Podophyllotoxin, Imiquimod is associated with lower recurrence but displays several side effects such as severe local inflammation and dermatitis, so tolerance can be an issue.[36]

An extensive body of research has focused on the development of specific viral treatments to be used either alone or in combination with current therapies. Examples include inhibitors targeting the viral proteins E6, E7, E1 and E2 or therapeutic vaccines, nevertheless no treatments have been yet approved.[36]

Together, the absence of effective treatments and the incomplete coverage offered by vaccine prophylaxis underlines the necessity for the development of novel and more effective therapies.

1.2 Pyrrole-Imidazole Polyamides

1.2.1 History of Pyrrole-Imidazole Polyamides

Pyrrole-imidazole polyamides (PAs) are synthetic ligands developed from Distamycin A and Netropsin[38], two natural antibiotics with antitumor and antibacterial properties that bind A·T sequences in the minor groove of DNA.[39-41] Distamicin A consists of three N-methylpyrrole rings connected through amide bonds whereas Netropsin contains only two N-methylpyrrole rings and a guanidinium tail.

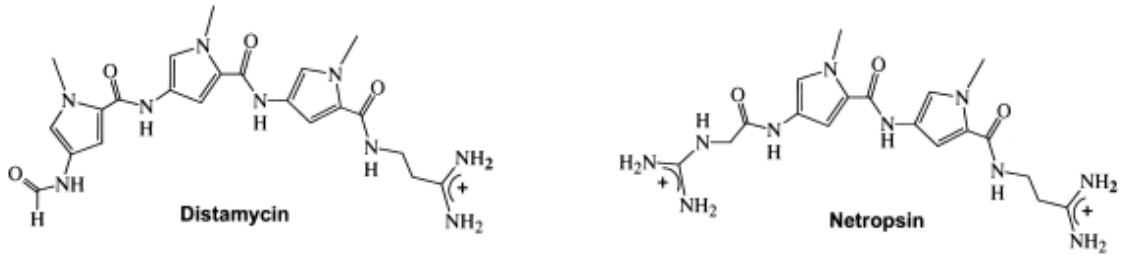


Figure 1.6 Distamycin and Netropsin Chemical Structures[3] (Permission number: 4093271317657)

The interaction between the aforementioned ligands and the DNA occurs through H-bonds between the polyamide amide hydrogens and the edges of the base pairs in the minor groove.[42] While both ligands can form 1:1 complex with the DNA, Distamycin can also form 2:1 ligand-DNA complexes. The peculiar crescent-shape of Distamycin and Netropsin and the specificity of the H-bonds account for their high affinity and specificity toward DNA. The ability of Netropsin and Distamycin to recognize also G·C DNA sequences was predicted by Dickerson[42] and achieved by Lown[43] with the replacement of N-methyl pyrrole by N-methyl imidazole (Im), so that the extra imidazole nitrogen could interact with the exocyclic $-NH_2$ of guanine (which protrudes into the minor groove).

Polyamides are synthetic analogs of Distamycin containing repeating units of the aromatic scaffold N-methylimidazole (Im) and N-methylpyrrole (Py) linked through amidic bonds. Similarly to the natural antibiotics, these compounds bind to the minor groove with high affinity, but do not exhibit significant cellular toxicity. Over the years polyamides have been optimized to enhance DNA binding affinity and sequence specificity. For instance, β-alanine is often incorporated to relax the curvature of the ligand and allow the synthesis of longer polyamides without losing the ability of the ligand to match the pitch of DNA.[44] The introduction of an γ-aminobutyric acid (GABA)

unit between two antiparallel polyamide strands afforded hairpin polyamides which display a binding affinity 100-fold higher than the unlinked strands.[45,46] While other

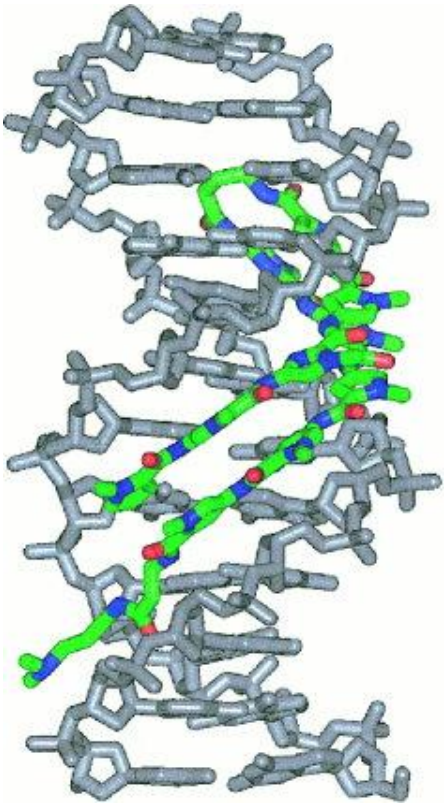


Figure 1.7 Hairpin Polyamide bound to DNA[3] (Permission number: 4093271317657)

structural variations have been explored such as H-Pin polyamides and tandem hairpin polyamides to increase binding affinity or target longer sites, hairpins polyamides are by far the most studied and utilized polyamides.

A further step in expanding the recognition code included the pairing of hydroxypyrrole (HPy) to Py with the purpose of distinguishing between A and T bases[47] though HPy resulted to be chemically unstable over time, especially in cell culture media.

Thus, a set of "polyamides pairing rules" were defined and refer to the ability of short polyamides to distinguish between the four DNA bases, with Im/Py pair recognizing G·C and Hp/Py recognizing T·A (**Figure 1.8**).

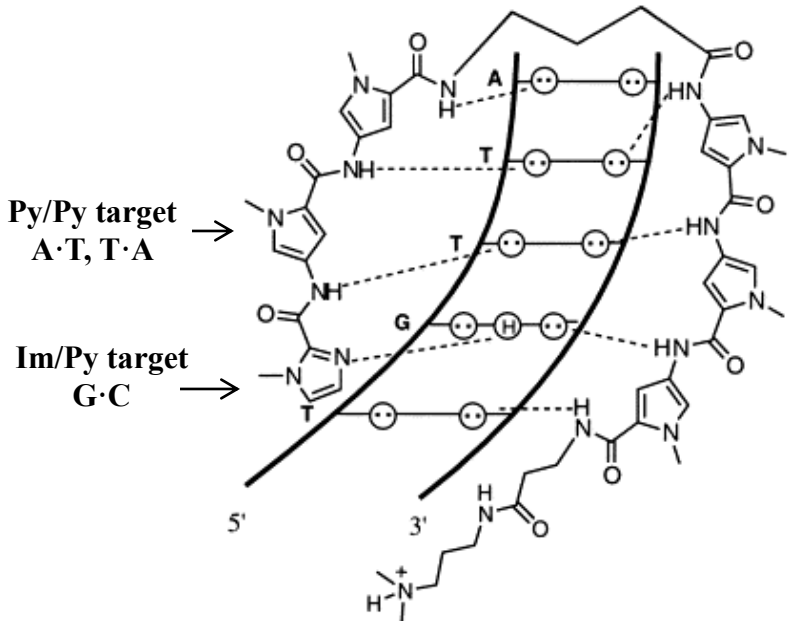


Figure 1.8 Pyrrole-Imidazole Polyamides Pairing Rules (Permission Number 4093290313925)[6]

1.2.2 Pyrrole-Imidazole Polyamides Applications

Small polyamides can be programmed to bind specific DNA sequences with specificity and affinity similar to naturally occurring DNA-binding proteins.

Because polyamides are cell and nucleus permeable, when applied to living systems, they can enter the nucleus and interfere with gene transcription disrupting protein-DNA interactions. There are numerous reports exploring the applications of PAs as gene expression modulators.[48,49] Sugiyama and Dervan's groups have shown how polyamides can act as transcription repressors by impeding the binding of transcription

factors to their DNA target sequences. Examples include the transcription inhibition of HIV-1[50] and the repression of NF-kB, an important mediator of inflammation.[51] In contrast, polyamides can also be used as gene transcription activators by conjugating the polyamide to a peptide activator. The polyamide binds the DNA sequence of interest acting as a DNA-binding domain while the peptide activator recruits the transcription factors necessary to initiate transcription.[52]

By conjugation to a fluorophore, polyamides can also be employed for *in vivo* imaging.[53] Polyamide-fluorophore conjugates, for instance, have been utilized to detect repeating units in DNA,[54] to visualize telomerases[55] and to stain chromosomes while avoiding a denaturation step.

1.2.3 Pyrrole-Imidazole Polyamides as Potential Anti-HPV Drugs

Studies conducted in our lab and in collaboration with biologists at NanoVir have shown that large polyamides exhibit anti-HPV activity in epithelial cell and tissue cultures, without evidence of cytotoxicity up to their solubility limit (150-200 μM).[56] These minor groove binders enter the HPV-infected cells, triggering the degradation and elimination of the HPV DNA. The binding of PA probably induces a structural alteration in the viral DNA, eliciting the DDR pathway[57,58] which mediates the antiviral activity of PAs, although the exact mechanism of DNA degradation is still under investigation. PAs specifically target the degradation of the viral DNA in the episomal form, since the DDR pathway alteration, and all antiviral activity, is solely observed in cells bearing HPV episomes and not in cells having HPV DNA integrated in the host genome.

More recently, the antiviral activity of polyamides was demonstrated against Simian virus 40 (SV40) and BK virus, viruses belonging to the polyomavirus family, thus expanding the spectrum of activity to other double-stranded small DNA tumor viruses.[59]

To date, a library of more than 100 compounds has been built. It includes PAs that differ in length, building-block sequence, C-terminal tail and N-terminal substitution. Although we currently know that longer PAs (minimum DNA binding site of 10 bp) and N-terminal substitution with guanidinium group (TMG) positively affect the antiviral efficacy, no clear structure-activity relationship (SAR) has yet emerged. Even small variations in PA chemical structure lead to significant changes in antiviral efficacy. The work described in this thesis focuses on the investigation of some of the aspects that may be important for antiviral activity. Specifically, we explored the interactions occurring between polyamides and HPV DNA sequences as well as the cellular uptake features of large polyamides. The latter investigation was important because traditional structure-activity relationships have not become clear even after extensive analysis based on thermodynamic binding constants and antiviral data.

Pyrrole-imidazole polyamides represent a nontraditional approach toward the development of antiviral medications for the treatment of HPV infections, for which a cure is currently not available.

1.3 References

(1) zur Hausen, H. *Nat Rev Cancer* **2002**, *2*, 342.

(2) Woodman, C. B.; Collins, S. I.; Young, L. S. *Nat Rev Cancer* **2007**, *7*, 11.

(3) Wemmer, D. E. *Biopolymers* **1999**, *52*, 197.

(4) Doorbar, J. *Clinical Science* **2006**, *110*, 525.

(5) Crow, J. M. *Nature* **2012**, *488*, S2.

(6) Dervan, P. B. *Bioorg Med Chem* **2001**, *9*, 2215.

(7) Malik, H.; Khan, F. H.; Ahsan, H. *Archives of Virology* **2014**, *159*, 199.

(8) Bernard, H.-U.; Burk, R. D.; Chen, Z.; van Doorslaer, K.; Hausen, H. z.; de Villiers, E.-M. *Virology* **2010**, *401*, 70.

(9) Bzhalava, D.; Guan, P.; Franceschi, S.; Dillner, J.; Clifford, G. *Virology* **2013**, *445*, 224.

(10) Psyrri, A.; DiMaio, D. *Nat Clin Prac Oncol* **2008**, *5*, 24. (11) Bosch, F. X.; Burchell, A. N.; Schiffman, M.; Giuliano, A. R.; de Sanjose, S.; Bruni, L.; Tortolero-Luna, G.; Kjaer, S. K.; Muñoz, N. *Vaccine* **2008**, *26, Supplement 10*, K1. (12) Muñoz , N.; Bosch , F. X.; de Sanjosé , S.; Herrero , R.; Castellsagué , X.; Shah , K. V.; Snijders , P. J. F.; Meijer , C. J. L. M. *New England Journal of Medicine* **2003**, *348*, 518.

(13) Stanley, M. *Infectious Agents and Cancer* **2010**, *5*, 19.

(14) Insinga, R. P.; Perez, G.; Wheeler, C. M.; Koutsky, L. A.; Garland, S. M.; Leodolter, S.; Joura, E. A.; Ferris, D. G.; Steben, M.; Brown, D. R.; Elbasha, E. H.; Paavonen, J.; Haupt, R. M. *Cancer Epidemiology Biomarkers & Prevention* **2010**, *19*, 1585.

(15) Williams, V. M.; Filippova, M.; Soto, U.; Duerksen-Hughes, P. J. *Future Virol* **2011**, *6*, 45.

(16) Doorbar, J.; Egawa, N.; Griffin, H.; Kranjec, C.; Murakami, I. *Reviews in Medical Virology* **2015**, *25*, 2.

(17) Cheng, S.; Schmidt-Grimminger, D. C.; Murant, T.; Broker, T. R.; Chow, L. T. *Genes Dev* **1995**, *9*, 2335.

(18) Fehrmann, F.; Laimins, L. A. *Oncogene* **2003**, *22*, 5201.

(19) Jeon, S.; Allen-Hoffmann, B. L.; Lambert, P. F. *J Virol* **1995**, *69*, 2989.

(20) Moody, C. A.; Laimins, L. A. *Nat Rev Cancer* **2010**, *10*.

(21) Zheng, Z.-M.; Baker, C. C. *Frontiers in bioscience : a journal and virtual library* **2006**, *11*, 2286.

(22) Stünkel, W.; Bernard, H.-U. *J Virol* **1999**, *73*, 1918.

(23) Conway, M. J.; Meyers, C. *Journal of Dental Research* **2009**, *88*, 307.

(24) Angeletti, P. C.; Kim, K.; Fernandes, F. J.; Lambert, P. F. *J Virol* **2002**, *76*.

(25) Aydin, I.; Schelhaas, M. *Traffic* **2016**, *17*, 327.

(26) Ghittoni, R.; Accardi, R.; Hasan, U.; Gheit, T.; Sylla, B.; Tommasino, M. *Virus Genes* **2010**, *40*, 1. (27) Yim, E.-K.; Park, J.-S. *Cancer Research and Treatment : Official Journal of Korean Cancer Association* **2005**, *37*, 319.

(28) Doorbar, J. *Virology* **2013**, *445*, 80.

(29) Müller, M.; Prescott, E. L.; Wasson, C. W.; Macdonald, A. *Future Virol* **2015**, *10*, 27.

(30) Venuti, A.; Paolini, F.; Nasir, L.; Corteggio, A.; Roperto, S.; Campo, M. S.; Borzacchiello, G. *Molecular Cancer* **2011**, *10*, 140.

(31) (CDC), C. f. D. C. a. P. *Morbidity and Mortality Weekly Report* **2010**, *59*, 626.

(32) (CDC), C. f. D. C. a. P. *Morbidity and Mortality Weekly Report* **2010**, *59*, 630.

(33) McCormack, P. L. *Drugs* **2014**, *74*, 1253.

(34) Joura, E. A.; Giuliano, A. R.; Iversen, O.-E.; Bouchard, C.; Mao, C.; Mehlsen, J.; Moreira, E. D.; Ngan, Y.; Petersen, L. K.; Lazcano-Ponce, E.; Pitisuttithum, P.; Restrepo, J. A.; Stuart, G.; Woelber, L.; Yang, Y. C.; Cuzick, J.; Garland, S. M.; Huh, W.; Kjaer, S. K.; Bautista, O. M.; Chan, I. S. F.; Chen, J.; Gesser, R.; Moeller, E.; Ritter, M.; Vuocolo, S.; Luxembourg, A. *New England Journal of Medicine* **2015**, *372*, 711.

(35) Roteli-Martins, C. M.; Naud, P.; De Borba, P.; Teixeira, J. C.; De Carvalho, N. S.; Zahaf, T.; Sanchez, N.; Geeraerts, B.; Descamps, D. *Human Vaccines & Immunotherapeutics* **2012**, *8*, 390.

(36) Stern, P. L.; van der Burg, S. H.; Hampson, I. N.; Broker, T.; Fiander, A.; Lacey, C. J.; Kitchener, H. C.; Einstein, M. H. *Vaccine* **2012**, *30*, F71.

(37) Yanofsky, V. R.; Patel, R. V.; Goldenberg, G. *The Journal of Clinical and Aesthetic Dermatology* **2012**, *5*, 25.

(38) Chen, X.; Ramakrishnan, B.; Rao, S. T.; Sundaralingam, M. *Nat Struct Mol Biol* **1994**, *1*, 169.

(39) Dervan, P. B.; Doss, R. M.; Marques, M. A. *Current Medicinal Chemistry -Anti-Cancer Agents* **2005**, *5*, 373.

(40) Dervan, P. B.; Edelson, B. S. *Current Opinion in Structural Biology* **2003**, *13*, 284.

(41) Harki, D. A.; Satyamurthy, N.; Stout, D. B.; Phelps, M. E.; Dervan, P. B. *Proceedings of the National Academy of Sciences* **2008**, *105*, 13039.

(42) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proceedings of the National Academy of Sciences of the United States of America* **1985**, *82*, 1376.

(43) Lown, J. W.; Krowicki, K.; Bhat, U. G.; Skorobogaty, A.; Ward, B.; Dabrowiak, J. C. *Biochemistry* **1986**, *25*, 7408.

(44) Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1998**, *120*, 6219.

(45) Mrksich, M.; Parks, M. E.; Dervan, P. B. *Journal of the American Chemical Society* **1994**, *116*, 7983.

(46) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559.

(47) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468.

(48) Melander, C.; Burnett, R.; Gottesfeld, J. M. *Journal of Biotechnology* **2004**, *112*, 195.

(49) Yasuda, A.; Noguchi, K.; Minoshima, M.; Kashiwazaki, G.; Kanda, T.; Katayama, K.; Mitsuhashi, J.; Bando, T.; Sugiyama, H.; Sugimoto, Y. *Cancer Sci* **2011**, *102*, 2221.

(50) Dickinson, L. A.; Gulizia, R. J.; Trauger, J. W.; Baird, E. E.; Mosier, D. E.; Gottesfeld, J. M.; Dervan, P. B. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95*, 12890.

(51) Dickinson, L. A.; Trauger, J. W.; Baird, E. E.; Dervan, P. B.; Graves, B. J.; Gottesfeld, J. M. *Journal of Biological Chemistry* **1999**, *274*, 12765.

(52) Dervan, P. B.; Poulin-Kerstien, A. T.; Fechter, E. J.; Edelson, B. S. In *DNA Binders and Related Subjects*; Waring, M. J., Chaires, J. B., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2005, p 1.

(53) Vaijayanthi, T.; Bando, T.; Pandian, G. N.; Sugiyama, H. *Chembiochem* **2012**, *13*, 2170.

(54) Fujimoto, J.; Bando, T.; Minoshima, M.; Uchida, S.; Iwasaki, M.; Shinohara, K.-i.; Sugiyama, H. *Bioorg Med Chem* **2008**, *16*, 5899.

(55) Kawamoto, Y.; Bando, T.; Kamada, F.; Li, Y.; Hashiya, K.; Maeshima, K.; Sugiyama, H. *Journal of the American Chemical Society* **2013**, *135*, 16468.

(56) Edwards, T. G.; Koeller, K. J.; Slomczynska, U.; Fok, K.; Helmus, M.; Bashkin, J. K.; Fisher, C. *Antiviral Research* **2011**, *91*, 177.

(57) Edwards, T. G.; Vidmar, T. J.; Koeller, K.; Bashkin, J. K.; Fisher, C. *PLoS One* **2013**, *8*, e75406.

(58) Fisher, C. *Journal of Clinical Medicine* **2015**, *4*, 204.

(59) Bashkin, J. K.; Edwards, T. G.; Fisher, C.; Harris, G. D.; Koeller, K. J.; Google Patents: 2016.

CHAPTER 2

BINDING STUDIES OF PYRROLE-IMIDAZOLE POLYAMIDES NV1028 AND NV1042 WITH NATURAL HPV DNA SEQUENCES

2.1 Introduction

Pyrrole-imidazole polyamides containing a minimum of ten rings actively eliminate the DNA of HPV and other viruses from infected cell.[1-5] The aim of this study is to identify and characterize the interactions occurring between antiviral pyrrole-imidazole polyamides and natural HPV DNA sequences by means of DNase I footprinting and affinity cleavage methods.

DNase-I footprinting is a widely used technique employed to determine the interactions between DNA and proteins or other DNA-binding ligands.[6,7] A scheme of the experiment workflow is reported in **Figure 2.1**. In a typical footprinting experiment, the ligand is incubated with the DNA and the deoxyribonuclease cleaves the phosphodiester bonds that are not protected by the ligand. The resulting DNA fragments are subsequently analyzed by gel-electrophoresis and the

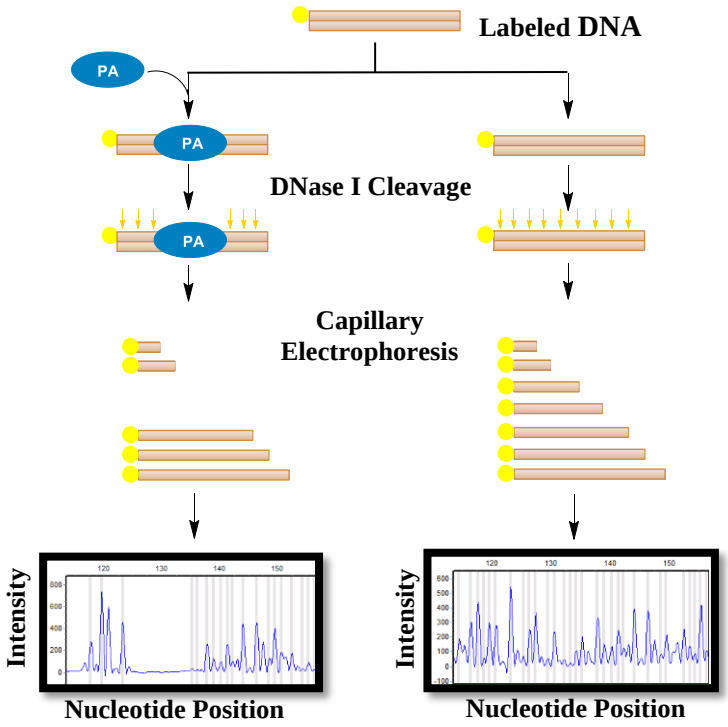


Figure 2.1. Workflow of a DNase I footprinting experiment. Electropherograms represent raw capillary electrophoresis (CE) data. Figure a*dapted from* "The Cell: A Molecular Approach", 2nd edition, Cooper GM. Sunderland (MA): Sinauer Associates, 2000.

fragmentation pattern compared to the cleavage pattern obtained in a control sample (DNA cleavage in absence of a ligand) to determine the binding sites. While gel-electrophoresis is the most common down-stream technique to analyze the fragments, capillary electrophoresis can also be employed if the DNA is labeled with a fluorescent dye.[8,9] Advantages of capillary electrophoresis include the elimination of the radioactive label as well high throughput. Furthermore, footprinting is extensively used to evaluate the binding strength of a DNA-ligand interaction.[10,11] In this case the DNA is titrated with increasing concentrations of ligand **(Figure 2.2)** and the observed cleavage in inversely proportional to the ligand concentration. Binding constants are obtained by plotting the fraction bound as a function of concentration and fitting the data to a Hill or Langmuir model.

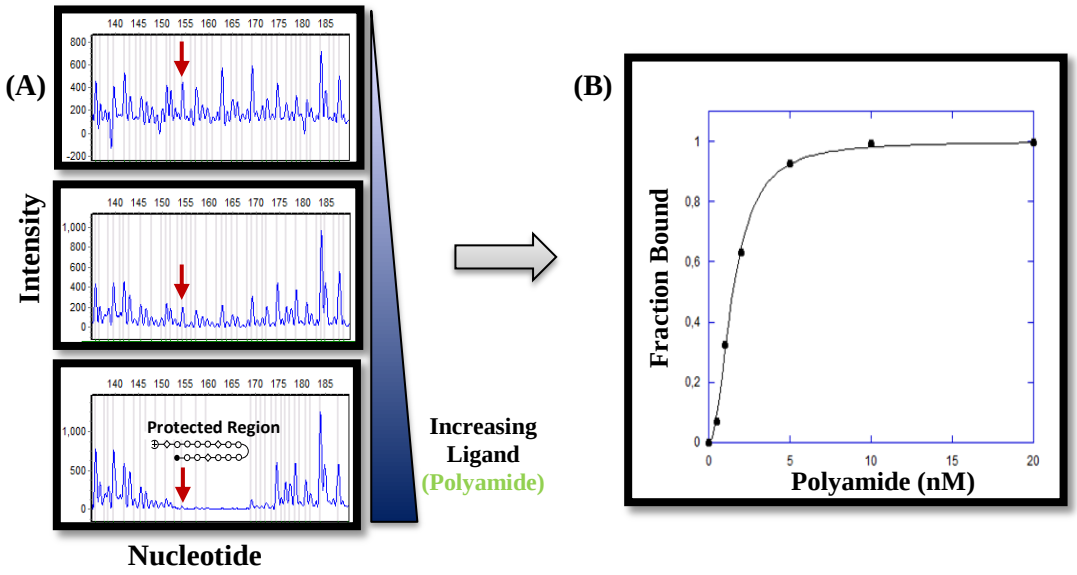


Figure 2.2 (A) Representative footprinting electropherograms showing the cleavage pattern generated by quantitative DNase I footprinting titration in the presence of increasing ligand concentration; (B) Hill binding titration isotherm obtained from DNase I data

While footprinting can be useful to locate the ligand position along the DNA and to evaluate the energetics of the interaction, affinity cleavage is a complementary technique that provides the orientation of the polyamide along a DNA fragment.[12] This method requires the conjugation of the chelating agent ethylenediaminetetraacetic acid (EDTA) to the C-terminal amino group of a polyamide, resulting in a PA-EDTA complex able to chelate Fe (II). In presence of H_2O_2, Fe (II) is oxidized to Fe (III) and hydroxyl radicals are generated by reduction of the peroxide. The radicals cleave DNA at the end of the polyamide binding site by abstracting H· from the deoxyribose in a distance-dependent manner from the C-terminus of the polyamide.

Here, I identified and characterized the binding sites of two hairpin polyamides, NV1028 and NV1042 **(Figure 2.3)**, in the long control region (LCR) of Human Papillomavirus type 18 (HPV18, Genebank No.: X05015) using DNase I footprinting and affinity cleavage coupled with capillary electrophoresis. Quantitative DNase I footprinting experiments were carried out to determine the dissociation constants of each compound. The LCR was chosen because it represents the regulatory region in the HPV genome and contains c*is*-responsive elements regulating the transcription and replication of HPV.[13]

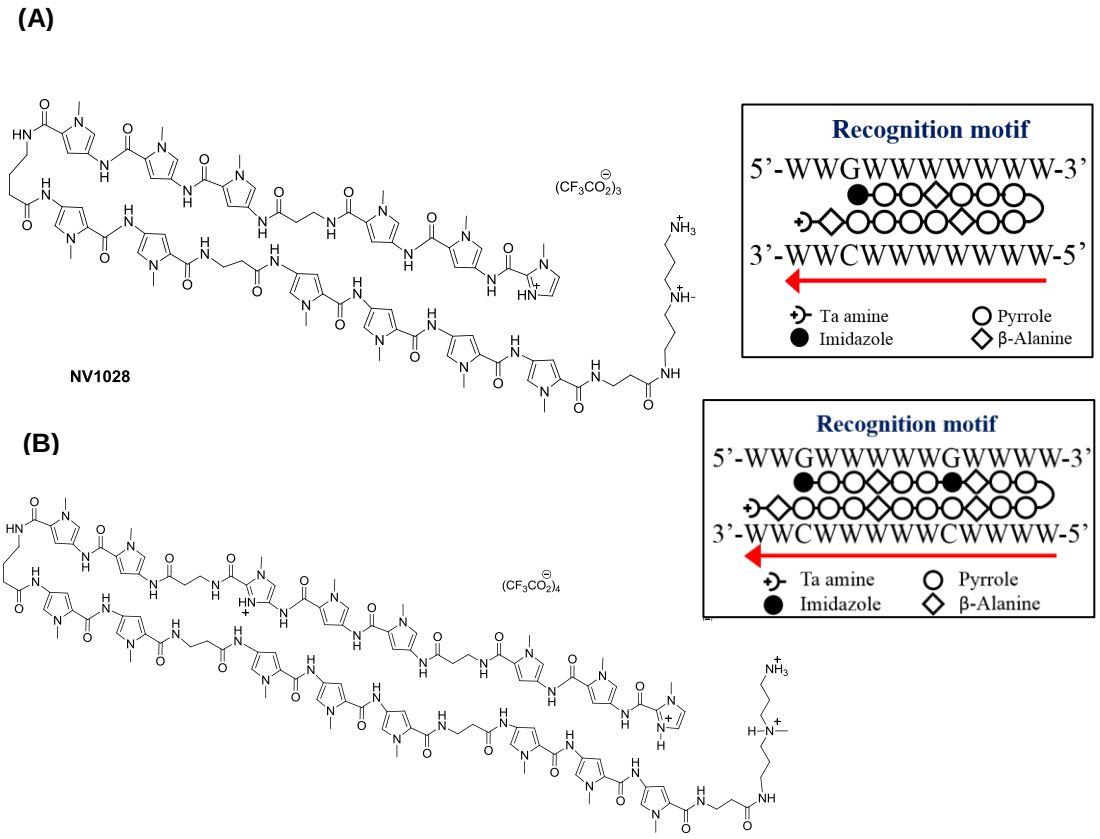


Figure 2.3. NV1028 (A) and NV1042 (B) chemical structures and correspondent DNA recognition motifs

2.2 Methods

2.2.1 Materials

RNA primers were ordered from Integrated DNA Technologies (Coralville, Iowa), Taq Polymerase was purchased from New England Biolabs (Ipswich, MA) and $Fe(NH4)_2(SO4)_2$, $MgCl_2$, $CaCl_2$, Dithiothreitol (DTT) were acquired from Sigma Aldrich (St. Louis, MO). Thermo Sequenase Cycle Sequencing Kit and 3-[(3-Cholamidopropyl)

dimethylammonio]-1-propanesulfonate (CHAPS) were obtained from Affymetrix (Santa Clara, CA) and DNase I was purchased from Promega (Madison, WI).

2.2.2 DNA and Polyamide Concentration Measurements

Polyamide concentration was determined by UV-Vis spectrometry. Briefly, 0.1 mg of lyophilized polyamide was dissolved in 100 µl of DMSO and the absorbance at 305 nm measured by UV-Vis using a Thermo Scientific Evolution 260 BIO UV spectrophotometer. The extinction coefficients used were determined by mass and were equal to 88235 $M^{-1}cm^{-1}$ for NV1028 and 147400 $M^{-1}cm^{-1}$ for NV1042.[2] The DNA concentration was quantified by UV-Vis spectrometry, measuring the absorbance at 260 nm.

2.2.3 385-bp DNA Fragment Preparation (Fragment I)

The fluorescently labeled DNA fragment (HPV18 7179-7563 bp) was generated by polymerase chain reaction (PCR) using the HPV18-pBR322 plasmid as a template and primers 5' labeled with 6-carboxyfluorescein (FAM) on the top strand and 6-carboxy-2',4,4',5',7,7' – hexachlorofluorescein (HEX) on the bottom strand. Primers were designed using Primer-3 and their physical properties analyzed using the Oligoanalyzer 3.1 website. The forward and reverse primer sequences were 5'-Fam-TATGTCCTGTGTTTGTGTTT-3' and 5'-Hex-CAAAATATGTAGGAGCAGTG-3' respectively. The PCR cocktail was prepared as described in **Table 2.1** and amplification was performed over 32 cycles at the conditions shown in **Table 2.2.**

Reagents	Vol (µL)
H_2O	41 µL
10X standard Taq Buffer with $MgCl_2$	5 µL
Template DNA	1 µL
10 µM Reverse Primer	1 µL
10 µM Forward Primer	1 µL
10 mM dNTPs	0.5 µL
Taq Polymerase	0.5 µL

Table 2.1. PCR Cocktail for a 50 µl reaction

PCR Step	Temperature	Time
Denaturing	95 ºC	30 sec
Annealing	55.4 ºC	30 sec
Extension	68 ºC	1 min

Table 2.2 Optimized PCR conditions for the HPV18 7179-7563 DNA fragment amplification

The PCR product was separated by gel electrophoresis (1% agarose gel in Tris-acetate-EDTA buffer), the band correspondent to the desired product was excised from the gel and purified using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). The DNA was also sent to the DNA core for Sanger sequencing,[14] to ensure that the sequence was correct.

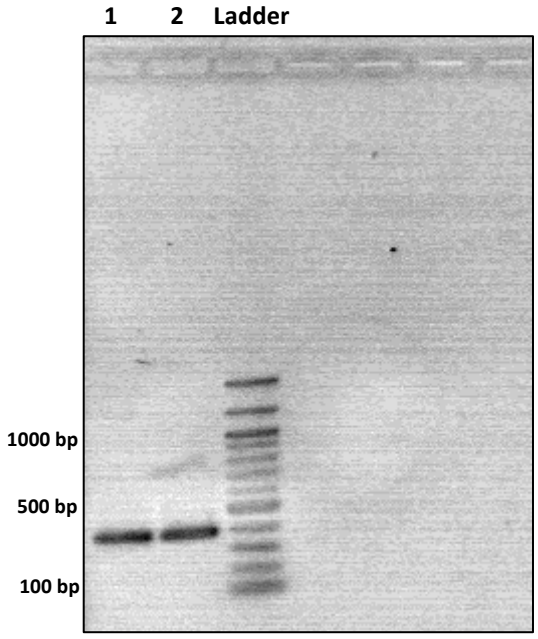


Figure 2.4. Representative agarose gel of the 385 bp DNA fragment. Lane 1 and Lane 2 correspond to the desired 385 bp DNA fragment

2.2.4 305-bp DNA Fragment Preparation (Fragment II)

The doubly labeled 305 bp fragment (HPV18 7479-7783) was generated following the procedure described in section 2.2.3. The forward and reverse primer sequences were 5'-Fam-CTTATGTCTGTGGTTTTCTG-3' and 5'-Hex-CAGAAAACCACAGACATAAG-3' respectively**.** PCR was performed over 32 cycles at the conditions shown in **Table 2.3.**

PCR Step	Temperature	Time
Denaturing	95 ºC	30 sec
Annealing	48 ºC	30 sec
Extension	68 ºC	30 sec

Table 2.3. Optimized PCR conditions for the HPV18 7479-7783 DNA fragment amplification

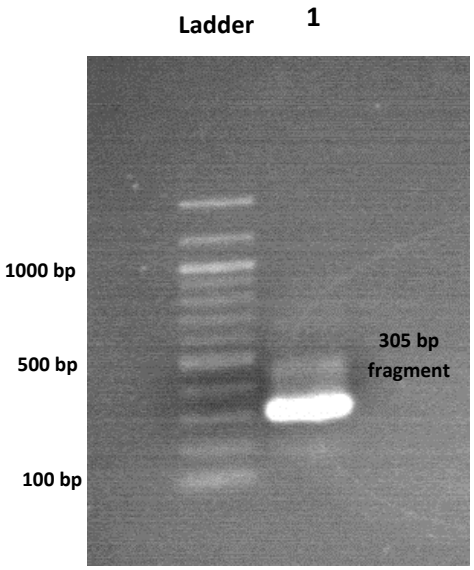


Figure 2.5. Representative agarose gel of the 305 bp DNA fragment
Lane 1 corresponds to the desired 305 bp DNA fragment

2.2.5 USB and Maxam-Gilbert Sequencing (Indexing)

Conjugation of a fluorescent dye to the 5' end of a DNA fragment alters the DNA mobility, affecting elution time and size calling during capillary electrophoresis (CE). Furthermore, the type of chemistry applied to generate DNA cleavage can also affect electrophoretic mobility since different cleavage products are generated.[15] In order to correlate the nucleotide position observed in the CE electropherogram with the real nucleotide position in the genome, we used Sanger chemistry[14] (USB, Affymetrix, Santa Clara, CA, USA) and Maxam and Gilbert[16] chemistry to map DNase-I footprinting and affinity cleavage results, respectively. Sanger sequencing involves the use of dideoxynucleotides (ddNTPs), nucleotides containining hydrogen on the 3' carbon instead of a hydroxyl group (OH). Once integrated into a sequence by the

Thermosequenase enzyme during PCR, ddNTPs prevent the addition of further nucleotides. By amplifying four different reaction mixtures, each containing one of the four dideoxynucleotides (ddATP, ddTTP, ddCTP, ddGTP), Thermosequenase and all dNTPs and then subjecting each mixture to fragment analysis, it is possible to determine the sequence and the position of each nucleotide along the DNA sequence. The results of the sequencing are provided in **Table SI 1A** (HPV 7169-7563) and **SI 1B** (7479-7789).

Maxam and Gilbert sequencing is used to map the results obtained by affinity cleavage since the chemistry involved best resembles the hydroxyl radical cleavage produced by Fenton chemistry. The correlation between capillary electrophoresis position and real position is provided in **Table SI 2**.

The difference between Sanger and Maxam–Gilbert positions versus the real position **(Figure 2.6)** was determined for the 385 bp fragment (HPV18 7179-7563). These results were consistent with those obtained and previously published by our group.[15]

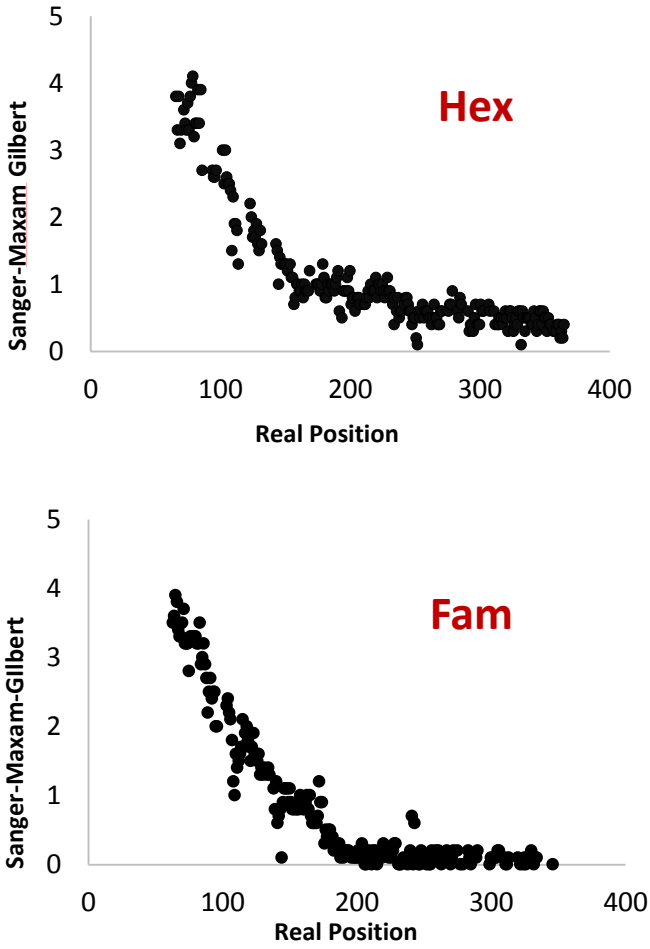


Figure 2.6. Difference between Sanger and Maxam–Gilbert positions versus the real position

2.2.6 DNase I footprinting

Briefly, samples (250 µl) of labeled DNA (final concentration: 200 pM) in TKMC buffer (10 mM Tris, 10 mM KCl, 5 mM $MgCl_2$ and 5 mM $CaCl_2$), 2 % DMSO and 10 mM CHAPS were incubated with polyamide (ranging from 0-40 nM) at 37 ºC for approximately 4 hours (incubation time can vary from 2 h to overnight). The mixture was digested with 0.03-0.09 AU of RQ1 RNase-Free DNase I (Promega, Madison, WI 53711)

for 5 min at 37 ºC, stopped by adding 10 µL of 200 mM EDTA. The DNase I digested product was purified with Qiagen PCR purification kit (Qiagen, Hilden, Germany). Footprinting results were mapped using USB sequencing and K_d values determined by fitting the results to a Hill model.

2.2.7 Affinity Cleavage

The polyamide-EDTA-Fe conjugate was prepared by adding 0.8 equivalents of $Fe(NH_4)_2(SO_4)_2$ to a solution of polyamide-EDTA in DMSO. Samples of the labeled DNA (final concentration: 1 nM) in 100 mM Tris-HCl buffer, 2% DMSO and 10 mM CHAPS were incubated with the polyamide-EDTA-Fe^{++} conjugate at 37 ºC for 4 hours. The polyamide-EDTA conjugate concentration ranged from 5 nM to 200 nM. Cleavage reactions were initiated by adding 5 µL of freshly made 100 mM DTT to the sample and incubating the reaction mixture for ~1 h. The reaction was quenched and products purified using a Qiagen PCR purification kit. Fragment analysis was conducted by CE. Affinity cleavage results were mapped using Maxam-Gilbert sequencing.

2.2.8 Capillary Electrophoresis (CE)

Samples of fragmented DNA were sent and processed at the University of Missouri DNA Core facility using an ABI 3730xl DNA Analyzer 96-capillary CE instrument. Low salt concentration and complete ethanol removal during DNA purification were essential for optimal results. DNA concentration, injection time and size standard quantity are also

important parameters to optimize. Data were processed using GeneMarker version 1.97 software (SoftGenetics, State College, PA, USA). Isotherms were generated using Kaleidagraph 4.1.3 software (Reading, PA, USA).

2.3 Results and Discussion

DNA binding studies were carried out on DNA fragments correspondent to the Long Control Region (LCR) of HPV18 (**Figure 2.7**). Because by capillary electrophoresis only DNA sequences shorter than ~400 bp can be accurately analyzed for footoprinting purposes, three different DNA fragments spanning the whole LCR were prepared as shown in Figure 2.7.

The pyrrole-imidazole polyamides used for these studies were NV1028 and NV1042. Polyamide NV1028 is a 12-ring polyamide that, according to the Dervan's rules,[17] should recognize the DNA sequence 5'-WWGWWWWWWW-3' (where W indicates adenine or thymine). NV1042 belongs to a class of larger polyamides containing 16 rings that should recognize the motif 5'-WWGWWWWGWWWWW-3'. The chemical structures of both compounds are reported in **Figure 2.3.**

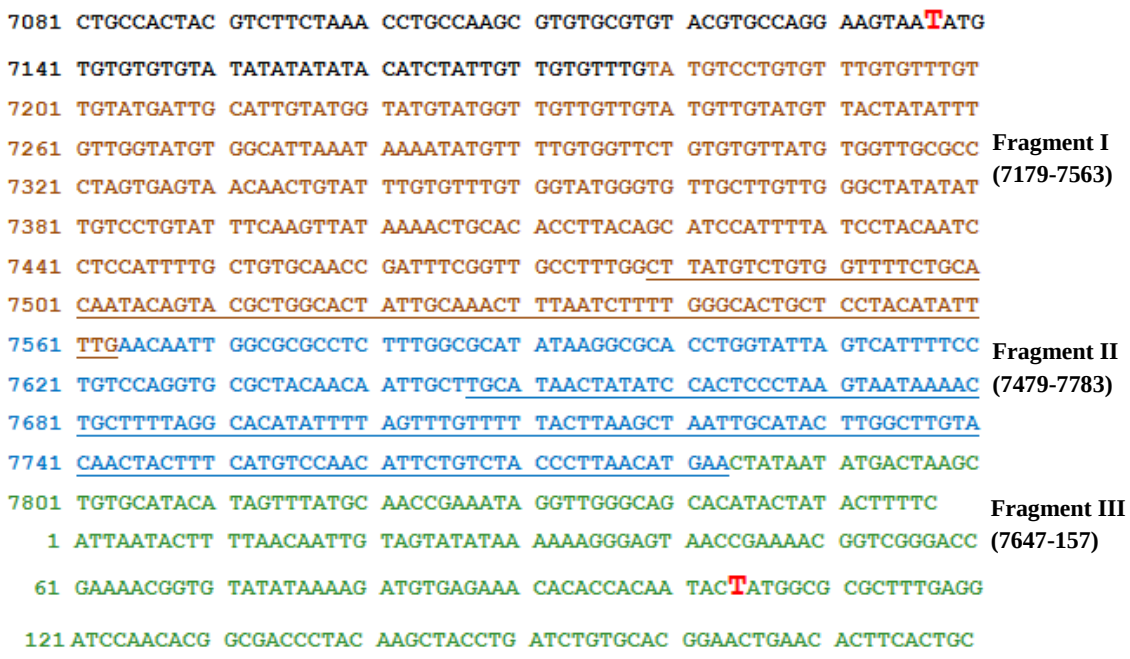


Figure 2.7. HPV18 Long Control region (LCR). Representation of the HPV18 LCR and the three different fragments (Fragment I: brown, Fragment II: blue, Fragment III: green) used to carry out the binding experiments. Red nucleotides indicate begin and end of the LCR, the underlined text indicates the nucleotides shared by two adjacent fragments

2.3.1 Binding of polyamide NV1028 on a 385 bp fragment within the LCR of HPV18 (7179-7563 bp – Fragment I)

Affinity cleavage experiments were performed in triplicate according to the procedure described in section 2.2.6 and results are mapped in **Figure 2.8**. Affinity cleavage is depicted by vertical arrows located above and below the DNA sequence, with black arrows representing affinity cleavage observed at 50 nM of NV1028 (high affinity) and red arrows representing affinity cleavage sites appearing only at 100 nM polyamide (low affinity).

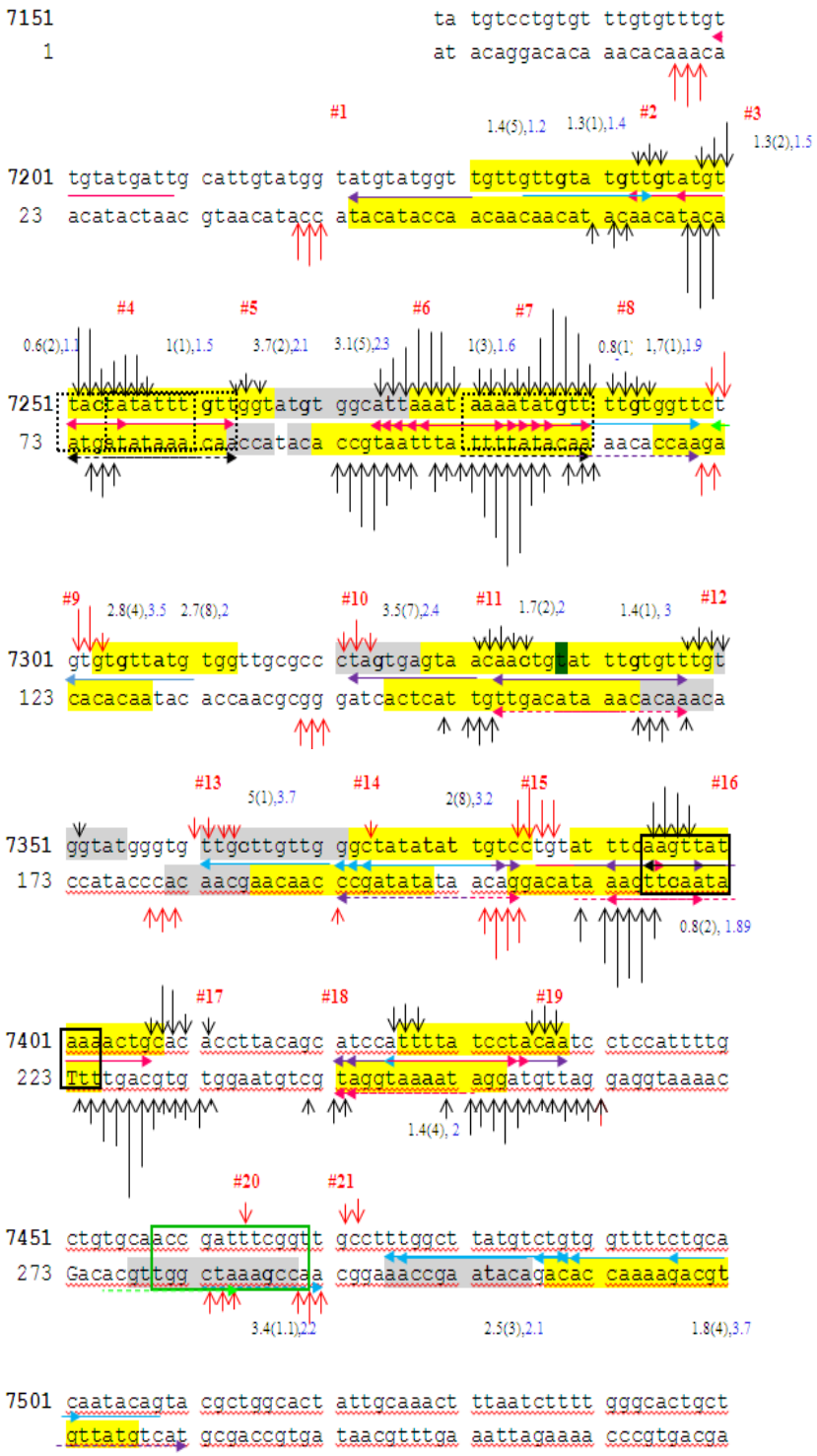
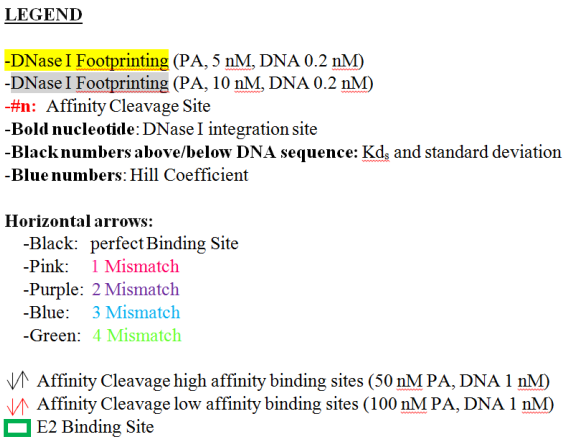


Figure 2.8. Map of affinity cleavage and footprinting of HPV18 (7189-7563) with NV1028 (overleaf)



The map provides a representation of observed binding sites along with zero or one mismatch sites that are predicted but not observed (not observed: site 7247-7256, site 7391-7400 and site 7398-7389). A mismatch site refers to any site within the DNA sequence bound to the polyamide that the reported binding rules fail to predict.[17,18] All the observed high affinity cleavage sites can be rationalized by 0, 1, or 2 mismatches, with the exception of site **18**, which can be better explained by a 3-mismatch site. At sites **4, 5, 8, 9, 11, 12, 15** the polyamide could bind in either a forward or a reverse orientation, while sites **6, 7, 9, 18, 19**, due to the resolution of our experimental

techniques, cannot unambiguously be assigned to forward, reverse or mismatch number. Interestingly, there are two perfect reverse sites to which NV1028 binds only modestly: site 5 (7263-7254) and site 8 (7290-7281). For site 8, a plausible explanation for the poor binding may be the context in which the perfect binding site is found. Adjacent to it, there is in fact a long (13 nucleotide) poly-W sequence corresponding to a very high and broad affinity cleavage. Therefore, this W-rich sequence is preferred to a perfect binding site having a guanine at the imidazole position. More likely, what is preferred is the characteristic local shape of the poly-W tract characterized by a narrow minor groove (or its low electrostatic potential[19]). A prediction of DNA hydroxyl radical cleavage for the sequence 7239-7358 carried out using ORChID (Tullius Lab., http://dna.bu.edu/orchid/) supports the peculiar shape of this W-rich sequence **(Figure 2.9).** Hydroxyl radical

cleavage provides an indirect representation of the groove shape. Narrow minor grooves are characterized by low hydroxyl radical cleavage, while where the minor groove is wide, and therefore deoxyribose hydrogens are more exposed, cleavage is high.[20] A significant hollow is observed in the plot **(Figure 2.9)** corresponding to the W-rich tract to which the polyamide binds.

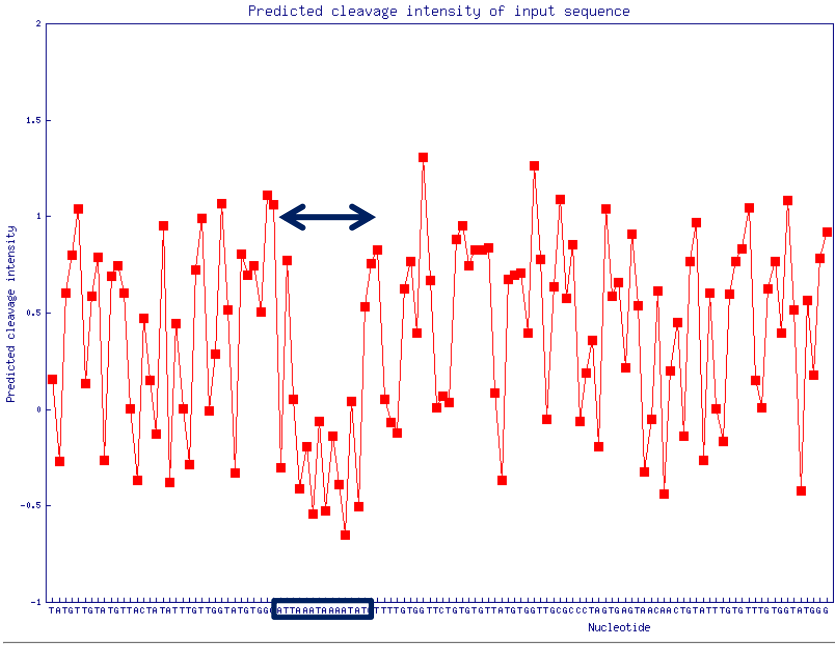


Figure 2.9. Hydroxyl radical cleavage of the sequence 7239-7358 predicted by ORChID. Low cleavage sites correlates with narrow minor grooves. The AT rich sequence correspondoing to Sites 6-7 is marked in blue.

When the DNA is incubated with a higher polyamide concentration (100 nM), new binding sites are observed and are depicted in **Figure 2.8** with red vertical arrows. The majorities of these sites usually correspond to a higher number of mismatches and higher K_d values and are characterized by sigmoidal binding curves. These sites can be described as low affinity sites that are occupied only when all the high affinity locations

are saturated. The binding at these sites occurs once the polyamide binds to the high affinity sites, probably by inducing a conformational change in the DNA structure, making these site more accessible than previously.

For quantitative DNase I footprinting experiments, DNA (200 pM final concentration) was titrated with NV1028 at a concentration ranging from zero to 40 nM. The footprinting sites match overall with the affinity cleavage results. Isotherms were obtained using data from either the HEX channel or the FAM channel depending on the location of the binding site within the fragment. Footprinting is represented in **Figure 2.8** by yellow highlighted (footprinting at 5 nM) and grey highlighted text (footprinting at 20 nM polyamide). Most of the high affinity sites show Kd_s between 0.6 and 4.9 nM. Almost all of the dissociation constants determined for the low-affinity sites show sigmoidal binding curves indicating cooperative binding. A summary of all the determined dissociation constants is provided in **Table SI 3A** and representative isotherms are provided in **Figure SI 2.**

2.3.2 Binding of polyamide NV1042 on a 385 bp fragment within the LCR of HPV18 (7179-7563- Fragment I)

Binding analysis of NV1042 was carried out on a 385 bp DNA sequence corresponding to nucleotides 7179-7563 of the HPV18 LCR. **Figure 2.10** illustrates all

the observed binding sites and one and two mismatch sites that were predicted by the pairing rules but not observed.

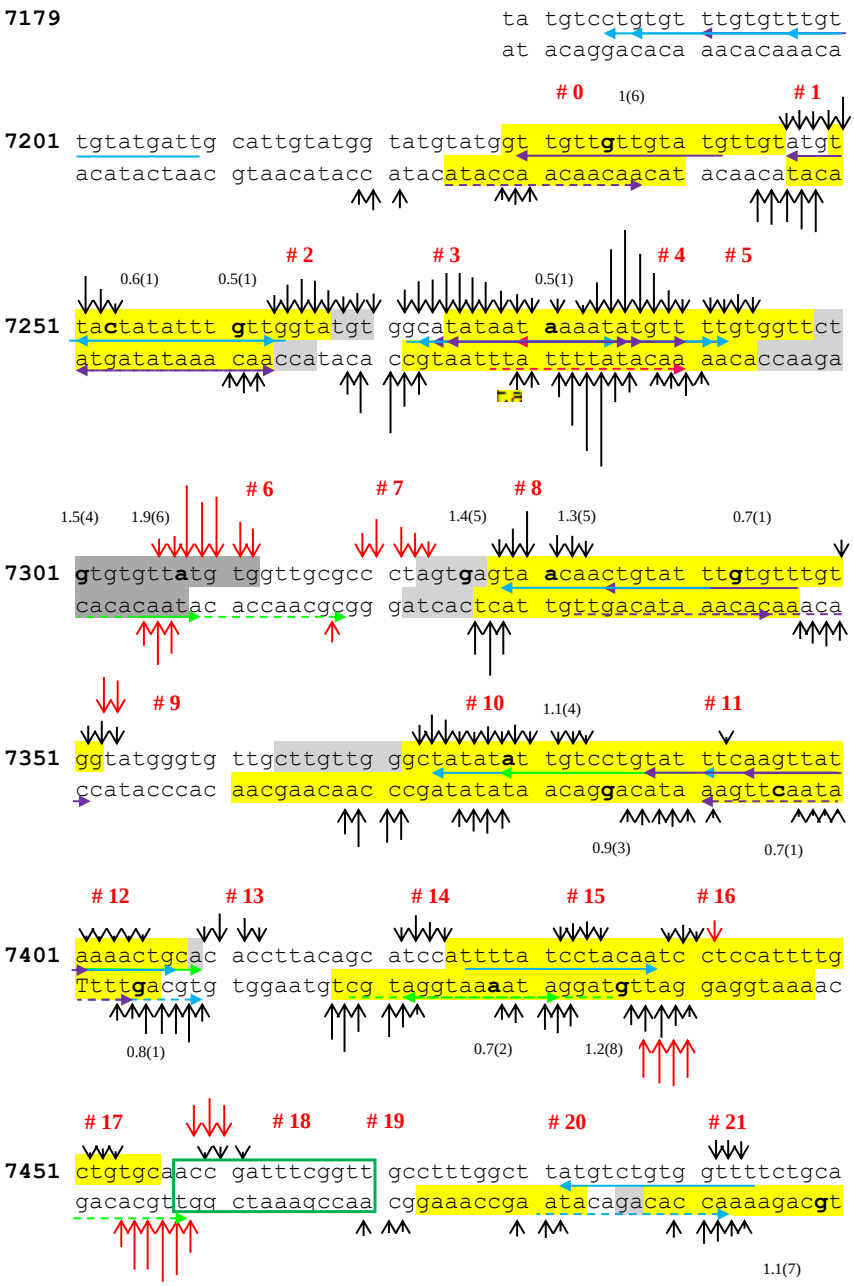
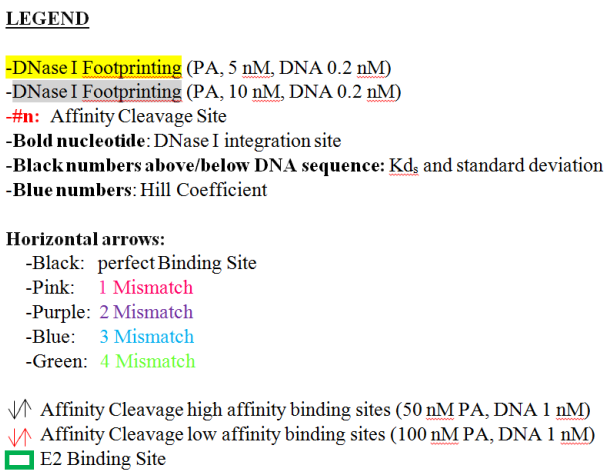


Figure 2.10. Map of affinity cleavage and footprinting of HPV18 (7189-7563) with NV1042



In contrast to NV1028, for NV1042 there are no predicted perfect binding sites in the analyzed sequence. Nevertheless, 21 binding events were observed, most of which can be explained only by a higher number of mismatch compared to NV1028 (at least two mismatches, mostly three mismatches). Conversely, the only two predicted 1-mismatch sites are not observed. Although, for this compound, the data are easier to discriminate between forward and reverse binding compared to NV1028, it is more difficult to assign the exact location of the binding sites since NV1042 can bind to a considerably higher number of mismatch sites than NV1028.

Sites **0, 1, 2, 8, 9** can be assigned almost unambiguously to two and three mismatch sites. An interesting region is the sequence corresponding to nucleotides 7274-7292 also mentioned for NV1028 in section 2.3.1, where the only two predicted 1-mismatch sites are not observed and instead the adjacent 3-mismatch site is preferred (Site **4**). As described in the previous section, the reason behind this observation likely is because these bound sites are found within the distinctive W-rich stretch, which perhaps better accommodates the polyamide regardless of the number of mismatches (as mentioned before both the shape and the electrostatic potential of DNA affect ligand binding). As described for NV1028, also for NV1042, low-affinity binding sites are observed when the

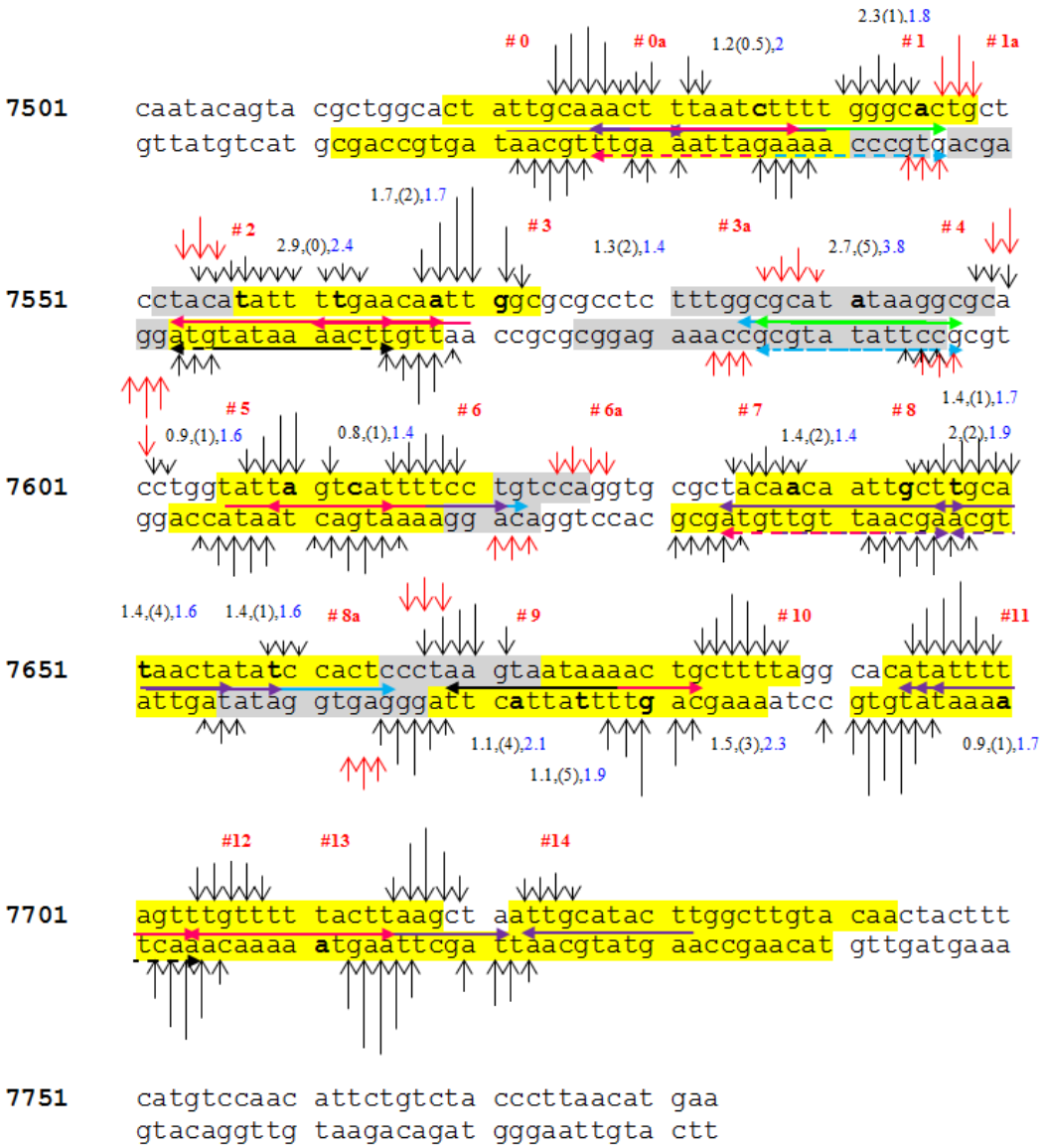
DNA is incubated with 100 nM ligand. Some of these sites can be rationalized with a mismatch number up to six.

For quantitative DNase I footprinting experiments, 200 pM of DNA were initially titrated with NV1042 at a ligand concentration ranging from zero to 40 nM, but afterward the polyamide concentration was lowered to 0-20 nM since saturation was reached at lower concentration than NV1028. The dissociation constants ranged between 0.5 nM and 1.8 nM, and, on average, are lower than those observed for NV1028. It is possible that we have reached the lower limit of detection for NV1042 with our methodology and therefore we may not have converged on the true K_d values for NV1042 yet - they could be lower. A table containing all the K_d values is provided in Table SI 3B and representative isotherms are reported **Figure SI3**.

2.3.3. Binding of polyamide NV1028 on a 305 bp fragment within the LCR of HPV18 (7479-7783 - Fragment II)

Affinity cleavage experiments were performed in triplicate in order to determine the binding sites and orientation NV1028 within the 305 bp DNA sequence. The results are shown in **Figure 2.11,** a map a representing the observed binding sites.

Overall, only four perfect binding sites (represented as black horizontal arrows) and 15 one mismatch sites are predicted for NV1028 in this DNA fragment. The map illustrates also zero or one mismatch sites that are not observed (site 2: 7553-7562, site 2-3: 7561-7570, site 3: 7565-7556, site 11: 7700-7709, site 12: 7708-7699).



LEGEND

-DNAse I Footprinting (PA, 5 nM, DNA 0.2 nM)
-DNAse I Footprinting (PA, 10 nM, DNA 0.2 nM)
-**#n:** Affinity Cleavage Site
-**Bold nucleotide**: DNase I integration site
-**Black numbers above/below DNA sequence:** K_ds and standard deviation
-**Blue numbers**: Hill Coefficient

Horizontal arrows:
-Black: perfect Binding Site
-Pink: 1 Mismatch
-Purple: 2 Mismatch
-Blue: 3 Mismatch
-Green: 4 Mismatch

Affinity Cleavage high affinity binding sites (50 nM PA, DNA 1 nM)
Affinity Cleavage low affinity binding sites (100 nM PA, DNA 1 nM)

Figure 11. Map of affinity cleavage and DNAse I footprinting of HPV18 7479-7783 with NV1028

Sites **0a, 3, 5, 6, 9, 10 and 13** match very well with the affinity cleavage and can be assigned to a single binding mode. Conversely, sites **1, 8** and **11** cannot unambiguously be explained by a single binding mode due to technique resolution. In site **11** given the shape of the affinity cleavage of the bottom strand, probably more than one polyamide is binding. Similarly, for sites **7, 8**, and **12** is not possible distinguish if the polyamide is binding in the forward or the reverse orientation. Of note, there are no data to indicate NV1028 binding to the predicted perfect binding sites 7753-7762 bp and 7565-7556 bp, or the binding is, at best, unusually weak for a perfect site relative to observed mismatch bindings. Although some affinity cleavage is observed, it is in the noise and does not show the characteristic bell shaped curve of a true affinity cleavage site. It may be possible that NV1028 is binding to all those sites in the ensemble of DNA molecules in solution and consequently provides a low and broad cleavage. NV1028 however, binds to one of these perfect binding site (7753-7762 bp) at a higher polyamide concentration (100 nM), while no binding is observed for the second site (7565-7556 bp) at the higher concentration. The reasons behind this are still not clear. The same sequences are not present in any other fragment analyzed so far, so is not possible to make a comparison and determine if the polyamide does not bind those particular positions because of the peculiar sequence or because the sequence is not in a favorable context.

The description provided above refers to affinity cleavage experiments and high affinity binding sites (observed at 50 nM polyamide concentration) which can all be explained by 0, 1 or 2 mismatch sites. As mentioned in section 2.3.1, when the DNA is incubated with a higher polyamide concentration (100 nM), new binding sites are

observed by affinity cleavage and are depicted with red vertical arrows. Most of these sites usually correspond to a higher number of mismatches, excluding the anomalous perfect binding site 2. Looking also at the sigmoidal isotherms generated at these positions, these sites can be described cooperatively as low affinity sites occurring only when all the high affinity sites are saturated.

In order to determine the binding affinity of NV1028 with the 305 bp DNA sequence, quantitative DNase I footprinting experiments were performed in triplicate according to the procedure described in section 2.2.6. DNA (200 pM final concentration) was titrated with NV1028 at a concentration ranging from 0 to 40 nM. Footprint sites are well defined at 5 nM polyamide and an increase of polyamide concentration leads to the appearance of more binding sites up to the point (40 nM) where the footprint covers most of the fragment. This protection may be due to polyamide binding or to the inability of DNase I to induce DNA cleavage at high polyamide concentration because of poisoning or because there not enough space is left for DNase I to cleave between adjacent binding sites. Overall, the footprinting data support the affinity cleavage data.

Table SI 3C provides all the K_d values determined for the fragment 7479-7783 bp. Kd_s range between 0.8 nM and 2.7 nM, and therefore within the acceptable ratio between the DNA concentration and the K_d (K_d must be at least five times the DNA concentration). Sites 5, 6 and 11, show lower K_d values (K_d= 4 x DNA concentration), but the range is still acceptable. Representative isotherms at three binding sites are reported in **Figure SI 1**.

2.3 Conclusion

The results obtained in this study show that polyamides NV1028 and NV1042 bind strongly to DNA, exhibiting binding affinity in the low nanomolar range (K_d: 0.5-5 nM). Furthermore, both compounds bind to HPV sequences in a more indiscriminate way than predicted by the reported binding rules as previously observed by our group for HPV16 DNA sequences.[2,5,21,22] NV1028 and NV1042, in fact, bind to most of the predicted 1-mismatch sites as well as to many 2-mismatch and 3-mismatch sites without significant loss in binding affinity. Of note, there are a few perfect sites where no polyamide binding is observed, and these may represent sites where the microstructure of the DNA becomes an important factor to take in consideration when predicting polyamide binding.

NV1042, in particular, exhibits a greater tolerance for a higher number of mismatch: most of the binding events are explained in fact, by at least 3-mismatch sites. The only predicted 1-mismatch sites in the fragment analyzed are not observed and the adjacent 3-mismatch sites are instead preferred. At these sites, it is likely that something beyond the sequence affects the binding. One possibility is that, the readout of the local DNA shape contributes to binding recognition. Readout of minor groove shape and electrostatic potential may in fact play a role in the recognition as observed for many DNA-specific proteins.[19] The best example supporting this idea is given in the region within Fragment I where NV1028 prefers to bind to a 1-mismatch instead of a perfect site. In this case the polyamide prefers having an A or T at the imidazole, demonstrating how the insertion of an imidazole to recognize G does not guarantee sequence specificity

because other factors play a role. This is corroborated by the fact that NV1042 also binds exactly to the same sequence, which corresponds to a 3-mismatch site in that case, although an adjacent single mismatch site is available. This observation indicates that probably the local DNA shape is driving the binding at this position regardless of the DNA sequence or the number of mismatches tolerated.

Although the binding affinities of NV1028 at all the determined binding sites are comparable to those determined for NV1042, the K_d values for NV1042 are on average lower than NV1028, although not statistically significant. Furthermore, it is worthwhile to mention that we may not have converged on the true K_d values for NV1042 yet (we have reached the lower limit of DNA concentration for DNase I footprinting with the technique currently employed).

The experiments described in this chapter provide a representation of the interactions occurring *in vitro* between polyamides and short DNA fragments. Mapping of polyamide along the entire HPV genome in its native conformation and in cells have recently been performed in our lab and will certainly provide an even more accurate depiction of the interactions occurring between large polyamides and DNA.

2.5 References

(1) Bashkin, J. K.; Edwards, T. G.; Fisher, C.; Harris, G. D.; Koeller, K. J.; Google Patents: 2016.

(2) Castaneda, C. H.; Scuderi, M. J.; Edwards, T. G.; Harris Jr, G. D.; Dupureur, C. M.; Koeller, K. J.; Fisher, C.; Bashkin, J. K. *MedChemComm* **2016**, *7*, 2076.

(3) Edwards, T. G.; Koeller, K. J.; Slomczynska, U.; Fok, K.; Helmus, M.; Bashkin, J. K.; Fisher, C. *Antiviral Research* **2011**, *91*, 177.

(4) Edwards, T. G.; Vidmar, T. J.; Koeller, K.; Bashkin, J. K.; Fisher, C. *PLoS One* **2013**, *8*, e75406.

(5) Koeller, K. J.; Harris, G. D.; Aston, K.; He, G.; Castaneda, C. H.; Thornton, M. A.; Edwards, T. G.; Wang, S.; Nanjunda, R.; Wilson, W. D.; Fisher, C.; Bashkin, J. K. *Med Chem (Los Angeles)* **2014**, *4*, 338.

(6) Brenowitz, M.; Senear, D. F.; Kingston, R. E. In *Current Protocols in Molecular Biology*; John Wiley & Sons, Inc.: 2001.

(7) Hampshire, A. J.; Rusling, D. A.; Broughton-Head, V. J.; Fox, K. R. *Methods* **2007**, *42*, 128.

(8) Mitra, S.; Shcherbakova, I. V.; Altman, R. B.; Brenowitz, M.; Laederach, A. *Nucleic Acids Research* **2008**, *36*, e63.

(9) Sivapragasam, S.; Pande, A.; Grove, A. *Anal Biochem* **2015**, *481*, 1.

(10) Connaghan-Jones, K. D.; Moody, A. D.; Bain, D. L. *Nat. Protocols* **2008**, *3*, 900.

(11) Dabrowiak, J. C.; Goodisman, J.; Ward, B. In *Drug-DNA Interaction Protocols*; Fox, K. R., Ed.; Humana Press: Totowa, NJ, 1997, p 23.

(12) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *Journal of the American Chemical Society* **1996**, *118*, 8198.

(13) Stünkel, W.; Bernard, H.-U. *J Virol* **1999**, *73*, 1918.

(14) Sanger, F.; Nicklen, S.; Coulson, A. R. *Proceedings of the National Academy of Sciences* **1977**, *74*, 5463.

(15) He, G.; Vasilieva, E.; Bashkin, J. K.; Dupureur, C. M. *Anal Biochem* **2013**, *439*, 99.

(16) Maxam, A. M.; Gilbert, W. In *Methods in Enzymology*; Academic Press: 1980; Vol. Volume 65, p 499.

(17) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468.

(18) Dervan, P. B.; Edelson, B. S. *Current Opinion in Structural Biology* **2003**, *13*, 284.

(19) Rohs, R.; West, S. M.; Sosinsky, A.; Liu, P.; Mann, R. S.; Honig, B. *Nature* **2009**, *461*, 1248.

(20) Bishop, E. P.; Rohs, R.; Parker, S. C. J.; West, S. M.; Liu, P.; Mann, R. S.; Honig, B.; Tullius, T. D. *ACS chemical biology* **2011**, *6*, 1314.

(21) He, G.; Vasilieva, E.; Harris Jr, G. D.; Koeller, K. J.; Bashkin, J. K.; Dupureur, C. M. *Biochimie* **2014**, *102*, 83.

(22) Vasilieva, E.; Niederschulte, J.; Song, Y.; Harris Jr, G. D.; Koeller, K. J.; Liao, P.; Bashkin, J. K.; Dupureur, C. M. *Biochimie* **2016**, *127*, 103.

2.6 Supporting Information

Table SI 1A. *USB Sequencing Results on HPV18 7479-7783*

HEX

Genome Position	Fragment Position	Sequence	CE postion		Genome Position	Fragment Position	Sequence	CE postion		Genome Position	Fragment Position	Sequence	CE postion
7783	1	T			7663	121	G	114.9		7543	241	C	234.7
7782	2	T			7662	122	T	115.8		7542	242	C	235.6
7781	3	C			7661	123	G	116.9		7541	243	C	236.5
7780	4	A			7660	124	G	117.9		7540	244	A	237.5
7779	5	T			7659	125	A	118.7		7539	245	A	238.5
7778	6	G			7658	126	T	119.8		7538	246	A	239.5
7777	7	T			7657	127	A	120.5		7537	247	A	240.5
7776	8	T			7656	128	T	121.6		7536	248	G	241.8
7775	9	A			7655	129	A	122.4		7535	249	A	262.6
7774	10	A			7654	130	G	123.7		7534	250	T	244
7773	11	G			7653	131	T	124.6		7533	251	T	244.8
7772	12	G			7652	132	T	125.6		7532	252	A	245.7
7771	13	G			7651	133	A	126.4		7531	253	A	246.7
7770	14	T			7650	134	T	127.5		7530	254	A	247.7
7769	15	A			7649	135	G	128.8		7529	255	G	248.9
7768	16	G			7648	136	C	129.5		7528	256	T	249.9
7767	17	A			7647	137	A	130.3		7527	257	T	250.9
7766	18	C			7646	138	A	131.1		7526	258	T	251.8
7765	19	A			7645	139	G	132.5		7525	259	G	252.9
7764	20	G			7644	140	C	133		7524	260	C	253.8
7763	21	A			7643	141	A	134.1		7523	261	A	254.7
7762	22	A			7642	142	A	135		7522	262	A	255.6
7761	23	T			7641	143	T	136.1		7521	263	T	256.7
7760	24	G			7640	144	T	137.2		7520	264	A	257.6
7759	25	T			7639	145	G			7519	265	G	258.8
7758	26	T			7638	146	T			7518	266	T	259.7
7757	27	G			7637	147	T	140.4		7517	267	G	260.9
7756	28	G			7636	148	G	141.8		7516	268	C	261.6
7755	29	A			7635	149	T	142.6		7515	269	C	262
7754	30	C			7634	150	A	143.5		7514	270	A	263.4
7753	31	A			7633	151	G	144.8		7513	271	G	264.6
7752	32	T			7632	152	C	145.5		7512	272	C	265.4
7751	33	G			7631	153	G	146.8		7511	273	G	266.6
7750	34	A			7630	154	C	147.4		7510	274	T	267.4
7749	35	A			7629	155	A	148.3		7509	275	A	268.4
7748	36	A			7628	156	C	149.2		7508	276	C	269.4
7747	37	G			7627	157	C	150		7507	277	T	270.5
7746	38	T			7626	158	T	151.2		7506	278	G	271.6
7745	39	A			7625	159	G	152.5		7505	279	T	272.7
7744	40	G			7624	160	G	153.7		7504	280	A	273.6
7743	41	T			7623	161	A	154.7		7503	281	T	274.7
7742	42	T			7622	162	C	155.7		7502	282	T	275.8
7741	43	G			7621	163	A	156.6		7501	283	G	276.9
7740	44	T			7620	164	G	157.8		7500	284	T	277.9
7739	45	A			7619	165	G	159		7499	285	G	279
7738	46	C			7618	166	A	160		7498	286	C	279.8
7737	47	A			7617	167	A	160.8		7497	287	A	280.7
7736	48	A			7616	168	A	162		7496	288	G	281.8
7735	49	G			7615	169	A	162.6		7495	289	A	283
7734	50	C			7614	170	T	163.8		7494	290	A	283.6
7733	51	C			7613	171	G	164.9		7493	291	A	285
7732	52	A			7612	172	A	165.7		7492	292	A	285.5
7731	53	A			7611	173	C	166.7		7491	293	C	289
7730	54	G			7610	174	T	167.8		7490	294	C	288
7729	55	T			7609	175	A	168.6		7489	295	A	288.3
7728	56	A	46.4		7608	176	A	170		7488	296	C	289.4
7727	57	T	47.9		7607	177	T	170.8		7487	297	A	290.2
7726	58	G	48.8		7606	178	A	171.7		7486	298	G	291
7725	59	C	49.6		7605	179	C	172.7		7485	299	A	292.3
7724	60	A	50.5		7604	180	C	173.5		7484	300	C	293.4
7723	61	A	51.2		7603	181	A	174.4		7483	301	A	294.3
7722	62	T	52.7		7602	182	G	175.7		7482	302	T	295.3
7721	63	T	53.9		7601	183	G	177		7481	303	A	296.3

7720	64	A	54.7		7600	184	T	178		7480	304	A	
7719	65	G	56.4		7599	185	G	179.1		7479	305	G	
7718	66	C	57.1		7598	186	C	179.7					
7717	67	T	58.4		7597	187	G	181					
7716	68	T	59.5		7596	188	C	181.7					
7715	69	A	60.3		7595	189	C	182.6					
7714	70	A	61.4		7594	190	T	183.6					
7713	71	G	63.1		7593	191	T	185					
7712	72	T	64		7592	192	A	185.7					
7711	73	A	64.9		7591	193	T	187					
7710	74	A	65.7		7590	194	A	187.8					
7709	75	A	66.7		7589	195	T	189					
7708	76	A	67.6		7588	196	G	190.3					
7707	77	A	68.6		7587	197	C	191.1					
7706	78	C	69.5		7586	198	G	192.3					
7705	79	A	70.6		7585	199	C	193					
7704	80	A	71.5		7584	200	C	193.9					
7703	81	A	72.6		7583	201	A	194.8					
7702	82	C	73.5		7582	202	A	196					
7701	83	T	74.1		7581	203	A	196.9					
7700	84	A	74.9		7580	204	G	198.1					
7699	85	A	75.8		7579	205	A	198.9					
7698	86	A	76.9		7578	206	G	200.2					
7697	87	A			7577	207	G	201.2					
7696	88	T			7576	208	C	202					
7695	89	A			7575	209	G	203.1					
7694	90	T			7574	210	C	203.6					
7693	91	G			7573	211	G	204.7					
7692	92	T			7572	212	C	205.6					
7691	93	G	85.9		7571	213	C	206.5					
7690	94	C	86.5		7570	214	A	207.5					
7689	95	C	87.4		7569	215	A	208.5					
7688	96	T	87.8		7568	216	T	210					
7687	97	A	89.5		7567	217	T	210.8					
7686	98	A	90.5		7566	218	G	212					
7685	99	A	91.6		7565	219	T	213					
7684	100	A	92.5		7564	220	T	214					
7683	101	G	93.9		7563	221	C	215					
7682	102	C	94		7562	222	A	215.8					
7681	103	A	95.3		7561	223	A	216.8					
7680	104	G	96.6		7560	224	A	217.8					
7679	105	T	97.5		7559	225	A	218.8					
7678	106	T	98.6		7558	226	T	220					
7677	107	T	99.7		7557	227	A	220.8					
7676	108	T	100.8		7556	228	T	222					
7675	109	A	101.7		7555	229	G	223					
7674	110	T	103.1		7554	230	T	224.1					
7673	111	T	104.2		7553	231	A	224.9					
7672	112	A	105.1		7552	232	G	226					
7671	113	C	106		7551	233	G	227.2					
7670	114	T	107.2		7550	234	A	228					
7669	115	T	108		7549	235	G	229.2					
7668	116	A	109.2		7548	236	C	229.9					
7667	117	G	110.5		7547	237	A	230.8					
7666	118	G	111.8		7546	238	G	232					
7665	119	G	112.9		7545	239	T	232.9					
7664	120	A	113.7		7544	240	G	234					

FAM

Genome Position	Fragment Position	Sequence	CE postion		Genome Position	Fragment Position	Sequence	CE postion		Genome Position	Fragment Position	Sequence	CE postion
7479	1	C			7599	121	C	116.1		7719	241	C	236.9
7480	2	T			7600	122	A	117.1		7720	242	T	237.8
7481	3	T			7601	123	C	117.8		7721	243	A	238.7
7482	4	A			7602	124	C	118.7		7722	244	A	240
7483	5	T			7603	125	T	119.7		7723	245	T	240.9
7484	6	G			7604	126	G	120.8		7724	246	T	241.9
7485	7	T			7605	127	G	122		7725	247	G	243.1
7486	8	C			7606	128	T	123.3		7726	248	C	244
7487	9	T			7607	129	A	124.2		7727	249	A	245
7488	10	G			7608	130	T	125.2		7728	250	T	245.9
7489	11	T			7609	131	T	126.2		7729	251	A	247
7490	12	G			7610	132	A	127.2		7730	252	C	248
7491	13	G			7611	133	G	128.2		7731	253	T	249
7492	14	T			7612	134	T	129.2		7732	254	T	250.1
7493	15	T			7613	135	C	130.2		7733	255	G	251.3
7494	16	T			7614	136	A	131		7734	256	G	252.3
7495	17	T			7615	137	T	132		7735	257	C	253.3
7496	18	C			7616	138	T	133		7736	258	T	254.2
7497	19	T			7617	139	T	134		7737	259	T	255.1
7498	20	G			7618	140	T	135		7738	260	G	256.3
7499	21	C			7619	141	C	136		7739	261	T	257.2
7500	22	A			7620	142	C	136.9		7740	262	A	258.3
7501	23	C			7621	143	T	138		7741	263	C	259.3
7502	24	A			7622	144	G	139.1		7742	264	A	260.1
7503	25	A			7623	145	T	140.2		7743	265	A	261
7504	26	T			7624	146	C	141.2		7744	266	C	262
7505	27	A			7625	147	C	142		7745	267	T	262.9
7506	28	C			7626	148	A	143		7746	268	A	264
7507	29	A			7627	149	G	144.1		7747	269	C	265.1
7508	30	G			7628	150	G	145		7748	270	T	266
7509	31	T			7629	151	T	146.2		7749	271	T	267
7510	32	A			7630	152	G	147.4		7750	272	T	268.2
7511	33	C			7631	153	C	148.3		7751	273	C	269.2
7512	34	G			7632	154	G			7752	274	A	270.2
7513	35	C			7633	155	C	150.2		7753	275	T	271.2
7514	36	T			7634	156	T	151		7754	276	G	272.5
7515	37	G			7635	157	A	152		7755	277	T	
7516	38	G			7636	158	C	153.1		7756	278	C	274.4
7517	39	C			7637	159	A	154		7757	279	C	275.5
7518	40	A			7638	160	A	155.1		7758	280	A	276.1
7519	41	C			7639	161	C	156.1		7759	281	A	277.1
7520	42	T			7640	162	A	157.1		7760	282	C	278.2
7521	43	A			7641	163	A	158		7761	283	A	279.2
7522	44	T			7642	164	T	159.1		7762	284	T	280
7523	45	T			7643	165	T	160.2		7763	285	T	281.2
7524	46	G			7644	166	G	161.4		7764	286	C	282.3
7525	47	C			7645	167	C	162.4		7765	287	T	283.2
7526	48	A			7646	168	T	163.3		7766	288	G	284.4
7527	49	A			7647	169	T	164.3		7767	289	T	285.3
7528	50	A			7648	170	G	165.4		7768	290	C	286.5
7529	51	C	40.5		7649	171	C	166.4		7769	291	T	287.3
7530	52	T	42		7650	172	A	167.3		7770	292	A	288.4
7531	53	T	43.2		7651	173	T	168.3		7771	293	C	289.3
7532	54	T	44.5		7652	174	A	169.3		7772	294	C	290.2
7533	55	A			7653	175	A	170.3		7773	295	C	291.1
7534	56	A			7654	176	C	171.3		7774	296	T	292
7535	57	T	48		7655	177	T	172.3		7775	297	T	293.7
7536	58	C	49		7656	178	A	173.3		7776	298	A	294.3
7537	59	T			7657	179	T	174.4		7777	299	A	295.3
7538	60	T			7658	180	A	175.5		7778	300	C	296.3
7539	61	T	54.1		7659	181	T	176.4		7779	301	A	297.3
7540	62	T	55.2		7660	182	C	177.4		7780	302	T	298.2
7541	63	G			7661	183	C	178.3		7781	303	G	

7542	64	G			7662	184	A	179.3		7782	304	A	
7543	65	G			7663	185	C	180.2		7783	305	A	
7544	66	C	59.3		7664	186	T	181.2					
7545	67	A	60.2		7665	187	C	182.3					
7546	68	C	60.5		7666	188	C	183.2					
7547	69	T	61.7		7667	189	C	184.1					
7548	70	G	62.9		7668	190	T	185.1					
7549	71	C	63.9		7669	191	A	186.4					
7550	72	T	65		7670	192	A	187.4					
7551	73	C	65.9		7671	193	G	188.5					
7552	74	C	66.8		7672	194	T	189.5					
7553	75	T	68		7673	195	A	190.6					
7554	76	A	69.3		7674	196	A	191.4					
7555	77	C	70.2		7675	197	T	192.5					
7556	78	A	71.5		7676	198	A	193.5					
7557	79	T	72.5		7677	199	A	194.5					
7558	80	A	73.7		7678	200	A	195.4					
7559	81	T	74.7		7679	201	A	196.4					
7560	82	T	75.9		7680	202	C	197.4					
7561	83	T	77		7681	203	T	198.4					
7562	84	T	78.2		7682	204	G	199.5					
7563	85	G	79.3		7683	205	C	200.5					
7564	86	A			7684	206	T	201.5					
7565	87	A			7685	207	T	202.5					
7566	88	C	82.2		7686	208	T	203.5					
7567	89	A	83.2		7687	209	T	204.6					
7568	90	A			7688	210	A	205.6					
7569	91	T			7689	211	G	206.8					
7570	92	T			7690	212	G	207.9					
7571	93	G			7691	213	C	208.8					
7572	94	G			7692	214	A	209.6					
7573	95	C			7693	215	C	210.4					
7574	96	G			7694	216	A	211.5					
7575	97	C			7695	217	T	212.3					
7576	98	G			7696	218	A	213.5					
7577	99	C			7697	219	T	214.5					
7578	100	C			7698	220	T	215.6					
7579	101	T			7699	221	T	216.7					
7580	102	C	96.1		7700	222	T	217.7					
7581	103	T	97.3		7701	223	A	218.7					
7582	104	T	98.5		7702	224	G	219.8					
7583	105	T	99.8		7703	225	T	220.9					
7584	106	G	101.1		7704	226	T	221.9					
7585	107	G	102.2		7705	227	T	221.9					
7586	108	C	103.3		7706	228	G	223.9					
7587	109	G	104.1		7707	229	T	225					
7588	110	C	104.6		7708	230	T	226					
7589	111	A	105.6		7709	231	T	226.8					
7590	112	T	106.7		7710	232	T	228					
7591	113	A	107.9		7711	233	T	228.9					
7592	114	T	109		7712	234	A	229.9					
7593	115	A	110.2		7713	235	C	230.8					
7594	116	A	111.2		7714	236	T	231.9					
7595	117	G	112.5		7715	237	T	232.8					
7596	118	G	113.1		7716	238	A	233.8					
7597	119	C	114.4		7717	239	A	234.8					
7598	120	G	115.3		7718	240	G	235.9					

Table SI 1B. *USB Sequencing Results on HPV18 7179-7563*

HEX

Genome position	Fragment Position	Sequence	CE Position		Genome position	Fragment Position	Sequence	CE Position		Genome position	Fragment Position	Sequence	CE Position
7563	1	C			7434	130	A	123		7305	259	C	251.9
7562	2	A			7433	131	G	124.1		7304	260	A	252.8
7561	3	A			7432	132	G	125.1		7303	261	C	253.7
7560	4	A			7431	133	A			7302	262	A	254.7
7559	5	A			7430	134	T			7301	263	C	255.5
7558	6	T			7429	135	A			7300	264	A	256.6
7557	7	A			7428	136	A			7299	265	G	257.7
7556	8	T			7427	137	A			7298	266	A	258.7
7555	9	G			7426	138	A			7297	267	A	259.6
7554	10	T			7425	139	T			7296	268	C	260.5
7553	11	A			7424	140	G			7295	269	C	261.4
7552	12	G			7423	141	G	133.9		7294	270	A	262.5
7551	13	G			7422	142	A	134.8		7293	271	C	263.5
7550	14	A			7421	143	T	135.9		7292	272	A	
7549	15	G			7420	144	G	136.9		7291	273	A	
7548	16	C			7419	145	C	137.6		7290	274	A	
7547	17	A			7418	146	T	138.7		7289	275	A	
7546	18	G			7417	147	G	139.7		7288	276	C	268.4
7545	19	T			7416	148	T	140.9		7287	277	A	269.4
7544	20	G			7415	149	A	141.9		7286	278	T	270.5
7543	21	C			7414	150	A	142.8		7285	279	A	271.9
7542	22	C			7413	151	G	143.7		7284	280	T	oroad peal
7541	23	C			7412	152	G	144.7		7283	281	T	
7540	24	A			7411	153	C			7282	282	T	
7539	25	A			7410	154	G	147		7281	283	T	275.7
7538	26	A			7409	155	T	148		7280	284	A	276.6
7537	27	A			7408	156	G	149.1		7279	285	T	277.8
7536	28	G			7407	157	C	149.8		7278	286	T	278.8
7535	29	A			7406	158	A	150.7		7277	287	T	
7534	30	T			7405	159	G	151.8		7276	288	A	
7533	31	T			7404	160	T	152.9		7275	289	A	
7532	32	A			7403	161	T	153.9		7274	290	T	
7531	33	A			7402	162	T	154.9		7273	291	G	283.7
7530	34	A			7401	163	T	156		7272	292	C	284.6
7529	35	G			7400	164	A	156.9		7271	293	C	285.5
7528	36	T			7399	165	T	158		7270	294	A	286.4
7527	37	T			7398	166	A	159		7269	295	C	287.3
7526	38	T			7397	167	A	159.9		7268	296	A	288.4
7525	39	G			7396	168	C	160.8		7267	297	T	289.5
7524	40	C			7395	169	T	162		7266	298	A	290.4
7523	41	A			7394	170	T			7265	299	C	291.3
7522	42	A			7393	171	G			7264	300	C	292.3
7521	43	T			7392	172	A			7263	301	A	293.3
7520	44	A			7391	173	A			7262	302	A	294.3
7519	45	G	33.5		7390	174	A	167		7261	303	C	295.3
7518	46	T	34.6		7389	175	T	168		7260	304	A	
7517	47	G	36.2		7388	176	A	169		7259	305	A	
7516	48	C			7387	177	C	169.9		7258	306	A	
7515	49	C			7386	178	A	170.9		7257	307	T	299.3
7514	50	A	38.3		7385	179	G	172.1		7256	308	A	300.3
7513	51	G	39.9		7384	180	G	173.2		7255	309	T	
7512	52	C	40.3		7383	181	A	174.1		7254	310	A	
7511	53	G	42.1		7382	182	C	175		7253	311	G	303.3
7510	54	T	43		7381	183	A	176		7252	312	T	304.3
7509	55	A	44.3		7380	184	A	176.9		7251	314	A	305.3
7508	56	C	45.5		7379	185	T	178		7250	315	A	306.2
7507	57	T	47.6		7378	186	A	179		7249	315	C	307.1
7506	58	G	48.8		7377	187	T	180		7248	316	A	308.1
7505	59	T	49.9		7376	188	A	181		7247	317	T	309.1
7504	60	A	50.8		7375	189	T	182.1		7246	318	A	310.1
7503	61	T	52		7374	190	A	183.2		7245	319	C	311
7502	62	T	53.2		7373	191	G	184.3		7244	320	A	312.1
7501	63	G	54.6		7372	192	C	185		7243	321	A	313
7500	64	T	55.6		7371	193	C	185.9		7242	322	C	313.9
7499	65	G	57		7370	194	C	186.7		7241	323	A	315
7498	66	C	57.5		7369	195	A	187.9		7240	324	T	316.1
7497	67	A	58.5		7368	196	A	188.9		7239	325	A	317.1
7496	68	G	59.9		7367	197	C	189.9		7238	326	C	318.1
7495	69	A	60.5		7366	198	A	191		7237	327	A	319
7494	70	A	61.6		7365	199	A	191.9		7236	328	A	320.1
7493	71	A			7364	200	G	193.1		7235	329	C	321.1

FAM

Genome position	Fragment Position	Sequence	CE Position		Genome position	Fragment Position	Sequence	CE Position		Genome position	Fragment Position	Sequence	CE Position
7179	1	T			7308	130	A	127.4		7437	259	A	257.9
7180	2	A			7309	131	T	128.3		7438	260	A	258.8
7181	3	T			7310	132	G	129.5		7439	261	T	259.8
7182	4	G			7311	133	T	130.5		7440	262	C	260.7
7183	5	T			7312	134	G	131.6		7441	263	C	261.7
7184	6	C			7313	135	G	132.7		7442	264	T	262.6
7185	7	C			7314	136	T	133.6		7443	265	C	263.6
7186	8	T			7315	137	T	134.5		7444	266	C	264.6
7187	9	G			7316	138	G	135.6		7445	267	A	265.7
7188	10	T			7317	139	C	136.4		7446	268	T	266.6
7189	11	G			7318	140	G	137.5		7447	269	T	267.7
7190	12	T			7319	141	C	138.1		7448	270	T	268.8
7191	13	T			7320	142	C	138.9		7449	271	T	269.9
7192	14	T			7321	143	C	139.8		7450	272	G	271.1
7193	15	G			7322	144	T	140.9		7451	273	C	272
7194	16	T			7323	145	A	142		7452	274	T	272.9
7195	17	G			7324	146	G	143.2		7453	275	G	274.2
7196	18	T			7325	147	T	144.4		7454	276	T	275.1
7197	19	T			7326	148	G	145.5		7455	277	G	276.3
7198	20	T			7327	149	A	146.5		7456	278	C	277
7199	21	G			7328	150	G	147.6		7457	279	A	277.9
7200	22	T			7329	151	T	148.6		7458	280	A	278.7
7201	23	T			7330	152	A	149.5		7459	281	C	279.5
7202	24	G			7331	153	A	150.4		7460	282	C	280.5
7203	25	T			7332	154	C	151.2		7461	283	G	281.7
7204	26	A			7333	155	A	152.1		7462	284	A	282.8
7205	27	T			7334	156	A	153.1		7463	285	T	283.8
7206	28	G			7335	157	C	154		7464	286	T	284.9
7207	29	A			7336	158	T	155.1		7465	287	T	285.9
7208	30	T			7337	159	G	156.3		7466	288	C	286.9
7209	31	T			7338	160	T	157.4		7467	289	G	288
7210	32	G			7339	161	A	158.4		7468	290	G	289
7211	33	C			7340	162	T	159.5		7469	291	T	290
7212	34	A			7341	163	T	160.6		7470	292	T	291
7213	35	T			7342	164	T	161.6		7471	293	G	292
7214	36	T			7343	165	G	162.8		7472	294	C	292.5
7215	37	G			7344	166	T	163.8		7473	295	C	293.5
7216	38	T			7345	167	G	165		7474	296	T	294.6
7217	39	A			7346	168	T	165.9		7475	297	T	295.8
7218	40	T			7347	169	T	166.9		7476	298	T	296.8
7219	41	G			7348	170	T	167.9		7477	299	G	298
7220	42	G			7349	171	G	169		7478	300	G	299.1
7221	43	T			7350	172	T	170.6		7479	301	C	299.7
7222	44	A			7351	173	G	171.3		7480	302	T	300.8
7223	45	T			7352	174	G	172.4		7481	303	T	301.8
7224	46	G			7353	175	T			7482	304	A	303
7225	47	T			7354	176	A	174		7483	305	T	304
7226	48	A			7355	177	T	175		7484	306	G	305.1
7227	49	T			7356	178	G	176.1		7485	307	T	306.1
7228	50	G			7357	179	G	177.2		7486	308	C	307
7229	51	G			7358	180	G	178.3		7487	309	T	308
7230	52	T			7359	181	T			7488	310	G	309
7231	53	T			7360	182	G	180.3		7489	311	T	310.1
7232	54	G			7361	183	T	181.3		7490	312	G	311.2
7233	55	T			7362	184	T	182.3		7491	313	G	312.3
7234	56	T			7363	185	G	183.3		7492	314	T	313.3
7235	57	G			7364	186	C	184.1		7493	315	T	314.2
7236	58	T	52.2		7365	187	T	185.2		7494	316	T	315.3
7237	59	T	53.3		7366	188	T	186.1		7495	317	T	316.2
7238	60	G	54.5		7367	189	G	187.2		7496	318	C	317.2
7239	61	T	55.5		7368	190	T	188.3		7497	319	T	318.2
7240	62	A	56.4		7369	191	T	189.4		7498	320	G	319.4
7241	63	T	57.5		7370	192	G	190.5		7499	321	C	320.4
7242	64	G	58.7		7371	193	G	191.6		7500	322	A	321.4
7243	65	T	59.7		7372	194	G	192.8		7501	323	C	322.2
7244	66	T	60.8		7373	195	C	193.6		7502	324	A	323.3
7245	67	G	62		7374	196	T	194.5		7503	325	A	324.2
7246	68	T	63.1		7375	197	A	195.5		7504	326	T	325.2
7247	69	A	64		7376	198	T	196.5		7505	327	A	326.4
7248	70	T	65		7377	199	A	197.6		7506	328	C	327.3
7249	71	G	66.3		7378	200	T	198.5		7507	329	A	328.3

7250	72	T	67.3		7379	201	A	199.6		7508	330	G	329.3
7251	73	T	68.4		7380	202	T	200.5		7509	331	T	330.3
7252	74	A	69.3		7381	203	T	201.6		7510	332	A	331.4
7253	75	C	69.9		7382	204	G	202.8		7511	333	C	332.2
7254	76	T	71.2		7383	205	T	203.8		7512	334	G	333.4
7255	77	A	72.3		7384	206	C	204.7		7513	335	C	334.2
7256	78	T	73.4		7385	207	C	205.5		7514	336	T	335.1
7257	79	A	74.5		7386	208	T	206.5		7515	337	G	336.3
7258	80	T	75.7		7387	209	G	207.7		7516	338	G	337.3
7259	81	T	76.8		7388	210	T	208.8		7517	339	C	338.3
7260	82	T	77.9		7389	211	A	209.8		7518	340	A	339.4
7261	83	G	79.2		7390	212	T	210.9		7519	341	C	340.1
7262	84	T	80.2		7391	213	T	211.9		7520	342	T	341.1
7263	85	T	81.3		7392	214	T	212.9		7521	343	A	342.3
7264	86	G	82.6		7393	215	C	213.9		7522	344	T	343.2
7265	87	G	83.7		7394	216	A	214.8		7523	345	T	344.3
7266	88	T	84.8		7395	217	A	215.7		7524	346	G	345.5
7267	89	A	85.5		7396	218	G	216.8		7525	347	C	346.3
7268	90	T	86.5		7397	219	T	217.8		7526	348	A	347.5
7269	91	G	87.8		7398	220	T	218.8		7527	349	A	
7270	92	T	88.7		7399	221	A	219.8		7528	350	A	349.1
7271	93	G	89.9		7400	222	T	220.9		7529	351	C	350.1
7272	94	G	91.1		7401	223	A	222		7530	352	T	351.2
7273	95	C	91.7		7402	224	A	222.8		7531	353	T	352.2
7274	96	A	92.5		7403	225	A	223.8		7532	354	T	353.2
7275	97	T			7404	226	A	224.7		7533	355	A	354.4
7276	98	T			7405	227	C	225.6		7534	356	A	355.3
7277	99	A			7406	228	T	226.6		7535	357	T	356.4
7278	100	A			7407	229	G	227.8		7536	358	C	357.3
7279	101	A	97.6		7408	230	C	228.7		7537	359	T	358.3
7280	102	T			7409	231	A	229.6		7538	360	T	359.4
7281	103	A	99.6		7410	232	C	230.5		7539	361	T	360.4
7282	104	A	100.7		7411	233	A	231.5		7540	362	T	361.5
7283	105	A	101.5		7412	234	C	232.4		7541	363	G	362.7
7284	106	A	102.5		7413	235	C	233.4		7542	364	G	363.7
7285	107	T	103.7		7414	236	T	233.1		7543	365	G	364.9
7286	108	A	104.6		7415	237	T			7544	366	C	365.7
7287	109	T	105.6		7416	238	A	236.6		7545	367	A	366.6
7288	110	G	106.9		7417	239	C	237.6		7546	368	C	367.3
7289	111	T	108		7418	240	A	238.6		7547	369	T	368.3
7290	112	T	109		7419	241	G	239.7		7548	370	G	369.5
7291	113	T	110		7420	242	C	240.6		7549	371	C	370.3
7292	114	T	111		7421	243	A	241.6		7550	372	T	371.4
7293	115	G	112.3		7422	244	T	242.6		7551	373	C	372.3
7294	116	T	113.3		7423	245	C	243.6		7552	374	C	373.3
7295	117	G	114.4		7424	246	C	244.5		7553	375	T	374.4
7296	118	G	115.6		7425	247	A	245.5		7554	376	A	375.4
7297	119	T	116.6		7426	248	T	246.7		7555	377	C	376.2
7298	120	T	117.5		7427	249	T	247.8		7556	378	A	377.4
7299	121	C	118.2		7428	250	T	248.9		7557	379	T	378.3
7300	122	T	119.2		7429	251	T	250		7558	380	A	379.5
7301	123	G	120.4		7430	252	A	251		7559	381	T	380.4
7302	124	T	121.4		7431	253	T	252		7560	382	T	381.4
7303	125	G	122.5		7432	254	C	252.9		7561	383	T	
7304	126	T	123.5		7433	255	C	253.8		7562	384	T	
7305	127	G	124.6		7434	256	T	254.8		7563	385	G	
7306	128	T	125.5		7435	257	A	255.9					
7307	129	T	126.5		7436	258	C	256.8					

Table SI 2. *Maxam-Gilbert Sequencing on HPV18 7179-7563*

HEX

Genome Position	Real Position	Sequence	CE position	Genome Position	Real Position	Sequence	CE position	Genome Position	Real Position	Sequence	CE position
7563	1	C		7436	128	G	119.2	7306	258	A	250.4
7562	2	A		7435	129	T	120.5	7305	259	C	251
7561	3	A		7434	130	A	121.5	7304	260	A	252.2
7560	4	A		7433	131	G	122.3	7303	261	C	253.2
7559	5	A		7432	132	G	123.5	7302	262	A	254.2
7558	6	T		7431	133	A	124.6	7301	263	C	255.1
7557	7	A		7430	134	T	125.4	7300	264	A	256.1
7556	8	T		7429	135	A	126.4	7299	265	G	257
7555	9	G		7428	136	A	127.3	7298	266	A	258.2
7554	10	T		7427	137	A	128	7297	267	A	259.1
7553	11	A		7426	138	A	129	7296	268	C	260.1
7552	12	G		7425	139	T	129.9	7295	269	C	261.1
7551	13	G		7424	140	G	131.1	7294	270	A	261.9
7550	14	A		7423	141	G	132	7293	271	C	262.9
7549	15	G		7422	142	A	133.5	7292	272	A	263.9
7548	16	C		7421	143	T	134	7291	273	A	264.9
7547	17	A		7420	144	G	135.4	7290	274	A	265.9
7546	18	G		7419	145	C	136.6	7289	275	A	266.9
7545	19	T		7418	146	T	137.7	7288	276	C	267.8
7544	20	G		7417	147	G	138.4	7287	277	A	268.8
7543	21	C		7416	148	T	139.6	7286	278	T	269.8
7542	22	C		7415	149	A	140.5	7285	279	A	271
7541	23	C		7414	150	A	141.5	7284	280	T	272
7540	24	A		7413	151	G	142.4	7283	281	T	273
7539	25	A		7412	152	G	143.5	7282	282	T	274
7538	26	A		7411	153	C	144.7	7281	283	T	275
7537	27	A		7410	154	G	145	7280	284	A	276.1
7536	28	G		7409	155	T	146.9	7279	285	T	277.1
7535	29	A		7408	156	G	147.9	7278	286	T	278
7534	30	T		7407	157	C	149.1	7277	287	T	
7533	31	T		7406	158	A	149.9	7276	288	A	280.2
7532	32	A		7405	159	G	150.9	7275	289	A	281.2
7531	33	A		7404	160	T	152	7274	290	T	282.1
7530	34	A		7403	161	T	153	7273	291	G	283.1
7529	35	G		7402	162	T	154	7272	292	C	284.3
7528	36	T		7401	163	T	155	7271	293	C	
7527	37	T		7400	164	A	156.1	7270	294	A	286
7526	38	T		7399	165	T	157	7269	295	C	287
7525	39	G		7398	166	A	158.1	7268	296	A	288
7524	40	C		7397	167	A	159	7267	297	T	289
7523	41	A		7396	168	C	159.9	7266	298	A	290
7522	42	A		7395	169	T	160.8	7265	299	C	290.9
7521	43	T		7394	170	T	161.9	7264	300	C	291.8
7520	44	A		7393	171	G	163	7263	301	A	292.7
7519	45	G		7392	172	A	164.2	7262	302	A	293.7
7518	46	T		7391	173	A	165.1	7261	303	C	294.7
7517	47	G		7390	174	A	166	7260	304	A	295.7
7516	48	C		7389	175	T	167	7259	305	A	296.7
7515	49	C		7388	176	A	168	7258	306	A	297.7
7514	50	A		7387	177	C	169	7257	307	T	298.6
7513	51	G		7386	178	A	169.9	7256	308	A	299.7
7512	52	C		7385	179	G	170.8	7255	309	T	300.8
7511	53	G		7384	180	G	172.1	7254	310	A	301.7
7510	54	T		7383	181	A	173	7253	311	G	302.7
7509	55	A		7382	182	C	174.2	7252	312	T	303.9
7508	56	C		7381	183	A	175	7251	314	A	304.8
7507	57	T		7380	184	A	176	7250	315	A	305.8
7506	58	G		7379	185	T	177	7249	315	C	306.6
7505	59	T		7378	186	A	178	7248	316	A	307.7
7504	60	A		7377	187	T	179	7247	317	T	308.6
7503	61	T		7376	188	A	180.1	7246	318	A	309.7
7502	62	T		7375	189	T	181.1	7245	319	C	310.5
7501	63	G		7374	190	A	182.1	7244	320	A	311.5
7500	64	T		7373	191	G	183	7243	321	A	312.5
7499	65	G		7372	192	C	184	7242	322	C	313.6
7498	66	C	53.7	7371	193	C		7241	323	A	314.6
7497	67	A	54	7370	194	C	186	7240	324	T	315.5
7496	68	G	55	7369	195	A	187	7239	325	A	316.6
7495	69	A	56	7368	196	A	188	7238	326	C	317.8
7494	70	A	57	7367	197	C	189	7237	327	A	318.8
7493	71	A	58	7366	198	A	189.9	7236	328	A	319.6
7492	72	A	59	7365	199	A	191	7235	329	C	320.7
7491	73	C	60.7	7364	200	G	191.9	7234	330	A	321.5

7490	74	C	61.7		7363	201	C	193.2		7233	331	A	322.5
7489	75	A	62.5		7362	202	A	194.1		7232	332	C	323.9
7488	76	C	63.7		7361	203	A	195.1		7231	333	A	324.4
7487	77	A	64.5		7360	204	C	196		7230	334	A	325.5
7486	78	G	65.5		7359	205	A	196.9		7229	335	C	326.6
7485	79	A	66.5		7358	206	C	198		7228	336	C	327.4
7484	80	C	68.4		7357	207	C	198.9		7227	337	A	328.5
7483	81	A	69.3		7356	208	C	199.8		7226	338	T	329.6
7482	82	T	70.5		7355	209	A	200.8		7225	339	A	330.6
7481	83	A	71.7		7354	210	T	201.8		7224	340	C	331.6
7480	84	A	72.7		7353	211	A	203.1		7223	341	A	332
7479	85	G	73.8		7352	212	C	204		7222	342	T	333
7478	86	C	75.2		7351	213	C	204.9		7221	343	A	334.6
7477	87	C	76.2		7350	214	A	205.8		7220	344	C	335.7
7476	88	A			7349	215	C	206.9		7219	345	C	336.7
7475	89	A	77.9		7348	216	A	207.8		7218	346	A	337.3
7474	90	A	78.9		7347	217	A	208.8		7217	347	T	338.4
7473	91	G	79.7		7346	218	A	209.8		7216	348	A	339.5
7472	92	G	81.4		7345	219	C	210		7215	349	C	340.5
7471	93	C	82.7		7344	220	A	211.7		7214	350	A	341.5
7470	94	A	93.7		7343	221	C	212.8		7213	351	A	342.5
7469	95	A	84.1		7342	222	A	213.7		7212	352	T	343.5
7468	96	C	85.1		7341	223	A	214.7		7211	353	G	344.5
7467	97	C	86		7340	224	A	215.8		7210	354	C	345.5
7466	98	G	87		7339	225	T	216.8		7209	355	A	346.6
7465	99	A	88		7338	226	A	217.9		7208	356	A	347.7
7464	100	A	89.6		7337	227	C	218.9		7207	357	T	348.6
7463	101	A	90.5		7336	228	A	219.8		7206	358	C	
7462	102	T	91.5		7335	229	G	220.8		7205	359	A	350.5
7461	103	C	92.6		7334	230	T	222.1		7204	360	T	351.5
7460	104	G	93.5		7333	231	T	223.1		7203	361	A	352.5
7459	105	G	95		7332	232	G	224.2		7202	362	C	353.6
7458	106	T	96.2		7331	233	T	225.4		7201	363	A	354.5
7457	107	T	97.2		7330	234	T	226.6		7200	364	A	355.5
7456	108	G	98.4		7329	234	A	227.5		7199	365	C	356.4
7455	109	C	99.7		7328	236	C	228.3		7198	366	A	357
7454	110	A	100		7327	237	T	229.3		7197	367	A	
7453	111	C	101.2		7326	238	C	230.3		7196	368	A	359.5
7452	112	A	102.5		7325	239	A	231.2		7195	369	C	
7451	113	G	103.8		7324	240	C	232		7194	370	A	
7450	114	C	105		7323	241	T	233		7193	371	C	363.4
7449	115	A	105.7		7322	242	A	234		7192	372	A	364.8
7448	116	A	106.5		7321	243	G	235.1		7191	373	A	
7447	117	A	107.5		7320	244	G	236		7190	374	A	
7446	118	A	108.5		7319	245	G	237.4		7189	375	C	368.4
7445	119	T	109.5		7318	246	C	238.5		7188	376	A	
7444	120	G	110.6		7317	247	G	239.4		7187	377	C	
7443	121	G	112		7316	248	C	240.5		7186	378	A	
7442	122	A	113.1		7315	249	A	241.4		7185	379	G	
7441	123	G	114		7314	250	A	242.3		7184	380	G	
7440	124	G	115.2		7313	251	C	243.3		7183	381	A	
7439	125	A	116.4		7312	252	C	244.1		7182	382	C	
7438	126	T	117		7311	253	A	245.1		7181	383	A	
7437	127	T	118.2		7310	254	C	264.2		7180	384	T	
					7309	255	A	247.2		7179	385	A	
					7308	256	T	248.2					
					7307	257	A	249.4					

FAM

Genome Position	Real Position	Sequence	CE position		Genome Position	Real Position	Sequence	CE position		Genome Position	Real Position	Sequence	CE position
7179	1	T			7308	130	A	126.1		7437	259	A	257.7
7180	2	A			7309	131	T	126.9		7438	260	A	258.7
7181	3	T			7310	132	G	129.7		7439	261	T	259.6
7182	4	G			7311	133	T	129.2		7440	262	C	260.8
7183	5	T			7312	134	G	130.2		7441	263	C	261.7
7184	6	C			7313	135	G	131.4		7442	264	T	262.7
7185	7	C			7314	136	T	132.5		7443	265	C	
7186	8	T			7315	137	T	133,,4		7444	266	C	
7187	9	G			7316	138	G	134.5		7445	267	A	265.5
7188	10	T			7317	139	C	135.6		7446	268	T	266.5
7189	11	G			7318	140	G	136.3		7447	269	T	267.6
7190	12	T			7319	141	C	137.5		7448	270	T	268.8
7191	13	T			7320	142	C	138.2		7449	271	T	269.8
7192	14	T			7321	143	C	139		7450	272	G	270.9
7193	15	G			7322	144	T	149.8		7451	273	C	272.1
7194	16	T			7323	145	A	141.1		7452	274	T	
7195	17	G			7324	146	G	142.1		7453	275	G	274
7196	18	T			7325	147	T	143.3		7454	276	T	275.2
7197	19	T			7326	148	G	144.4		7455	277	G	276.3
7198	20	T			7327	149	A	145.6		7456	278	C	277.3
7199	21	G			7328	150	G	146.5		7457	279	A	277.9
7200	22	T			7329	151	T	147.7		7458	280	A	278.8
7201	23	T			7330	152	A	148.7		7459	281	C	279.6
7202	24	G			7331	153	A	149.6		7460	282	C	280.6
7203	25	T			7332	154	C	150.4		7461	283	G	281.5
7204	26	A			7333	155	A	151.3		7462	284	A	282.7
7205	27	T			7334	156	A	152.3		7463	285	T	283.8
7206	28	G			7335	157	C	153.2		7464	286	T	285
7207	29	A			7336	158	T	154.1		7465	287	T	286
7208	30	T			7337	159	G	155.4		7466	288	C	287
7209	31	T			7338	160	T	156.6		7467	289	G	287.9
7210	32	G			7339	161	A	157.6		7468	290	G	288.8
7211	33	C			7340	162	T	158.6		7469	291	T	290.2
7212	34	A			7341	163	T	159.6		7470	292	T	291.1
7213	35	T			7342	164	T	160.8		7471	293	G	292.1
7214	36	T			7343	165	G	161.8		7472	294	C	293.3
7215	37	G			7344	166	T	163.1		7473	295	C	294.5
7216	38	T			7345	167	G	164.4		7474	296	T	295.7
7217	39	A			7346	168	T	165.3		7475	297	T	296.8
7218	40	T			7347	169	T	166.3		7476	298	T	
7219	41	G			7348	170	T	167.3		7477	299	G	298
7220	42	G			7349	171	G	168.3		7478	300	G	299
7221	43	T			7350	172	T	169.4		7479	301	C	300.3
7222	44	A			7351	173	G	170.4		7480	302	T	301
7223	45	T			7352	174	G	171.5		7481	303	T	301.7
7224	46	G			7353	175	T	172.7		7482	304	A	302.9
7225	47	T			7354	176	A	173.7		7483	305	T	303.8
7226	48	A			7355	177	T	174.5		7484	306	G	304.9
7227	49	T			7356	178	G	175.7		7485	307	T	305
7228	50	G			7357	179	G	176.7		7486	308	C	306
7229	51	G			7358	180	G	177.8		7487	309	T	307.1
7230	52	T			7359	181	T	178.9		7488	310	G	308.9
7231	53	T			7360	182	G	179.9		7489	311	T	310.1
7232	54	G			7361	183	T	181.1		7490	312	G	311.2
7233	55	T	43.4		7362	184	T	182		7491	313	G	312.2
7234	56	T	44.6		7363	185	G	183		7492	314	T	313.5
7235	57	G	45.9		7364	186	C	184.2		7493	315	T	314.5
7236	58	T	47.2		7365	187	T	184.9		7494	316	T	
7237	59	T	48.4		7366	188	T	186		7495	317	T	
7238	60	G			7367	189	G	187		7496	318	C	
7239	61	T			7368	190	T	188.2		7497	319	T	
7240	62	A			7369	191	T	189.2		7498	320	G	319.3
7241	63	T	54		7370	192	G	190.3		7499	321	C	320.6
7242	64	G	55.1		7371	193	G	191.4		7500	322	A	321.4
7243	65	T	55.8		7372	194	G	192.6		7501	323	C	322.4
7244	66	T	57		7373	195	C	193.7		7502	324	A	323.3
7245	67	G	58.6		7374	196	T	194.4		7503	325	A	324.1
7246	68	T	59.8		7375	197	A	195.6		7504	326	T	325.2
7247	69	A	60.7		7376	198	T	196.4		7505	327	A	326.3
7248	70	T	61.5		7377	199	A	197.5		7506	328	C	327.2
7249	71	G	62.6		7378	200	T	198.4		7507	329	A	328.2
7250	72	T	64.1		7379	201	A	199.4		7508	330	G	329.1
7251	73	T	65.2		7380	202	T	200.4		7509	331	T	330.4

7251	73	T	65.2		7380	202	T	200.4		7509	331	T	330.4
7252	74	A	66.1		7381	203	T	201.4		7510	332	A	331.4
7253	75	C	67.1		7382	204	G	202.5		7511	333	C	332.4
7254	76	T	67.9		7383	205	T	203.7		7512	334	G	333.3
7255	77	A	69		7384	206	C	204.7		7513	335	C	334.7
7256	78	T	70.1		7385	207	C	205.5		7514	336	T	
7257	79	A	71.2		7386	208	T	206.4		7515	337	G	336.4
7258	80	T	72.4		7387	209	G	207.5		7516	338	G	337.5
7259	81	T	73.6		7388	210	T	208.7		7517	339	C	338.6
7260	82	T	74.7		7389	211	A	209.8		7518	340	A	339.6
7261	83	G	75.7		7390	212	T	210.7		7519	341	C	340.5
7262	84	T	77.3		7391	213	T	211.8		7520	342	T	
7263	85	T	78.3		7392	214	T	212.7		7521	343	A	342.4
7264	86	G	79.4		7393	215	C	213.9		7522	344	T	343.4
7265	87	G	80.8		7394	216	A	215.7		7523	345	T	344.5
7266	88	T	82.1		7395	217	A			7524	346	G	345.5
7267	89	A	83.3		7396	218	G			7525	347	C	346.7
7268	90	T	84		7397	219	T	217.7		7526	348	A	
7269	91	G	85.1		7398	220	T	218.5		7527	349	A	
7270	92	T	86.3		7399	221	A	219.9		7528	350	A	349.4
7271	93	G	87.4		7400	222	T	220.7		7529	351	C	350.6
7272	94	G	88.6		7401	223	A	221.8		7530	352	T	
7273	95	C	89.7		7402	224	A	222.7		7531	353	T	
7274	96	A	90.5		7403	225	A	223.6		7532	354	T	
7275	97	T	91.3		7404	226	A	224.6		7533	355	A	355
7276	98	T	92.3		7405	227	C	225.4		7534	356	A	
7277	99	A	93.3		7406	228	T	226.3		7535	357	T	
7278	100	A	94.3		7407	229	G	227.5		7536	358	C	
7279	101	A	95.3		7408	230	C	228.8		7537	359	T	
7280	102	T	96.1		7409	231	A	229.6		7538	360	T	
7281	103	A	97.3		7410	232	C	230.5		7539	361	T	
7282	104	A	98.3		7411	233	A	231.5		7540	362	T	
7283	105	A	99.3		7412	234	C	232.5		7541	363	G	
7284	106	A	100.4		7413	235	C	233.3		7542	364	G	
7285	107	T	101.9		7414	236	T	234.2		7543	365	G	
7286	108	A	103.4		7415	237	T	235.3		7544	366	C	
7287	109	T	104.6		7416	238	A	236.6		7545	367	A	
7288	110	G	105.3		7417	239	C	237.5		7546	368	C	
7289	111	T	106.6		7418	240	A	238.4		7547	369	T	
7290	112	T	107.5		7419	241	G	239		7548	370	G	
7291	113	T	108.4		7420	242	C	240.6		7549	371	C	
7292	114	T	109.3		7421	243	A	241		7550	372	T	
7293	115	G	110.2		7422	244	T	242.5		7551	373	C	
7294	116	T	111.6		7423	245	C	243.7		7552	374	C	
7295	117	G	112.5		7424	246	C	244.5		7553	375	T	
7296	118	G	113.6		7425	247	A	245.5		7554	376	A	
7297	119	T	114.8		7426	248	T	246.5		7555	377	C	
7298	120	T	115.8		7427	249	T	247.7		7556	378	A	
7299	121	C	116.7		7428	250	T	248.8		7557	379	T	
7300	122	T	117.5		7429	251	T	249.9		7558	380	A	
7301	123	G	118.5		7430	252	A	251		7559	381	T	
7302	124	T	119.8		7431	253	T	251.9		7560	382		
7303	125	G	120.9		7432	254	C	252.9		7561	383		
7304	126	T	122		7433	255	C	253.8		7562	384		
7305	127	G	123		7434	256	T	254.6		7563	385		
7306	128	T	124.2		7435	257	A	255.8					
7307	129	T	125.1		7436	258	C	256.7					

Table SI 3A. *Binding affinity data of NV1028 on HPV 7179-7563 bp.*

# Mismatch (MM)	NV1028 WWGWWWWWWW (FW) WWCWWWWWWW (RV) DNA Sequence (5'-3')	FORWARD (F) REVERSE (R)	Site	Genome Position	Hill Coefficient	Kds (Hill Model)	Integration Peak (for Kd)
0 MM	TATGT TACTATATTT GTTGG	R	3	7251-7260	1.08	0.64	7252
	ATACC AACAAATATA GTAAC	R	5	7263-7254	1.48	0.96	7261
	CACAA AACATATTTT ATTTA	R	8	7290-7281	2.03	0.79	7288
	ATTTC AAGTTATAAA ACTGC	F	16	7394-7403	1.89	3.05	7401
1 MM	GTTTG TTGTATGATT GCATT	F	0	7200-7209		/	/
	GTATG TTGTATGTTA CTATA	F	2	7243-7252	1.51	1.32	7245
	TATGT TACTATATTT GTTGG	F	3	7251-7260	1.08	0.64	7252
	AAATA TAGTAACATA TACAA	F	4	7255-7246	1.08	0.64	7252
	ATACC AACAAATATA GTAAC	F	5	7263-7254	1.48	0.96	7261
	GTGGC ATTAAATAAA ATATG	F	6	7274-7283	1.62	1.15	7281
	TGGCA TTAAATAAAA TATGT	F	6	7275-7284	1.62	1.15	7281
	GGCAT TAAATAAAAT ATGTT	F	6	7276-7285	1.62	1.15	7281
	GCATT AAATAAAATA TGTTT	F	6	7277-7286	1.62	1.15	7281
	CATTA AATAAAATAT GTTTT	F	6	7278-7287	1.62	1.15	7281
	CATAT TTTATTTAAT GCCAC	F	7	7283-7274	1.62	1.15	7281
	ACATA TTTTATTTAA TGCCA	F	7	7284-7275	1.62	1.15	7281
	AACAT ATTTTATTTA ATGCC	F	7	7285-7276	1.62	1.15	7281
	AAACA TATTTTATTT AATGC	F	7	7286-7277	1.62	1.15	7281
	AAAAC ATATTTTATT TAATG	F	7	7287-7278	1.62	1.15	7281
	CACAA AACATATTTT ATTTA	F	8	7290-7281	2.03	0.79	7288
	GTAAC AACTGTATTT GTGTT	R	11	7333-7342	2.05	1.66	7335
	CCACA AACACAAATA CAGTT	R	12	7347-7338	3.04	1.24	7343
	TACAG GACAATATAT AGCCC	R	15	7384-7375		2.14	7380
	ATAAC TTGAAATACA GGACA	F	16	7395-7386	1.89	3.05	7401
	GTGTG CAGTTTTATA ACTTG	F	17	7407-7396	1.89	0.87	7396
	ACAGC ATCCATTTTA TCCTA	R	18	7421-7430	2.20	1.37	7281
	CAGCA TCCATTTTAT CCTAC	R	18	7422-7431	2.20	1.37	7281
	ATTGT AGGATAAAAT GGATG	F	19	7434-7325	2.20	1.37	7281
	GATTG TAGGATAAAA TGGAT	F	19	7435-7326	2.20	1.37	7281
2 MM	ATGGT ATGTATGGTT GTTGT	F	1	7222-7231		/	/
	GTTAC TATATTTGTT GGTAT	F	4	7254-7263	1.48	0.96	7261
	CACAG AACCACAAAA CATAT	R	9	7298-7289	1.88	1.69	7293
	CGCCC TAGTGAGTAA CAACT	F	10	7322-7331	2.42	3.52	7324
	GTAAC AACTGTATTT GTGTT	F	11	7333-7342	2.05	1.66	7335
	CCACA AACACAAATA CAGTT	F	12	7347-7338	3.04	1.24	7343
	TGTTG GGCTATATAT TGTCC	R	14	7371-7380	3.24	2.14	7380
	TGGGC TATATATTGT CCTGT	F	14	7374-7383	3.24	2.14	7380
	ACAGG ACAATATATA GCCCA	F	15	7383-7374	3.24	2.14	7380
	TACAG GACAATATAT AGCCC	F	15	7384-7375	3.24	2.14	7380
	ACAGC ATCCATTTTA TCCTA	F	18	7421-7430	2.20	1.37	7281
	CAGCA TCCATTTTAT CCTAC	F	18	7422-7431	2.20	1.37	7281
	GAGGA TTGTAGGATA AAATG	F	19	7427-7438	2.20	1.37	7281
3 MM	CATAC AACATACAAC AACAA	F	2	7244-7235	1.88	1.36	7238
	CACAG AACCACAAAA CATAT	F	9	7298-7289	1.88	1.69	7293
	TGGTT CTGTGTGTTA TGTGG	F	9	7299-7308	3.47	2.38	7305
	GTTCT GTGTGTTATG TGGTT	F	9	7301-7310	3.47	2.38	7305
	GGGTG TTGCTTGTTG GGCTA	F	13	7361-7370	3.64	4.37	7364
	TGTTG GGCTATATAT TGTCC	F	14	7371-7380	3.24	2.14	7380
	GTTGG GCTATATATT GTCCT	F	14	7372-7381	3.24	2.14	7380
	TTGGG CTATATATTG TCCTG	F	14	7373-7382	3.24	2.14	7380
	GTAGG ATAAAATGGA TGCTG	R	19	7431-7422	2.20	1.37	7281
	AAGGC AACCGAAATC GGTTG	R	21	7470-7461	3.74	4.86	7466
4 MM	GGTTC TGTGTGTTAT GTGGT	F	9	7300-7309	3.47	2.38	7305

Red nucleotide indicate mismatch position

Table SI 3B. *Binding affinity data of NV1042 on HPV 7179-7563 bp.*

# Mismatch (MM)	NV1042 WWGWWWWWGWWWW (FW) WWCWWWWWCWWWW (RV) DNA Sequence	FORWARD (F) REVERSE (R)	Site	Genome Position	Hill Coefficient Avg.	Kd (Hill model) ε=147400 M x cm^{-1}	Integration Peak (for Kd)
1 MM	TATAA TAAAATATGTTTT GTGGT	F	3/4	7280-7292	1.74	0.54	7281
	CACAA AACATATTTTATT TAATG	R	5	7290-7278	1.74	0.54	7281
2 MM	TATGG TTGTTGTTGTATG TTGTA	F	0	7230-7242	2.47	1.18	7245
	GTTGT ATGTTACTATATT TGTTG	F	1	7247-7259	1.72	0.77	0
	TATGT TACTATATTTGTT GGTAT	R	1	7251-7263	1.72	0.77	0
	ATACC AACAAATATAGTA ACATA	R	2	7263-7251	1.78	0.54	7261
	GTGGC ATATAATAAAATA TGTTT	F	3	7274-7286	1.74	0.54	7281
	TGGCA TATAATAAAATAT GTTTT	F	3	7275-7287	1.74	0.54	7281
	TTTGT ATAAAATAATATA CGGTG	F	4	7286-7274	1.74	0.54	7281
	TTTTG TATAAAATAATAT ACGGT	F	4	7287-7275	1.74	0.54	7281
	CACAA AACATATTTTATT TAATG	F	5	7290-7278	1.74	0.54	7281
	CATAC CACAAACACAAAT ACAGT	R	9	7351-7339	2.34	0.71	7343
	TCCTG TATTTCAAGTTAT AAAAC	F	11	7388-7400	3.56	0.71	7396
	TGTAT TTCAAGTTATAAA ACTGC	R	11	7391-7403	3.56	0.71	7396
	AGTTT TATAACTTGAAAT ACAGG	F	12	7401-7389	3.56	0.71	7396
3 MM	TATGT TACTATATTTGTT GGTAT	F	1	7251-7263	1.72	0.77	0
	ATACC AACAAATATAGTA ACATA	F	2	7263-7251	1.78	0.54	7261
	TGTGG CATATAATAAAAT ATGTT	F	3	7273-7285	1.74	0.54	7281
	AACAT ATTTTATTTAATG CCACA	F	4	7285-7273	1.74	0.54	7281
	CCACA AAACATATTTTAT TTAAT	F	5	7291-7279	1.74	0.54	7281
	ACCAC AAAACATATTTTA TTTAA	F	5	7292-7280	1.74	0.54	7281
	GTGAG TAACAACTGTATT TGTGT	F	8	7329-7341	2.04	1.36	7331
	TGGGC TATATATTGTCCT GTATT	F	10	7374-7386	2.43	1.09	7379
	TGTAT TTCAAGTTATAAA ACTGC	F	11	7391-7403	3.56	0.71	7396
	GTGTG CAGTTTTATAACT TGAAA	F	13	7407-7395	3.56	0.71	7396
	AGGTG TGCAGTTTTATAA CTTGA	R	13	7409-7397	3.56	0.71	7396
	GAGGA TTGTAGGATAAAA TGGAT	F	16	7438-7336	2.94	1.13	7436
	GGCTT ATGTCTGTGGTTT TCTGC	F	20	7482-7494	/	/	/
	AGAAA ACCACAGACATAA GCCAA	R	21	7492-7480	/	/	/
4 MM	TATAT ATTGTCCTGTATT TCAAG	F	10	7379-7391	2.43	1.09	7379
	AGGTG TGCAGTTTTATAA CTTGA	F	13	7409-7397	1.74	0.54	7281
	CAGCA TCCATTTTATCCT ACAAT	R	14	7422-7434	2.26	0.74	7428
	GTAGG ATAAAATGGATGC TGTAA	R	15	7431-7419	2.26	0.74	7428
	ATCGG TTGCACAGCAAAA TGGAG	R	18	7458-7446	/	/	/
	CGCCC TAGTGAGTAACAA CTGTA	F	7	7322-7335	2.39	1.65	7326
5 MM	CACAT AACACACAGAACC ACAAA	F	6	7307-7295	3.13	1.77	7308

Red nucleotide indicate mismatch position

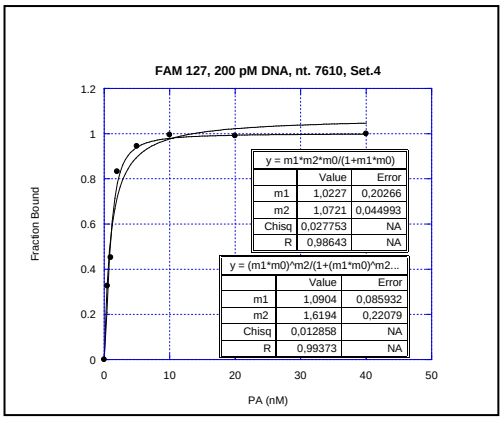
Table SI 3C. *Binding affinity data of NV1028 on HPV 7479-7783 bp.*

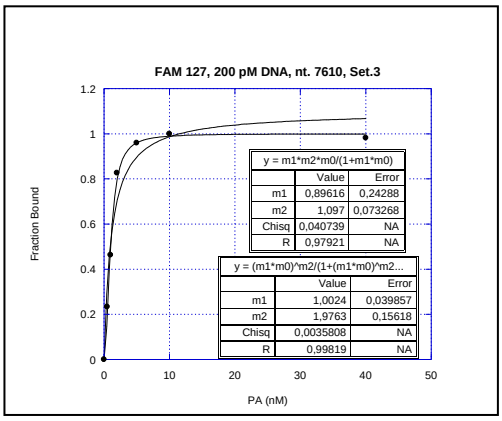
# Mismatch (MM)	NV1028 WWGWWWWWWW (FW) WWCWWWWWWW (RV) DNA Sequence (5'-3')	FORWARD (F) REVERSE (R)	Site	Genome Position	Hill Coefficient	Kds (Hill Model) ε=113300 M^{-1} cm^{-1}	Integration Peak (for Kd)
0 MM	GCTCC TACATATTTT GAACA	R	2	7553-7562	2.41	2.87	7557
	AATTG TTCAAAATAT GTAGG	R	3	7565-7557	1.84	0.92	7563
	TCCCT AAGTAATAAA ACTGC	F	9	7669-7678	2.11	1.08	7672
	AAACA AACTAAAATA TGTGC	R	12	7704-7695	1.73	0.79	7700
1 MM	CCCAA AAGATTAAAG TTTGC	F	1	7538-7529	2.07	1.23	7536
	GCTCC TACATATTTT GAACA	F	2	7553-7562	2.41	2.87	7557
	AATTG TTCAAAATAT GTAGG	F	3	7565-7557	1.84	0.92	7563
	GCCAA TTGTTCAAAA TATGT	F	3	7568-7559	1.74	1.74	7568
	GGTAT TAGTCATTTT CCTGT	F	5	7609-7618	1.59	0.92	7610
	GGAAA ATGACTAATA CCAGG	F	6	7615-7606	1.40	0.78	7613
	AAAAG CAGTTTTATT ACTTA	F	10	7682-7673	2.27	1.49	7680
	AAACA AACTAAAATA TGTGC	F	12	7704-7695	1.73	0.79	7700
	TTAGT TTGTTTTTAC TTAAG	F	12	7704-7713	CHC	CHC	CHC
	AGCTT AAGTAAAAAC AAACT	F	13	7715-7706	CHC	CHC	CHC
	ACGCG ATGTTGTTAA CGAAC	R	7	7634-7645	1.40	1.40	7638
1 MM don't bind	AACGT TTGAAATTAG AAAAC	R	0	7527-7536	/	/	/
	ATATT TTGAACAATT GGCGC	F	2-3	7561-7570	2.41	2.87	7557
	TATTT TAGTTTGTTT TTACT	F	11	7700-7709			
	GTAAA AACAAACTAA AATAT	R	12	7708-7697			
2 MM	ATTGC AAACTTTAAT CTTTT	F	0	7526-7535	/	/	/
	TTGCA AACTTTAATC TTTTG	F	0	7527-7536	/	/	/
	AAACT TTAATCTTTT GGGCA	F	0a	7531-7540	2.07	1.23	7536
	GCCCA AAAGATTAAA GTTTG	F	1	7540-7531	2.07	1.23	7536
	CCTAC ATATTTTGAA CAATT	F	2	7556-7565	2.41	2.87	7557
	TGCGC TACAACAATT GCTTG	F	7	7634-7645	1.40	1.40	7638
	ATTGC TTGCATAACT ATATC	F	8	7646-7655	1.74	1.37	7647.7651
	TATGC AAGCAATTGT TGTAG	F	8	7647-7638	1.89	1.96	7644.7647
	ATGCA AGCAATTGTT GTAGC	R	8	7646-7637	1.89	1.96	7644
	TTGCT TGCATAACTA TATCC	R	8	7647-7656	1.56	1.35	7651
	GGATA TAGTTATGCA AGCAA	F	8a	7656-7647	1.54	1.35	7659
	GGCAC ATATTTTAGT TTGTT	F	11	7694-7703	1.73	0.79	7700
	GCACA TATTTTAGTT TGTTT	F	11	7695-7704	1.73	0.79	7700
	CACAT ATTTTAGTTT GTTTG	F	11	7696-7705	1.73	0.79	7700
	GCTAA TTGCATACTT GGCTT	F	14	7723-7732	1.73	0.79	7700
3 MM	GAGCA GTGCCCAAAA GATTA	R	1a	7546-7537	1.77	2.28	7545
	CTTTG GCGCATATAA GGCGC	F	3a	7585-7594	3.77	2.74	7591
	AAACC GCGTATATTC CGCGT	R	3a	7586-7595	3.77	2.74	7591
	GGTGC GCCTTATATG CGCCA	R	4	7597-7588	3.77	2.74	7591
	CTGGA CAGGAAAATG ACTAA	F	6a	7622-7613			
	TTAGG GAGTGGATAT AGTTA	F	8a	7665-7656	1.54	1.35	7659
4 MM	GAGCA GTGCCCAAAA GATTA	F	1a	7546-7537	1.77	2.28	7545

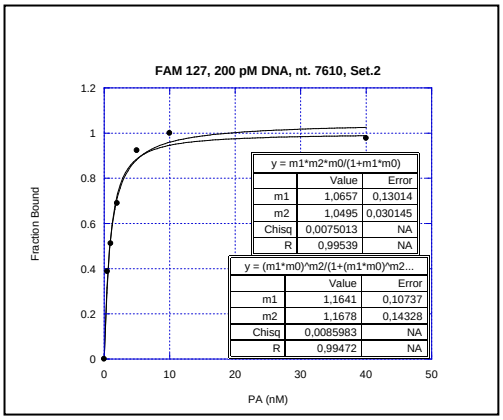
Red nucleotide indicate mismatch position

Figure SI 1. *Representative binding isotherms for NV1028 on HPV 7479-7783*

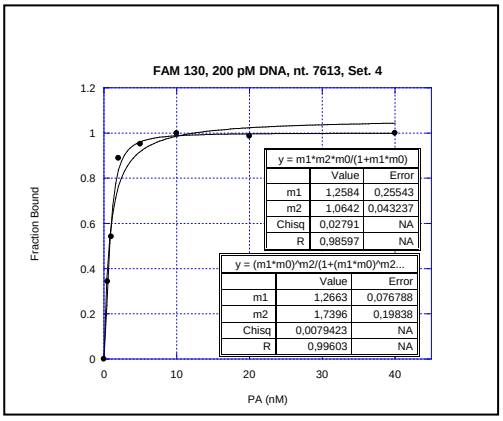
Site 7610

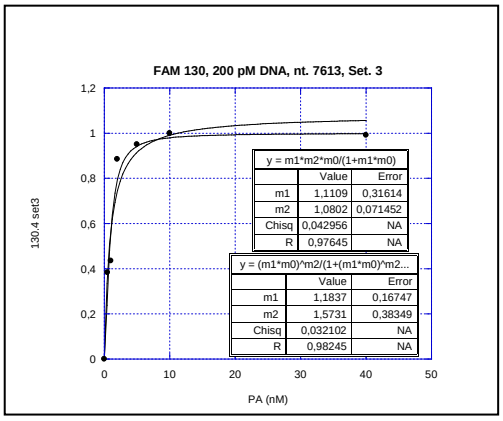


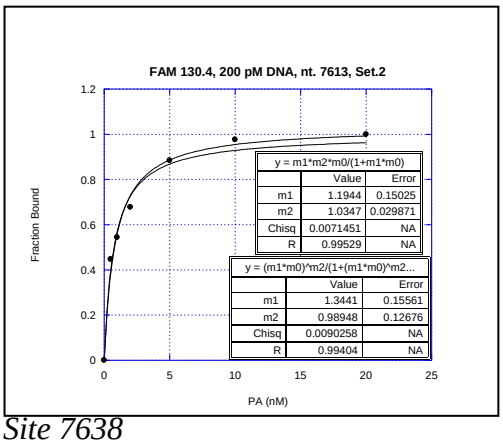




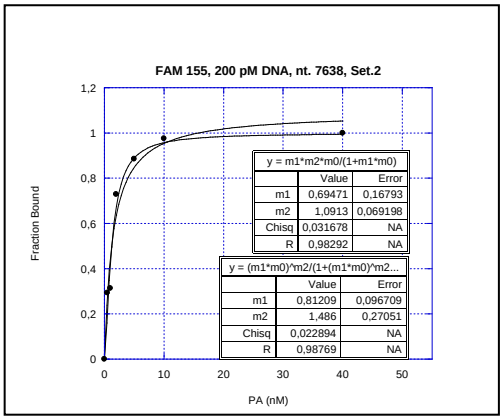
Site 7613

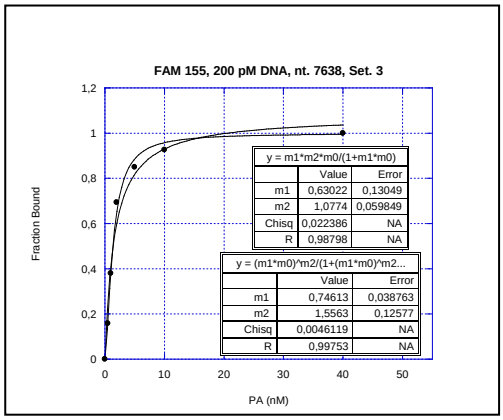






Site 7638





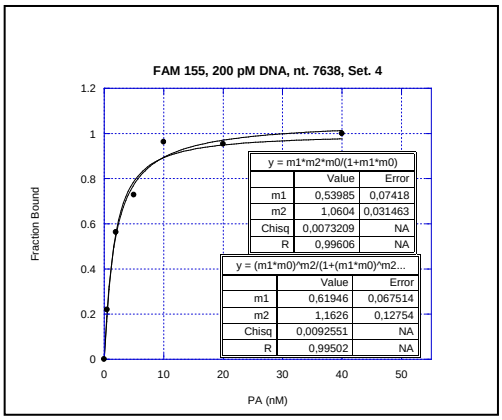
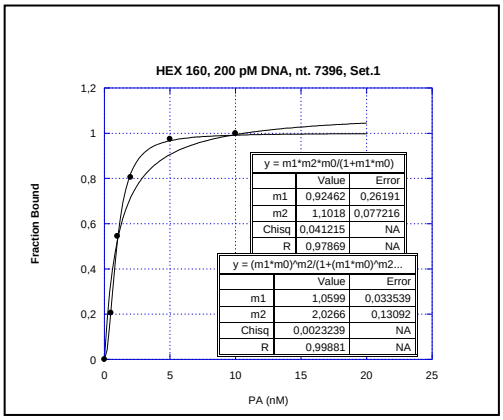
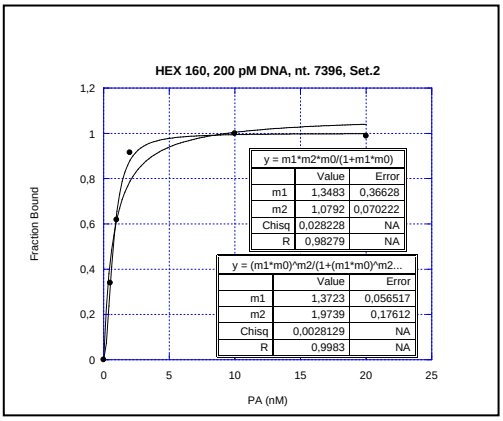
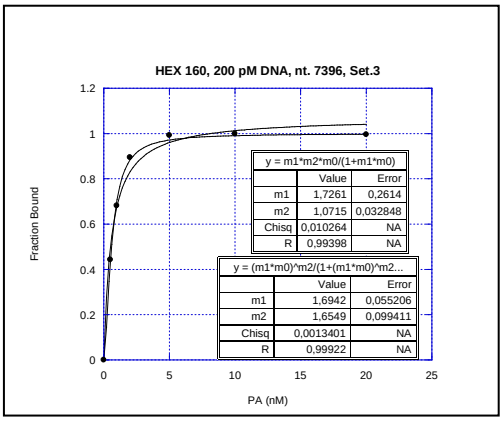


Figure SI 2. *Representative binding isotherms for NV1028 on HPV 7179-7563*

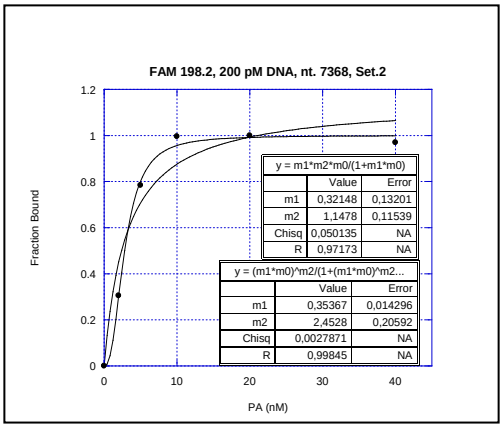
Site 7396

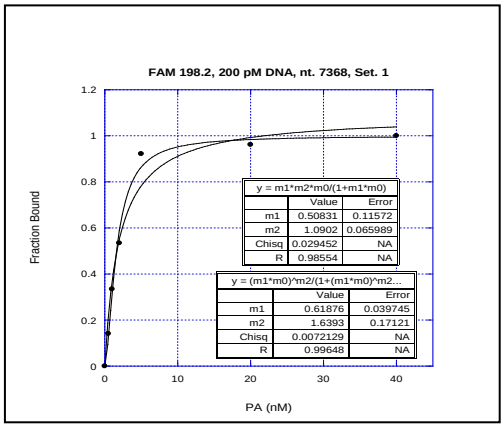






Site 7368





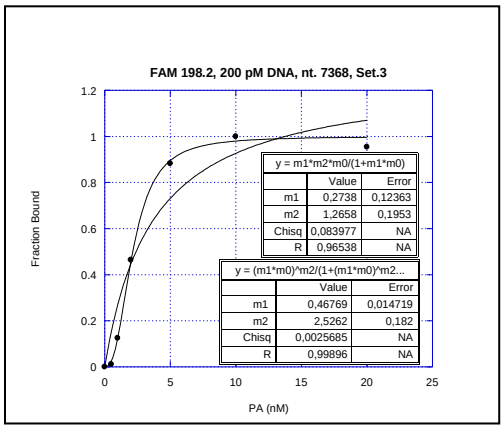
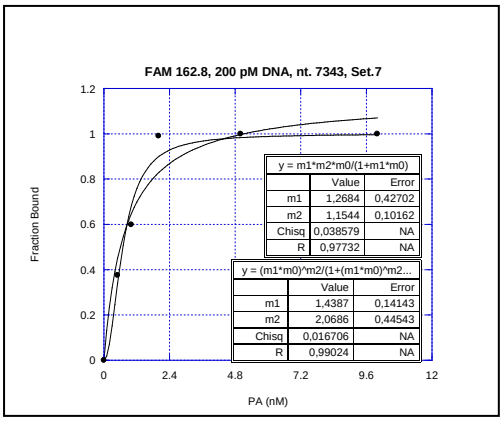
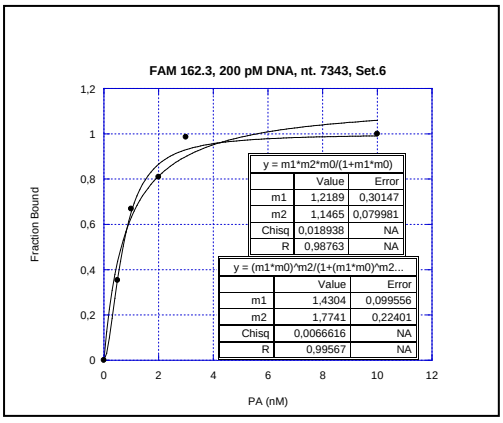
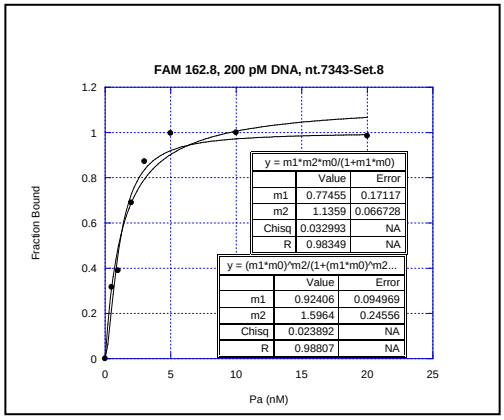
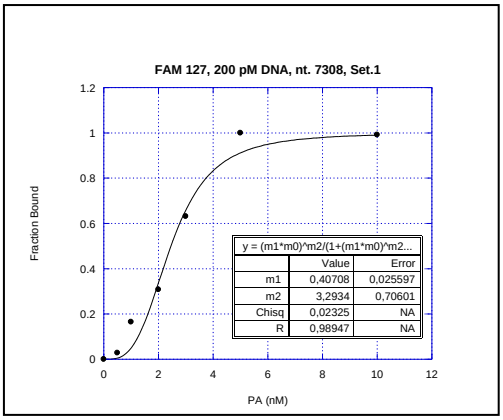


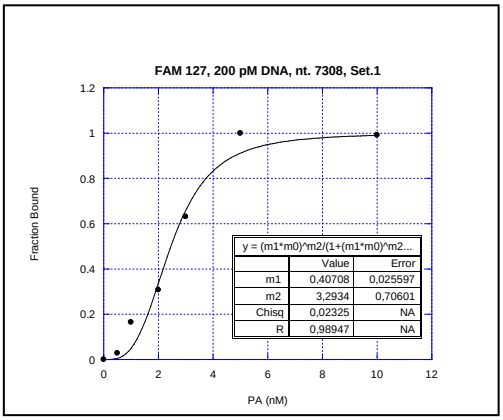
Figure SI 3. *Representative binding isotherms for NV1042 on HPV 7179-7563 Site 7343*

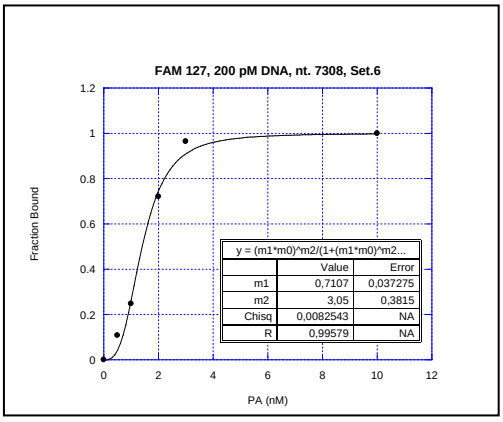












CHAPTER 3

IMPROVED ANTIVIRAL POLYAMIDES WITH BROAD-SPECTRUM ACTIVITY AGAINST HIGH-RISK HUMAN PAPILLOMAVIRUS

*The work presented in this chapter is based on the published paper "**Improved Antiviral Polyamides with Broad-Spectrum Activity Against High-Risk Human Papillomavirus**" from MedChemComm, 2016,* **7**, 2076-2082

3.1 Abstract

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

3.2 Introduction

N-methylpyrrole (Py) and N-methylimidazole (Im) hairpin polyamides (PAs) are synthetic homologs of the natural antibiotics distamycin A and netropsin which bind AT-rich regions of DNA in the minor groove.[1] The antiparallel nature of hairpin PAs allows for the side-by-side pairing of building blocks and recognition of DNA base pairs (bp) within a double strand (dsDNA). It has been extensively reported that Py/Py and Py/β-alanine (β) pairs recognize A/T and T/A base pairs, while an Im/Py pair distinguishes G/C from C/G.[2] Furthermore, introduction of the flexible β linker in place of heterocycles resets the heterocycle-base pair register and allows targeting of longer DNA sequences.[3]

In numerous reports, hairpin PAs distinguish DNA sequences by binding to patterns of hydrogen-bond donors and acceptors in the minor groove of DNA,[4] so synthetic PAs are attractive for many biomedical applications. These applications, for polyamides themselves and related minor groove binders, include the regulation of gene expression,[5] biological imaging,[6] and antimicrobial/antiviral applications.[7] Antiviral polyamides include the long polyamides reported here which, as a class, access a new biological mechanism: they actively degrade the genome of dsDNA tumour viruses in cells while avoiding acute cytotoxicity. The DNA degradation is accomplished by inducing the DNA Damage Response (DDR) significantly, only in infected cells carrying small, circular, dsDNA genomes known as episomes.[8a,b] As

discussed below, another characteristic of this new class of antiviral polyamides is their failure to follow the reported DNA recognition rules described above.

The process by which the DDR works in concert with the polyamides is not completely established at the molecular level, but many details have emerged as a result of a variety of experiments, a portion of which are summarized here. For example, 224 DDR genes were knocked down individually using siRNA, and we checked to see if the knockdowns affected polyamide activity in any way; 21 genes were found to affect PA activity, and the results were validated by additional experiments using QPCR (rather than relying on the PCR array data alone).[8a,b] Some gene knockdowns enhanced PA activity and others opposed PA activity. Additional experiments used Mirin, a small molecule inhibitor of DDR enzyme Mre11. Mre11 has many activities alone and in protein complexes. For example, Mre11 is a member of the MRN complex, which participates in double-strand DNA break repair and telomere maintenance as a member of the ATM branch of the DDR.[8c] As a standalone enzyme, one activity associated with MRE11 is that of an endonuclease. MRE11 inhibition experiments were carried out in the presence and absence of polyamides.[8a] 100 μM Mirin had no effect on its own, but it sensitized cells to PA25, making the compound 4× more active in terms of lower IC_{50} and IC_{90} values. Part of the activation of the DDR was also followed by phosphorylation of Rad9, a protein that is phosphorylated upon DNA damage as part of the ATR pathway of the DDR, one of the two major DDR branches (the other being ATM).[8a,b,d] Treatment of HPV16 episome bearing W12E cells with active polyamide causes Rad9 phosphorylation within 2–4 hours, but HPV-free

cervical cancer cell line C33A shows no phosphorylation of Rad9 upon treatment with polyamide.[8a,b]

We previously reported a library of long hairpin PAs, including **PA1** and **PA25**, which exhibits antiviral activity against three high-risk (oncogenic) HPVs.[9] HPV is an AT-rich dsDNA virus[10] that causes cervical cancer, among other diseases, and subtypes HPV16 and 18 are responsible for 70% of cervical cancer in most of the world.[11,12] Although specific HPV vaccines are available,[13] they are effective prophylactics, not therapies. **PA1**, our first preclinical lead, exhibits anti-HPV activity with no cytotoxicity up to its solubility limit in cell culture media of 200 μM.[9] A characteristic of these active anti-HPV polyamides is that they bind at least ten bp, or one full turn of DNA.

Here, we report the antiviral activities and DNA-binding properties of two PAs of a new structural class, **PA30**, and **PA31**, where the des-amino N-terminus of the parent compound **PA1** has been substituted with either tetramethylguanidinium (TMG) or guanidinium (Guan), respectively. **PA30** and **PA31** are biomimetic in that netropsin has an N-terminal guanidinium, though **PA30**, **31** are not exact structural analogs of netropsin. While others have explored guanidinylated DNA-binding ligands, no biophysical studies on guanidinylated hairpin PAs have been undertaken, to our knowledge.[14] We found that these substitutions at the N-terminus, which are small relative to the overall molecular size, can significantly improve anti-HPV activity. However, improvement in antiviral efficacy cannot be attributed to altered DNA binding strength, as will be shown. Nevertheless, **PA30** and **PA31** display noticeable

differences in binding distribution on a fragment of native HPV18 DNA in comparison with parent **PA1**.

3.3 Methods

3.3.1 Polyamides Synthesis and Characterization

PA1 and its EDTA conjugate were synthesized and characterized as previously described.[9, 18] **PA30** and **31** were synthesized by virtually identical solid-phase methods up to the Im group (4-amino-N-methyl imidazole-2-carboxamide) and its 4-N-attached guanidinyl substituent. For **PA30**, HATU[28] was utilized as the final building block to incorporate the TMG group onto the H_2N-Im moiety. For **PA31**, in parallel to the literature,[29] treatment with N,N'-Bis (tert-butoxycarbonyl)-1H-pyrazole-1-carboxamidine and diisopropylethylamine for 24 h, followed by TFA treatment to remove the BOC groups produced the desired, unsubstituted Guan group. The EDTA conjugate of **PA30** was prepared via the same method as previously described for **PA1**'s conjugate.[18] Briefly, to a vigorously-stirred mixture of EDTA dianhydride (15.8 mg, 0.060 mmol, 22 eq) in N,N-diisopropylethylamine (DIEA) (0.5 mL), DMF (0.25 mL) and DMSO (0.25 mL) at 55 °C was added a mixture of **PA30** (5.9 mg, 0.0027 mmol, 1 eq) in DIEA (0.5 mL) and DMF (0.5 mL) dropwise over a period of 30 min. The reaction mixture was stirred at 55 °C for 30 min. Following addition of 0.1 N NaOH (1.03 mL), the mixture was stirred at 55 °C for an additional 20 min. The bottom layer of the biphasic mixture was removed from the reaction vessel and neutralized with formic acid (0.013 mL). The mixture was diluted with 0.5 mL DMSO, filtered through a 20 μm polyethylene filter, and purified by reversed-

phase HPLC using a Phenomenex Luna 250 x 30 mm, 5 μm, 100 Å, C_{18} (version 2) column maintained at 25 ˚C. Mobile phases were 0.2% formic acid in water (A) and 100 % methanol (B). The gradient was 35% B for 5 min followed by a ramp to 70% B over 40 min at 50 mL/min. Concentration of pooled fractions, followed by lyophilization, gave the **PA30-EDTA** conjugate (1.0 mg, 16% yield) as a white, fluffy solid: exact mass $[M+H]^+$= 2282.0657, experimental (ESI) $[M+H]^+$ = 2282.1257. The EDTA conjugate of **PA31** was prepared *via* a virtually identical procedure to the one described for **PA30**, but with the following changes. **PA31** (4.0 mg) was substituted for **PA30**. The preparative HPLC purification parameters were identical to those described above except mobile phase A consisted of 0.2% trifluoroacetic acid in water. Concentration of pooled fractions, followed by lyophilization gave the **PA31-EDTA** conjugate (2.3 mg, 55% yield) as an off-white fluffy solid: exact mass $[M+2H+Na]^+$ = 2250.0007, experimental (ESI) $[M+2H+Na]^+$ = 2249.9562. Analytical HPLC characterization was performed using a Phenomenex Jupiter Proteo (C12), 4.6 x 50 mm, 4 μm, 90 Å maintained at 40˚C. Mobile phases consisted of 0.1% formic acid in water (A) and 100 % acetonitrile (B). The gradient consisted of 5% B for 0.75 min followed by a ramp to 60% B over 3.25 min at 2 mL/min. Retention times were 2.641 min, 2.486 min and 2.437 min for **PA1**, **PA30** and **PA31** respectively.

PA1 Im-Py-Py-β-Py-Py-Py-γ-Py-Py-β-Py-Py-Py-Py-β-Ta (3 TFA):

^{1}H NMR (600 MHz, DMSO-d_6) δ = 10.46 (s, 1 H), 9.93 (s, 1 H), 9.91 (s, 4 H), 9.90 - 9.87 (m, 2 H), 9.86 (s, 1 H), 9.83 (s, 1 H), 9.63 (br. s., 2 H), 8.12 - 8.02 (m, 5 H), 7.98 - 7.92 (m, 1 H), 7.88 (br. s., 3 H), 7.40 (s, 1 H), 7.28 (d, *J* = 1.8 Hz, 1 H), 7.24 - 7.21 (m, 3 H), 7.21 - 7.18 (m, 4 H), 7.18 - 7.16 (m, 3 H), 7.15 (d, *J* = 1.8 Hz, 1 H), 7.08 (d,

J = 1.8 Hz, 1 H), 7.07 (d, *J* = 1.8 Hz, 1 H), 7.06 (s, 1 H), 7.04 (d, *J* = 1.8 Hz, 1 H), 6.90 (d, *J* = 1.8 Hz, 1 H), 6.89 (d, *J* = 1.8 Hz, 1 H), 6.89 (s, 2 H), 6.85 (s, 2 H), 6.84 (d, *J* = 1.2 Hz, 1 H), 3.99 (s, 3 H), 3.85 (s, 3 H), 3.85 (s, 3 H), 3.84 (s, 9 H), 3.82 (s, 6 H), 3.81 (s, 3 H), 3.81 (s, 9 H), 3.49 - 3.42 (m, 4 H), 3.41 - 3.35 (m, 4 H), 3.25 - 3.16 (m, 4 H), 3.16 - 3.10 (m, 2 H), 3.10 - 2.98 (m, 4 H), 2.87 (br. s, 2 H), 2.74 (d, *J* = 4.7 Hz, 3 H), 2.55 - 2.51 (m, 2 H), 2.36 (t, *J* = 7.0 Hz, 2 H), 2.28 (t, *J* = 7.3 Hz, 2 H), 1.96 - 1.87 (m, 2 H), 1.82 - 1.74 (m, 2 H)

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

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

HPLC purity: 98%. (See **Figure 1a**).

PA30 TMG-Im-Py-Py-β-Py-Py-Py-γ-Py-Py-β-Py-Py-Py-Py-β-Ta (4 TFA):

^{1}H NMR (600 MHz, DMSO-d_6) δ = 10.19 (s, 1 H), 10.08 (s, 1 H), 9.94 (s, 1 H), 9.92 (s, 3 H), 9.91 (s, 1 H), 9.90 (s, 1 H), 9.87 (s, 1 H), 9.84 (s, 1 H), 9.45 (br. s., 1 H), 8.13 - 8.03 (m, 5 H), 7.86 (br. s., 1 H), 7.80 (br. s., 3 H), 7.29 (d, *J* = 1.8 Hz, 1 H), 7.23 (d, *J* = 1.2 Hz, 2 H), 7.22 - 7.21 (m, 2 H), 7.21 - 7.18 (m, 4 H), 7.16 (s, 3 H), 7.15 (d, *J* = 1.8 Hz, 1 H), 7.08 (d, *J* = 1.8 Hz, 1 H), 7.07 (d, *J* = 1.8 Hz, 1 H), 7.05 (d, *J* = 1.2 Hz, 1 H), 6.91 (d, *J* = 1.8 Hz, 1 H), 6.89 (d, *J* = 1.8 Hz, 2 H), 6.88 (s, 1 H), 6.85 (s, 2 H),

6.84 (d, *J* = 1.2 Hz, 1 H), 6.56 (br. s., 2 H), 3.99 (s, 3 H), 3.85 (s, 3 H), 3.85 (s, 3 H), 3.84 (s, 3 H), 3.84 (s, 3 H), 3.84 (s, 6 H), 3.81 (s, 6 H), 3.81 (s, 3 H), 3.80 (s, 6 H), 3.50 - 3.26 (m, 6 H), 3.24 - 3.16 (m, 3 H), 3.16 - 3.09 (m, 2 H), 3.09 - 2.98 (m, 3 H), 2.98 - 2.80 (m, 14 H), 2.74 (d, *J* = 5.3 Hz, 3 H), 2.55 - 2.51 (m, 4 H), 2.36 (t, *J* = 7.3 Hz, 2 H), 2.27 (t, *J* = 7.3 Hz, 2 H), 1.94 - 1.85 (m, 2 H), 1.82 - 1.73 (m, 4 H)

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

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

HPLC purity: 99%.(See **Figure 1b**).

PA31 Guan-Im-Py-Py-β-Py-Py-Py-γ-Py-Py-β-Py-Py-Py-Py-β-Ta (4 TFA):

^{1}H NMR (600 MHz, DMSO-d_6) δ = 10.46 (s, 1 H), 10.42 (s, 1 H), 9.94 (s, 1 H), 9.93 (s, 1 H), 9.92 (s, 3 H), 9.91 (s, 1 H), 9.90 (s, 1 H), 9.87 (s, 1 H), 9.84 (s, 1 H), 9.46 (br. s., 1 H), 8.13 - 8.07 (m, 4 H), 8.06 (t, *J* = 5.6 Hz, 2 H), 7.86 (br. s., 1 H), 7.81 (br. s., 3 H), 7.28 (d, *J* = 1.8 Hz, 1 H), 7.23 (d, *J* = 1.2 Hz, 2 H), 7.22 (d, *J* = 1.2 Hz, 1 H), 7.21 - 7.18 (m, 5 H), 7.17 (s, 3 H), 7.14 (d, *J* = 1.8 Hz, 1 H), 7.08 (d, *J* = 1.2 Hz, 1 H), 7.07 (d, *J* = 1.8 Hz, 1 H), 7.04 (d, *J* = 1.8 Hz, 1 H), 6.91 (d, *J* = 1.8 Hz, 1 H), 6.89 (s, 2 H), 6.88 (d, *J* = 1.8 Hz, 1 H), 6.87 - 6.85 (m, 2 H), 6.84 (d, *J* = 1.8 Hz, 1 H), 6.69 (s, 1 H), 6.56 (br. s., 2 H), 4.00 (s, 3 H), 3.86 (s, 3 H), 3.85 (s, 3 H), 3.84 (s, 3 H), 3.84 (s, 3 H),

3.84 (s, 6 H), 3.81 (s, 6 H), 3.81 (s, 3 H), 3.80 (s, 6 H), 3.51 - 3.32 (m, 6 H), 3.24 - 3.16 (m, 3 H), 3.16 - 3.09 (m, *J* = 6.2, 12.5, 12.5 Hz, 2 H), 3.09 - 2.98 (m, 3 H), 2.90 - 2.81 (m, 2 H), 2.74 (d, *J* = 4.7 Hz, 3 H), 2.55 - 2.51 (m, 4 H), 2.36 (t, *J* = 7.3 Hz, 2 H), 2.27 (t, *J* = 7.3 Hz, 2 H), 1.94 - 1.85 (m, 2 H), 1.82 - 1.75 (m, 4 H)

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

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

HPLC purity: 98%. (See **Figure 1c**).

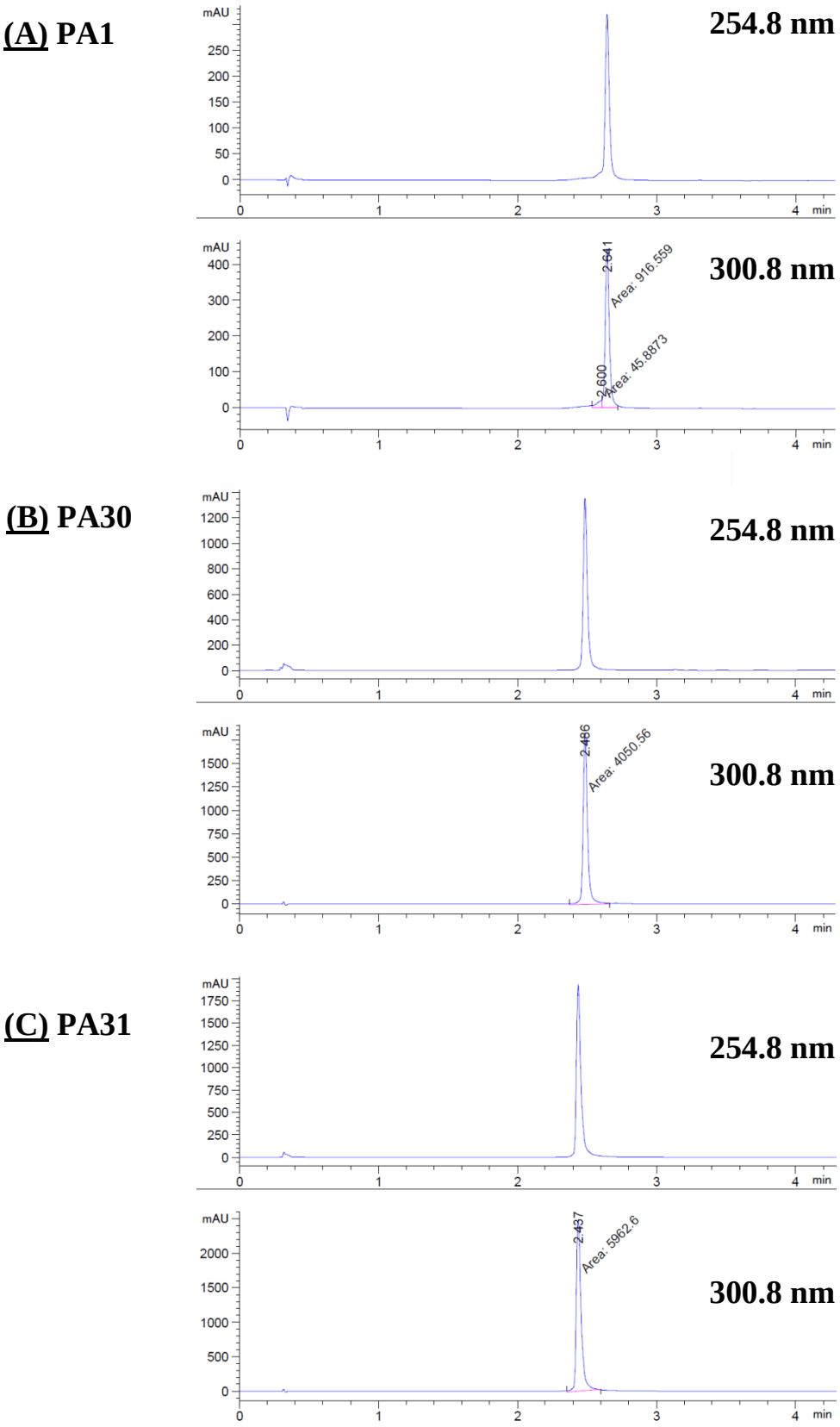


Figure 3.1 HPLC Purity of PA1, PA30 and PA31 with λ monitored.

3.3.2 Compound efficacy testing

The antiviral activities were calculated by measuring the ability of PAs to decrease the viral DNA (or episomal) load in monolayer keratinocyte cultures. The episomal load was determined by Q-PCR as previously described.[9] **PA1** was already reported, along with extensive discussion of how antiviral parameters are defined for this new mechanism of antiviral action in which the viral genome is degraded while the host cells are unharmed as measured by standard MTT and LDH assays and observations of numbers of floating vs. adherent cells, etc.[8,9,10] Typical assays were run for 48 h though much longer assays have also been run, and ref. 9 also describes organotypic, differentiated keratinocyte tissue culture assays as well as cell monolayer assays. Because **PA1** has been central to our work for over a decade, it has been used as an internal standard to benchmark new compound series, so even though we cite the prior study, **PA1** was studied side-by-side with **PA30** and **PA31** for the current work. The new **PA1** results were indistinguishable from the dose-response curves previously reported. Dose-response curves for **PA30** and **PA31** are given below in **Figure 2a** and **2b.**

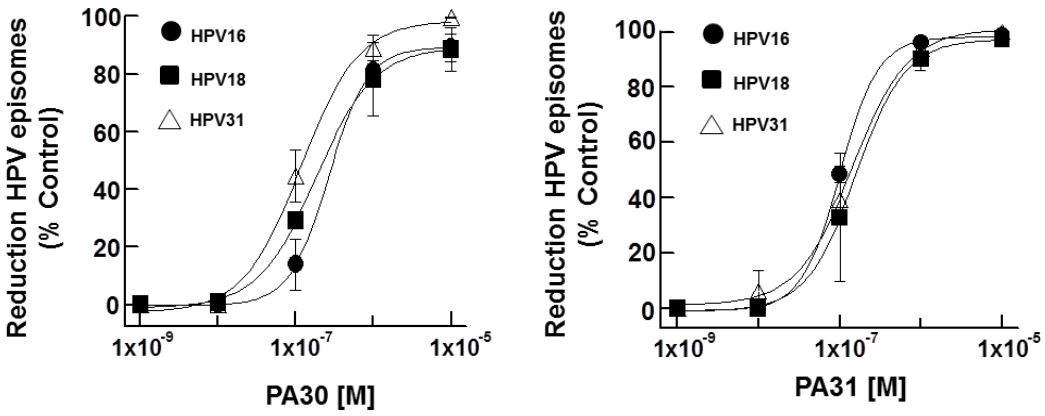


Figure 3.2. Dose-response curves for PA30 and PA31 against HPV16, 18 and 31 from which IC_{50} and IC_{90} values were determined as polyamide concentrations where 50 and 90% of viral DNA was undetectable relative to vehicle control. (a) Results for PA30. **(b)** Results for PA31.

Note that the plots show reduction, or decrease, in HPV DNA as a % of control, so 100% reduction corresponds to elimination of viral DNA on the scale plotted. Vehicle control was 0.1% DMSO in water, which is well tolerated by the keratinocytes.[8,9,10] Differences in efficacy between **PA30** and **PA31** are striking when HPV16 and 18 are considered: **PA30** (**Figure 2a**) achieves close to 100% reduction in viral DNA only for HPV31, a high-risk but minor form of the disease that is still clinically important. **PA31** is highly-active and essentially indistinguishably-active against all three high-risk HPVs, on the scale plotted in **Figure 2b**.

3.3.3 Quantitative DNase I footprinting

The molar extinction coefficient for **PA1** was determined as ε = 88,235 $M^{-1}cm^{-1}$ by fluorescence as previously described.[30] The molar extinction coefficients for **PA30** and **PA31** were determined as ε = 91,164 $M^{-1}cm^{-1}$, ε = 89,263 $M^{-1}cm^{-1}$ by measuring the absorbance at 305 nm of solutions with known polyamide concentrations in the range 5 x 10^{-7} M to 7 x 10^{-5} M.

Quantitative DNase I footprinting experiments were performed as described previously.[18] Briefly, a 305 bp DNA fragment corresponding to the nucleotides 7479-7783 in the Long Control Region (LCR) of the HPV18 genome (Genbank accession number X05015) was PCR amplified and fluorescently-labeled at both 5' ends using the following primers: (top strand) 5'-FAM-CT TAT GTC TGT GGT TTT CTG and (bottom strand) 5'-HEX-TT CAT GTT AAG GGT AGA CAG. The DNA fragment was diluted to 200 pM in TKMC buffer with 10 mM CHAPS and incubated with different concentrations of each polyamide (0.5-40 nM) for at least 4 h at 37 °C. The polyamide solutions were always added to DNA solutions not *vice versa*. At the end of this incubation period, each solution was subjected to ~0.08 U of DNase I for 5 min. The reactions were then quenched with EDTA and the DNA fragments were purified using a QIAquick PCR purification kit. Each experiment was performed at least three times.

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

Table 3.1. K_d Values Determined by Hill Equation for Sites 1-3 with Hill Coefficients and Overall Fit

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

Table 3.2. K_d Values Determined by Langmuir Equation for Sites 1-3 with Overall Fit

3.3.4 Affinity cleavage

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

The samples from quantitative DNase I footprinting and affinity cleavage experiments were analyzed using capillary electrophoresis and indexed using Sanger and Maxam-Gilbert sequencing, respectively.[31] Dissociation constant (K_d) values were calculated by normalizing the area from a peak in the footprint region to the area of a peak not sensitive to PA concentration. These values were plotted and fitted to Hill and Langmuir binding isotherms (as appropriate) with KaleidaGraph in methods reported in detail elsewhere. [18,31.32]

3.4 Results and Discussion

Polyamides were synthesized by Boc solid-phase methods[15] as reported[9] (**PA1**) or followed by TMG (**PA30**) or Guan (**PA31**) conjugation. After cleavage from the resin, polyamides were purified by reverse-phase HPLC and characterized with ^{1}H and ^{13}C NMR and HR mass spectrometry. Structures of

PA1, **PA30** and **PA31** and their EDTA-conjugates used in affinity cleavage maps are shown in **Figure 3.**

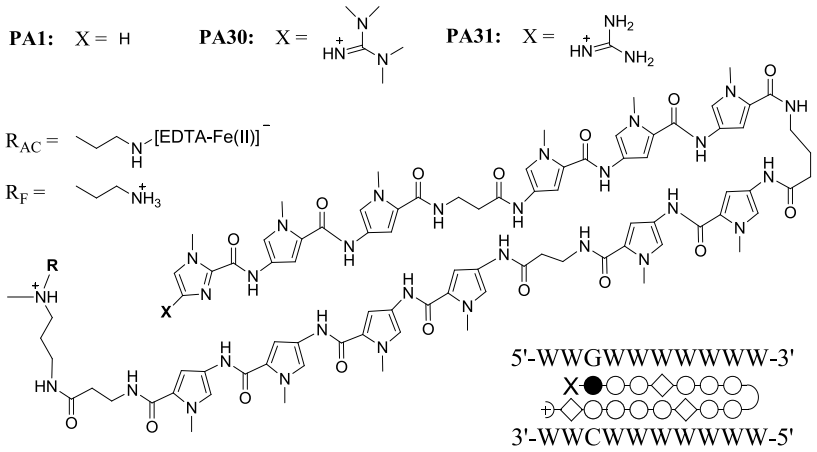


Figure 3.3. Chemical structure of PA1, PA30 and PA31. R_F refers to the substituent used for DNase I footprinting experiments, whereas R_{AC} indicates the EDTA-Fe conjugate for affinity cleavage experiments. Inset: cartoon representation of the PA with building blocks represented by open circles for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

Anti-HPV activities and MTT cell toxicity assays (shown for PA1, Figure SI 1A) were determined as previously described.[9a] However, new side-by-side, concentration-dependent assays were performed for previously reported **PA1**.[9a] **Table 3** provides the resulting pseudo IC_{50} and IC_{90} values for **PA1**, **PA30** and **PA31** corresponding to the PA concentration that causes a 50% or 90% decrease in viral DNA concentration. **Figure 2** gives dose-response curves for **PA30** and **PA31** while data are given for **PA1** and HPV16 and HPV31 in **Figure SI 1;** data for **PA1** were the same as previously reported.[26,9a]

Although IC_{50} values for the compounds are comparable against HPV16 and -31, **PA30** and **PA31** afforded somewhat improved IC_{50} values against HPV18. In contrast, significant improvement in IC_{90} for HPV16 and -18 was observed with **PA31**. However, PA30 activity plateaus just below a 90% decrease in viral DNA and, in two days, does not quite reach an IC90, an important measure of antiviral activity. We will explore longer time intervals, but higher concentrations do not seem promising based on the dose–response curves, and would make PA30 a poor competitor with the more-active PA31.

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

Table 3.3 IC_{50} and IC_{90} values of PA1,[a] PA30, and PA31 against high-risk HPV16, -18 and -31 types. The numbers in parentheses are sample standard deviations; n is the number of independent measurements. [a] IC_{50} values for **PA1** were first reported in reference 9; we re-measured them and found them unchanged.

Because only **PA31** showed an improved IC_{90} against HPV18, we investigated the DNA-binding properties of these hairpin PAs on a natural HPV18 sequence and compared them to the parent compound, **PA1**. Based on well-established PA-DNA pairing rules,[4a] **PA1**, **PA30**, and **PA31** were expected to target the ten base-pair DNA sequence WWGWWWWWWW, where W=A or T (**Figure 3** inset). A PA-DNA mismatch refers to any deviation from this recognition motif, i.e. where a PA building block pair binds to a noncanonical Watson-Crick base pair: one not predicted by the pairing rules.[1a] We used quantitative DNase I footprinting[16] to obtain dissociation

constants (K_d) and affinity cleavage experiments to assign binding orientations of the compounds on a 305 bp DNA fragment corresponding to nucleotides (nt) 7479 to 7783 in the Long Control Region of HPV18, a viral genomic region containing regulatory elements for replication.[17] We determined K_d values and orientation for all PA binding events along the entire 305 bp DNA fragment, and representative dissociation constants at three binding sites are given in **Table 4**. Similar results were observed in the rest of the DNA fragment (**Figure S1 2a-2b**). K_d was determined by the Hill or Langmuir equation, based on the magnitude of the derived Hill coefficient and overall fit (see **Table 1** and **2**).[18]

There is no significant difference in DNA-binding affinities of **PA1**, **PA30**, and **PA31**, indicating that the addition of a potential H-bond donor and a positive charge for **PA30** and **PA31** *vs.* **PA1** does not explain enhanced antiviral activity. It is noteworthy that since we are sampling an ensemble of different binding conformations in multiple linear, cell-free DNA molecules, binding affinities are the sum of these binding events at a particular binding site. Furthermore the negative supercoiling of the viral genome, and bound viral and host proteins in vivo, probably affect binding (see Conclusions). We are currently examining these matters experimentally. All three PAs bind to Sites 1 and 2 (single-mismatch sites) with K_d values ranging from 0.9-1.2 nM, and it is noteworthy that they all tolerate a triple base-pair mismatch at Site 3 with K_d values ranging from 2.2-3.2 nM (**Table 4**). The high Hill coefficients at this site are explained by previously-reported polyamide binding cooperativity,[4d] in which the binding of one polyamide can preorganize the minor groove for the binding of a second (and more) polyamides, for example by widening the minor groove. This cooperativity would be most expected in areas of

high polyamide density on the DNA, which is what we found in the Long Control Region. We need to point out that we used a single-site binding model to analyze binding thermodynamics. If we had parameterized binding to allow multiple Hill coefficients to float simultaneously, we would more accurately have accounted for the statistics of multi-site binding. The results would likely have been that some Hill Kd values and coefficients would have decreased, and others would have increased. However, we were not convinced that our data would hold up to the introduction of so many additional parameters (a Hill coefficient per binding site and orientation), and did not believe that the outcome would change significantly in a scientific sense with a more precise handling of the statistical question. We will revisit this with more complete data sets for multiple compounds to test whether our approximation, which is normal in polyamide literature,[1-4] is suitable or not.

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

Table 3.4. Representative PA binding sites on nt 7479-7783 of the HPV18 LCR as determined by quantitative DNase I footprinting and affinity cleavage. DNA sequences are illustrated so that the middle block represents the PA binding site along with 5 flanking nt. In each sequence, the underlined base is bound by imidazole/pyrrole (or pyrrole/imidazole) pair of the polyamide. The bold bases in column 2 correspond to mismatches according to recognition "rules." K_d values were calculated from DNase I experiments performed at least in triplicate (**Table 1 and 2**).

Thus, large antiviral hairpin PAs tolerate reportedly unfavourable DNA interactions without a significant decrease in binding affinity, which agrees with our prior work on HPV16.[18] In fact, in that prior work we found for 14-

ring **PA1** that numerous "mismatch" sites were bound with higher affinity than so-called canonical sites. Also, up to four mismatches were tolerated with 2 nM dissociation constants identical to numerous "perfect-match" sites. For a larger, 20-ring antiviral polyamide (**PA25**), four mismatches were even more common than with **PA1**, leading to a complete blanketing of the HPV16 long control region when only sparse binding, at best, was predicted.[18b]

Data for HPV18-polyamide binding to a portion of the LCR is reported here and allow us to expand our understanding (polyamide-HPV18 binding was previously reported for the E2-protein binding sites).[7a] We hypothesize that a deviation from established binding rules occurs for our unusually large hairpin PAs because they form many H-bonds with DNA and significant internal π-stacking compared to the smaller PAs typically found in the literature.[19] Thus, a few unfavourable enthalpic interactions are likely insignificant. The entropy of binding is likely very favourable for long PAs from the loss of many water molecules previously resident in the minor groove and from loss of exposed nonpolar surface area upon π-stacking. Entropy and enthalpy are being determined by the CM Dupureur group. Since we determined polyamide-DNA binding properties in the context of a large, natural DNA sequences instead of a small artificial, DNA sequences designed to examine sequence preferences at a single location, our results are much more complex than those found in the literature. Published binding rules are not obeyed with significant fidelity, and "mismatches" are very well tolerated (see also He et al.[18] and Qiao et al.[20]).

Binding sites for **PA1**, **PA30**, and **PA31** in a region of HPV18 DNA spanning nt 7601-7620 are shown in **Figure 4A**. In this part of the viral genome, there are two predicted single-mismatch binding sites, sequence **(i)** 5'-

TATTAGTCAT-3' (7606-7615) and sequence **(ii)** 5'-TAGTCATTTT-3' (7609-7618). Our data indicate that **PA1** and **PA31** bind both of these sites, while **PA30** only binds to **(i)**.

Raw DNase I footprinting data of dsDNA (top strand) is shown in **Figures 4B** (control) and **4C** (5 nM PA). Inspection of the footprints generated by the three PAs reveals subtle differences in the protected regions. For instance, the peaks at the left-hand edge of the protected region corresponding to nt 7606-09 (**Figure 4*** shows 7606) display increased nuclease accessibility in the case of **PA1** vs. **PA30** and **PA31**. At the right-hand edge of the protected region, the peak corresponding to nt 7619 (**Figure 4C**[†]) is less protected from DNase I by **PA30** (**Figure 4C** middle panel) than by **PA1** (**Figure 4C** left panel) at equivalent PA concentrations. In contrast, **PA31** affords more-extended nuclease protection than either **PA1** or **PA30**, with a footprint including extra nt 7619-7626; this extra nuclease protection corresponds to the double-mismatch site **(iii)** at 7612-7621. For steric reasons, the TMG group of **PA30** may not bind to DNA, decreasing productive enthalpic interactions compared to **PA1** and **PA31**. Footprinting differences are supported by affinity cleavage (AC) results.[21]

Patterns generated by **PA1-**, **PA30-** and **PA31-EDTA** conjugates are given in **Figure 4D**. For **PA1** and **PA31** there are two binding events, each corresponding to a distinct AC pattern, agreeing with DNase I data that both **PAs** bind at both available single-mismatch sites, **(i)** and **(ii)**.

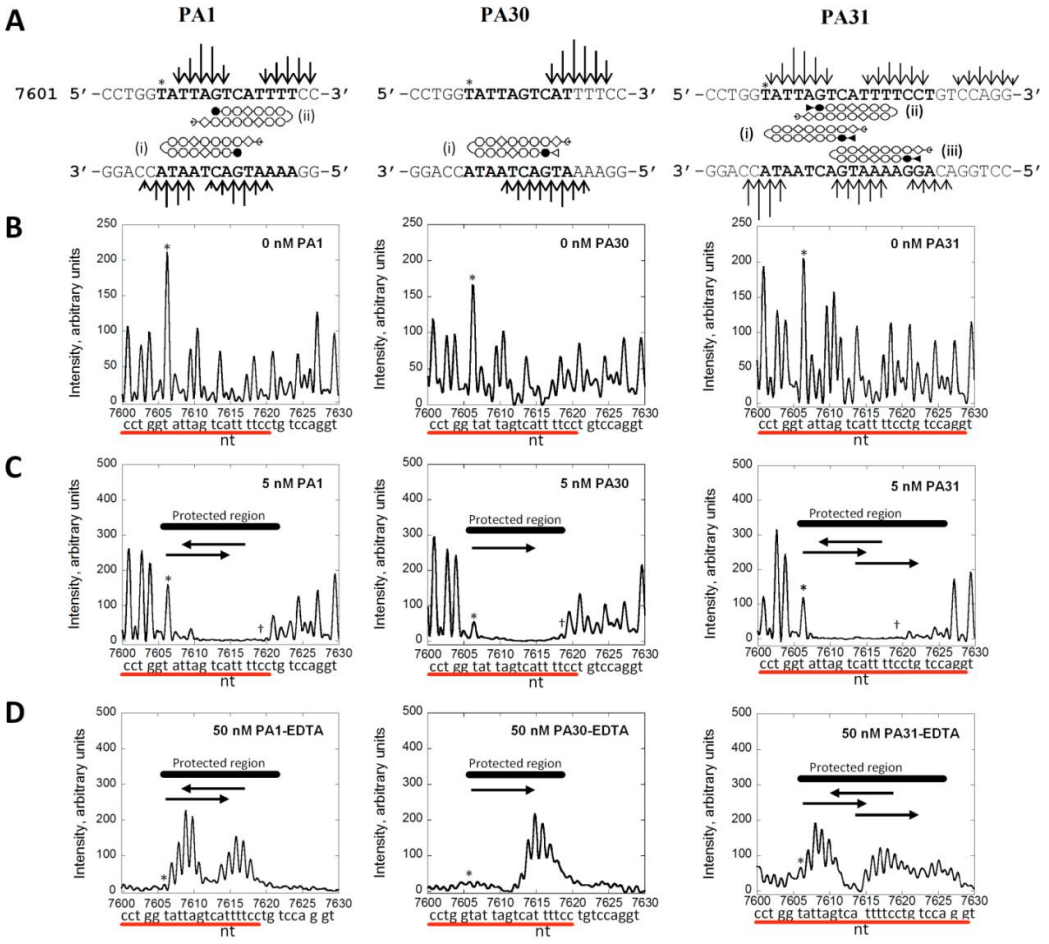


Figure 3.4. (A) Binding sites for PA1, PA30, and PA31 from a section of HPV18 LCR corresponding to 7600-7620 as determined by capillary electrophoresis, DNase I footprinting and AC. The relative heights above the DNA sequence correspond to observed AC intensities. The remaining panels show the top strand of ten more base pairs than panel **A**, 7600–7630. Representative electropherograms showing the top strand for **(B)** DNase I footprinting controls (0 nM PA), **(C)** DNase I-protected regions upon the addition of 5 nM of PA and **(D)** AC patterns at 50 nM PA. The red line below each plot corresponds to the DNA sequence shown in panel **(A)**. The * and † symbols mark nt 7606 and 7619. Arrows inside the plots show the PA binding direction, with the arrowhead in the position of the PA tail. PA building blocks as defined in Fig. 1; TMG and Guan groups are open and filled triangles.

However, an extra AC pattern was observed for PA31 corresponding to the double mismatch site **(iii)**. Conversely, only one of the predicted single-mismatch sites is occupied by **PA30**: a single AC pattern corresponding to site **(i)** was observed. The AC pattern for site **(ii)** did not appear even upon increasing the concentration of **PA30** from 50 nM to 200 nM (data not shown).

Our results with these 14-ring PAs agree with previous 8-ring PAs studies reporting that substitution at the N-terminal imidazole can modulate orientation and DNA-binding preferences of hairpin polyamides.[22]

3.5 Conclusions

In summary, we show that N-terminal substitution of anti-HPV polyamide **PA1** improves antiviral activity against the two most high-risk HPV types that are most prevalent (HPV16 and -18). N-terminal substitution with TMG leads to improvement in anti-HPV activity (IC_{50}) against HPV18. N-terminal substitution with guanidinium itself enhanced anti-HPV activity even further relative to **PA1** by decreasing IC_{50} against HPV18 and IC_{90} for both HPV16 and 18. For antiviral agents in general, a dosage above IC_{90} decreases viral load, the magnitude, and frequency of viral rebound,[23] and the probability of viral mutation, so improvements in IC_{90} are significant for therapy, as they are for EC_{90}.[24]

Although the dissociation constants for the reported PAs are similar and cannot account for the differences in antiviral activity, differences in PA-DNA binding distribution were observed: **PA31** binds more sites on the LCR of HPV18 than **PA1** or **PA30**. Binding kinetics may be more important for activity[25] and remain under investigation. Polyamide uptake is also under investigation. We note that **PA1** and its analogs **PA30** and **PA31** are among the smallest molecule in a library of 75+ active compounds that reach nearly twice

the molecular weight of **PA1**. We have also prepared a series of compounds based on the highly-active **PA25** and other active series, and they will be described subsequently. **PA25** was our second preclinical lead because of its IC_{50}/IC_{90} data (μM) as follows: HPV16, 0.036/0.35; HPV18, 0.056/1.5; HPV31, 0.030/0.51.[9,26] The first biophysical studies of **PA25** were reported recently.[18b] However, **PA25** is a more difficult synthetic and purification challenge than **PA1** and its derivatives, though we have scaled up both **PA1** and **PA25** to 10 g, and the cost-of-goods is much higher for the 20-ring **PA25** than the 14-ring **PA1**, so we were interested to see if modifications to **PA1** could increase its activity significantly. That goal was successful with **PA31**.

Long hairpin polyamides of this general class function by a novel polyamide mechanism that invokes the host DNA Damage Response to destroy the viral genome.[8,10] Their activity is not limited to HPV but is also found for other small, double-stranded DNA tumour viruses, including polyomaviruses.[26] This mechanism is unrelated to any prior polyamide literature that we can determine, even though our work was initially inspired by early PA research. **PA30/31** represent an improved class of long polyamide that invokes this new mechanism to combat cancer-causing HPV. More recently, Dervan has reported a similar, non-sequence-specific, polyamide-induced DNA Damage Response.[27]

Why might this nonspecific binding, or poorly-specific binding, be so relevant? We have come to hypothesize and are performing experiments to test the idea, that the interaction of large polyamides with small, circular, double-stranded DNA genomes is biologically important more from physical than chemical reasons.

Hydrogen bonding does go on in the minor groove, albeit at moderate selectivity, but of great importance is the binding of many hundreds or more of rigid, crescent-shaped molecules to supercoiled DNA targets. This must result in considerable strain on the supercoiled viral genomes, and could in the most extreme case result in single-stranded DNA regions appearing, which would be rapidly noticed and acted upon by the DNA Damage Response machinery. However, the process may be simpler: the dsDNA viruses evolved to use parts of the DDR for their life cycles but otherwise to hide, to avoid having their genomes recognized as foreign or damaged in any way. When the supercoiled genomes are distorted by the effect of many bound polyamides, and 50% of HPV DNA is bound by active polyamides, at least for the linearized pieces we have reported on to date,[18,9c] the viral genome may no longer remain hidden. The strain imposed by all of the bound polyamides may change the shape enough so that the DDR naturally notices unwanted additions to the chromatin and is automatically activated. Support for this hypothesis is found in our recently published patent application, where, as mentioned above, we showed that the same polyamides active against HPV are also active against three polyomaviruses having little in common with HPV except a small, dsDNA genome that is supercoiled.[26] The DNA Damage Response is also activated for these disparate viruses.

3.6 Notes

We thank Jacqui Niederschulte for extinction coefficient determination of **PA1**, the DNA Core at the University of Missouri for DNA CE fragment analysis, the Danforth Plant Science Center (NSF DBI 0922879], Prof. B. Bythell for

HRMS and Prof. P. Lambert for providing W12E cells. The Agilent 600 MHz NMR spectrometer was obtained using funds from NSF (#0959360). This work was supported by NIH-NIAID: AI083803, AI062182, and AI068159.

3.7 References

1 (*a*) P. B. Dervan, B. S. Edelson, *Curr. Opin. Struct. Biol.*, 2003, **13**, 284; (*b*) X. Chen, B. Ramakrishnan, S.T. Rao, M. Sundaralingam, *Nat. Struct. Mol. Biol.*, 1994, **1**, 169; (*c*) L. A. Marky, K. J. Breslauer, *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 4359; (*d*) A. C. Finlay, F. A. Hochstein, B. A. Sobin, F. X. Murphy, *J. Am. Chem. Soc.*, 1951, **73**, 341.

2 *(a)* S. White, E. E. Baird, P. B. Dervan, *Biochemistry*, 1996, **35**, 12532. *(b)* S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan, *Nature*, 1998, **391**, 468-471.

3 (*a*) J. W. Trauger, E. E. Baird, M. Mrksich, P. B. Dervan, *JACS* 1996, **118**, 6160; (*b*) C.C.C. Wang, U. Ellervik; P. B. Dervan, *Bioorg. Med. Chem.*, 2001, **9**, 653.

4 (*a*) S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan, *Nature*, 1998, **391**, 468; (*b*) M. Mrksich, W. S. Wade, T. J. Dwyer, B. H. Geierstanger, D. E. Wemmer, P. B. Dervan, *Proc. Natl. Acad. Sci., USA* 1992, **89**, 7586; (*c*) M. Mrksich, M. E. Parks, P. B. Dervan, *J. Am. Chem. Soc.*, 1994, **116**, 7983. *(d)* C. A. Hawkins, E. E. Baird, P. B. Dervan and D. E. Wemmer, J. Am. Chem. Soc., 2002, **124**, 12689.

5 (*a*) A. K. Mapp, A. Z. Ansari, M. Ptashne, P. B. Dervan, *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 3930; (*b*) X. Wang, H. Nagase, T. Watanabe, H. Nobusue, T. Suzuki, Y. Asami, Y. Shinojima, H. Kawashima, K. Takagi, R. Mishra, J. Igarashi, M. Kimura, T. Takayama, N. Fukuda, H. Sugiyama, *Cancer. Sci.*, 2010, **101**, 759; (*c*) J. A. Raskatov, J. L. Meier, J. W. Puckett, F. Yang, P. Ramakrishnan, P. B. Dervan, *Proc. Natl. Acad. Sci. USA*, 2012, **109**, 1023.

6 K. Maeshima, S. Janssen, U. K. Laemmli, *EMBO J.*, 2001, **20**, 3218.

7 (*a*) T. D. Schaal, W. G. Mallet, D. L. McMinn, N. V. Nguyen, M. M. Sopko, S. John, B. S. Parekh, *Nucleic Acids Res.*, 2003, **31**, 1282; (*b*) A. Yasuda, K. Noguchi, M. Minoshima, G. Kashiwazaki, T. Kanda, K. Katayama, J. Mitsuhashi, T. Bando, H. Sugiyama, Y. Sugimoto, *Cancer Sci.*, 2011, **102**, 2221; (*c*) L. A. Dickinson, R. J. Gulizia, J. W. Trauger, E. E. Baird, D. E. Mosier, J. M. Gottesfeld, P. B. Dervan, *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 12890; *(d)* R. W. Burli, D. McMinn, J. A. Kaizerman, W. Hu, Y. Ge, Q. Pack, V. Jiang, M. Gross, M. Garcia, R. Tanaka, H. E. Moser, *Bioorg. Med. Chem. Lett.*, 2004, **14**, 1253-1257; *(e)* R. W. Burli, P. Jones, D. McMinn, Q. Le, J.-X. Duan, J. A. Kaizerman, S. Difuntorum, H. E. Moser, *Bioorg. Med. Chem. Lett.*, 2004, **14**, 1259-1263; *(f)* M. P. Barrett, C. G. Gemmell, C. J. Suckling, *Pharmacol. Ther.*, 2013, **139**, 12-23; *(g)* M. N. C. Soeiro, K. Werbovetz, D. W. Boykin, W. D. Wilson, M. Z. Wang, A. Hemphill, *Parasitology*, 2013, **140**, 929-951.

8 (*a*) T. G. Edwards, T. J. Vidmar, K. Koeller, J. K. Bashkin, C. Fisher, *PLoS One*, 2013, **8**, e75406. (*b*) C. Fisher, *J. Clin. Med.*, 2015, **4**, 204. *(c)* B. J.

Lamarche, N. I. Orazio and M. D. Weitzman, FEBS Lett., 2010, **584,** 3682–3695; *(d)* J. J. Lin and A. Dutta, J. Biol. Chem., 2007, **282**, 30357–30362.

9 *(a)* T. G. Edwards, K. J. Koeller, U. Slomczynska, K. Fok, M. Helmus, J. K. Bashkin, C. Fisher, *Antiviral Res.*, 2011, **91**, 177; *(b)* G. He, J. K. Bashkin, *Future Med. Chem.*, 2015, **7**, 1953-1955; *(c)* K. J. Koeller, G. D. Harris, K. Aston, G. He, C. H. Castaneda, M. A. Thornton, T. G. Edwards, S. Wang, R. Nanjunda, W. D. Wilson, C. Fisher, J. K. Bashkin, *Med Chem (Los Angeles)*, 2014, **4**, 338-344.

10 T. G. Edwards, M. J. Helmus, K. Koeller, J. K. Bashkin, C. Fisher, *J. Virol.*, 2013, **87**, 3979.

11 J. M. M. Walboomers, M. V. Jacobs, M. M. Manos, F. X. Bosch, J. A. Kummer, K. V. Shah, P. J. F. Snijders, J. Peto, C. J. L. M. Meijer, N. Muñoz, *J. Pathol.*, 1999, **189**, 12.

12 G. Clifford, S. Franceschi, M. Diaz, N. Munoz, L. L. Villa, *Vaccine*, 2006, **24** *(Suppl 3)*, S26.

13 (*a*) C. Dochez, J. J. Bogers, R. Verhelst, H. Rees, *Vaccine*, 2014, **32**, 1595; (*b*) E. A. Joura, A. R. Giuliano, O.E. Iversen, C. Bouchard, C. Mao, J. Mehlsen, E. D. Moreira, Y. Ngan, L. K. Petersen, E. Lazcano-Ponce, P. Pitisuttithum, J. A. Restrepo, G. Stuart, L. Woelber, Y. C. Yang, J. Cuzick, S. M. Garland, W. Huh, S. K. Kjaer, O. M. Bautista, I. S. F. Chan, J. Chen, R. Gesser, E. Moeller, M. Ritter, S. Vuocolo, A. Luxembourg, *N. Engl. J. Med.*, 2015, **372**, 711.

14 (*a*) P. S. Nagle, C. McKeever, F. Rodriguez, B. Nguyen, W. D. Wilson, I. Rozas, *J. Med. Chem.*, 2014, **57**, 7663; (*b*) R. K. Arafa, M. A. Ismail, M. Munde, W. D. Wilson, T. Wenzler, R. Brun, D. W. Boykin, *Eur. J. Med.*

Chem., 2008, **43**, 2901; (*c*) F. Rodríguez, I. Rozas, M. Kaiser, R. Brun, B. Nguyen, W. D. Wilson, R. N. García, C. Dardonville, *J. Med. Chem.*, 2008, **51**, 909.

15 *(a)* E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.*, 1996, **118**, 6141. *(b)* J. W. Puckett, J. T. Green, P. B. Dervan, *Org. Lett.*, 2012, **14**, 2774-2777; *(c)* note: in the hands of a microwave peptide synthesis machine manufacturer over an extended period, microwave-assisted synthesis was not successful for PA1, failing after the γ turn.

16 (*a*) D. J. Galas, A. Schmitz, *Nucleic Acids Res.*, 1978, **5**, 3157; (*b*) W. Yindeeyoungyeon, M. A. Schell, *Biotechniques*, 2000, **29**, 1034.

17 *(a)* W. Stünkel, H.-U. Bernard, *J. Virol.*, 1999, **73**, 1918.

18 *(a)* G. He, E. Vasilieva, G. D. Harris Jr, K. J. Koeller, J. K. Bashkin, C. M. Dupureur, *Biochimie*, 2014, **102**, 83; *(b)* E. Vasilieva, J. Niederschulte, Y. Song, J. George D Harris, K. J. Koeller, P. Liao, J. K. Bashkin, C. Dupureur, *Biochimie*, 2016, **127**, 103-114.

19 J. M. Turner, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.*, 1997, **119**, 7636.

20 H. Qiao, C. Ma, X. Zhang, X. Jing, C. Li, and Y. Zhao, *Bioconjugate Chem.*, 2015, **26,** 2054.

21 S. E. Swalley, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.*, 1996, **118**, 8198.

22 (*a*) S. White, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.*, 1997, **119**, 8756; (*b*) C. A. Hawkins, R. P. de Clairac, R. N. Dominey, E. E. Baird, S. White, P. B. Dervan, D. E. Wemmer, *J. Am. Chem. Soc.*, 2000, **122**, 5235.

23 J. M. Pawlotsky, *Hepatology*, 2011, **53**, 1742.

24 S. Duffy, L. A. Shackelton, E. C. Holmes, *Nat. Rev. Genet.*, 2008, **9**, 267.

25 (*a*) M. Goyal, M. Rizzo, F. Schumacher, C. F. Wong, *J. Med. Chem.*, 2009, **52**, 5582; (*b*) R. A. Copeland, D. L. Pompliano, T. D. Meek, *Nat. Rev. Drug Discov.*, 2006, **5**, 730.

26 J. K. Bashkin, T. G. Edwards, C. Fisher, G. D. Harris, Jr., K. J. Koeller, US Patent Office, Patent number: 20150329596, Appln. Number: 14/818881, November 19, 2015.

27 T. F. Martinez, J. W. Phillips, K. K. Karanja, P. Polaczek, C. Wang, B. C. Li, J. L. Campbell, P. B. Dervan, *Nucleic Acids Res.*, 2014, **42**, 11546.

28 M. del Fresno, A. El-Faham, L. A. Carpino, M. Royo, F. Albericio, *Org. Lett.*, 2000, **2**, 3539.

29 S. Robinson, E. J. Roskamp, *Tetrahedron*, 1997, **53**, 6697.

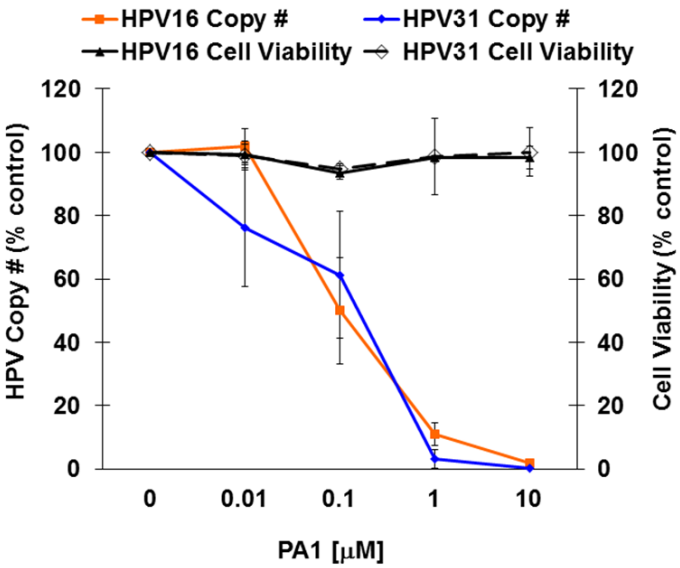
30 C. M. Dupureur, J. K. Bashkin, K. Aston, K. J. Koeller, K. R. Gaston, G. He, *Anal. Biochem.*, 2012, **423**, 178.

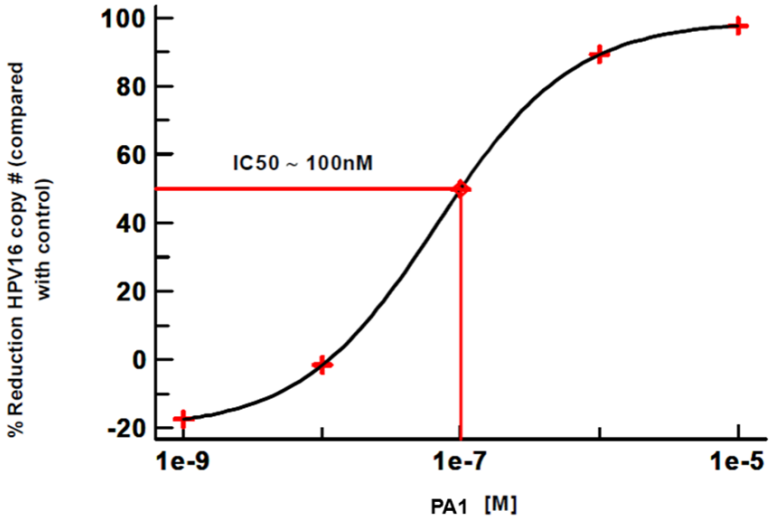
31 G. He, E. Vasilieva, J. K. Bashkin, C. M. Dupureur, *Anal. Biochem.*, 2013, **439**, 99.

32 K. J. Koeller, G. D. Harris, J., K. Aston, G. He, C. H. Castaneda, M. A. Thornton, T. G. Edwards, S. Wang, R. Nanjunda, W. D. Wilson, W. D., C. Fisher, J. K. Bashkin, *Med. Chem.*, 2014, **4**, 338.

3.8 Supporting Information

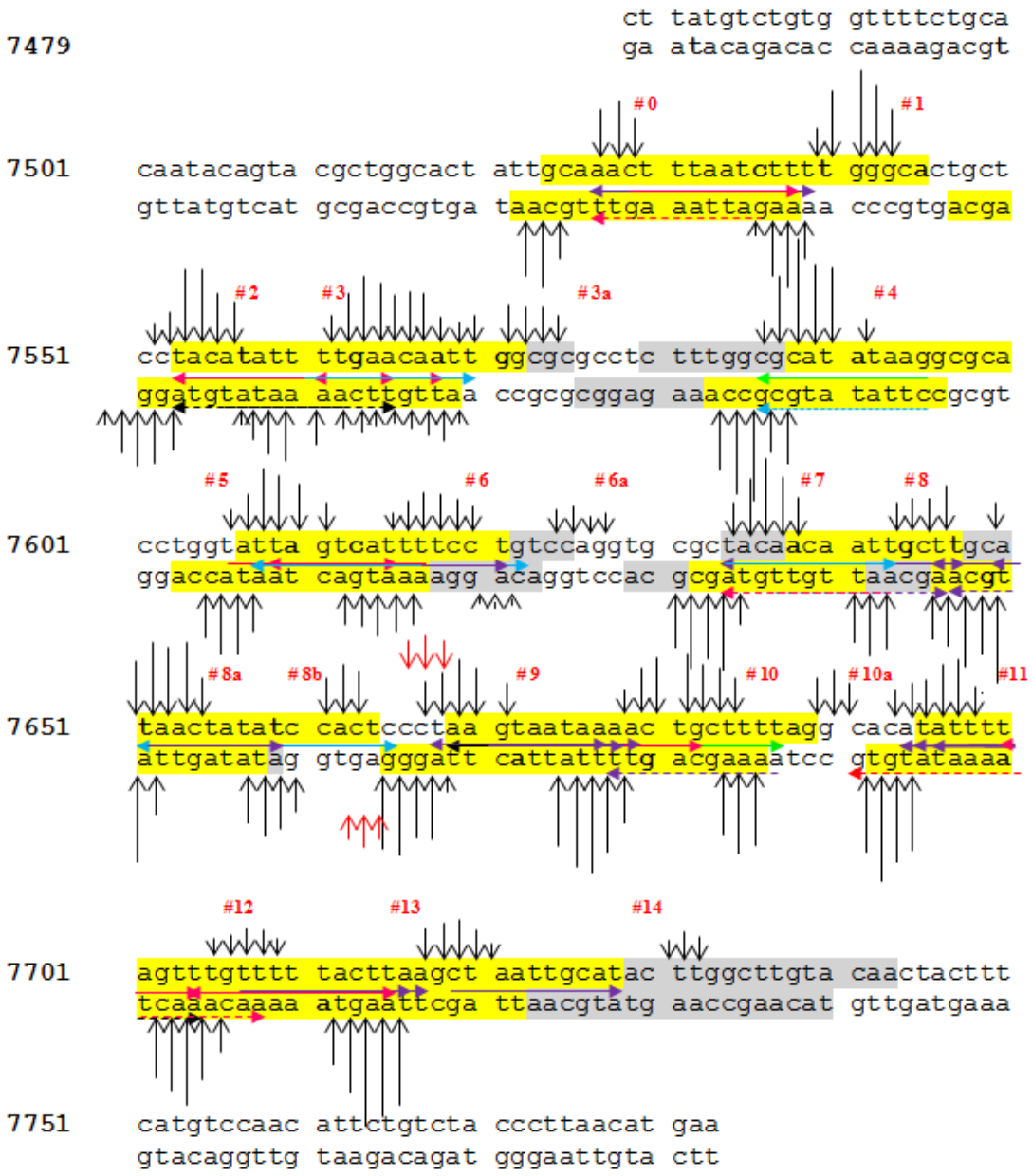
Figure SI 3.1. Dose-response curves for PA1 against HPV16 and 31 from which IC50 and IC90 values were determined. (A) Results for HPV16 and 31 with cell toxicity data. **(B)** Expanded detail of results for HPV16, plotted as the inverse of the other results (as the percent decrease of viral DNA vs. control,
reaching 100% decrease at maximum efficacy, instead of the simple decrease in DNA vs. control).





Q-PCR; n = 4 independent cell culture experiments in the presence of increasing concentrations of PA1.

Figure SI 3.2a. Map of NV1115 binding sites on HPV18 7479-7563

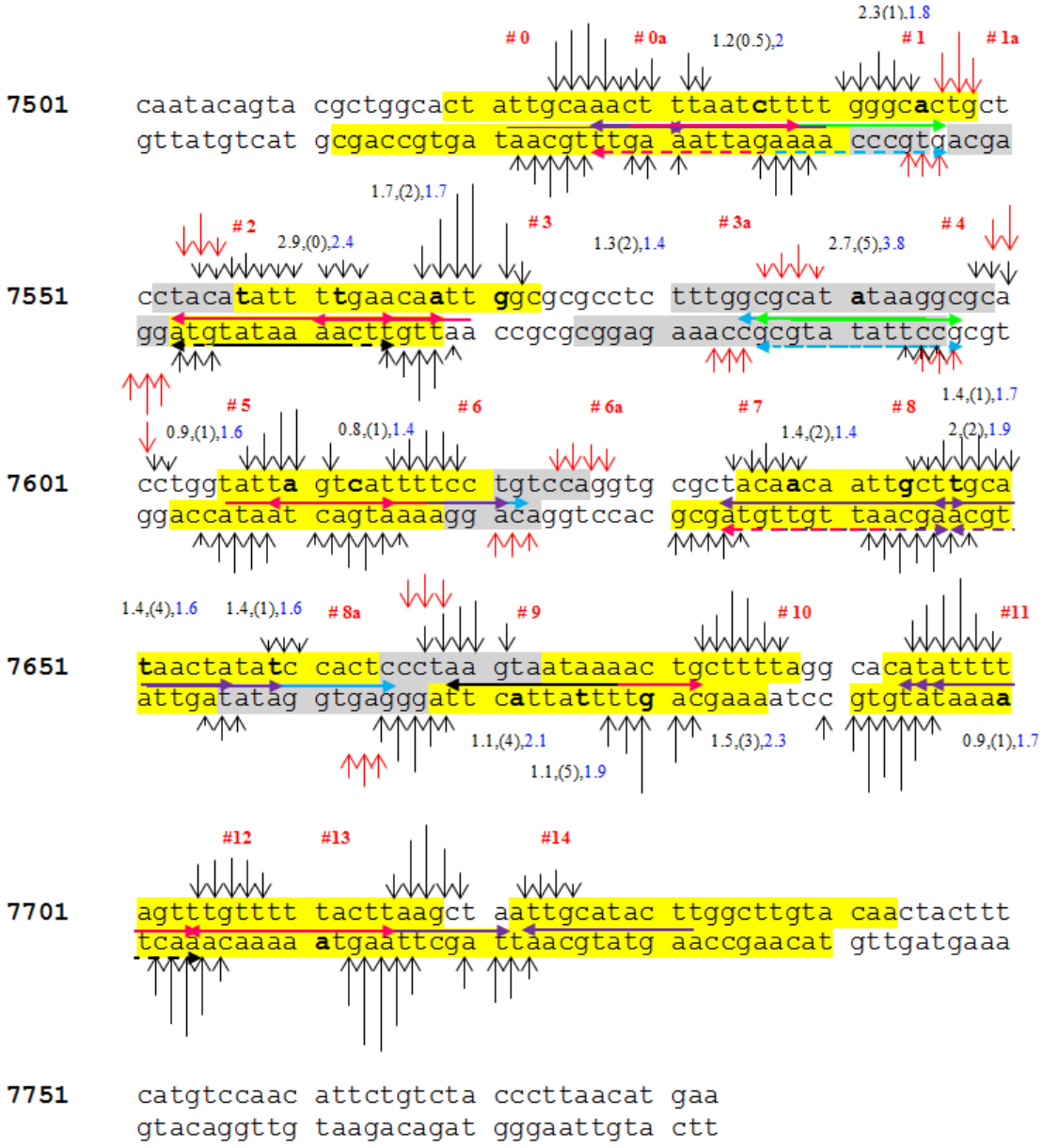


LEGEND

-DNAse I Footprinting (PA, 5 nM, DNA 0.2 nM)
-DNAse I Footprinting (PA, 10 nM, DNA 0.2 nM)
-#n: Affinity Cleavage Site
-**Bolded nucleotide**: DNaseI integration site
-**Black numbers above/below DNA sequence**: K_d and standart deviation
-**Blue numbers**: Hill Coefficient (represented only when >2)

Orizzontal arrows:
-Black: perfect Binding Site
-Pink: 1 Mismatch
-Purple: 2 Mismatch
-Blue: 3 Mismatch
-Green: 4 Mismatch

Figure SI 3.2b. Map of NV1128 binding sites on HPV18 7479-7563



LEGEND

-DNAse I Footprinting (PA, 5 nM, DNA 0.2 nM)
-DNAse I Footprinting (PA, 10 nM, DNA 0.2 nM)
-#n: Affinity Cleavage Site
-**Bolded nucleotide**: DNaseI integration site
-**Black numbers above/below DNA sequence**: Kd_s and standart deviation
-**Blue numbers**: Hill Coefficient (represented only when >2)

Orizzontal arrows:
-Black: perfect Binding Site
-Pink: 1 Mismatch
-Purple: 2 Mismatch
-Blue: 3 Mismatch
-Green: 4 Mismatch

Table SI 3.1. Kd values determined for NV1115 on HPV 7479-7783 (Kd values for NV1028 are reported in the supplementary material in Chapter 1).

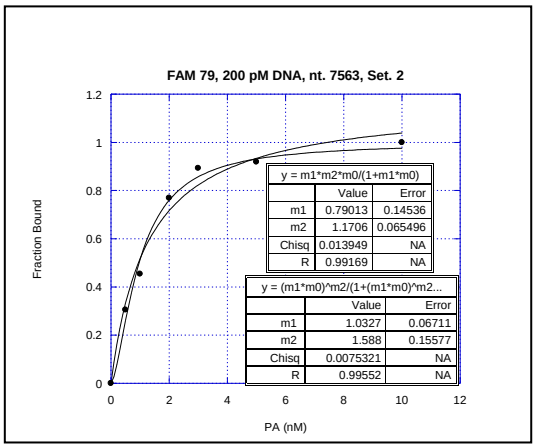
Site	Sequence	Position	Site type	Binding orientation	Integration nt	Kd Dnase- I (nM)
0	TTGCA AACTTTAATC TTTTG	7527-7536	Double Single	Forward Reverse	/	/
1	CCCAA AAGATTAAAG TTTGC	7538-7529	Single	Forward	/	/
2	CGAGG ATGTATAAAA CTTGT	7553-7562	Perfect Single	Reverse Forward	7557	1.48 (0.34)
3	ATATT TTGAACAATT GGCGC	7561-7570	Single	Forward	7563	1.01 (0.13)
3a	CGCGG TTAACAAGTT TTATA	7570-7561	Triple	Forward	7568	1.46 (0.20)
4	AAACC GCGTATATTC CGCGT	7586-7595	Triple Quadruple	Reverse Forward	7591	2.20 (0.23)
5	GGTAT TAGTCATTTT CCTGT	7609-7618	Single	Forward	7610	0.99 (0.05)
6	GGAAA ATGACTAATA CCAGG	7615-7606	Single	Forward	7613	1.00 (0.14)
6a	TGGAC AGGAAAATGA CTAAT	7621-7612	Double	Forward	7613	1.00 (0.14)
7	ACGCG ATGTTGTTAA CGAAC	7634-7645	Single Double	Forward Reverse	7638	2.70 (0.27)
8	GTTCG TTAACAACAT CGCGT	7693-7684	Triple	Forward		
8a	CTTGC ATAACTATAT CCACT	7650-7659	Double	Forward	7651	1.14 (0.15)
	TTGCA TAACTATATC CACTC	7651-7660	Triple	Forward		
8b	TCACC TATATCAATA CGTTC	7659-7650	Double	Forward	7659	0.78 (0.15)
9	TCCCT AAGTAATAAA ACTGC	7669-7678	Perfect	Forward	7672	0.88 (0.22)
10	CGTCA AAATAATGAA TCCCT	7678-7669	Double	Forward	7680	1.22 (0.28)
10a	ACGGA TTTTCGTCAA AATAA	7687-7678	Quadruple	Forward	7680	1.22 (0.28)
11	GGCAC ATATTTTAGT TTGTT	7694-7703	Double	Forward	7700	0.61 (0.32)
	GCACA TATTTTAGTT TGTTT	7695-7704	Double	Forward	7700	0.61 (0.32)
	CACAT ATTTTAGTTT GTTTG	7696-7705	Double	Forward	7700	0.61 (0.32)
12	TTAGT TTGTTTTTAC TTAAG	7704-7713	Single	Forward	CHC*	CHC
	TTTGT TTGATTTTAT ACACG		Single Perfect	Forward Reverse	7700	0.61 (0.32)
13	AGCTT AAGTAAAAAC AAACT	7715-7706	Single	Forward	CHC*	CHC

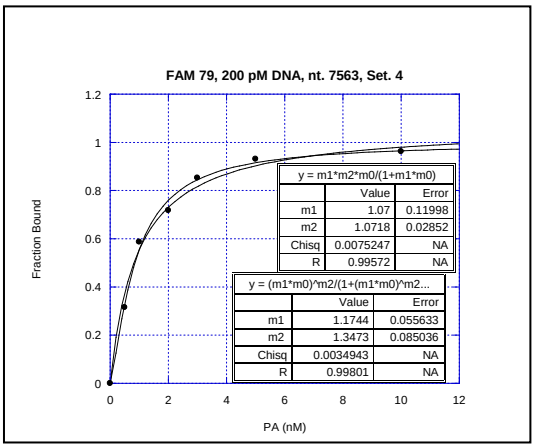
* Kd values at this site were determined on a different DNA fragment by Carlos Castaneda

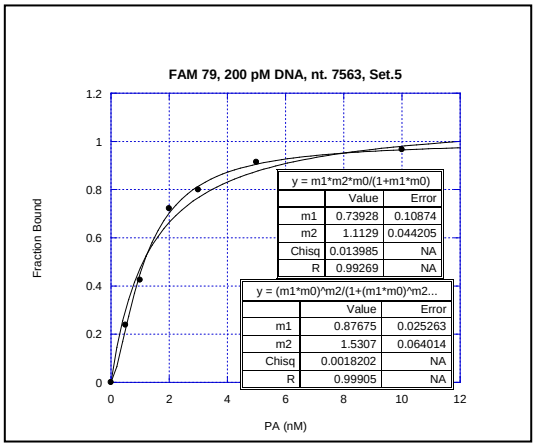
Red nucleotides indicate mismatch positions

Figure SI 3.3. Representative isotherms of NV1115 binding on HPV18 7479-7783

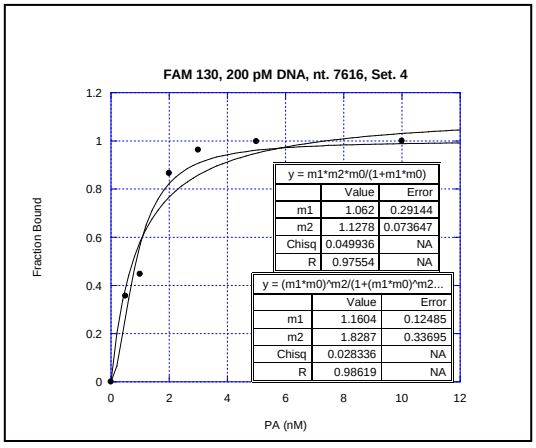
Site 7563

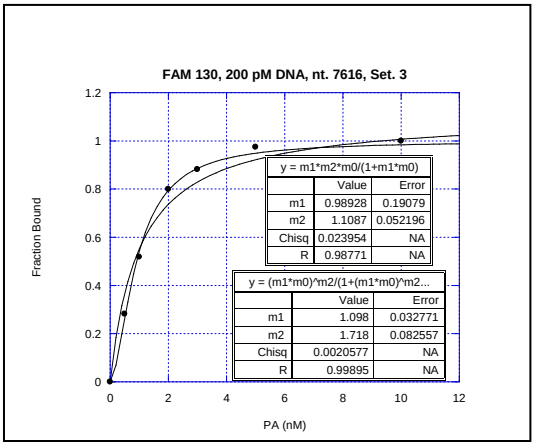


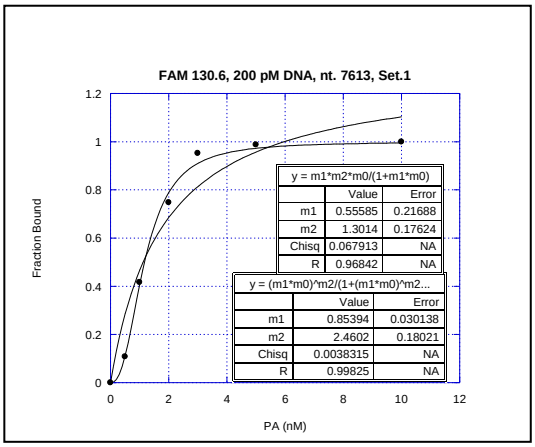




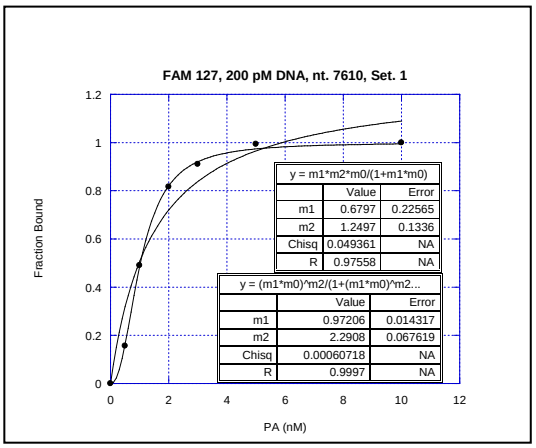
Site 7616

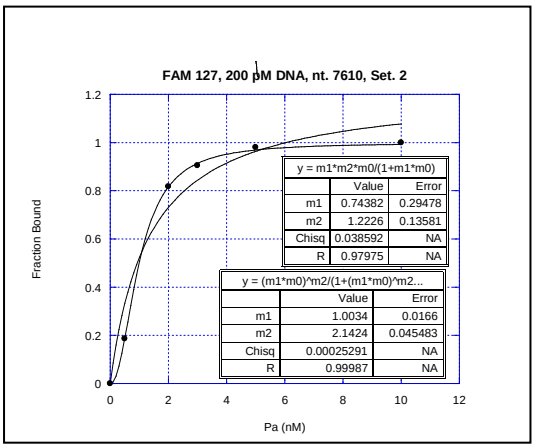


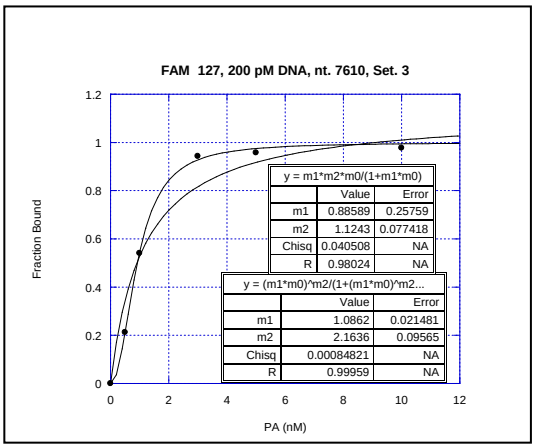




Site 7610







CHAPTER 4

SOLUBILITY AND CELL UPTAKE STUDIES OF GUAN- AND TMG-SUBSTITUTED PYRROLE-IMIDAZOLE POLYAMIDES

4.1 Abstract

Our group has previously reported that substitution at the polyamide N-terminus with a guanidinium (Guan) or tetramethylguanidiunium (TMG) group can positively affect antiviral activity against Human Papillomavirus (HPV). Since a guanidinium moiety has often been used to enhance cellular uptake and solubility of biomolecules, here we investigate the effect of this motif on polyamide solubility and cellular uptake in HPV-infected keratinocytes.

4.2 Introduction

Minor groove binders represent a class of agents with disparate potential applications, spanning from biotechnology to clinic (*e.g.* DNA detection, antitumor and antimicrobial activity, gene expression modulation, *etc.*).[1-4] In order to be functional, however, their ability to permeate the cell membrane and reach the nucleus to interact with the DNA are of primary importance. Compared to other minor groove binders, pyrrole-imidazole polyamides do not need a carrier to be delivered into cells, but instead have the peculiarity to easily access many live cells despite their high molecular weight (although the entrance mechanism has not yet been fully understood).

Uptake studies have been undertaken by other groups,[5-7] mainly to elucidate the relationship between polyamide structure and cell accessibility. The outcomes of these studies have shown that polyamide uptake is cell-line dependent and that the N-methylimidazole position and β-alanine content may affect cellular permeability.[8,9] While some studies reported a correlation between uptake and molecular size,[10] no dependency on polyamide length has been found in other studies.[5] Furthermore small polyamides can accumulate in cytoplasmic vesicles, a process dependent on polyamide charge.[6] Nevertheless an exhaustive structure-cell permeability correlation was not deduced and rather uptake appears to be very complex and not easily predictable.

The techniques employed so far to evaluate cell uptake include confocal microscopy, flow cytometry and cell-based reporter assays.[5,7] Fluorescence microscopy and flow cytometry require the conjugation of a dye, normally to the C-terminus of the

polyamide (or less frequently to other positions), which reportedly influences the mechanisms of transport and accumulation inside cells. Furthermore, labeling of large polyamides with a dye has been hindered by solubility issues (our research group, unpublished data). Cell-based reporter assays or PCR-based assays overcome these limitations but polyamide quantification is based on indirect measurements that may not correlate linearly with concentration.

Here we investigate the cellular uptake of large polyamides in live cells by means of High Pressure Liquid Chromatography (HPLC), which provides a direct approach to quantify intracellular polyamide concentration and does not require structure alteration of the analyte. Specifically, we aim to study the uptake of guanidine- (PA31) or TMG- (PA30) substituted polyamides (**Figure 4.1**) which have been previously reported to have greater anti-HPV activity compared to their unsubstituted analog (PA1).[11] Given the known use of the guanidinium moiety to facilitate the cellular delivery of peptides and drugs,[12,13] we aim to evaluate the effect of this substitution on polyamides cell permeability and solubility.

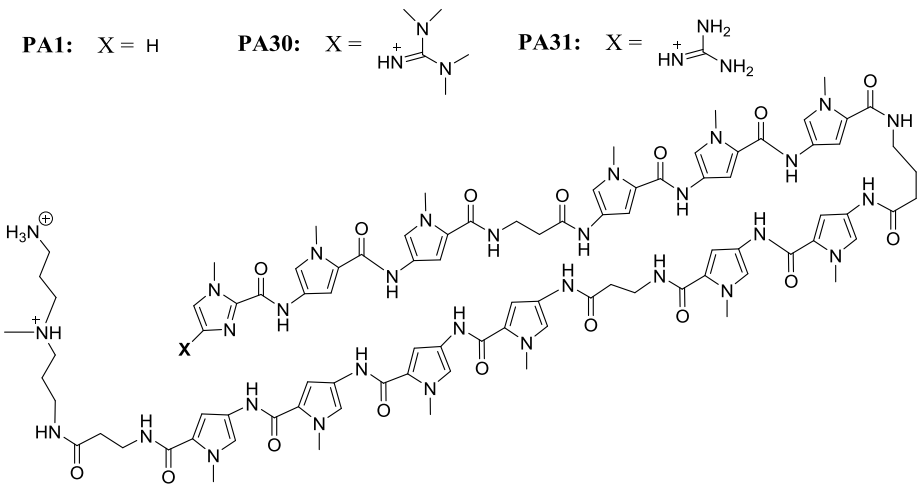


Figure 4.1. PA, PA30 and PA31 Chemical Structures[11]

4.3 Methods

4.3.1 Materials

Dimethyl sulfoxide (DMSO), acetonitrile (ACN) dimethylformamide (DMF), culture dishes, insulin, hydrocortisone, adenine, transferrin, penicillin-streptomycin, cholera toxin, 3′,5-triiodo-L-thyronine sodium salt, Hepes and Tris·HCl (pH 7.5) buffers were purchased from Sigma Aldrich (St. Louis, MO). Mitomycin-C, Epidermal Grow Factor, Dulbecco's modified Eagle medium, Ham's Nutrient Mixture F-12 (F12), Hoechst® 33342, trypsin, were obtained from Thermo Fisher Scientific (Waltham, MA). Phosphate Buffer Saline (PBS) was purchased from GE-Healthcare Bio-Sciences (Pittsburgh, PA).

4.3.2 Cell Culture

J2-3T3 fibroblasts and W12E keratinocytes were a kind gift of the Paul Lambert lab at The University of Wisconsin, Madison Medical School.

The embryonic mouse fibroblasts J2-3T3, used as feeder cells for the keratinocytes, were maintained in Dulbecco's modified Eagle medium (DMEM), supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. Cells were incubated at 37 ˚C in a 5% CO_2 incubator and passaged at 70-80% confluence.

W12E cells, cervical keratinocytes harboring HPV16 episomes (~ 300-500 copies/cells), were cultured in Dulbecco's modified Eagle medium (DMEM)/ Ham's Nutrient Mixture F-12 (F12) supplemented with 5% fetal bovine serum (FBS), 5 µg/mL insulin, 100 U/mL penicillin, 100 µg/mL streptomycin, 0.4 µg/mL hydrocortisone, 10 ng/mL cholera toxin, 20 pM 3,3',5-triiodo-thyronine (T3), 5 µg/mL transferrin, 24 µg/mL adenine and 5 ng/mL epidermal grow factor (EGF). Cells were cultured on top of mitomycin-C (Mit-C) treated 3T3 cells and grown at 37 ˚C in a 5% CO_2 incubator (**Figure 4.2**). Cells used for 24 h uptake experiments were from passage 14 through 16.

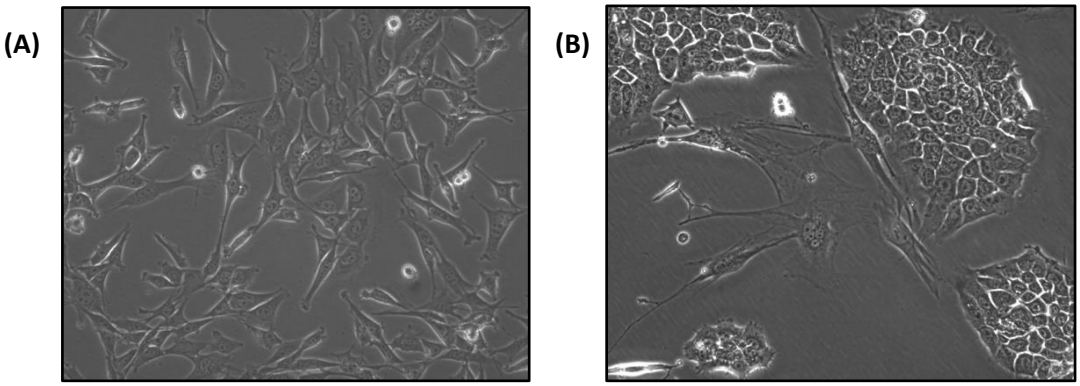


Figure 4.2. (A) 3T3 fibroblasts and (B) W12E keratinocytes cell cultures

4.3.3 Uptake Experiments

W12E cells were seeded onto 6-well culture dishes at a density of ~ $5x10^4$ cells/well on top of Myt-C treated fibroblasts and grown until reaching ~ 50% confluence. Lyophilized polyamide was dissolved in DMSO at a 4 mM concentration and further diluted to 0.4 mM in MQH_2O. Cells were incubated with polyamide diluted in fresh cell medium at a final concentration of 2.5 µM and 0.1% DMSO and incubated for 24 h. Following medium removal, cells were washed twice with phosphate buffer saline (PBS) and once with 0.02 % EDTA/PBS for 30 seconds. After 10 minutes incubation with 0.02 % EDTA/PBS at 37 ˚C to remove 3T3 fibroblasts, W12E were harvested by trypsinization and counted using a Nexcelom Bioscience Cellometer Auto T4 instrument. Cells were collected by centrifugation at 500 g for 5 min and polyamides extracted by treating the cell pellet with DMF. The solution was incubated for 20 min with frequent vortexing at high speed. The sample was then centrifuged at 16000 g for 20 min and the supernatant filtered with a 0.2 µM polytetrafluoroethylene (PTFE) filter and analyzed by HPLC.

All experiments were performed in triplicate and PA concentration determined based on comparison to standard curves. Total cellular uptake was normalized to cell number.

4.3.4 Active and Passive Uptake Studies

Active and passive uptake experiments were performed by incubating W12E cells with polyamides at different temperatures for 2 h. Cells were cultured in cell medium

supplemented with 20 mM Hepes at 4 ˚C to evaluate passive uptake or at 37 ˚C to evaluate both passive an active transport. After 2 h, the cells were treated as described in section 4.3.3.

4.3.5 Hoechst Treatment

A 1:2000 Hoechst® 33342 solution in cell medium was prepared by adding 1 µL of Hoechst (16 mM) to 1999 µL cell medium. This solution was added to the W12E cells and incubated for 30 minutes. Medium was removed and saved for analysis while the cells were treated as described in section 4.3.3 to extract and quantify the dye. A portion of the cell medium (200 µL) was mixed with 200 µL of ACN to precipitate proteins, centrifuged at 16000 g for 20 min and the supernatant analyzed by HPLC. A representative chromatogram showing the Hoechst dye peak (2.14 min) is reported in **Figure 4.3**. Both the medium and the cell dye concentrations were compared to standards made in the same conditions.

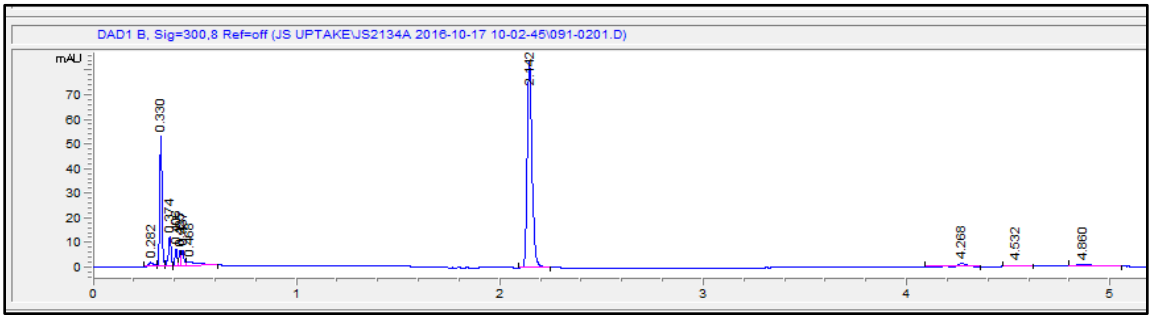


Figure 4.4. HPLC trace of Hoechst in W12E cells.

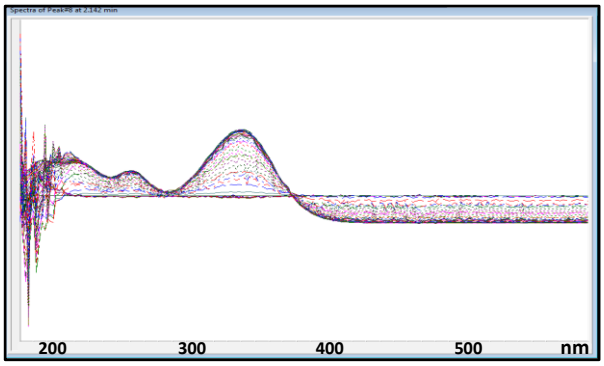


Figure 3.4. HPLC-DAD data for peak at 2.142 min; absorbance at 350 nm corresponds to the Hoechst maximum absorbance

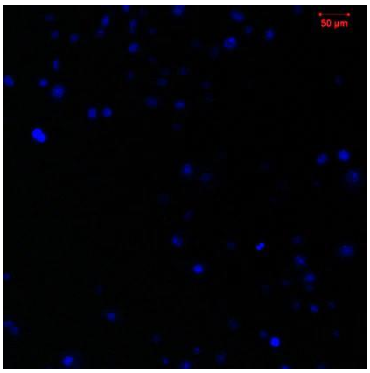
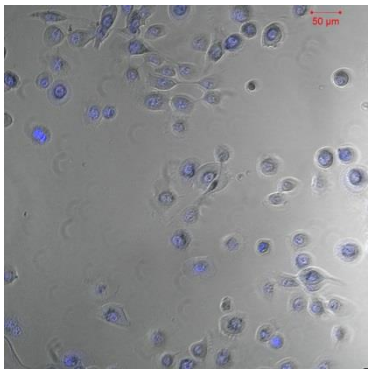
(A)  (B) 

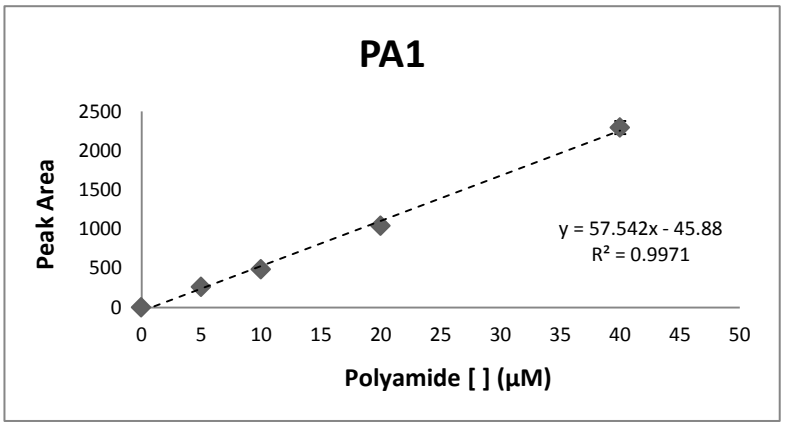
Figure 4.5. W12E nuclei stained with Hoechst® 33342 (A) Hoechst fluorescence; (B) overlap of bright and fluorescence field images.

4.3.6 HPLC Conditions and Calibration Curves

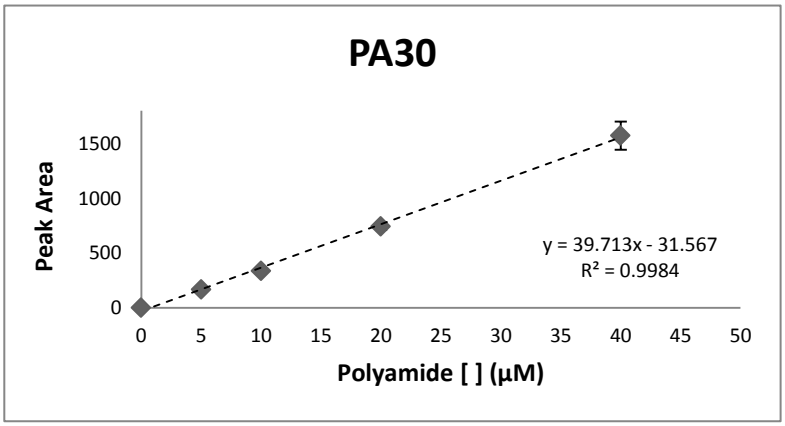
Chromatography was performed using a Jupiter Proteo (C12) column (90 Å pore size, 4 µm particle size, 4.6 x 50 mm, Phenomenex, Torrance, CA, USA) at 40 ˚C. Mobile phase consisted of 100 % acetonitrile (Solvent B) and 0.1% formic acid in water (Solvent A) and the gradient was as follow: 5% B (0-3 min, 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). Flow rate was maintained at 2.0 mL/min and total run was 6.5 min. Detection wavelength was 300.8 nm and injection volume was 20 µL. Data were processed with Agilent ChemStation software.

Calibration curve samples (**Figure 4.6**) were prepared in DMF after finding that measurement of polyamide concentration in DMF alone or in a W12E cell lysate was comparable by HPLC (to within a tolerance of 5%). Standards were made at 5.00, 10.0, 20.0, 40.0 µM polyamide concentration using DMSO stocks at 1.00, 2.50, 4.00 mM.

(A)



(B)



(C)

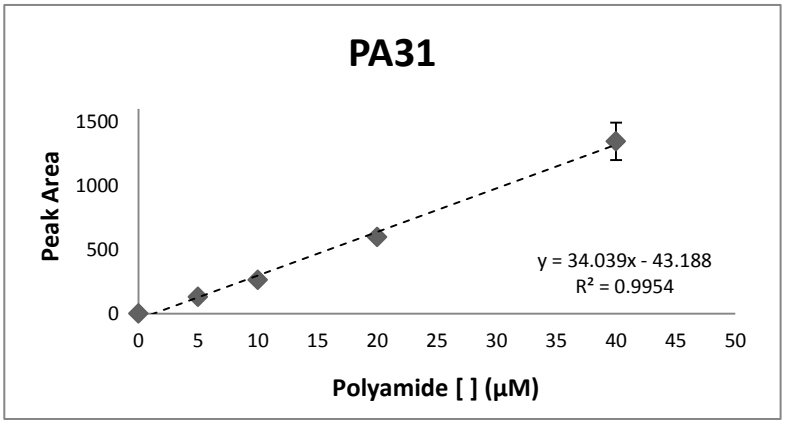


Figure 4.6. HPLC calibration standards for (A) PA1, (B) PA30 and (C) PA31. Each calibration was done in triplicate.

4.3.7 Solubility Assay

Polyamide stock solutions (4 mM) were prepared in DMSO using single-use 0.1 mg lyophilized aliquots. Work solutions were made by adding 0.5 µL of polyamide to 500 µL of the desired solvent. Samples were analyzed before and after centrifugation (16000 g, 20 min) by measuring the absorbance at 304 nm with an UV-Vis spectrometry (Thermo Scientific Evolution 260 BIO UV spectrophotometer). Molar extinction coefficients (ε) were 88,235 $M^{-1}cm^{-1}$ for PA1, 91,557 $M^{-1}cm^{-1}$ for PA30 and 89,263 $M^{-1}cm^{-1}$ for PA31.[11]

4.4 Results and Discussion

4.4.1 Solubility Studies

The guanidine group has often been employed to improve drugs and peptides solubility[12,14,15] especially in aqueous environment; therefore we investigated the solubility properties of these compounds in physiologically relevant solvents. The solubility was determined by measuring the concentration of each polyamide at a final concentration of 4 µM in H_2O, Phosphate buffer saline (PBS) and Tris·HCl buffer by UV-Vis spectroscopy. Absorbance was measured before and after a centrifugation step intended to remove aggregates.

In H_2O, all three polyamides (PAs) exhibit high solubility, but incorporation of a Guan group leads to a greater solubility than for TMG or the parent PA. Upon centrifugation, the concentrations of PA1 and PA31 remain more or less unchanged,

indicating that not many aggregates are formed, while for PA30 a slightly decrease in concentration was observed after centrifugation. Overall, the guanidinylated PA shows a better solubility in H_2O than the other two PA investigated. In PBS, the solubility was somehow lower than in water for all PAs and further decreased upon centrifugation, showing that probably some of the compound may be in the aggregated form. Interestingly, PA30 shows the highest solubility. A similar result was observed in Tris, where the TMG displays a greater solubility before centrifugation, and especially after centrifugation.

Pyrrole-imidazole polyamides solubility studies have been also undertaken at Caltech by Peter Dervan[16] with the goal of investigating the solubility of several short polyamides in PBS. The outcomes of these studies have shown that although chemical modifications such acetylation could slightly increase polyamide solubility, overall none of the polyamide analyzed was fully soluble in PBS, reaching a concentration below 50 % of the maximum estimated concentration. An improvement in polyamide solubility was achieved only upon addition of formulating agents such as cyclodextrins and other carbohydrate-based solubilizing agents. Similarly, the Sugiyama's lab has reported the poor solubility of polyamides in Milli-Q H_2O and a solubility improvement afforded by polyamide PEGylation or liposome incorporation.[17,18]

In summary, here we show that all the analyzed polyamides show good solubility properties when diluted in Milli–Q water. In particular, the guanidinylated substitution improves the solubility compared to the other analogs. In buffers, all the three

compounds show lower solubility than in H_2O upon centrifugation; surprisingly the TMG analog shows the highest solubility.

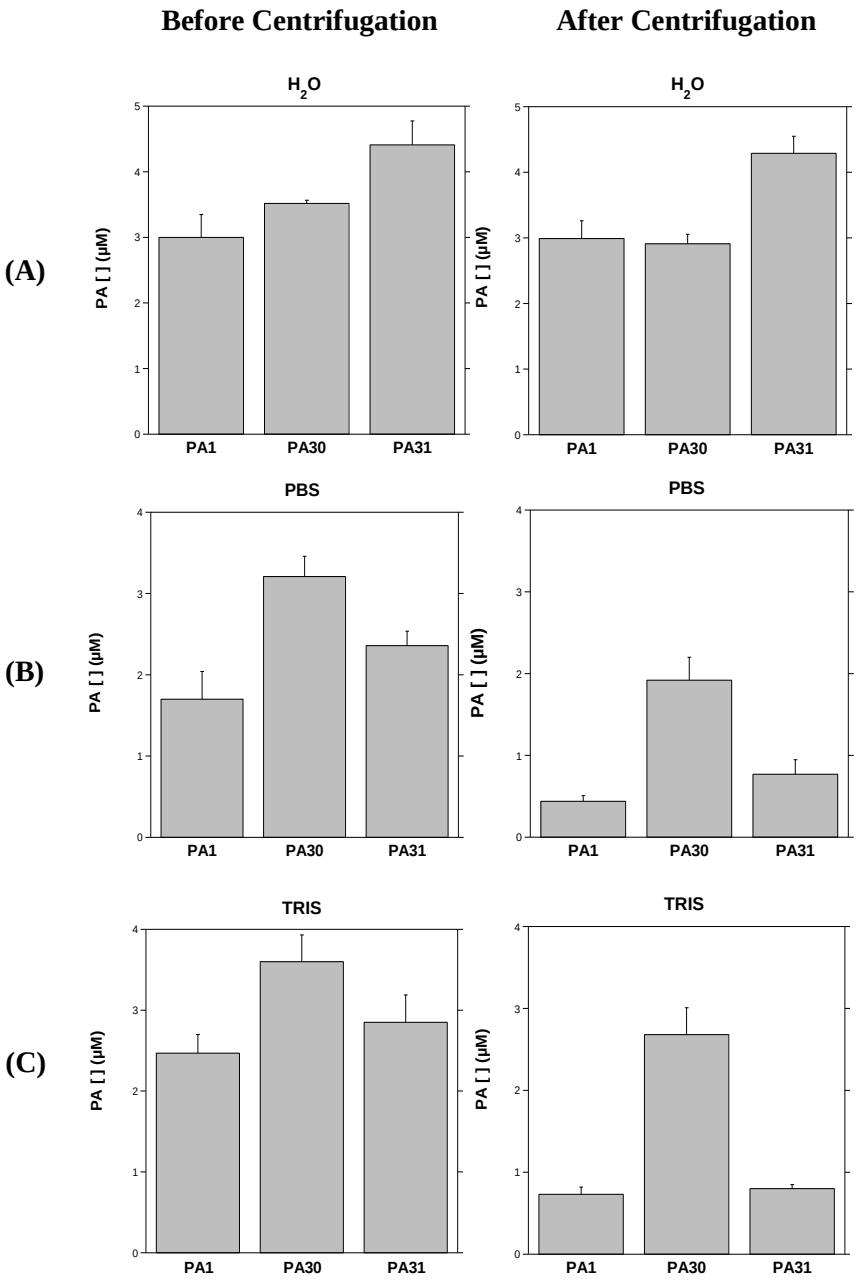


Figure 4.7 Polyamides solubility in (A) H_2O, (B) PBS and (C) Tris

4.4.2 Cellular uptake in W12 keratinocytes

Polyamide uptake was examined in W12E cells, cervical keratinocytes harboring the HPV genome in an extra-chromosomal form. Antiviral assays performed in these cells have previously shown that PA1, PA30 and PA31 are able to reduce the viral DNA load, with IC_{50} (polyamide concentration required to reduce viral DNA to 50 %) and IC_{90} (polyamide concentration required to eliminate 90 % of viral DNA) values in the nanomolar range.[11] Although all three PAs show good antiviral properties, the TMG and especially the Guan analog exhibit improved antiviral activity. Therefore we investigated how the substitution with a Guan or TMG group affects PA cellular uptake, using High Pressure Liquid Chromatography (HPLC). This study was followed up by HPLC/MS and HPLC/MS/MS studies as described in the following Chapter.

We first employed this method to detect the uptake of Hoechst® 33342 in W12E, a compound known to permeate cell membranes easily.[19,20] After 30 minutes incubation, the dye is already detected in the nucleus as shown by confocal microscopy in **Figure 4.5.** The response determined in both cell culture media and cell pellets was compared to standards prepared under the same conditions and reported in **Table 4.1**.

The amount of dye used to treat the cells was equal to 7.22 µg and the amount of dye found was 7.42 µg.

	STANDARD	SAMPLE
Peak Area in Media	34.37	27.81
Peak Area in Cells	393.0	85.93
% Recovered	102.8 %	

Table 4.1. Hoechst amount in cell pellet and medium

Next, W12E cells were incubated with PAs at 2.5 μM for 24 h and the concentration of cell-associated PA was determined by HPLC by comparing the response with that obtained from standard curves. The results were then normalized to cell number. Polyamide was not detected in untreated cells or after 30 minutes of treatment (at least for PA1), but it was detected after 2 h. As shown in **Figure 4.9**, the amount of polyamide associated with cells after 24 h incubation was greater for PA31 and PA30 than PA1. The uptake increase from 2 hours to 24 h for all three PAs is shown in **Figure 4.10.** For PA1, the amount accumulated at 2 h was already comparable to that at 24 h.

We also analyzed the uptake of a small polyamide, PA11, which was previously reported to be inactive and unable to decrease levels of HPV-episomes from infected cells.[21] The chemical structure is showed in **Figure 4.8.**

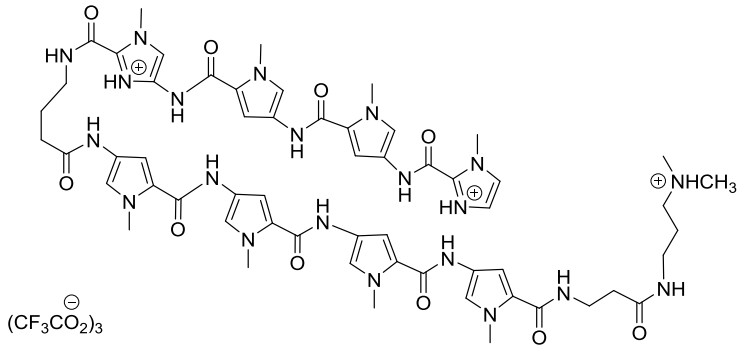


Figure 4.8. PA11 Chemical Structure

Surprisingly, almost no polyamide was detected in cells (less than 1 %), and instead, the bulk of it remained in the medium after 24 h. Pre-treatment of W12E cells with Verapamil (30, 100, 200 μM), a multi drug resistant inhibitor, did not have any

effect on polyamide accessibility. Only a treatment at 500 µM, a concentration that was toxic to the cells, leaded to PA concentration increase.

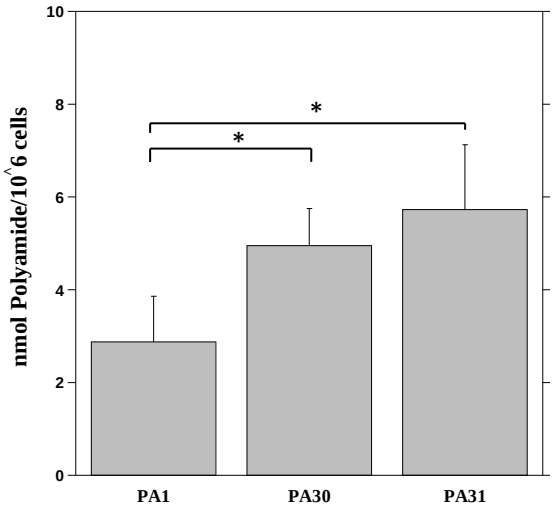


Figure 4.9. Uptake of PA1, PA30 and PA31 in W12E cells after 24 h (P < 0.05)

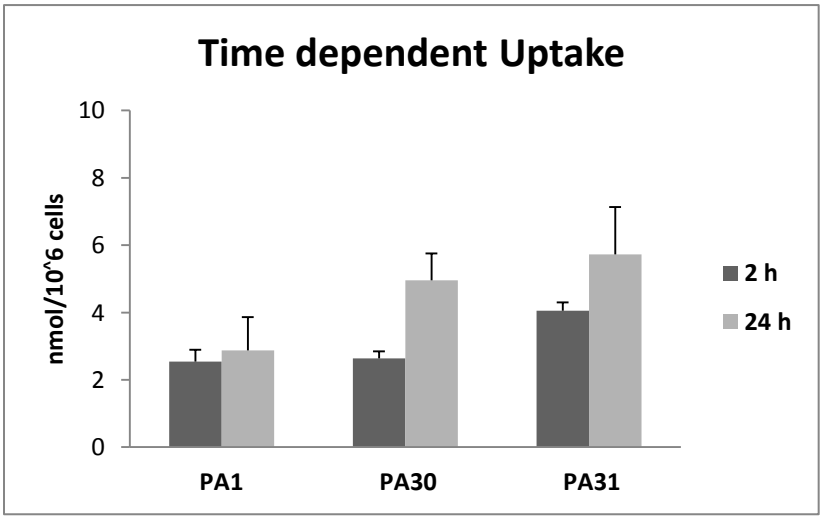


Figure 4.10. Uptake in W12E at 2h and 24h

4.4.3 Characterization of passive and active transport

Next, we wanted to evaluate if the tested polyamides are transported inside the cells by passive uptake or by an energy-dependent process. Previous studies suggest that uptake may be energy dependent,[22] but the general mode of uptake remains elusive. Although both active and passive pathways are affected by temperature variations, active transport is reduced more significantly than passive transport at temperatures below 22 ˚C and blocked at 4 ˚C.[23,24]

We investigated PA uptake in keratinocytes after 2 h incubation at 37 ˚C and 4 ˚C. **Figure 10** shows that for all three PAs a more or less pronounced decrease in uptake is observed at 4 ˚C relative to 37 ˚C. The major change was observed for PA30 with a ~ 81 % decrease, followed by PA31 and PA1. These data are in line with few reports indicating that an energy-dependent transport may be the primary mechanism of polyamide uptake[22]. It would be interesting to evaluate the effect of endocytosis inhibitors on polyamide uptake to confirm these results, since also passive uptake may be affected by temperature changes, although to a lesser extent.

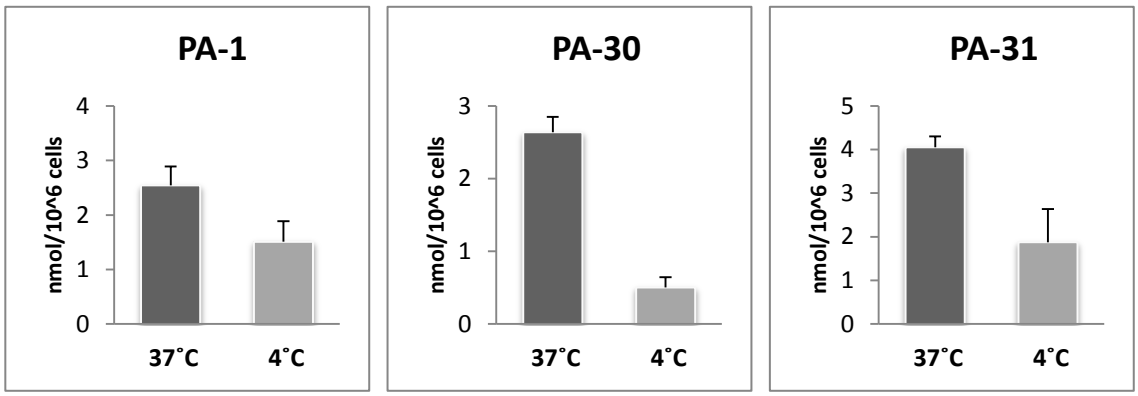


Figure 4.11. PA1 (A), PA30 (B) and PA31 (C) uptake in W12E at 37 ˚C and 4 ˚C after 2 h

4.5 Conclusions

This study aimed at investigating the uptake and solubility properties of a series of polyamides known to have antiviral properties but differing in their potency.

Overall the three analyzed polyamides show noteworthy high solubility in H_2O without the aid of solubilizing agents, with the TMG and Guan analogs displaying a greater solubility than the unsubstituted analog. In buffers, a more pronounced aggregation takes place as expected, and the TMG analog displays the better solubility.

Uptake studies show that PAs are detected in cultured cells after about 2 h and a greater amount of PA seems to be associated with the cells in the case of the TMG and Guan analogs after 24 h incubation. The transport seems to be energy-dependent since a remarkable decrease in uptake was observed for the Guan and TMG PAs, and a less extensive decrease for PA1. Interestingly, negligible uptake was observed for PA11, a small inactive polyamide. It would be of interest in the future to extend the uptake studies on other small inactive PAs and evaluate nuclear drug localization as well as the effect of

endocytosis blocking agents on uptake to gain insights on the polyamide mechanism of transport.

Safety: Because the experiments described in this proposal required handling cell lines containing HPV, all the laboratory procedures were conducted according to the Biosafety Level 2 (BSL-2) practices under the approval of the UMSL Biosafety Committee and the Environmental Health & Safety (EHS) Department.

4.6 References

(1) Nanjunda, R.; Wilson, W. D. *Curr Protoc Nucleic Acid Chem* **2012**, *Chapter 8*, Unit8 8.

(2) Mazur, S.; Tanious, F. A.; Ding, D.; Kumar, A.; Boykin, D. W.; Simpson, I. J.; Neidle, S.; Wilson, W. D. *Journal of Molecular Biology* **2000**, *300*, 321.

(3) Dervan, P. B.; Doss, R. M.; Marques, M. A. *Current Medicinal Chemistry-Anti-Cancer Agents* **2005**, *5*, 373.

(4) Mdachi, R. E.; Thuita, J. K.; Kagira, J. M.; Ngotho, J. M.; Murilla, G. A.; Ndung'u, J. M.; Tidwell, R. R.; Hall, J. E.; Brun, R. *Antimicrob Agents Chemother* **2009**, *53*, 953.

(5) Liu, B.; Kodadek, T. *Journal of medicinal chemistry* **2009**, *52*, 4604.

(6) Crowley, K. S.; Phillion, D. P.; Woodard, S. S.; Schweitzer, B. A.; Singh, M.; Shabany, H.; Burnette, B.; Hippenmeyer, P.; Heitmeier, M.; Bashkin, J. K. *Bioorganic & Medicinal Chemistry Letters* **2003**, *13*, 1565.

(7) Belitsky, J. M.; Leslie, S. J.; Arora, P. S.; Beerman, T. A.; Dervan, P. B. *Bioorg Med Chem* **2002**, *10*, 3313.

(8) Best, T. P.; Edelson, B. S.; Nickols, N. G.; Dervan, P. B. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 12063.

(9) Edelson, B. S.; Best, T. P.; Olenyuk, B.; Nickols, N. G.; Doss, R. M.; Foister, S.; Heckel, A.; Dervan, P. B. *Nucleic Acids Research* **2004**, *32*, 2802.

(10) Nishijima, S.; Shinohara, K.; Bando, T.; Minoshima, M.; Kashiwazaki, G.; Sugiyama, H. *Bioorg Med Chem* **2010**, *18*, 978.

(11) Castaneda, C. H.; Scuderi, M. J.; Edwards, T. G.; Harris Jr, G. D.; Dupureur, C. M.; Koeller, K. J.; Fisher, C.; Bashkin, J. K. *MedChemComm* **2016**, *7*, 2076.

(12) Wexselblatt, E.; Esko, J. D.; Tor, Y. *The Journal of Organic Chemistry* **2014**, *79*, 6766.

(13) Vazquez, O.; Blanco-Canosa, J. B.; Vazquez, M. E.; Martinez-Costas, J.; Castedo, L.; Mascarenas, J. L. *Chembiochem* **2008**, *9*, 2822.

(14) Hirano, A.; Arakawa, T.; Shiraki, K. *The Journal of Biochemistry* **2008**, *144*, 363.

(15) Li, J.; Garg, M.; Shah, D.; Rajagopalan, R. *The Journal of Chemical Physics* **2010**, *133*, 054902.

(16) Hargrove, A. E.; Raskatov, J. A.; Meier, J. L.; Montgomery, D. C.; Dervan, P. B. *Journal of medicinal chemistry* **2012**, *55*, 5425.

(17) Takagaki, T.; Bando, T.; Kitano, M.; Hashiya, K.; Kashiwazaki, G.; Sugiyama, H. *Bioorg Med Chem* **2011**, *19*, 5896.

(18) Kashiwazaki, G.; Bando, T.; Yoshidome, T.; Masui, S.; Takagaki, T.; Hashiya, K.; Pandian, G. N.; Yasuoka, J.; Akiyoshi, K.; Sugiyama, H. *Journal of medicinal chemistry* **2012**, *55*, 2057.

(19) Lalande, M. E.; Ling, V.; Miller, R. G. *Proceedings of the National Academy of Sciences of the United States of America* **1981**, *78*, 363.

(20) Parish, C. R. *Immunol Cell Biol* **1999**, *77*, 499.

(21) Edwards, T. G.; Vidmar, T. J.; Koeller, K.; Bashkin, J. K.; Fisher, C. *PLoS One* **2013**, *8*, e75406.

(22) Nickols, N. G.; Jacobs, C. S.; Farkas, M. E.; Dervan, P. B. *Nucleic Acids Research* **2007**, *35*, 363.

(23) Poirier, A.; Lave, T.; Portmann, R.; Brun, M. E.; Senner, F.; Kansy, M.; Grimm, H. P.; Funk, C. *Drug Metab Dispos* **2008**, *36*, 2434.

(24) Sugano, K.; Kansy, M.; Artursson, P.; Avdeef, A.; Bendels, S.; Di, L.; Ecker, G. F.; Faller, B.; Fischer, H.; Gerebtzoff, G.; Lennernaes, H.; Senner, F. *Nat Rev Drug Discov* **2010**, *9*, 597.

CHAPTER 5

LC-MS/MS UPTAKE AND SOLUBILITY STUDIES OF TWO DIFFERENT PYRROLE-IMIDAZOLE POLYAMIDE SALTS

5.1 Abstract

A Liquid Chromatography - Tandem Mass Spectrometry (LC-MS/MS) method was developed to determine pyrrole-imidazole polyamide uptake in cells cultures. This method was used to evaluate the cellular internalization in W12E keratinocytes of PA1 and $PA1_F$, two polyamides sharing the same chemical structure but paired with different counterions. The substitution results in a difference between the two polyamides in terms of antiviral efficacy against three HPV strains, with $PA1_F$ showing higher efficacy. Solubility studies were also undertaken to determine the effect of the different counterion.

5.2 Introduction

Antiviral assays in cervical keratinocytes (W12E) have shown that large pyrrole-imidazole polyamides, chemical agents used mostly to modulate gene expression, can elicit the elimination of viral DNA from infected cells and therefore have the potential to be used as antiviral agents.[1-3] Their activity is manifested against different viruses,[4] including Human Papillomavirus (HPV).

PA1, a twelve ring compound paired with a trifluoroacetate (TFA) counterion, is one of the first compounds reported by our group to show antiviral efficacy.[1] A simple counterion substitution, from a TFA to the formate anion, leads to $PA1_F$, which exhibits higher potency against three oncogenic HPV strains (unpublished data from T.G. Edwards, C. Fisher, NanoVir, LLC). The chemical structures of these compounds are shown in **Figure 5.1**. The concentrations of polyamides that induce 50% (IC_{50}) or 90% (IC_{90}) decrease in viral DNA are reported in **Table 5.1** for HPV16 and show that both values are more than three-fold lower for the formic analog. A better activity was also observed for other two oncogenic HPV strains (data not shown). Pairing a drug with different counterions is a common strategy in preformulation studies to improve the physicochemical or biopharmaceutical properties of the parent compound.[5,6] Very often, different salt forms are used specifically to overcome solubility issues in aqueous environments while avoiding major changes in structure.[7]

Here, we investigate the effect of the counterion on aqueous solubility and on cellular uptake in keratinocyte monolayers. With this aim we developed a novel Liquid

Chromatography - Tandem Mass Spectrometry method to quantitate the concentration of polyamides in W12E cells. This method provides the advantage of keeping the polyamide structure unaltered, while other methods previously used to study polyamide cellular uptake such as confocal microscopy or flow cytometry require labeling, which may alter the compound permeability. Furthermore, the coupling of mass spectrometry with chromatography provides the advantage of greatly increased sensitivity *vs.* HPLC alone, as well as high specificity and confidence in the analysis using a minimal amount of compound.[8,9] Solubility studies were performed using UV-Vis Spectroscopy as before (Chapter 4).

Polyamide	HPV16 IC_{50} [μM]	HPV16 IC_{90} [μM]
PA1	0.100 ± 0.015	1.113 ± 0.627
$PA1_F$	0.026 ± 0.004	0.299 ± 0.082

Table 5.1. Antiviral activity of PA1 and $PA1_F$ against HPV

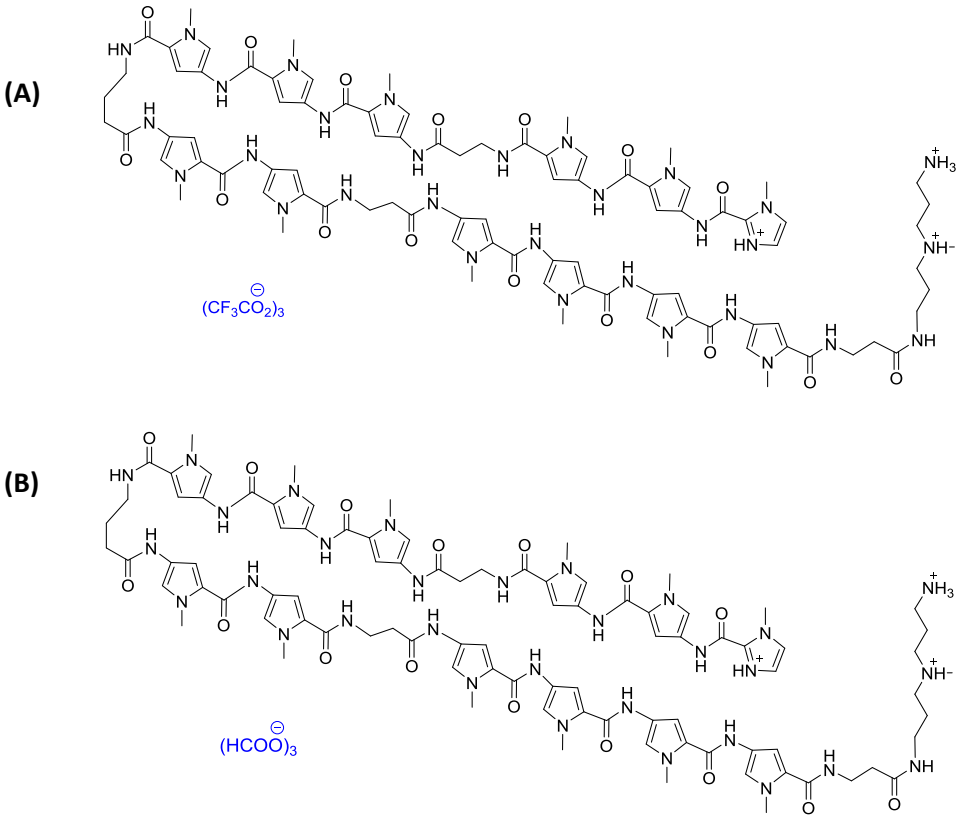


Figure 5.1. Polyamides (A) PA1 and (B) $PA1_F$ chemical structures

5.3 Methods

5.3.1 Materials

Dimethyl sulfoxide (DMSO), acetonitrile (ACN), dimethylformamide (DMF), HPLC grade water, culture dishes, insulin, hydrocortisone, adenine, transferrin, penicillin-streptomycin, cholera toxin, 3′,5-triiodo-L-thyronine sodium salt, HEPES, and Tris·HCl buffers were purchased from Sigma Aldrich (St. Louis, MO). Mitomycin-C, Epidermal Growth Factor, Dulbecco's modified Eagle medium, Ham's Nutrient Mixture

F-12 (F12), and trypsin, were obtained from Thermo Fisher Scientific (Waltham, MA). Phosphate Buffer Saline (PBS) was purchased from GE-Healthcare Bio-Sciences (Pittsburgh, PA).

5.3.2 Cell Culture

J2-3T3 fibroblasts and W12E keratinocytes were a kind gift of Paul Lambert lab at The University of Wisconsin, Madison. The embryonic mouse fibroblasts J2-3T3, used as feeder cells for the keratinocytes, were maintained in Dulbecco's modified Eagle medium (DMEM), supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. Cells were incubated at 37 ˚C in a 5% CO_2 incubator and passaged at 70% confluence.

W12E cells, cervical keratinocytes harboring HPV16 episomes (~ 300-500 copies/cells), were cultured in Dulbecco's modified Eagle medium DMEM/F12 supplemented with 5% fetal bovine serum (FBS), 5 µg/mL insulin, 100 U/mL penicillin, 100 µg/mL streptomycin, 0.4 µg/mL hydrocortisone, 10 ng/mL cholera toxin, 20 pM 3,3',5-triiodo-thyronine (T3), 5 µg/mL transferrin, 24 µg/mL adenine and 5 ng/mL epidermal grow factor (EGF). Cells were cultured on top of mitomycin-C treated 3T3 cells and grown at 37 ˚C in a 5% CO_2 incubator. Cells used for 24 h uptake experiments were from passage 10 throught 14.

5.3.3 Uptake Experiment and Sample Preparation

W12E cells were seeded onto 6-well culture dishes at a density of ~$5x10^4$ cells/well on top of Myt-C treated fibroblasts and grown until reaching ~50% confluence. Lyophilized polyamide was dissolved in DMSO at a 4 mM concentration and further diluted to 0.4 mM in Milli-Q H_2O. Cells were incubated with polyamide diluted in fresh cell medium at a final concentration of 2.5 µM and 0.1% DMSO and incubated for 24 h. Following medium removal, cells were washed twice with phosphate buffer saline (PBS) and once with 0.02 % EDTA/PBS for 30 seconds. After 10 minutes incubation with 0.02 % EDTA/PBS at 37 ˚C to remove 3T3 fibroblasts, W12E were harvested by trypsinization and counted using a Nexcelom Bioscience Cellometer Auto T4 instrument. Cells were collected by centrifugation at 500 g for 5 minutes and polyamides extracted by treating the cell pellet with a DMF solution containing the internal standard at a 5 µg/ml concentration. The mixture was incubated for 20 minutes and subjected to frequent vortexing at high speed. The sample was then centrifuged at 16000 g for 20 minutes and the supernatant filtered with a 0.2 µm polytetrafluoroethylene (PTFE) filter (Xpertek, Cobert Associates Lab Store, St. Louis) and analyzed by LC-MS/MS. All experiments were performed at least in triplicate.

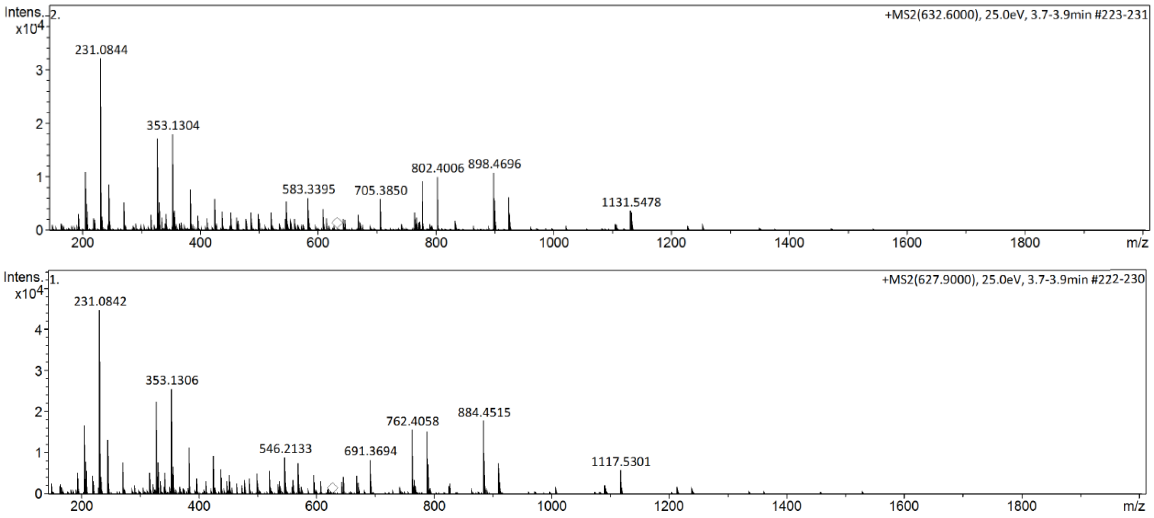
5.3.4 Chromatography and Mass Spectrometry Conditions

Chromatography was performed on a Dionex Ultima 3000 µHPLC system using a Zorbax Eclipse Plus C18 column (95 Å pore size, 1.8 µm particle size, 2.1 x 50 mm), Phenomenex, Torrance, CA, USA) maintained at 40 ˚C. Mobile phase consisted of 100 %

acetonitrile (Solvent B) and 0.1% formic acid in water (Solvent A) and the gradient consisted of 5% ACN followed by a ramp to 65 % ACN over 6 minutes.

Flow rate was maintained at 0.50 mL/min and total run time was 15 min. Masses were acquired on a Bruker MaXis plus quadrupole time-of-flight mass spectrometer using positive mode electro-spray ionization (ESI). The following settings were used for the analysis: the source temperature was 250 °C, capillary voltage was 3500 V, collision energy was 25 eV and N_2 gas flow was 12 L/min.

The mass spectrometer was operated in the multiple reaction monitoring mode (MRM) and quantitation was performed using the total ion abundance generated from the fragmentation of the $[M+3H]^{3+}$ parent ion. Data were analyzed using Bruker Daltonik Data Analysis software.



5.3.5 Stock Solution Preparation

Stock solutions of PA1 and $PA1_F$ (0.5 mg/ml) were prepared by dissolving a fresh lyophilized aliquot of polyamide (0.1 mg) in 200 µl DMF. The stock solution of Internal Standard (IS) PA57 (1 mg/ml) was prepared by dissolving a lyophilized 0.1 mg aliquot in 100 µL of DMSO. Analyte working solutions were made by further diluting each stock with DMF at the desired concentration.

5.3.6 Calibration Standards

Calibration curves were prepared fresh for each run by adding the appropriate working solution of analyte and internal standard to a W12E cell pellet previously treated with DMF, centrifuged at 16000 g for 20 minutes and filtered through a 0.2 µm PTFE filter (see section 5.3.3). Standards were made at concentrations of 0.5, 1, 5, 10, 17.5, 25, 35, 50 µg/ml analyte and 5 µg/ml IS. Quality control samples (QC) were made in the same conditions at 2 (low), 15 (medium), 45 (high) µg/ml concentrations.

5.3.7 Solubility Assay

Polyamide stock solutions (4 mM) were prepared in DMSO using single-use 0.1 mg lyophilized aliquots. Working solutions were made by adding 0.5 µL of polyamide to 500 µL of the desired solvent. Samples were analyzed before and after centrifugation (16000 g, 20 min) by measuring the absorbance at 304 nm with an UV-Vis spectrometry (Thermo Scientific Evolution 260 BIO UV spectrophotometer). The molar extinction coefficient (ε) used was 88,235 $M^{-1}cm^{-1}$.[3]

5.4 Results and Discussion

5.4.1 LC/MS/MS Analysis

For these studies, the internal Standard (IS) was the demethylated version of the analyte PA1, i.e. demethylated at both nitrogen atoms on the C-terminal (**Figure 5.3)**. Both compounds eluted at 3.9 min. Representative MS spectra are showed in **Figure 5.4** with the observed ions being $[M+2H]^{2+}$, $[M+3H]^{3+}$ and $[M+4H]^{4+}$. The triply charged ion $[M+3H]^{3+}$ was the dominant form, therefore was chosen as the precursor ion for fragmentation. $[M+3H]^{3+}$ ions were monitored at m/z 632.6 (PA1, $PA1_F$) and m/z 627.9 (IS), while product ions were monitored at m/z 898.4 for PA1 and m/z 884.4 for IS. Relative concentrations were determined using the peak area ratio of the analyte over the IS of the total ion abundance and the MRM chromatogram is shown in **Figure 5.5**.

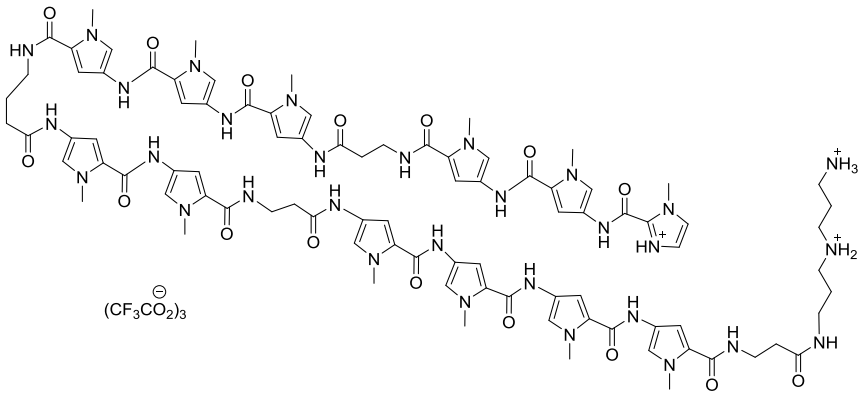


Figure 5.3. Internal Standard Chemical Structure

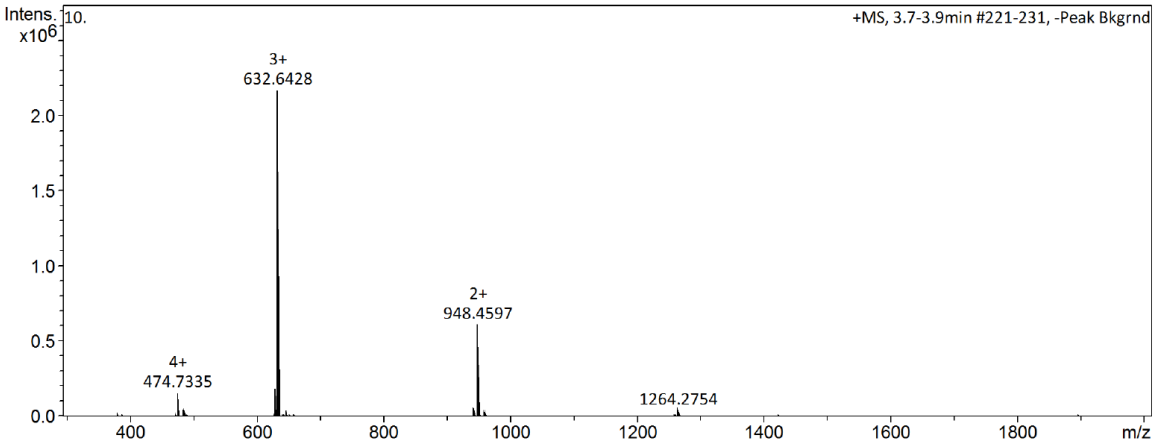


Figure 5.4 Representative MS spectra of PA1 parent ions.

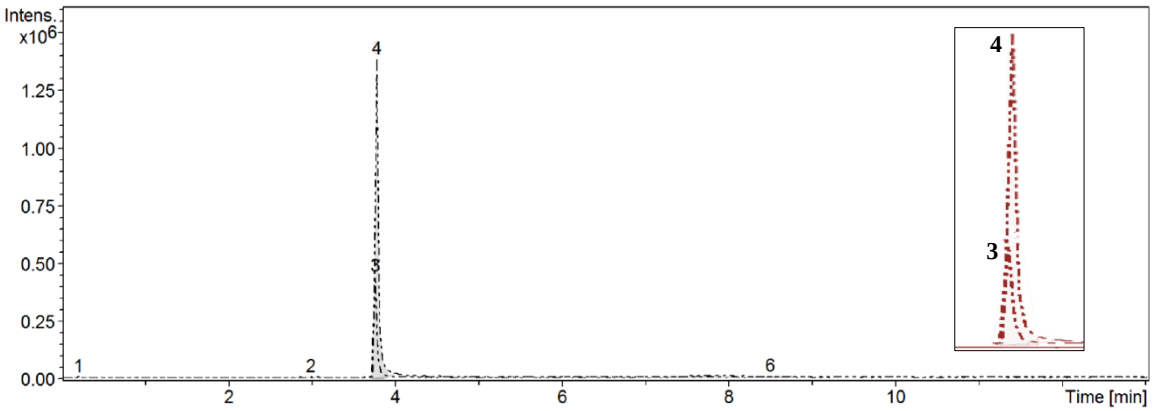


Figure 5.5 Representative MRM chromatogram of PA1 (Peak 4) and IS (Peak 3) from cell-extracted samples. Inset shows an expansion of peaks 3 and 4.

5.4.2 Selectivity

Selectivity was examined by comparing a blank cell culture matrix sample and matrix treated with PA1. As observed in **Figure 5.6** no peak interference was observed at the polyamide retention time.

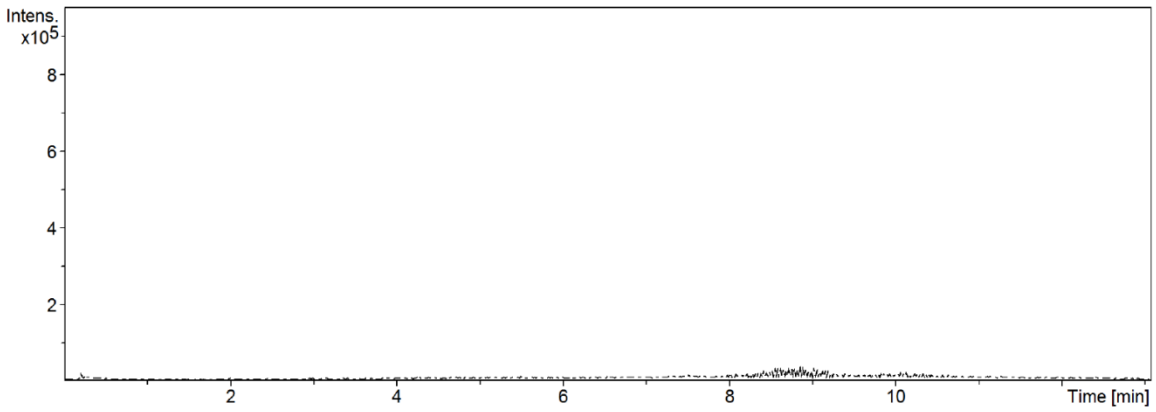
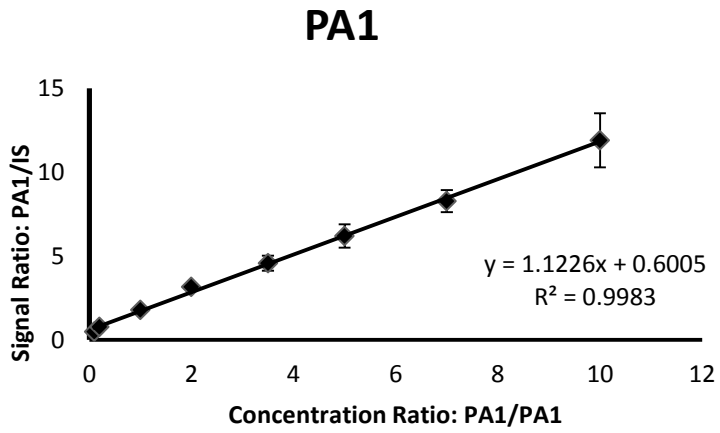


Figure 5.6. Blank sample (cell extracts without added polyamides). No peak interference was observed at the PA retention time.

5.4.3 Calibration Curves and Linearity

The eight point standard curves for PA1 and $PA1_F$ were generated by plotting peak area ratios against concentration ratios of analyte over internal standard. Curves showed a linear response over the concentration range of 0.5 to 50 µg/ml analyte and representative calibration curves are showed in **Figure 5.7.** R^2 values were 0.9983 and 0.9982 for PA1 and $PA1_F$ respectively.

(A)



(B)

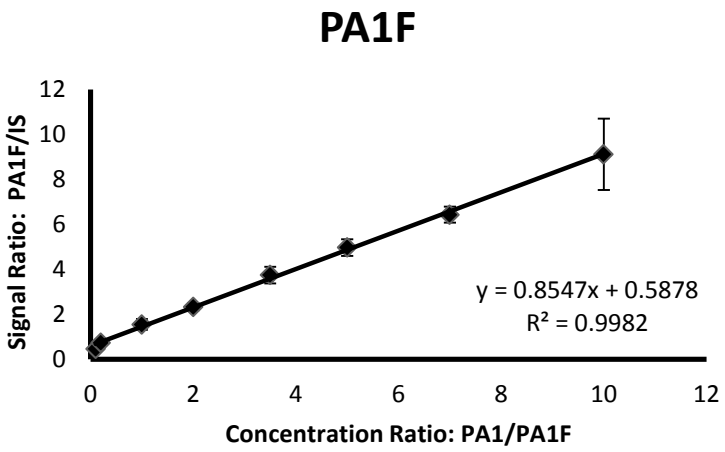


Figure 5.7. Representative calibration curves for (A) PA1 and (B) $PA1_F$ from triplicate experiments

5.4.4 Uptake Studies

The LC-MS/MS method was employed to study the uptake of two long pyrrole-imidazole antiviral polyamides, PA1 and $PA1_F$. The only structural difference between the two compounds is their counterion: PA1 was produced as a TFA salt while $PA1_F$ as a formate salt. $PA1_F$ was found to be three-fold more active than PA1 against HPV16 in antiviral assays. Therefore, the effect of counterion on uptake was evaluated in keratinocytes infected with HPV16 episomes. The percentages of polyamide associated with cells after 24 h were 51% (PA1) and 87% ($PA1_F$) of the total amount used to treat the cells (55929 ng/ml and 50829 ng/ml) (**Figure 5.8**). In order to evaluate the relationship between uptake and PA concentration, cells were also incubated with different PA concentrations (0.5 µM, 6.5 µM) and, as expected, an increase in concentration is associated with an increase in polyamide uptake (**Figure 5.9**).

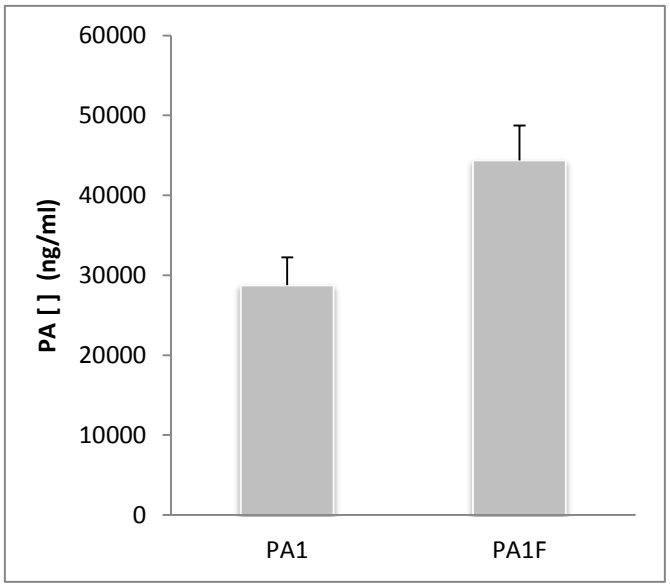


Figure 5.8. PA1 and PA1F uptake in W12E cells after 24 h

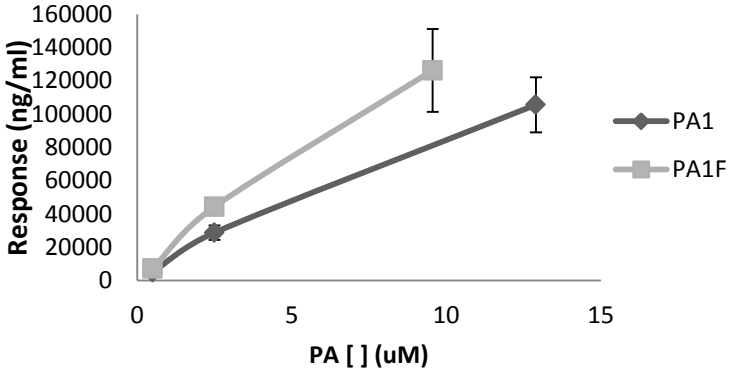


Figure 5.9. Concentration Dependent Uptake

5.4.5 Solubility Studies

Since the two polyamides differ only for their counterion, their solubility was an obvious property to investigate. The solubility was determined by measuring the concentration of each polyamide at a final concentration of 4 µM in H_2O, phosphate buffer saline (PBS) and Tris·HCl buffer by UV-Vis spectroscopy (**Figure 5.10**). Absorbance was measured before and after a centrifugation step intended to remove aggregates.

In H_2O, although both polyamides exhibited a good solubility, the formic salt analog showed a greater solubility after centrifugation (4.26±0.13 µM for $PA1_F$, 2.99±0.27 µM for PA1). In PBS (pH = 7.4) and Tris (pH = 7.5) buffers, the overall solubility of both compounds was reduced especially upon centrifugation, probably due to aggregation. While in PBS there was no appreciable difference between the two PAs, in Tris the formic salt analog showed a better solubility (1.73±0.39 µM for $PA1_F$ and 0.73±0.09 µM for PA1).

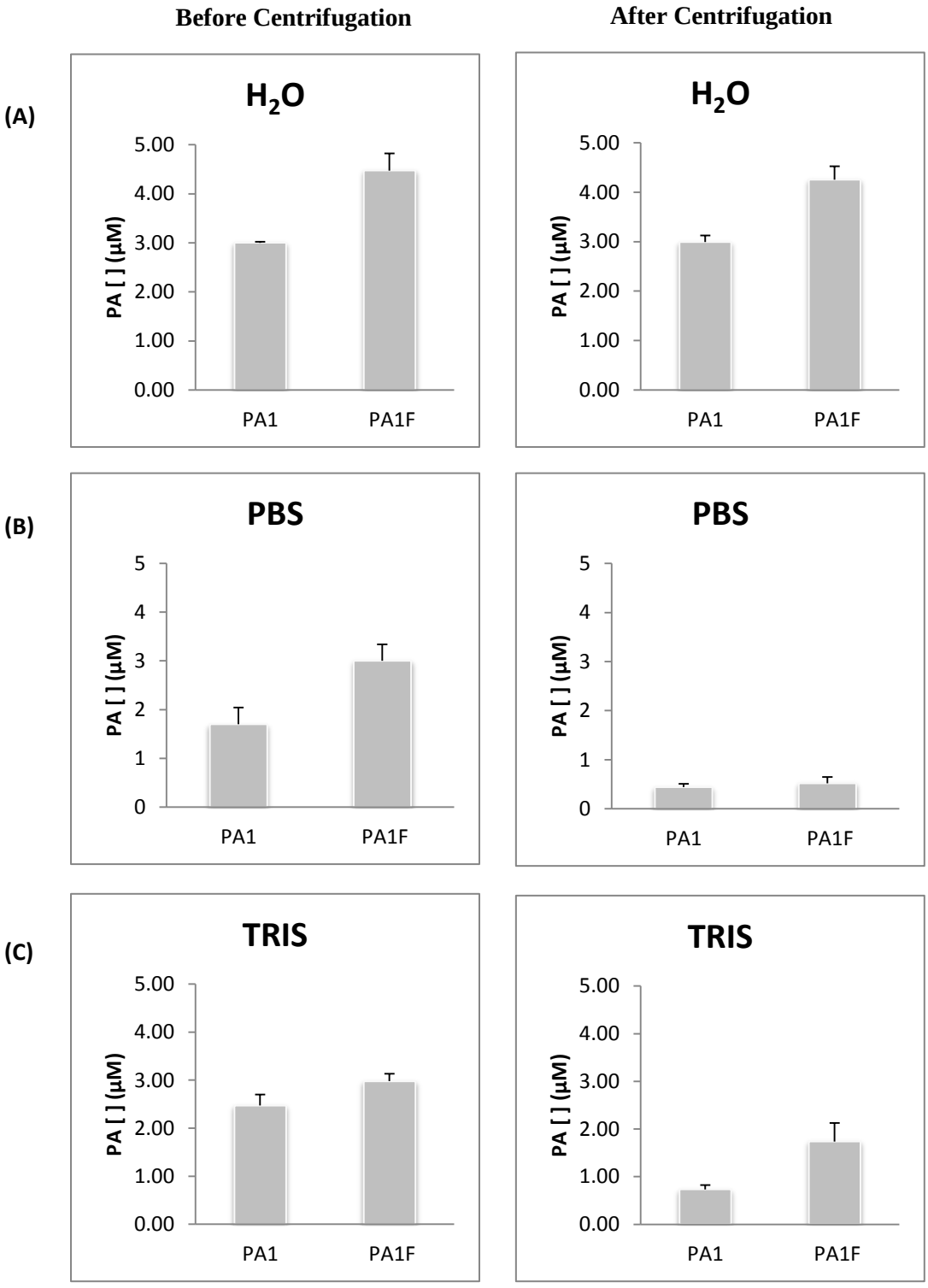


Figure 5.10. Solubility of PA1 and $PA1_F$ in (A) H_2O, (B) PBS and (C) Tris buffers

A major difference was however observed in their long term solubility. The solubility in DMSO was evaluated after one week from the stock preparation and analyzed as described above for the other solvents (**Figure 5.11**). In this case, the concentration of $PA1_F$ was almost double than PA1 and it was associated with a change in the UV-Vis spectrum profile, where a shoulder appears at ~365 nm for PA1 but not for $PA1_F$ (**Figure 5.12**). Furthermore, analysis of the same samples by High Pressure Liquid Chromatography (HPLC) shows only a negligible peak for PA1 (2.8 min), while the characteristic peak was observed for $PA1_F$. The altered chromatogram and UV spectrum of PA1 may be due to chemical instability of PA1 or to a different aggregation state in solution. The long term stability of PA1F may represent an advantage especially in terms of costs, since the high concentration stock may be utilized multiple times, as opposite as using a fresh aliquot for every experiment (**Figure**

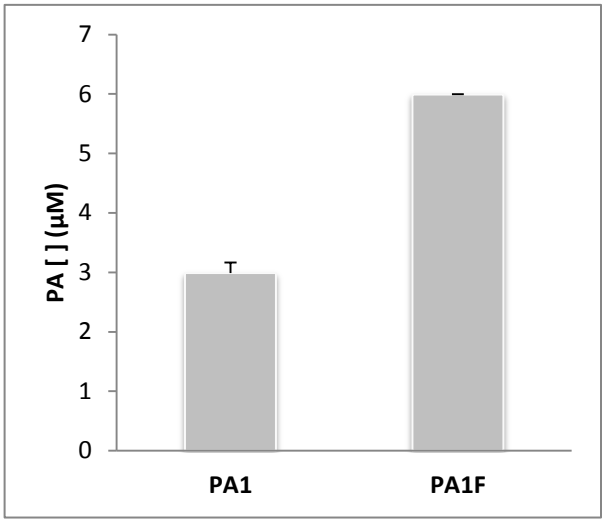


Figure 5.11. Solubility of PA1 and $PA1_F$ in DMSO after one week

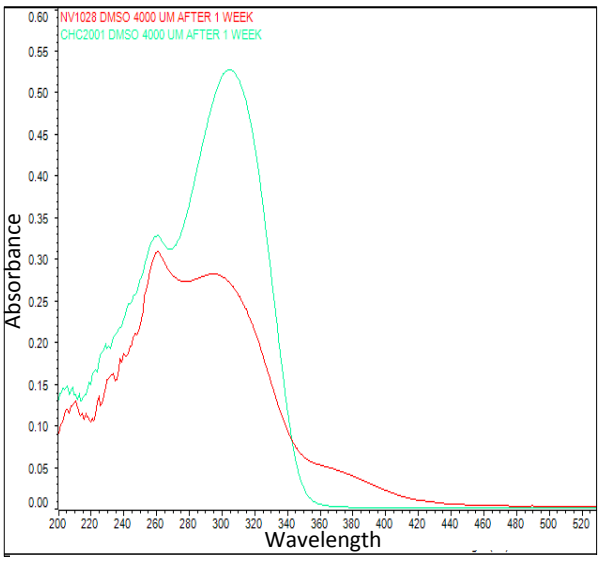


Figure 5.12. UV spectrum of PA1 and $PA1_F$

5.13). Even after one month, the UV-Vis spectrum of $PA1_F$ and its HPLC trace were unaltered.

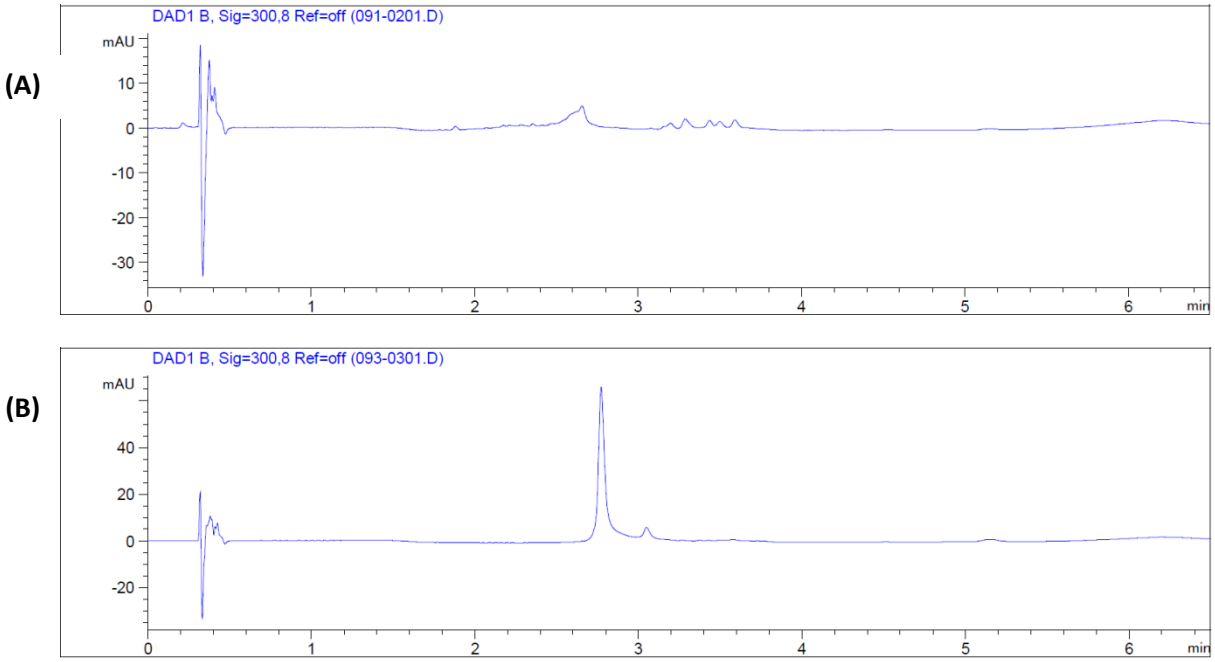


Figure 5.13. HPLC traces of (A) PA1 and (B) $PA1_F$ samples analyzed after one week of sample preparation

5.5 Conclusion

We developed an LC-MS/MS method to evaluate polyamide uptake in cell culture. We obtained a linear response within the concentration range of 0.5 to 50 µg/ml. The limit of quantification (LOQ) was 0.5 µg/ml, a fairly high value for standard LC/MS/MS methods. In fact, LOQ in the nano or picograms is usually obtained. However, the determined LOQ value was either lower or comparable to LOQ values determined by other groups for polyamides quantitation in plasma. The only method

reporting a lower LOQ value was developed by Sugiyama in plasma. However, no internal standard was used in the method. Furthermore our method was developed in accordance to the concentration of polyamide used to treat the cells, a concentration that was rather high. By lowering the range of polyamide concentration used to treat the cells along with a new set of calibration curves at lower concentration it will certainly be possible to increase the method sensitivity.

The developed method was used to evaluate uptake of PA1 and $PA1_F$ in HPV-infected keratinocytes after 24 h incubation. The results showed that cell-associated concentration was higher for the formic analog. The uptake increased when the cells were treated at higher polyamide concentration, as expected. Since the only structural change between the PAs is a counterion, it is likely that the improved uptake is due to improved solubility of the compound in cell culture media. Although we tried to evaluate the solubility in cell media by UV-Vis spectroscopy, the results were ambiguous, mostly due to the complexity in terms of components present in it and possible interaction between PAs and medium components. An interesting finding was the apparent improved stability of the formic salt compared to the TFA salt. The concentration measured in the primary stock in DMSO was stable for over a month. Probably, the formate salt does not confer immediate solubility advantages, but maybe it does confer advantages on the long term and most likely at high concentrations.

Further studies include the improvement of the method's sensitivity so that it could be utilized to detect lower polyamide concentrations and the investigation of the structural stability of these analogs by LC-MS/MS.

Safety: Because the experiments described in this proposal required handling cell lines containing HPV, all the laboratory procedures were conducted according to the Biosafety Level 2 (BSL-2) practices under the approval of the UMSL Biosafety Committee and the Environmental Health & Safety (EHS) Department.

5.6 References

(1) Edwards, T. G.; Koeller, K. J.; Slomczynska, U.; Fok, K.; Helmus, M.; Bashkin, J. K.; Fisher, C. *Antiviral Research* **2011**, *91*, 177.

(2) Edwards, T. G.; Vidmar, T. J.; Koeller, K.; Bashkin, J. K.; Fisher, C. *PLoS One* **2013**, *8*, e75406.

(3) Castaneda, C. H.; Scuderi, M. J.; Edwards, T. G.; Harris Jr, G. D.; Dupreur, C. M.; Koeller, K. J.; Fisher, C.; Bashkin, J. K. *MedChemComm* **2016**, *7*, 2076.

(4) Bashkin, J. K.; Edwards, T. G.; Fisher, C.; Harris, G. D.; Koeller, K. J.; Google Patents: 2016.

(5) Bastin, R. J.; Bowker, M. J.; Slater, B. J. *Organic Process Research & Development* **2000**, *4*, 427.

(6) Brittain, H. *Pharm. Tech.* **2007**, *31*, 78.

(7) Serajuddin, A. T. M. *Adv Drug Deliv Rev* **2007**, *59*, 603.

(8) Pitt, J. J. *The Clinical Biochemist Reviews* **2009**, *30*, 19.

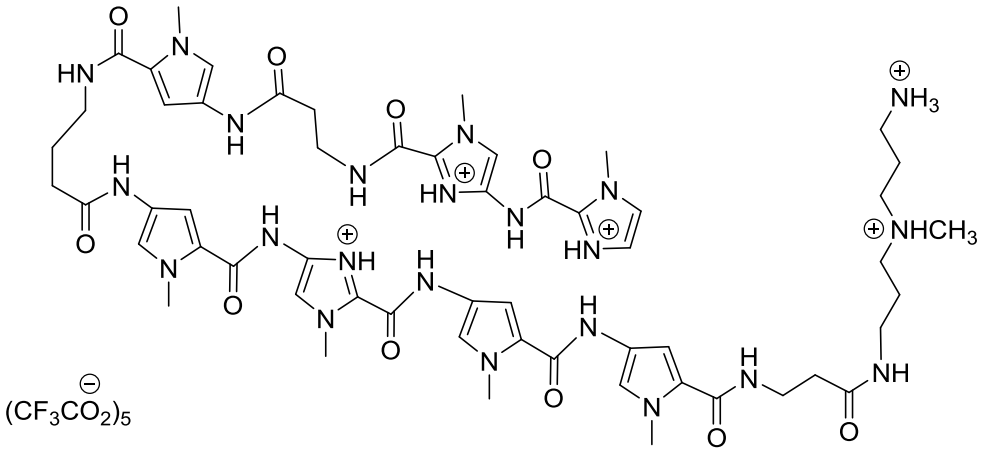
(9) Xu, R. N.; Fan, L.; Rieser, M. J.; El-Shourbagy, T. A. *J Pharm Biomed Anal* **2007**, *44*, 342.

6. Appendix

- **Preparation of MJS-KA1007-Ta** (DesImImβPyɣPyImPyPyβ-Ta)
- **Preparation of MJS-KA1002-Ta** (DesImImPyPyɣPyImPyPyβ-Ta)
- **Preparation of MJS-KA1011-Ta** (DesImImPyImɣPyPyβPyβ-Ta)
- **Preparation of MJS-KA1013-Ta** (DesImImβImɣPyPyPyPyβ-Ta)
- **Preparation of MJS-KA1033-Ta** (DesImImPyImɣPyPyPyPyβ-Ta)
- **Preparation of MJS-KA1055 (KA1041)-Ta** (DesImImPyPyɣPyImβPyβ-Ta)

- **Preparation of MJS-KA1007-EDTA** (DesImImβPyɣPyImPyPyβ-Ta-EDTA)
- **Preparation of MJS-KA1002-EDTA** (DesImImPyPyɣPyImPyPyβ-Ta-EDTA)
- **Preparation of MJS-KA1011-EDTA** (DesImImPyImɣPyPyβPyβ-Ta-EDTA)
- **Preparation of MJS-KA1013-EDTA** (DesImImβImɣPyPyPyPyβ-Ta-EDTA)
- **Preparation of MJS-KA1033-EDTA** (DesImImPyImɣPyPyPyPyβ-Ta-EDTA)
- **Preparation of MJS-KA1055 (KA1041)-EDTA** (DesImImPyPyɣPyImβPyβ-Ta-EDTA)

Preparation of MJS-KA1007 -Ta (DesImImβPyɣPyImPyPyβ-Ta)

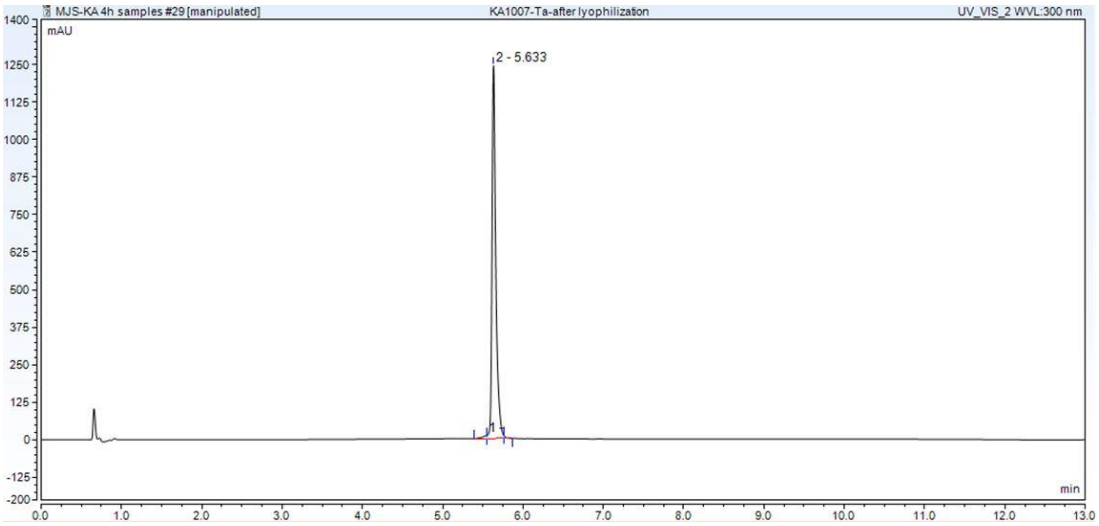


300 mg of resin bound to the polyamide (KA1007B) was placed in a 1 dram vial and incubated with 2 mL of 3,3'-Diamino-N-methyldipropylamine (Ta) at 60 ºC in a heating block, vortexing the mixture every 30 minutes.

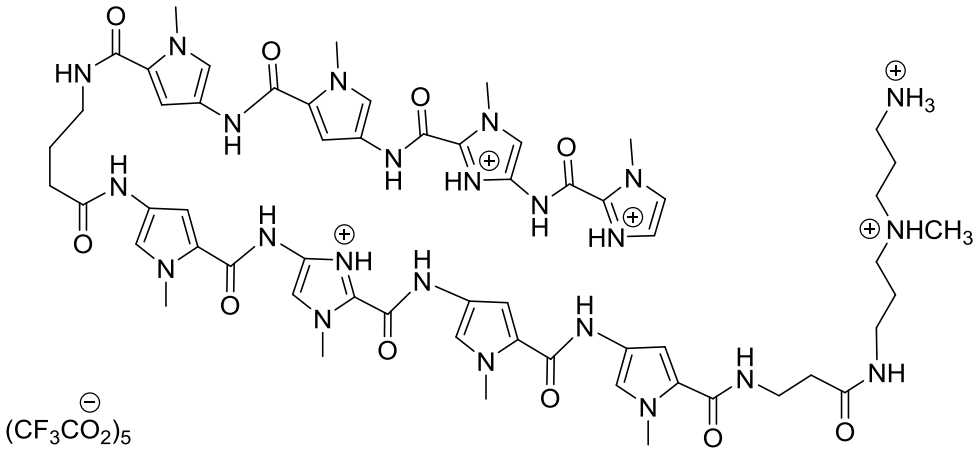
The reaction was monitored by High Pressure Liquid Chromatography (HPLC) and LC-MS (ESI) using an Agilent 1100 Series G1956B LC-MSD system. Briefly, 10 µL aliquots were withdrawn from the reaction mixture, diluted with 600 µL DMSO and filtered with a 0.2 µm syringe filter before analysis.

After 4 h, the reaction was stopped and the resin was removed by filtration through a 20 µM polyethylene filter. The vial was rinsed with 1 mL of DMSO which was added to the filtrate. After addition of 1 mL of water to the filtrate, the mixture was purified on a Dionex prep HPLC (Ultimate 3000) using a Phenomenex Jupiter C12 column (250 x 30 mm, 10 um particle size, 90A pore size). Mobile phases consisted of MeOH (0.1% TFA)/H_2O (0.1% TFA) and the gradient was from 25% to 85% methanol. Compound eluted at 25 min. Fractions were collected in a 250 mL round bottom flask, concentrated

under reduced pressure and transferred to a 20 mL scintillation vial, rinsing the flask with a 1:1 ACN:H_2O (0.1 % TFA) solution. The concentrated solution was frozen on dry ice and lyophilized overnight yielding 61.7 mg of MJS-KA1007-Ta (Purity: 99%, $[M+H]^+$ 1215.40 m/z).



Preparation of MJS-KA1002 (DesImImPyPyɣPyImPyPyβ-Ta)

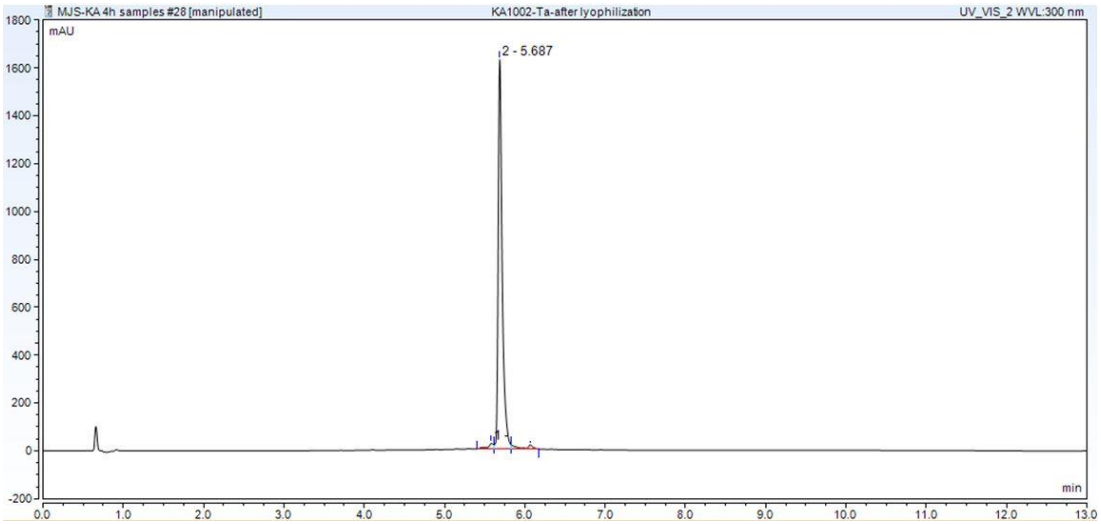


300 mg of resin bound to the polyamide (KA1002A) was placed in a 1 dram vial and incubated with 2 mL of 3,3'-Diamino-N-methyldipropylamine (Ta) at 60 ºC in a heating block, vortexing the mixture every 30 minutes.

The reaction was monitored by High Pressure Liquid Chromatography (HPLC) and LC-MS (ESI) using an Agilent 1100 Series G1956B LC-MSD system. Briefly, 10 µL aliquots were withdrawn from the reaction mixture, diluted with 600 µL DMSO and filtered with a 0.2 µm syringe filter before analysis.

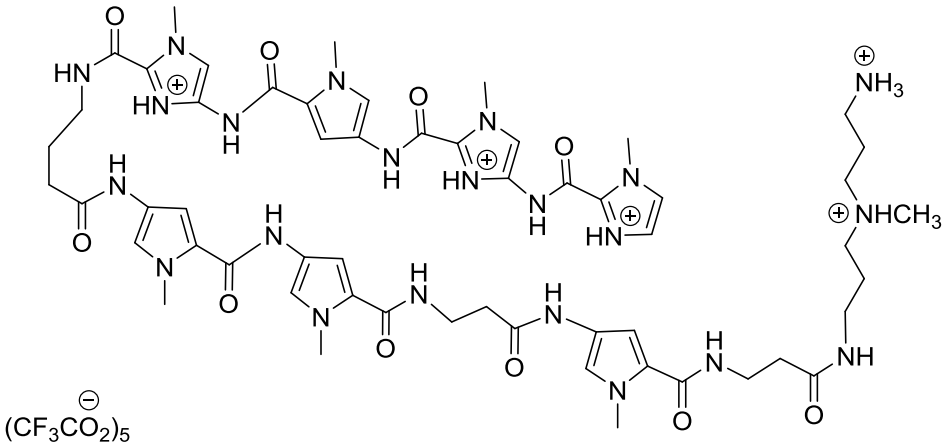
After 4 h, the reaction was stopped and the resin was removed by filtration through a 20 µM polyethylene filter. The vial was rinsed with 1 mL of DMSO which was added to the filtrate. After addition of 1 mL of water to the filtrate, the mixture was purified on a Dionex prep HPLC (Ultimate 3000) using a Phenomenex Jupiter C12 column (250 x 30 mm, 10 um particle size, 90A pore size). Mobile phases consisted of MeOH (0.1% TFA)/H_2O (0.1% TFA) and the gradient was from 25% to 85% methanol. Compound

eluted at 25 min. Fractions were collected in a 250 mL round bottom flask, concentrated under reduced pressure and transferred to a 20 mL scintillation vial, rinsing the flask with a 1:1 ACN:H_2O (0.1 % TFA) solution. The concentrated solution was frozen on dry ice and lyophilized overnight yielding 65.2 mg of MJS-KA1002-Ta (Purity: 95%, $[M+H]^+$



1267.60 m/z).

Preparation of MJS-KA1011-Ta (DesImImPyImγPyPyβPyβ-Ta)

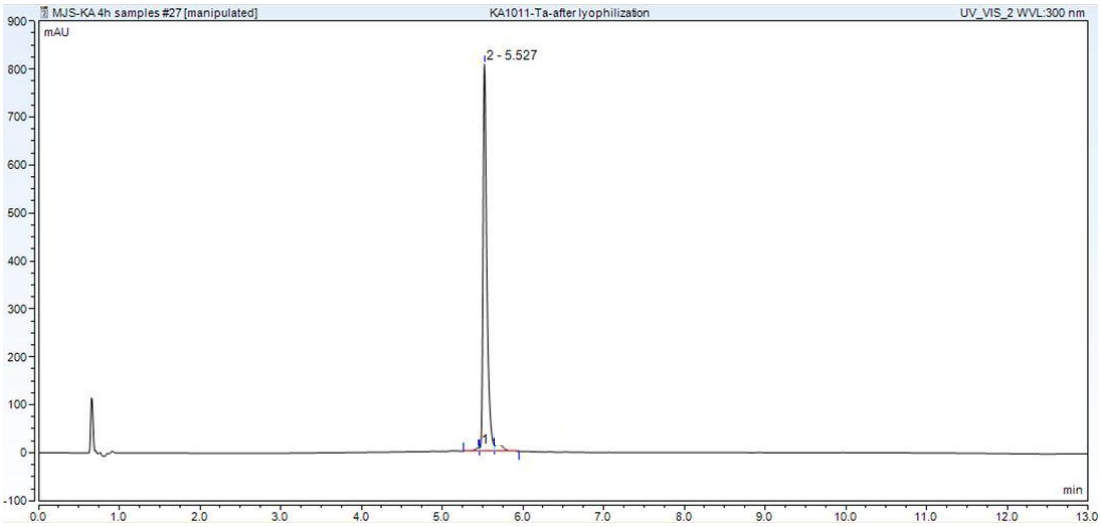


300 mg of resin bound to the polyamide (KA1011B) was placed in a 1 dram vial and incubated with 2 mL of 3,3'-Diamino-N-methyldipropylamine (Ta) at 60 ºC in a heating block, vortexing the mixture every 30 minutes.

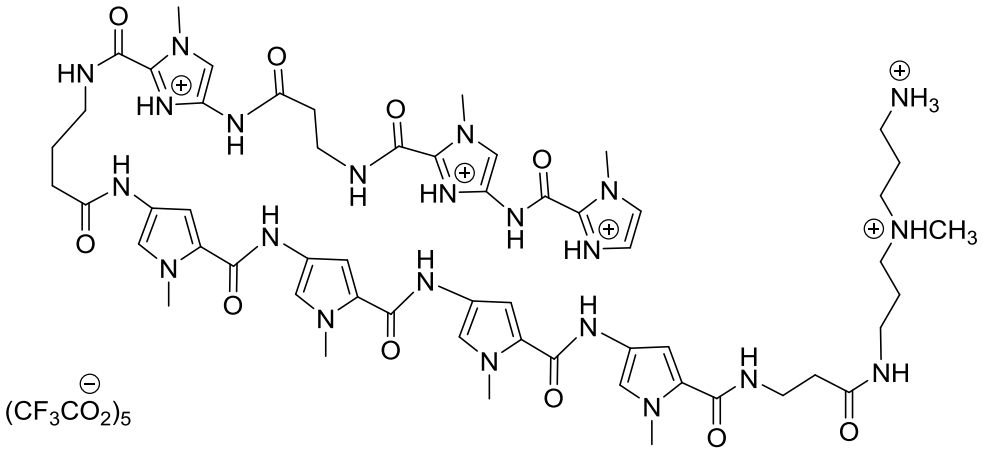
The reaction was monitored by High Pressure Liquid Chromatography (HPLC) and LC-MS (ESI) using an Agilent 1100 Series G1956B LC-MSD system. Briefly, 10 µL aliquots were withdrawn from the reaction mixture, diluted with 600 µL DMSO and filtered with a 0.2 µm syringe filter before analysis.

After 4 h, the reaction was stopped and the resin was removed by filtration through a 20 µM polyethylene filter. The vial was rinsed with 1 mL of DMSO which was added to the filtrate. After addition of 1 mL of water to the filtrate, the mixture was purified on a Dionex prep HPLC (Ultimate 3000) using a Phenomenex Jupiter C12 column (250 x 30 mm, 10 um particle size, 90A pore size). Mobile phases consisted of MeOH (0.1% TFA)/H_2O (0.1% TFA) and the gradient was from 25% to 85% methanol. Compound

eluted at 45 min. Fractions were collected in a 250 mL round bottom flask, concentrated under reduced pressure and transferred to a 20 mL scintillation vial, rinsing the flask with a 1:1 ACN:H_2O (0.1 % TFA) solution. The concentrated solution was frozen on dry ice and lyophilized overnight yielding 64.4 mg of MJS-KA1011-Ta (Purity: 95%, $[M+H]^+$ 1215.40 m/z).



Preparation of MJS-KA1013-Ta (DesImImβImγPyPyPyPyβ-Ta)

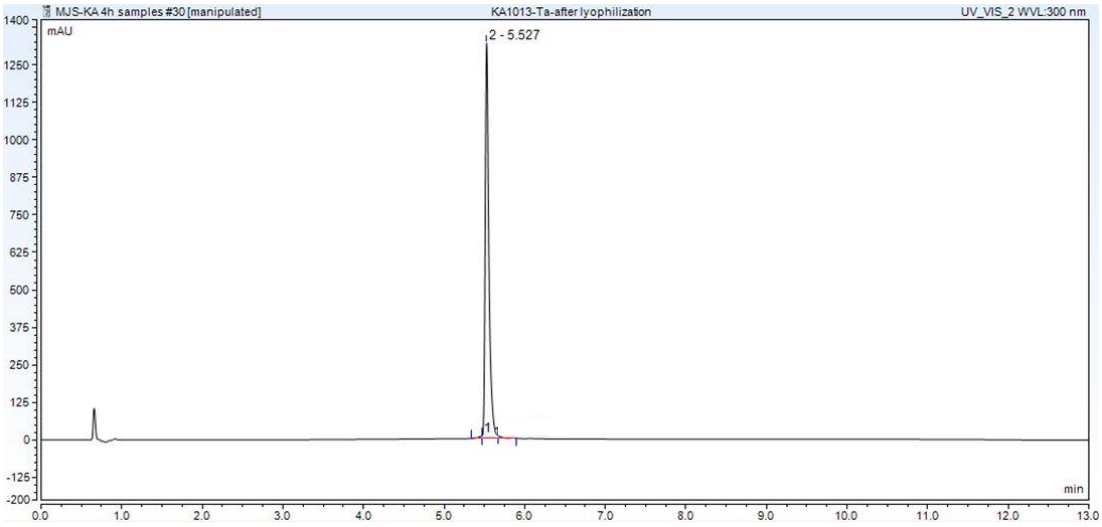


300 mg of resin bound to the polyamide (KA1013B) was placed in a 1 dram vial and incubated with 2 mL of 3,3'-Diamino-N-methyldipropylamine (Ta) at 60 ºC in a heating block, vortexing the mixture every 30 minutes.

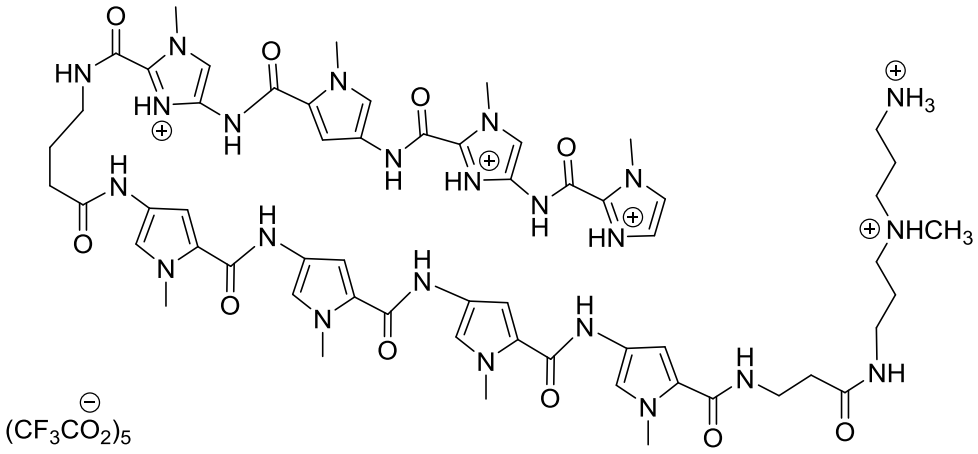
The reaction was monitored by High Pressure Liquid Chromatography (HPLC) and LC-MS (ESI) using an Agilent 1100 Series G1956B LC-MSD system. Briefly, 10 µL aliquots were withdrawn from the reaction mixture, diluted with 600 µL DMSO and filtered with a 0.2 µm syringe filter before analysis.

After 4 h, the reaction was stopped and the resin was removed by filtration through a 20 µM polyethylene filter. The vial was rinsed with 1 mL of DMSO which was added to the filtrate. After addition of 1 mL of water to the filtrate, the mixture was purified on a Dionex prep HPLC (Ultimate 3000) using a Phenomenex Jupiter C12 column (250 x 30 mm, 10 um particle size, 90A pore size). Mobile phases consisted of MeOH (0.1% TFA)/H_2O (0.1% TFA) and the gradient was from 25% to 85% methanol. Compound eluted at 27 min. Fractions were collected in a 250 mL round bottom flask, concentrated

under reduced pressure and transferred to a 20 mL scintillation vial, rinsing the flask with a 1:1 ACN:H_2O (0.1 % TFA) solution. The concentrated solution was frozen on dry ice and lyophilized overnight yielding 59.6 mg of MJS-KA1013-Ta (Purity: 99%, $[M+H]^+$ 1215.40 m/z).



Preparation of MJS-KA1033-Ta (DesImImPyImɣPyPyPyPyβ-Ta)

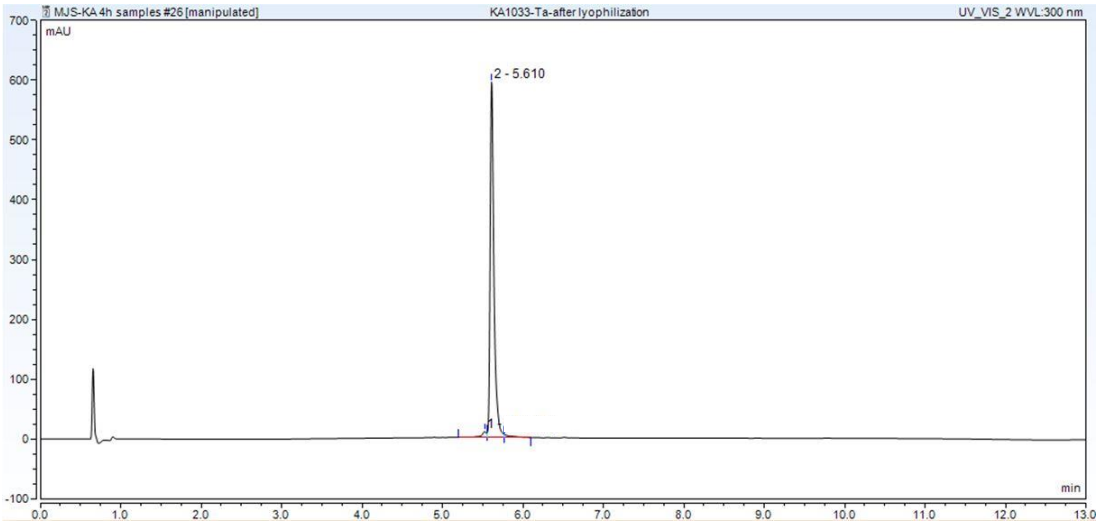


300 mg of resin bound to the polyamide (KA1033) was placed in a 1 dram vial and incubated with 2 mL of 3,3'-Diamino-N-methyldipropylamine (Ta) at 60 ºC in a heating block, vortexing the mixture every 30 minutes.

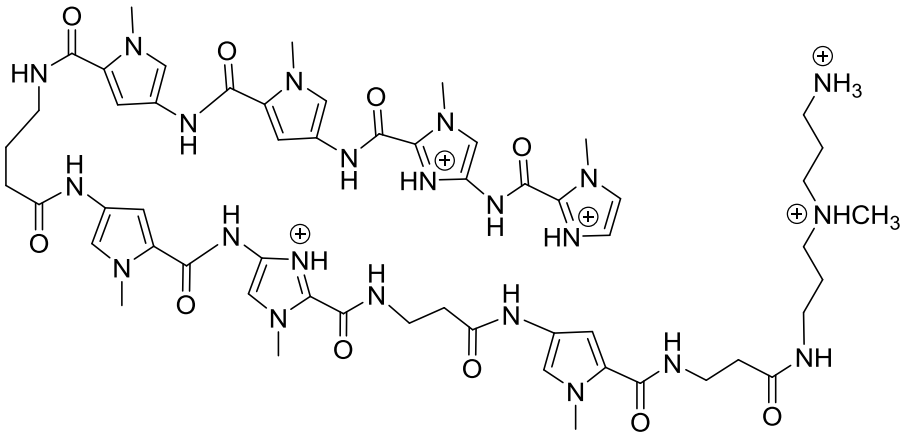
The reaction was monitored by High Pressure Liquid Chromatography (HPLC) and LC-MS (ESI) using an Agilent 1100 Series G1956B LC-MSD system. Briefly, 10 µL aliquots were withdrawn from the reaction mixture, diluted with 600 µL DMSO and filtered with a 0.2 µm syringe filter before analysis.

After 4 h, the reaction was stopped and the resin was removed by filtration through a 20 µM polyethylene filter. The vial was rinsed with 1 mL of DMSO which was added to the filtrate. After addition of 1 mL of water to the filtrate, the mixture was purified on a Dionex prep HPLC (Ultimate 3000) using a Phenomenex Jupiter C12 column (250 x 30 mm, 10 um particle size, 90A pore size). Mobile phases consisted of MeOH (0.1% TFA)/H_2O (0.1% TFA) and the gradient was from 25% to 85% methanol. Compound

eluted at 30 min. Fractions were collected in a 250 mL round bottom flask, concentrated under reduced pressure and transferred to a 20 mL scintillation vial, rinsing the flask with a 1:1 ACN:H_2O (0.1 % TFA) solution. The concentrated solution was frozen on dry ice and lyophilized overnight yielding 48 mg of MJS-KA1033-Ta (Purity: 97%, $[M+H]^+$ 1266.40 m/z).



Preparation of MJS-KA1055 (1041)-Ta (DesImImPyPyɣPyImβPyβ-Ta)

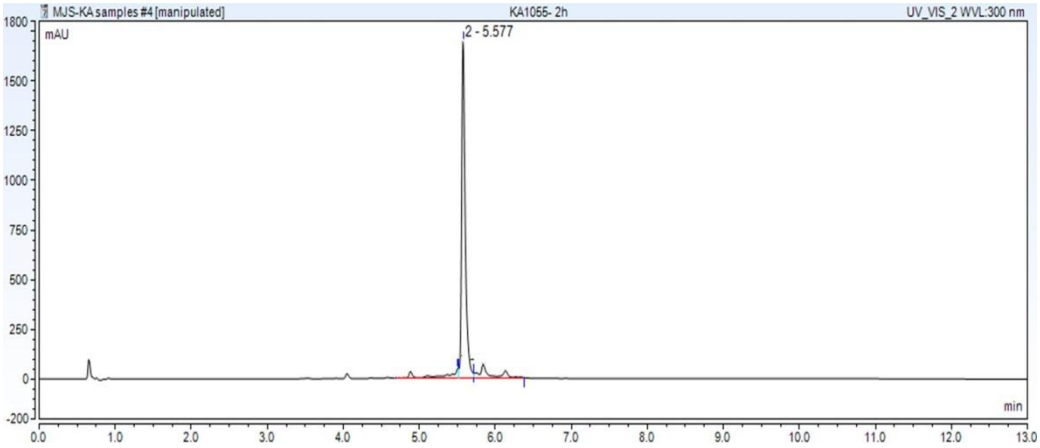


300 mg of resin bound to the polyamide (KA1041B) was placed in a 1 dram vial and incubated with 2 mL of 3,3'-Diamino-N-methyldipropylamine (Ta) at 60 ºC in a heating block, vortexing the mixture every 30 minutes.

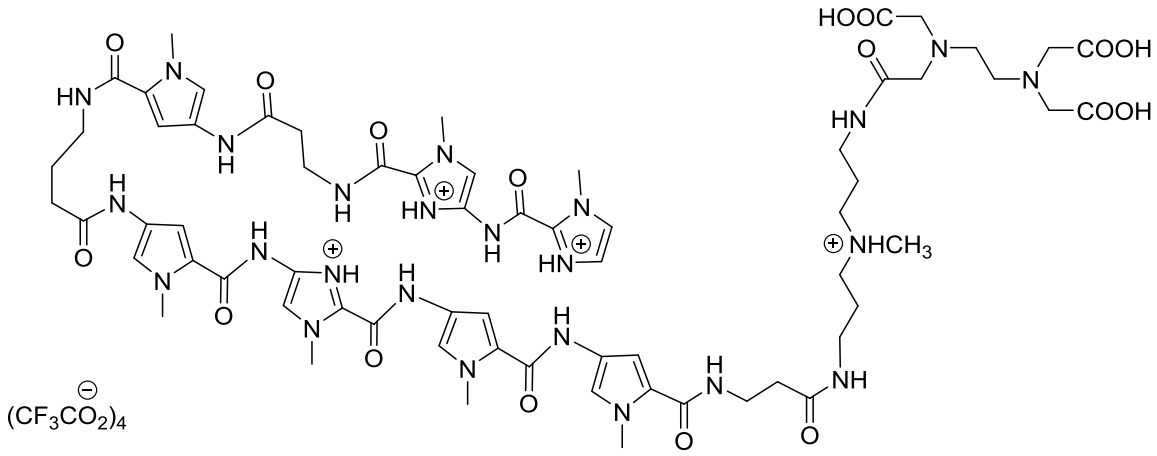
The reaction was monitored by High Pressure Liquid Chromatography (HPLC) and LC-MS (ESI) using an Agilent 1100 Series G1956B LC-MSD system. Briefly, 10 µL aliquots were withdrawn from the reaction mixture, diluted with 600 µL DMSO and filtered with a 0.2 µm syringe filter before analysis.

After 4 h, the reaction was stopped and the resin was removed by filtration through a 20 µM polyethylene filter. The vial was rinsed with 1 mL of DMSO which was added to the filtrate. After addition of 1 mL of water to the filtrate, the mixture was purified on a Dionex prep HPLC (Ultimate 3000) using a Phenomenex Jupiter C12 column (250 x 30 mm, 10 um particle size, 90A pore size). Mobile phases consisted of MeOH (0.1% TFA)/H_2O (0.1% TFA) and the gradient was from 25% to 85% methanol. Compound eluted at 25 min. Fractions were collected in a 250 mL round bottom flask, concentrated

under reduced pressure and transferred to a 20 mL scintillation vial, rinsing the flask with a 1:1 ACN:H_2O (0.1 % TFA) solution. The concentrated solution was frozen on dry ice and lyophilized overnight yielding 50.5 mg of MJS-KA1055 (1041)-Ta ($[M+H]^+$ 1215.40 m/z).



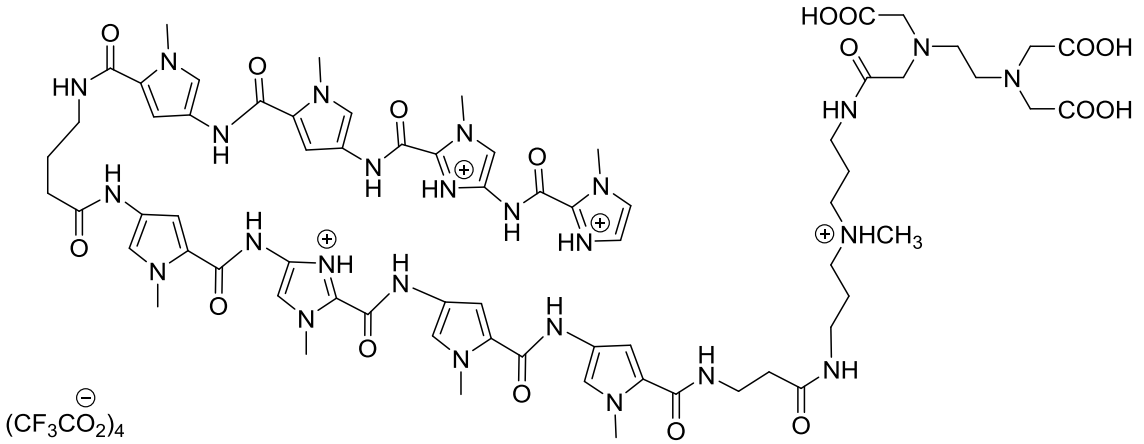
Preparation of MJS-KA1007 -Ta (DesImImβPyɣPyImPyPyβ-Ta-EDTA)



14.4 mg of EDTA dianhydride were placed in a 1 dram vial and added 0.5 mL DIEA, 0.25 mL DMF and 0.25 mL DMSO. After vigorous vortexing, the mixture was placed in a heating block and stirred at 55 ºC. To the EDTA mixture, a biphasic mixture of MJS-KA1007-Ta (5 mg) in 0.5 mL *N*,*N*-Diisopropylethylamine (DIEA) and 0.5 mL DMF was added dropwise *via* syringe over 15 min. The mixture was stirred for 30 min at 55 ºC and the reaction monitored by HPLC. Briefly, a 10 µL aliquot from the bottom phase (DMF/DMSO) was diluted with 100 µL of DMSO and analyzed by LC-MS. Following addition of 1.12 mL of 0.1 N NaOH, the mixture was then stirred at 55 ºC for 20 min and further analyzed by LC-MS. The bottom layer of the biphasic solution was transferred to a 20 mL vial and neutralized with 1.4 mL of 1 M trifluoroacetic acid (TFA). The mixture was diluted with 1 mL of DMSO and filtered through a 20 µm polyethylene filter. After purification by reverse-phase HPLC on a Dionex prep HPLC (Ultimate 3000), fractions containing the desired product were transferred into a 50 mL round bottom flask, concentrated under reduced pressure and transferred to a 20 mL vial,

rinsing the flask with a 1:1 ACN:H_2O (0.1% TFA) solution. The concentrated mixture was frozen on dry ice and lyophilized overnight to give 3.5 mg of MJS-KA1007-EDTA ($[M+H]^+$ 1489.61 m/z, purity: 85%).

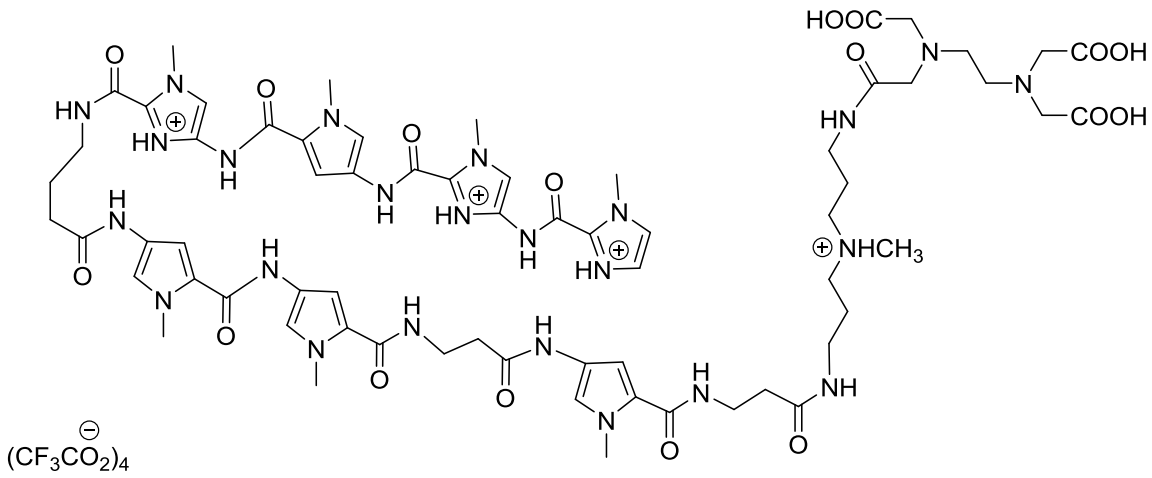
Preparation of MJS-KA1002-EDTA (DesImImPyPyɣPyImPyPyβ-Ta-EDTA)



13.95 mg of EDTA dianhydride were placed in a 1 dram vial and added 0.5 mL DIEA, 0.25 mL DMF and 0.25 mL DMSO. After vigorous vortexing, the mixture was placed in a heating block and stirred at 55 ºC. To the EDTA mixture, a biphasic mixture of MJS-KA1002-Ta (5 mg) in 0.5 mL *N*,*N*-Diisopropylethylamine (DIEA) and 0.5 mL DMF was added dropwise *via* syringe over 15 min. The mixture was stirred for 30 min at 55 ºC and the reaction monitored by HPLC. Briefly, a 10 µL aliquot from the bottom phase (DMF/DMSO) was diluted with 100 µL of DMSO and analyzed by LC-MS. Following addition of 1.09 mL of 0.1 N NaOH, the mixture was then stirred at 55 ºC for 20 min and further analyzed by LC-MS. The bottom layer of the biphasic solution was

transferred to a 20 mL vial and neutralized with 600 µL of 1 M trifluoroacetic acid (TFA). The mixture was diluted with 1 mL of DMSO and filtered through a 20 µm polyethylene filter. After purification by reverse-phase HPLC on a Dionex prep HPLC (Ultimate 3000), fractions containing the desired product were transferred into a 50 mL round bottom flask, concentrated under reduced pressure and transferred to a 20 mL vial, rinsing the flask with a 1:1 ACN:H_2O (0.1% TFA) solution. The concentrated mixture was frozen on dry ice and lyophilized overnight to give 2.9 mg of MJS-KA1002-EDTA ($[M+H]^+$ 1540.6 m/z, purity: 88%).

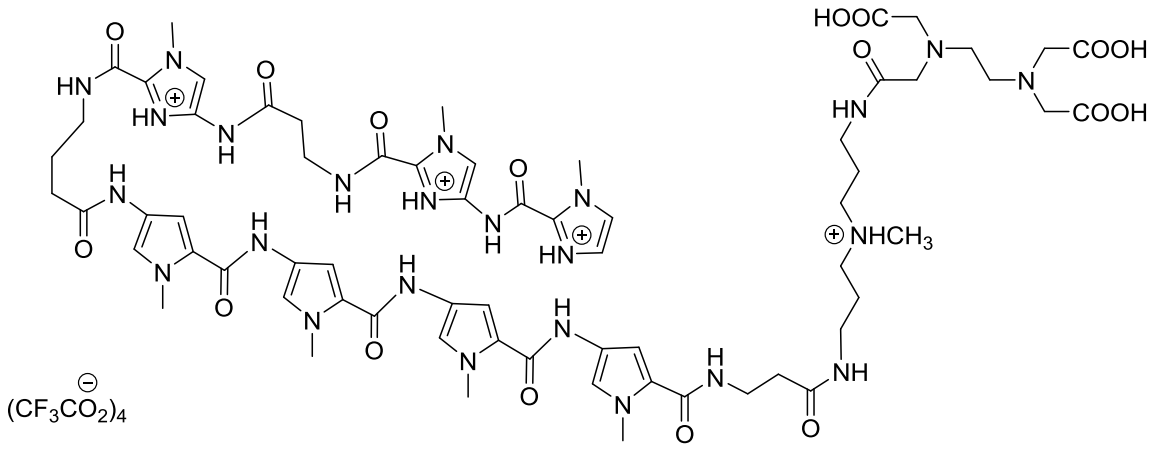
Preparation of MJS-KA1011-Ta-EDTA (DesImImPyImɣPyPyβPyβ-EDTA)



14.3 mg of EDTA dianhydride were placed in a 1 dram vial and added 0.5 mL DIEA, 0.25 mL DMF and 0.25 mL DMSO. After vigorous vortexing, the mixture was placed in a heating block and stirred at 55 ºC. To the EDTA mixture, a biphasic mixture of MJS-KA1011-Ta (5 mg) in 0.5 mL *N*,*N*-Diisopropylethylamine (DIEA) and 0.5 mL

DMF was added dropwise *via* syringe over 15 min. The mixture was stirred for 30 min at 55 ºC and the reaction monitored by HPLC. Briefly, a 10 µL aliquot from the bottom phase (DMF/DMSO) was diluted with 100 µL of DMSO and analyzed by LC-MS. Following addition of 1.12 mL of 0.1 N NaOH, the mixture was then stirred at 55 ºC for 20 min and further analyzed by LC-MS. The bottom layer of the biphasic solution was transferred to a 20 mL vial and neutralized with 450 µL of 1 M trifluoroacetic acid (TFA). The mixture was diluted with 1 mL of DMSO and filtered through a 20 µm polyethylene filter. After purification by reverse-phase HPLC on a Dionex prep HPLC (Ultimate 3000), fractions containing the desired product were transferred into a 50 mL round bottom flask, concentrated under reduced pressure and transferred to a 20 mL vial, rinsing the flask with a 1:1 ACN:H_2O (0.1% TFA) solution. The concentrated mixture was frozen on dry ice and lyophilized overnight to give 2.4 mg of MJS-KA1011-EDTA ($[M+H]^+$ 1489.4 m/z, purity: 88%).

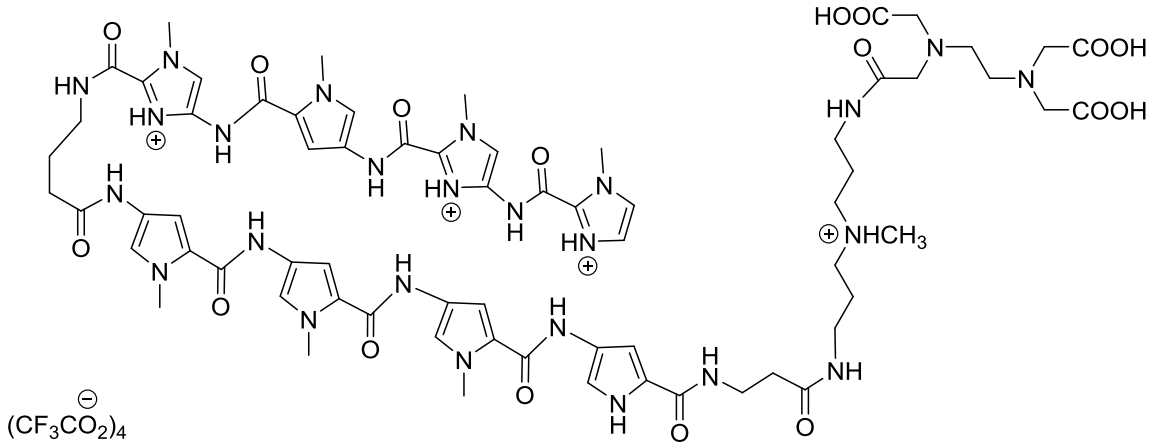
Preparation of MJS-KA1013-EDTA (DesImImβImγPyPyPyPyβ-Ta-EDTA)



14.3 mg of EDTA dianhydride were placed in a 1 dram vial and added 0.5 mL DIEA, 0.25 mL DMF and 0.25 mL DMSO. After vigorous vortexing, the mixture was placed in a heating block and stirred at 55 ºC. To the EDTA mixture, a biphasic mixture of MJS-KA1013-Ta (5 mg) in 0.5 mL *N*,*N*-Diisopropylethylamine (DIEA) and 0.5 mL DMF was added dropwise *via* syringe over 15 min. The mixture was stirred for 30 min at 55 ºC and the reaction monitored by HPLC. Briefly, a 10 µL aliquot from the bottom phase (DMF/DMSO) was diluted with 100 µL of DMSO and analyzed by LC-MS. Following addition of 1.12 mL of 0.1 N NaOH, the mixture was then stirred at 55 ºC for 20 min and further analyzed by LC-MS. The bottom layer of the biphasic solution was transferred to a 20 mL vial and neutralized with 650 µL of 1 M trifluoroacetic acid (TFA). The mixture was diluted with 1 mL of DMSO and filtered through a 20 µm polyethylene filter. After purification by reverse-phase HPLC on a Dionex prep HPLC (Ultimate 3000), fractions containing the desired product were transferred into a 50 mL

round bottom flask, concentrated under reduced pressure and transferred to a 20 mL vial, rinsing the flask with a 1:1 ACN:H_2O (0.1% TFA) solution. The concentrated mixture was frozen on dry ice and lyophilized overnight to give 2.8 mg of MJS-KA1013-EDTA ($[M+H]^+$ 1489.61 m/z, purity: 93%).

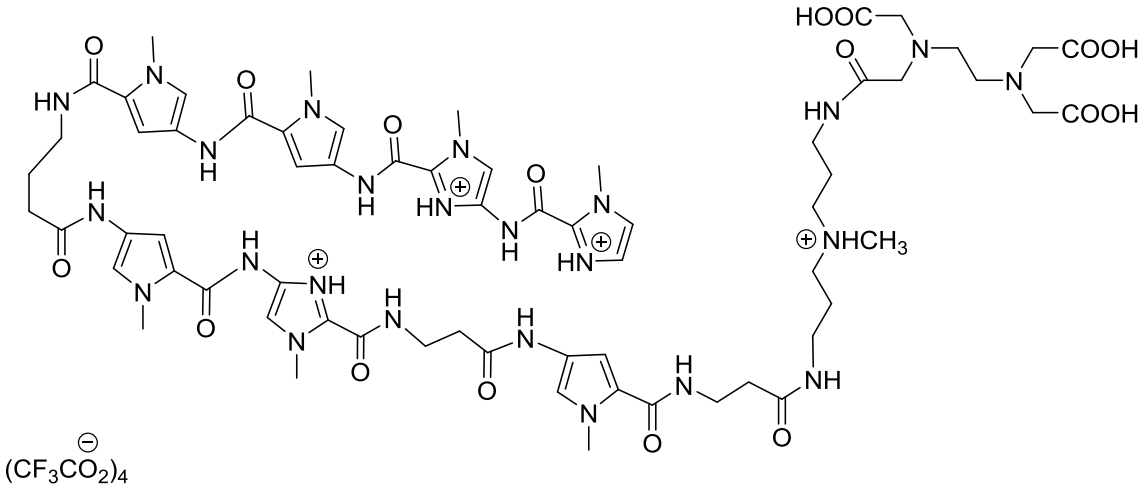
Preparation of MJS-KA1033-Ta-EDTA (DesImImPyImɣPyPyPyPyβ-Ta-EDTA)



21 mg of EDTA dianhydride were placed in a 1 dram vial and added 0.5 mL DIEA, 0.25 mL DMF and 0.25 mL DMSO. After vigorous vortexing, the mixture was placed in a heating block and stirred at 55 ºC. To the EDTA mixture, a biphasic mixture of MJS-KA1033-Ta (5 mg) in 0.5 mL *N*,*N*-Diisopropylethylamine (DIEA) and 0.5 mL DMF was added dropwise *via* syringe over 15 min. The mixture was stirred for 30 min at 55 ºC and the reaction monitored by HPLC. Briefly, a 10 µL aliquot from the bottom phase (DMF/DMSO) was diluted with 100 µL of DMSO and analyzed by LC-MS. Following addition of 1.62 mL of 0.1 N NaOH, the mixture was then stirred at 55 ºC for

20 min and further analyzed by LC-MS. The bottom layer of the biphasic solution was transferred to a 20 mL vial and neutralized with 350 µL of 1 M trifluoroacetic acid (TFA). The mixture was diluted with 1 mL of DMSO and filtered through a 20 µm polyethylene filter. After purification by reverse-phase HPLC on a Dionex prep HPLC (Ultimate 3000), fractions containing the desired product were transferred into a 50 mL round bottom flask, concentrated under reduced pressure and transferred to a 20 mL vial, rinsing the flask with a 1:1 ACN:H_2O (0.1% TFA) solution. The concentrated mixture was frozen on dry ice and lyophilized overnight to give 1.3 mg of MJS-KA1033-EDTA ($[M+H]^+$ 1540.4 m/z, purity: 80%).

Preparation of MJS-KA1055 (KA1041)-EDTA (DesImImPyPyɣPyImβPyβ-Ta-EDTA)



14.4 mg of EDTA dianhydride were placed in a 1 dram vial and added 0.5 mL DIEA, 0.25 mL DMF and 0.25 mL DMSO. After vigorous vortexing, the mixture was placed in a heating block and stirred at 55 ºC. To the EDTA mixture, a biphasic mixture

of MJS-KA1055 (KA1041)-Ta (5 mg) in 0.5 mL *N*,*N*-Diisopropylethylamine (DIEA) and 0.5 mL DMF was added dropwise *via* syringe over 15 min. The mixture was stirred for 30 min at 55 ºC and the reaction monitored by HPLC. Briefly, a 10 µL aliquot from the bottom phase (DMF/DMSO) was diluted with 100 µL of DMSO and analyzed by LC-MS. Following addition of 1.12 mL of 0.1 N NaOH, the mixture was then stirred at 55 ºC for 20 min and further analyzed by LC-MS. The bottom layer of the biphasic solution was transferred to a 20 mL vial and neutralized with 400 µL of 1 M trifluoroacetic acid (TFA). The mixture was diluted with 1 mL of DMSO and filtered through a 20 µm polyethylene filter. After purification by reverse-phase HPLC on a Dionex prep HPLC (Ultimate 3000), fractions containing the desired product were transferred into a 50 mL round bottom flask, concentrated under reduced pressure and transferred to a 20 mL vial, rinsing the flask with a 1:1 ACN:H_2O (0.1% TFA) solution. The concentrated mixture was frozen on dry ice and lyophilized overnight to give 3.3 mg of MJS-KA1055 (KA1041)-EDTA ($[M+H]^+$ 1489.4 m/z, purity: 86%).