

University of Missouri, St. Louis

IRL @ UMSL

---

Chemistry & Biochemistry Faculty Works

Chemistry and Biochemistry

---

January 2016

## Interactions of two large antiviral polyamides with the long control region of HPV16.

Elena Vasilieva

*University of Missouri–St. Louis*

Jacquelyn Niederschulte

*University of Missouri–St. Louis*

Yang Song

*University of Missouri–St. Louis*

George Harris

*University of Missouri–St. Louis*

Kevin Koeller

*University of Missouri–St. Louis*

*See next page for additional authors*

Follow this and additional works at: <https://irl.umsl.edu/chemistry-faculty>



Part of the [Biochemistry Commons](#), [Genetics Commons](#), and the [Molecular Biology Commons](#)

---

### Recommended Citation

Vasilieva, Elena; Niederschulte, Jacquelyn; Song, Yang; Harris, George; Koeller, Kevin; Liao, Puhong; Bashkin, James; and Dupureur, Cynthia, "Interactions of two large antiviral polyamides with the long control region of HPV16." (2016). *Chemistry & Biochemistry Faculty Works*. 11.

DOI: <https://doi.org/10.1016/j.biochi.2016.04.022>

Available at: <https://irl.umsl.edu/chemistry-faculty/11>

This Article is brought to you for free and open access by the Chemistry and Biochemistry at IRL @ UMSL. It has been accepted for inclusion in Chemistry & Biochemistry Faculty Works by an authorized administrator of IRL @ UMSL. For more information, please contact [marvinh@umsl.edu](mailto:marvinh@umsl.edu).

---

## Authors

Elena Vasilieva, Jacquelyn Niederschulte, Yang Song, George Harris, Kevin Koeller, Puhong Liao, James Bashkin, and Cynthia Dupureur

# Interactions of Two Large Antiviral Polyamides with the Long Control Region of HPV16

“Post Print”

Elena Vasilieva, Jacquelyn Niederschulte, Yang Song, George Davis Harris Jr., Kevin J. Koeller, Puhong Liao, James K. Bashkin, and Cynthia M. Dupureur

**ABSTRACT:** PA1 and PA25 are large hairpin polyamides that are effective in nearly eliminating HPV16 episomes (DNA) in cell culture, and PA25 has broad spectrum activity against three cancer-causing forms of HPV (Edwards, T. G., Koeller, K. J., Slomczynska, U., Fok, K., Helmus, M., Bashkin, J. K., Fisher, C., *Antiviral Res.* 91 (2011) 177–186). Described here are the interactions of these PAs with sequences in the long control region (LCR) of HPV16 (7348-122). Using an FeEDTA conjugate of PA1 (designed to recognize 5'-W2GW7-3'; W = A or T), 34 affinity cleavage (AC) patterns were detected for this fragment. These sites can be rationalized with sequences featuring perfect, single, double, triple and quadruple mismatches. Quantitative DNase I footprinting analysis indicates that perfect sites bind PA1 with Kds between 0.7 and 2.2 nM. Kds for single, double, triple and quadruple mismatch sites range from 1-3 nM–20 nM. Using AC and EDTA conjugates, we report that unlike smaller 8-ring hairpin PAs, introduction of a chiral turn in this large polyamide has no effect on binding orientation (forward vs. reverse). Despite its design to recognize 5'-W2GW5GW4-3' via two Im residues, a motif not represented in this HPV sequence, a PA25-EDTA conjugate yielded 31 affinity cleavage sites on the region. Low nM Kds for PA25 without EDTA indicates a high tolerance for triple and quadruple mismatches. While there is extensive coverage of the sequence examined, AC cleavage patterns for the two PAs show discrete binding events and do not overlap significantly. This indicates that within the context of A/T rich sequences, these PAs do not recognize a simple shared sequence-related feature of the DNA. These insights continue to inform the complex nature of large hairpin PA-DNA interactions and antiviral behavior.

Vasilieva, E., Niederschulte, J., Song, Y., Harris, G. D., Jr., Koeller, K. J., Liao, P., . . . Dupureur, C. M. (2016). Interactions of two large antiviral polyamides with the long control region of HPV16. *Biochimie*, 127, 103-114. doi:10.1016/j.biochi.2016.04.022

## 1. Introduction

Hairpin polyamides (PAs) are small polymers composed principally of pyrrole and imidazole. They are inspired by the crescent-shaped distamycin, a natural product that recognizes the hydrogen bonding patterns in AT rich sequences in the minor groove of DNA. Due to the elegant design but high toxicity, these natural products have evolved synthetically over a number of years to achieve high DNA affinity and reported sequence specificity [1e10] and have been subsequently applied to control the expression of genes [3,11]

---

*Abbreviations:* AC, affinity cleavage; CE, capillary electrophoresis; DIEA, N,N-Diisopropylethylamine; DMF, dimethylformamide; DTT, dithiothreitol; EDTA, ethylene diamine tetraacetic acid; LCR, long control region; TFA, trifluoroacetic acid.

\* Corresponding author.

E-mail address: [cdup@umsl.edu](mailto:cdup@umsl.edu) (C.M. Dupureur).

and as chemical probes of nucleic acid structure [12].

A number of larger hairpin PAs (10 or more rings) have been shown to practically eliminate the HPV genome in cell culture with low IC<sub>50</sub>s and no detectable toxicity [13,14]. The two lead compounds reported in this anti-HPV study are PA1 and PA25 (Fig. 1), compounds that exhibit IC<sub>50</sub>'s of 0.1 and 0.036 mM against HPV16, respectively. Recently, the mechanism of action of this class of anti-HPV compounds has been partly elucidated in several reports that show an important role of the DNA Damage Response (DDR) modulated by PA administration to cells bearing small, circular dsDNA viral genomes, or episomes [14e16]. The access of these PAs to a DDR-related mechanism was unprecedented for any polyamides. Since our initial work was published, Dervan has reported other significant noncanonical PA influences on the DDR [17], though the mechanism seems largely unrelated since it does not depend on the presence of small, supercoiled circular dsDNA

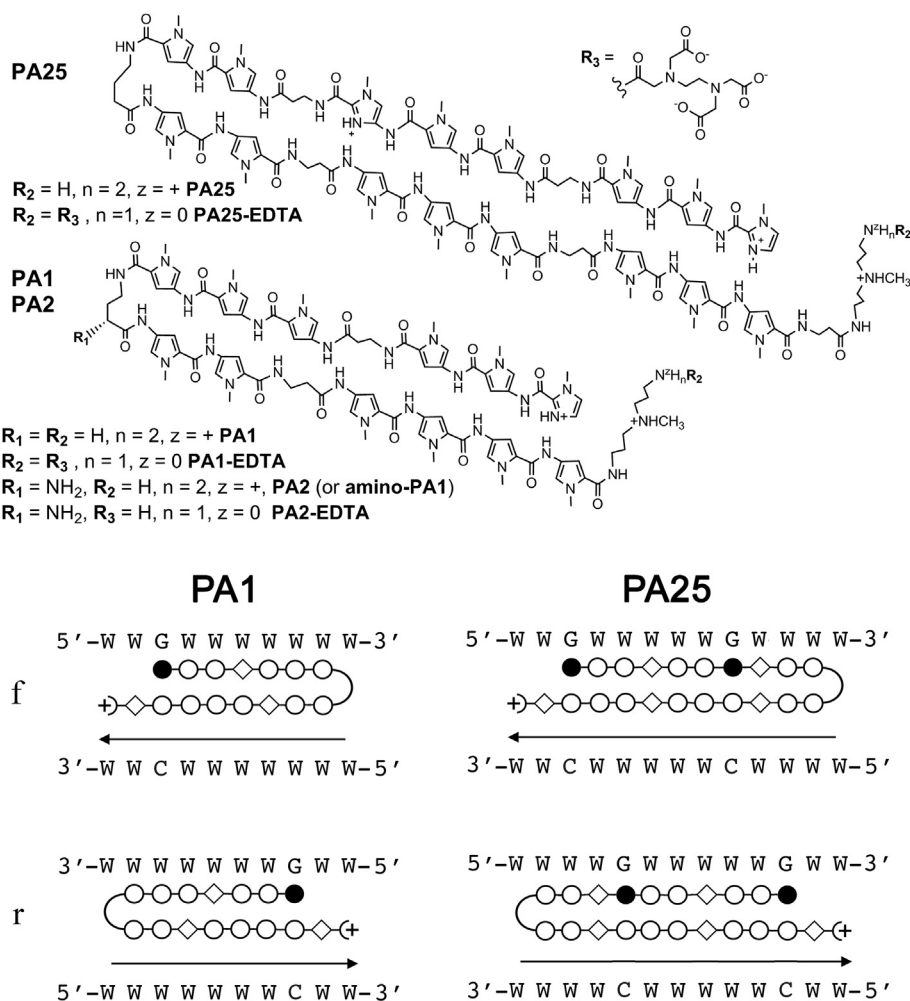


Fig. 1. Structures, predicted recognition sequences, and sample binding orientations for PA1 and PA25. Top: structures of PA1, PA25, amino-PA1 and relevant EDTA conjugates. Bottom: Target sequences for PA1 and PA25 and alternative binding orientations. Open circles  $\frac{1}{4}$  pyrrole; closed circles  $\frac{1}{4}$  imidazole; open diamond  $\frac{1}{4}$  b alanine; W  $\frac{1}{4}$  A/T. The arrowhead indicates the location of the polyamide tail and EDTA group. f refers to the forward binding orientation (amino terminus aligns with 5' end of target sequence); r refers to the reverse binding orientation (amino terminus aligns with the 3' end of the target sequence).

molecules.

In our recent studies of the interactions of PA1 with a 524 bp fragment of the E1 gene region of HPV16, we noted that even though only three perfect sites were predicted, 19 binding events were characterized using a combination of DNase I footprinting and affinity cleavage. All of these could be rationalized as perfect, single or double mismatch sites, with the two former types of sites having indistinguishable  $K_d$  values. Such tolerance for mismatch sequences is not universal among the more extensively characterized 8-ring hairpin PAs, as far as we know. For example, among 8-ring PAs, single mismatch sites are bound with  $K_d$  values  $1e7$  to  $26e55$  times weaker than cognate (perfect) sites [18]. The density and degree of overlap of sites can be attributed in part to the AT-rich nature of the HPV genome.

Our initial results led to additional questions. For example, given that the antiviral PAs are larger and distinct from their 8-ring hairpin cousins, how well do they tolerate sites with a greater number of mismatches? Another issue that emerged is the binding orientation. In theory, hairpin PAs can bind in either forward (the N terminal Im of the PA aligns with the 5' end of the duplex) or reverse orientation (the C terminal Im of the PA aligns with the 5' end of the duplex) (Fig. 1). Among the 8-ring PAs, the forward orientation is generally considered preferable ([19]; *vide infra*); the

combination of high AT density of the HPV genome and the apparently greater tolerance for mismatches indicate that larger PAs like PA1 might behave differently with respect to binding orientation. These issues could have significant implications for the mechanism of action of larger PAs (binding ten or more bp) that show significant anti-HPV activity and in some cases show broad-spectrum anti-HPV activity [13,20].

It is therefore imperative that we understand that mechanism on a molecular scale, approaching it from chemical, biophysical and biological perspectives. Here we address these issues with a comparative biophysical study of two antiviral PAs, PA1 and PA25, and their interactions with the long control region (LCR) region of HPV16.

## 2. Materials and methods

### 2.1. Polyamide preparation

PA1, dlM-PyPybPyPyPy-g-PyPybPyPyPyPyb-Ta, its EDTA conjugate (PA1-EDTA) and PA25 were prepared as previously described (PA1, PA25) [13,21]. Minimal details (HRMS only) were given to characterize PA1 and PA25 in a biology report [13], so their 600 MHz  $^1H$  and corresponding  $^{13}C$  NMR data are reported here.

The building blocks used were amino acids and amines with the following abbreviations: Im ¼ imidazole or N-methyl-2-carboxy-4-amino-imidazole, dIm ¼ desamino-imidazole or N-methyl-2-carboxy-imidazole, Py ¼ pyrrole or N-methyl-2-carboxy-4-amino-pyrrole, b-b-alanine, g-g-aminobutyric acid (NH<sub>2</sub>) (R)-2-Cbz-4-Boc-diaminobutyric acid yielding (R)-2,4-diaminobutyric acid incorporated into the PA via the acid and 4-amino group after all deprotection; Ta is the amino tail at PA C-terminus formed from NMe(CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>NH<sub>2</sub>)<sub>2</sub>. New compounds PA2, PA2-EDTA (or amino-PA1) and PA25-EDTA were prepared by analogy with PA1 and PA1-EDTA, based on literature [21,22]. For new polyamide PA2, purity was assessed by two orthogonal analytical HPLC methods, one using a Phenomenex Jupiter® Proteo C12 RP column (90 Å, 4 mm) with mobile phase of (A: 0.1% HCO<sub>2</sub>H in H<sub>2</sub>O; B: MeCN) and the a Phenomenex Synergi Polar-RP column (80 Å, 4 mm) with mobile phase (A: 0.1% NH<sub>4</sub>OH in H<sub>2</sub>O; B: MeCN). Analytical HPLC with in-line diode array and MS employed an Agilent 1100 or 1200 HPLC with thermostatted column compartment and Agilent 1956B MSD single quadrupole mass spec detector, *m/z* range up to 3000. HRMS was carried out at the proteomics center of the Donald J. Danforth Plant Sciences Center or by Prof. B. J. Bythell of UMSL.

PA stocks were prepared by dissolving in 100% DMSO and quantitated using an extinction coefficient of 92,600 M<sup>-1</sup> cm<sup>-1</sup> at 305 nm (PA1, PA2 and derivatives), or 91,700 M<sup>-1</sup> cm<sup>-1</sup> at 305 nm (PA25 and derivatives). Both values were obtained via fluorescence assay [23].

PA2 precursor dIm-PyPy-b-PyPyPy-g(NHCbz)-PyPy-b-PyPyPyPy-b-Ta and PA2 (i.e. amino-PA1) dIm-PyPy-b-PyPyPy-g(NH<sub>2</sub>)-PyPy-b-PyPyPyPy-b-Ta were prepared and characterized by methods analogous to the literature [13,22,24].

#### 21.1 PA1 ImPyPy-b-PyPyPy-g-PyPy-b-PyPyPyPy-b-Ta (3 TFA)

<sup>1</sup>H NMR (600 MHz, DMSO-d<sub>6</sub>) d ¼ 10.46 (s, 1 H), 9.93 (s, 1 H), 9.91 (s, 4 H), 9.90e9.87 (m, 2 H), 9.86 (s, 1 H), 9.83 (s, 1 H), 9.63 (br. s., 2 H), 8.12e8.02 (m, 5 H), 7.98e7.92 (m, 1 H), 7.88 (br. s., 3 H), 7.40 (s, 1 H), 7.28 (d, J ¼ 1.8 Hz, 1 H), 7.24e7.21 (m, 3 H), 7.21e7.18 (m, 4 H), 7.18e7.16 (m, 3 H), 7.15 (d, J ¼ 1.8 Hz, 1 H), 7.08 (d, J ¼ 1.8 Hz, 1 H), 7.07 (d, J ¼ 1.8 Hz, 1 H), 7.06 (s, 1 H), 7.04 (d, J ¼ 1.8 Hz, 1 H), 6.90 (d, J ¼ 1.8 Hz, 1 H), 6.89 (d, J ¼ 1.8 Hz, 1 H), 6.89 (s, 2 H), 6.85 (s, 2 H), 6.84 (d, J ¼ 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), 3.49e3.42 (m, 4 H), 3.41e3.35 (m, 4 H), 3.25e3.16 (m, 4 H), 3.16e3.10 (m, 2 H), 3.10e2.98 (m, 4 H), 2.87 (br. s., 2 H), 2.74 (d, J ¼ 4.7 Hz, 3 H), 2.55e2.51 (m, 2 H), 2.36 (t, J ¼ 7.0 Hz, 2 H), 2.28 (t, J ¼ 7.3 Hz, 2 H), 1.96e1.87 (m, 2 H), 1.82e1.74 (m, 2 H).

<sup>13</sup>C NMR (151 MHz, DMSO-d<sub>6</sub>) d ¼ 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, 122.0, 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.

#### 21.2 PA2 (Amino-PA1) dImPyPy-b-PyPyPy-g(NH<sub>2</sub>)-PyPy-b-PyPyPyPy-b-Ta (4 TFA)

<sup>1</sup>H NMR (600 MHz, DMSO-d<sub>6</sub>) d ¼ 10.56 (s, 1 H), 10.45 (s, 1 H), 9.93 (s, 2 H), 9.91 (s, 5 H), 9.89 (s, 1 H), 9.61e9.53 (m, 1 H), 8.35e8.25 (m, 3 H), 8.23 (t, J ¼ 5.3 Hz, 1 H), 8.12e8.01 (m, 4 H), 7.91 (br. s., 2 H), 7.85 (br. s., 3 H), 7.39 (s, 1 H), 7.28 (d, J ¼ 1.8 Hz, 1 H), 7.24 (d, J ¼ 1.2 Hz, 1 H), 7.22 (s, 2 H), 7.21 (d, J ¼ 1.8 Hz, 1 H), 7.20e7.17 (m, 4 H), 7.16 (d, J ¼ 1.8 Hz, 1 H), 7.15 (d, J ¼ 1.8 Hz, 1 H), 7.11 (s, 1 H), 7.09 (d, J ¼ 1.2 Hz, 1 H), 7.07 (d, J ¼ 1.2 Hz, 1 H), 7.06 (d, J ¼ 1.2 Hz, 1 H), 7.05 (d, J ¼ 1.2 Hz, 1 H), 7.03 (s, 1 H), 6.98 (d, J ¼ 1.8 Hz, 1 H), 6.93 (d, J ¼ 1.8 Hz, 1 H), 6.89 (d, J ¼ 1.8 Hz, 2 H), 6.86 (d, J ¼ 1.8 Hz, 1 H), 6.85 (d, J ¼ 1.8 Hz, 1 H), 3.99 (s, 3 H), 3.85 (s, 6 H), 3.84 (s, 3 H), 3.84 (s,

3 H), 3.84 (s, 9 H), 3.82 (s, 9 H), 3.80 (s, 3 H), 3.48e3.42 (m, 4 H), 3.42e3.35 (m, 2 H), 3.35e3.23 (m, 2 H), 3.21e3.16 (m, 1 H), 3.16e3.09 (m, 2 H), 3.09e2.98 (m, 4 H), 2.91e2.82 (m, 2 H), 2.74 (d, J ¼ 4.7 Hz, 3 H), 2.56e2.51 (m, 4 H), 2.36 (t, J ¼ 7.3 Hz, 2 H), 2.03e1.94 (m, 2 H), 1.94e1.86 (m, 2 H), 1.82e1.74 (m, 2 H).

<sup>13</sup>C NMR (151 MHz, DMSO-d<sub>6</sub>) d ¼ 171.0, 167.8, 167.8, 165.3, 161.9, 161.3, 161.3, 161.2, 158.5, 158.4, 158.3, 158.1, 158.1, 157.9, 157.6, 156.0, 138.7, 127.9, 126.8, 126.3, 123.2, 123.0, 122.8, 122.8, 122.7, 122.7, 122.7, 122.4, 122.2, 122.2, 122.2, 122.2, 122.1, 122.1, 121.9, 121.4, 120.5, 118.6, 118.5, 118.4, 118.4, 118.4, 118.3, 118.2, 118.0, 117.9, 117.9, 117.6, 115.6, 104.9, 104.8, 104.8, 104.3, 104.2, 104.2, 104.0, 104.0, 53.3, 52.1, 50.8, 40.4, 40.0, 36.2, 36.2, 36.1, 36.1, 36.0, 36.0, 35.9, 35.8, 35.6, 35.5, 35.5, 35.4, 35.1, 24.0, 21.8.

HRMS (ESI) calcd for C<sub>91</sub>H<sub>112</sub>N<sub>32</sub>O<sub>16</sub> [MH]<sup>+</sup>, 1909.9012, found, 1909.9018. HPLC purity: 95%.

The EDTA conjugate of PA2 was prepared via the same method as previously described for PA1-EDTA conjugate [21]. To a vigorously-stirred mixture of EDTA dianhydride (25 mg, 0.096 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 Cbz-protected PA2 (12 mg, 0.0055 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 (1.0 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 mm polyethylene filter, and purified by reversed-phase HPLC [A: H<sub>2</sub>O (0.1% TFA), B: MeOH] using a Phenomenex Jupiter 250 x 80 mm, 5 mm, 100 Å, C<sub>18</sub>(2) column. Concentration of pooled fractions checked by analytical HPLC/MS, followed by lyophilization gave the Cbz-protected PA2-EDTA conjugate (9.9 mg, 73% yield) as a white, fluffy solid.

The Cbz-protected PA2-EDTA conjugate (3.7 mg, 0.0015 mmol) was treated with 10% trifluoromethylsulfonic acid/trifluoroacetic acid (0.30 mL) at 25 °C for 30 min to remove the Cbz-protecting group [25]. The mixture was then flash frozen with liquid nitrogen, treated with N,N-dimethylformamide (0.30 mL) and allowed to warm to 25 °C. The mixture was diluted with N,N-dimethylformamide (0.30 mL) and water (0.30 mL), filtered through a 20 mm polyethylene filter, and purified by reversed-phase HPLC [A: H<sub>2</sub>O (0.2% formic acid), B: MeOH] using a Macherey Nagel

50 x 1 mm, 5 mm, C<sub>18</sub> column. Concentration of pooled fractions, followed by lyophilization gave the desired EDTA conjugate, dIm-PyPy-b-PyPyPy-g(NH<sub>2</sub>)-PyPy-b-PyPyPyPy-b-Ta-EDTA (2 HCO<sub>2</sub>H) (PA2-EDTA, i.e. amino-PA1-EDTA) (3.0 mg, 86% yield) as a white, fluffy solid: exact mass [MH]<sup>+</sup> ¼ 2183.9813, experimental (ESI) [MH]<sup>+</sup> ¼ 2183.9727.

#### 21.3 PA25 ImPyPy-b-PyPyIm-b-PyPy-g-PyPy-b-PyPyPyPy-b-PyPyPyPy-b-Ta (4 TFA)

<sup>1</sup>H NMR (600 MHz, DMSO-d<sub>6</sub>) d ¼ 10.46 (s, 1 H), 10.35 (s, 1 H), 9.95 (s, 1 H), 9.92 (s, 1 H), 9.91 (s, 4 H), 9.90e9.87 (m, 2 H), 9.86 (s, 1 H), 9.85 (s, 1 H), 9.82 (s, 1 H), 9.59 (br. s., 2 H), 8.11e8.01 (m, 7 H), 7.97e7.91 (m, 2 H), 7.87 (br. s., 3 H), 7.51 (s, 1 H), 7.39 (s, 1 H), 7.32 (d, J ¼ 1.8 Hz, 1 H), 7.28 (d, J ¼ 1.2 Hz, 1 H), 7.23e7.21 (m, 2 H), 7.20 (d, J ¼ 1.8 Hz, 1 H), 7.20e7.18 (m, 6 H), 7.17e7.16 (m, 2 H), 7.16 (d, J ¼ 1.8 Hz, 1 H), 7.15 (d, J ¼ 1.8 Hz, 1 H), 7.07 (d, J ¼ 1.2 Hz, 1 H), 7.06e7.05 (m, 2 H), 7.04 (d, J ¼ 1.2 Hz, 1 H), 6.89 (d, J ¼ 1.2 Hz, 1 H), 6.89e6.87 (m, 4 H), 6.87e6.84 (m, 4 H), 6.84 (s, 1 H), 3.99 (s, 3 H), 3.95 (s, 3 H), 3.84 (s, 12 H), 3.83 (s, 6 H), 3.82 (s, 3 H), 3.82 (s, 6 H), 3.81 (s, 6 H), 3.80 (s, 3 H), 3.79 (s, 6 H), 3.57e3.50 (m, 2 H), 3.48e3.42 (m, 4 H), 3.42e3.34 (m, 2 H), 3.24e3.16 (m, 2 H), 3.16e3.09 (m, 2 H), 3.09e2.98 (m, 4 H), 2.91e2.83 (m, 4 H), 2.74 (d, J ¼ 4.7 Hz, 3 H), 2.58e2.51 (m, 6 H), 2.36 (t, J ¼ 7.0 Hz, 2 H), 2.27 (t,

J ¼ 7.6 Hz, 2 H), 1.97 (m, 4 H), 1.78 (td, J ¼ 6.6, 13.8 Hz, 4 H).  
<sup>13</sup>C NMR (151 MHz, DMSO-d<sub>6</sub>) δ ¼ 171.0, 169.2, 167.9, 167.8, 167.8, 161.3, 161.3, 161.3, 158.8, 158.5, 158.5, 158.5, 158.4, 158.4, 158.3, 158.1, 157.9, 155.9, 138.7, 136.2, 133.7, 126.8, 126.3, 123.0, 122.9, 122.8, 122.8, 122.8, 122.7, 122.7, 122.7, 122.2, 122.1, 122.1, 122.1, 122.0, 122.0, 121.9, 121.9, 121.9, 121.9, 121.8, 121.4, 119.5, 118.6, 118.5, 118.2, 118.1, 117.9, 117.8, 117.5, 115.5, 114.4, 105.7, 104.9, 104.8, 104.7, 104.3, 104.3, 104.2, 104.2, 104.1, 104.0, 104.0, 103.9, 103.9, 53.3, 52.3, 52.1, 40.4, 40.0, 39.3, 38.2, 36.2, 36.2, 36.1, 36.1, 36.1, 36.0, 35.8, 35.8, 35.6, 35.6, 35.5, 35.4, 35.2, 35.1, 35.0, 34.9, 33.3, 25.7, 24.1, 21.8, 21.7.

HRMS (ESI) calcd for C<sub>120</sub>H<sub>144</sub>N<sub>42</sub>O<sub>22</sub> [MH]<sup>+</sup>, 2526.1447, found, 2526.1557, HPLC purity: 98%.

The EDTA conjugate of PA25 was prepared via the same method as previously described for PA1-EDTA conjugate [21]. 17 mg (0.067 mmol) of EDTA dianhydride were stirred vigorously at 55 °C in 0.5 mL DIEA, 0.25 mL DMF and 0.25 mL DMSO. To the anhydride, a biphasic mixture of 10 mg (0.0034 mmol) PA25, 0.5 mL DIEA, and 0.5 mL DMF over 15 min via syringe. After HPLC (Phenomenex Jupiter 250 ¼ 1 mm, C<sub>12</sub>, 4 m column, A: H<sub>2</sub>O 0.1% TFA, B: MeOH) and lyophilization, 3.5 mg (0.0012 mmol, 35% yield) of white fluffy solid were collected, PA25-EDTA ImPyPy-b-PyPyIm-b-PyPy-g-PyPy-b-PyPyPy-b-PyPyPy-b-Ta-EDTA (2 TFA), LCMS [M Na]<sup>+</sup> 12.11 found 1413.0, HPLC purity: 86%.

## 22. Preparation of LCR fragment

To study this region of the HPV16 genome (Accession # AF125673) quantitatively via CE, two fragments were generated. To generate the 7348-7798 bp fragment, two primer oligomers were used: (top strand) 5<sup>0</sup>-FAM-ATT GTG TTG TGG TTA TTC AT-3<sup>0</sup> and (bottom strand) 5<sup>0</sup>-HEX-CAT GTA TGA ACT AGG GTG AC-3<sup>0</sup>, where FAM indicates the fluorescein dye and HEX indicates the hexachlorofluorescein dye. To generate the 7662-122 bp fragment by PCR, two primers were used: (top strand) 5<sup>0</sup>-FAM-TAA ATC ACT ATG CGC CAA CGC-3<sup>0</sup> and (bottom strand) 5<sup>0</sup>-HEX-CCT GTG GGT CCT GAA ACA TTG-3<sup>0</sup>. All oligonucleotides were obtained from IDT (Coralville, Iowa).

## 23. Analysis of DNA-PA interactions

### 23.1. DNase I footprinting

DNase I footprinting experiments were performed in TKMC buffer (10 mM Tris, 10 mM KCl, 5 mM MgCl<sub>2</sub> and 5 mM CaCl<sub>2</sub>) with 10 mM CHAPS as previously described [26]. DNA concentrations were adjusted between 200 pM and 1 nM as needed. The reaction was initiated with 0.01 (0.1 U DNase I; the amount of DNase I was dependent on the age of the enzyme, the DNA fragment, and DNA concentration. Reaction products were purified using the Qiagen MiniElute purification kit. Capillary electrophoresis (CE) analyses of samples were conducted at the DNA Core Facility at University of Missouri. Data were processed using Genemarker V1.97 software (Softgenetics LLC, State College, PA). DNase I cleavage products were mapped using Sanger sequencing (USB, Affymetrix, Santa Clara, CA). Peaks in the footprint were normalized to a peak not sensitive to PA concentration and plotted as fraction bound vs. PA concentration; the data were fit to a binding isotherm as previously described [26].

### 23.2. Affinity cleavage

0.8 eq of Fe(NH<sub>4</sub>)<sub>2</sub>(SO<sub>4</sub>)<sub>2</sub> was added to the polyamide-EDTA-Fe conjugate and incubated for 5 min. The DNA fragment was then added to 1 nM and the resulting mixture incubated 4 (6 h at room temperature in 10 mM Tris, 10 mM CHAPS, pH 7.5. The PA-EDTA conjugate concentration was varied from 5 nM to 250 nM and

reported at 50 nM. To initiate the cleavage reaction, 5 mL of fresh 100 mM DTT was added to a 100 mL reaction and incubated 30 min to 4 h. The reaction was quenched and products purified using a Qiagen PCR purification kit. Fragments were analyzed by capillary electrophoresis. Affinity cleavage results were mapped using Maxam-Gilbert reactions on 5<sup>0</sup> FAM endlabeled DNA [27].

## 3. Results

### 3.1. Contexts for PA site studies

The first reported study of anti-HPV PAs involved tandem (covalently linked) hairpin PAs that targeted a 12 bp recognition sequence for the HPV transcription and replication factor E2 [28]. The binding of the tandem PA to the minor groove of the E2 recognition sequence (ACCN<sub>6</sub>GGT) prevented this protein from binding the major groove [28].

PA1 and PA25 (Fig. 1) both represent evolution from the above design. They are both single hairpins with a ten bp recognition sequence of 5<sup>0</sup>-W<sub>2</sub>GW<sub>7</sub>-3<sup>0</sup> for PA1 (where W ¼ A or T) and a 13 bp recognition sequence and 5<sup>0</sup>-W<sub>2</sub>GW<sub>5</sub>GW<sub>4</sub>-3<sup>0</sup> for PA25. As mentioned above, they have both been shown to nearly eliminate the HPV16 virus *in vivo* [13].

The LCR of HPV16 starts at bp 7158 and ends at bp 83 of the circular genome. It features a number of binding sites for the E2 proteins and the E1 hexamer helicase assembly site [29] that is typically involved in initiating replication of the viral genome [13]. In this study, we examine PA1, PA2 and PA25 binding behaviors toward the sequence comprising 7348-122 bp of the HPV16 genome. In this region, there are nine predicted PA1 recognition sites; five of these are concentrated in the region where E1 and E2 bind (positions commencing at nt 7903, 14, 15, 26, and 62 and indicated with bolded arrows in Fig. 2). There are no predicted perfect PA25 sites in this region, but it does feature 2 predicted single mismatch and almost 30 predicted double mismatch sites. Here "mismatch" refers to an alignment between PA residue and nt that does not conform to the reported PA-DNA recognition rules [30-36]. Typically this means a G or C appears where the PA residue targets W, or W aligns with an Im residue.

Since this fragment of the genome is too large for accurate studies with capillary electrophoresis (CE) [21,37-39], this sequence was divided into two overlapping fragments, 7348- 7798 bp (451 bp) and 7662-122 bp (365 bp).

### 3.2. Studies of PA1-HPV16 LCR interactions

#### 3.2.1. Affinity cleavage with PA1-EDTA

In the affinity cleavage (AC) technique, hydroxyl radical cleavage of DNA at the binding site is achieved using FeEDTA [40-42]. Since the EDTA is located on the tail of the PA (open end of the hairpin; Fig. 1), the location of the tail on the DNA sequence is indicated with the cleavage patterns. Information about PA binding orientation is also provided by AC data (*vide infra*).

Fig. 2 shows the affinity cleavage of PA1-EDTA with the 7348- 122 bp DNA fragment of HPV16. As noted previously, 40 (50 bp on the ends of the duplex are not accessible via CE [21]. 38 AC patterns were observed, and all AC sites have 3<sup>0</sup> shifts consistent with minor groove binders [43]. Some AC patterns are extended, suggestive of overlapping binding events (AC 5, 14, 22, and 24). Prediction of the PA1 binding sites according to the rules developed by the Dervan lab showed that one AC pattern can be rationalized by various possible predicted PA1 binding sites. Binding site assignments assume that forward sites are preferred over reverse ones (see below), and sites with fewer mismatches were preferred over those with a higher number of mismatches.



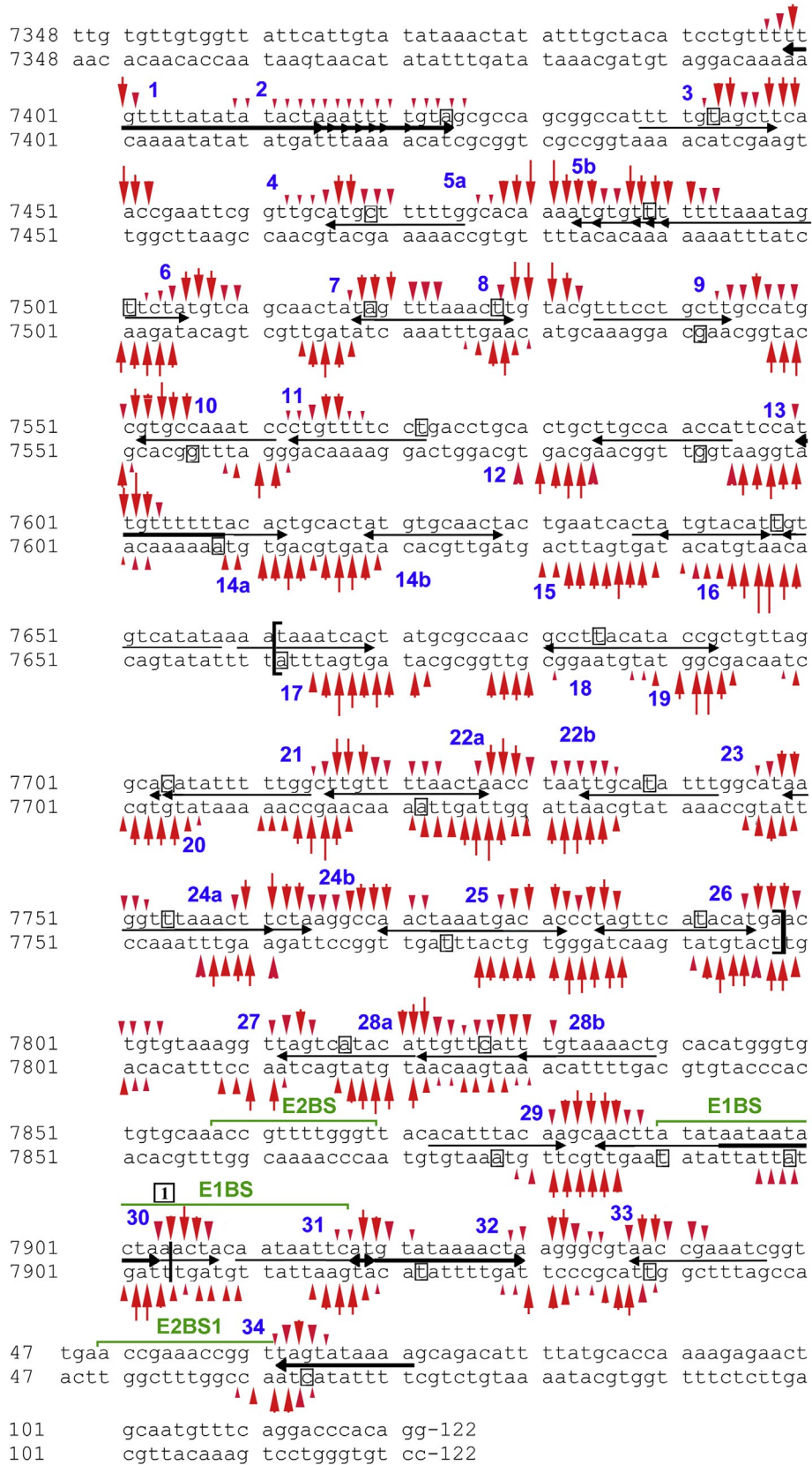


Fig. 2. Sequence map of PA1 affinity cleavage (AC) and footprinting in the 7348-122 LCR region of HPV16. Affinity cleavage was observed at 50 nM PA1-FeEDTA conjugate. Since signal-to-noise ratios vary with fragment length, red vertical arrows indicate intensities relative to other arrows in a site, not to other sites. Small arrow heads reflect particularly weak affinity cleavage. Solid horizontal arrows indicate PA1 sites consistent with AC data. Bolded horizontal arrows indicate perfect-match (canonical) sequences. Boxed nt indicates a site of integration. E protein binding sites are indicated in green. Brackets indicate where the two overlapping duplexes begin and end. See text for details. Possible reverse binding sites are omitted for clarity, as are forward double mismatches in regions where they overlap significantly with perfect or single mismatch sites.



Relative to the HPV16 fragment studied previously (2150e2622), the 7348-122 region is slightly more GC rich (36.4% vs. 33.2%) and 40% longer but harbors about twice the number of AC sites. This would suggest that either there are runs of more suitable PA1 sites or that there is a greater number of sites with multiple binding orientations. While there is a greater number of perfect sites in this region (7 vs. 3), they account for only a small fraction of the observed AC sites. As described below, the larger number of AC patterns result from sites with greater numbers of mismatches, and as a consequence of this, PA1 binding the same sequence in more than one orientation.

AC sites 1, 13, 30e32, and 34 can be rationalized by forward orientation perfect match binding sites only. Since single and double mismatch sites have already been shown to be abundant and generally of affinity similar to those observed for perfect mismatches [21], these sites will be tabulated but not discussed. Discussion is restricted to binding behaviors that are distinct from those observed earlier in the 2150-2672 fragment [21].

AC sites 11, 19, 20, and the 5<sup>0</sup> part of AC 25 can be rationalized by forward triple mismatch sites only. In AC sites 9 and 10, there are two peaks on the top strand; on the bottom strand, there is only one AC peak. This could be explained by differences in resolution (bottom strand vs. top strand). AC sites 9 and 10 are unusual cases because AC 9 can be explained by PA1 bound to a forward triple mismatch site while AC 10 can only be rationalized by a quadruple mismatch or by PA1 binding to the DNA in some other conformation (half hairpin or linear).

AC sites 18 and 33 can be rationalized by PA1 bound to a forward quadruple mismatch site or to a reverse triple mismatch site; these cannot be distinguished.

For AC 12, the top strand AC pattern is missing even though the AC pattern further down the sequence was present (AC 13) and the bottom strand pattern was present for this same site. We suspect resolution issues associated with CE analysis. This AC pattern could be rationalized only by a forward quadruple mismatch site.

In summary, for the 7348-122 bp region of the circular HPV16

Table 1  
PA1 sites on 7348-122 fragment of HPV16.<sup>a</sup>

AC	Sequence	Position	Site type	Integration nt	K <sub>a</sub> , nM
1	GTTTT TTGTTTTATA TATAC	7399-7408	Perfect		
2	AAATF TAGTATATAT AAAAC	7415-7406	Perfect	7424	1.8 ± 0.3
2	CCGCG ATGTTTTAAA TCATA	7424-7415	Perfect <sup>b</sup>		
3	GGTTG AAGCTACAAA ATGGC	7448-7439	Double	7443	3.4 ± 0.7
4	GTTGC ATGCTTTTTG GCACA	7466-7475	Double	7469	2.3 ± 1.0
5a,b	AAAAT GTGTTTTTTT AAATA	7485-7494	Single	7489	1.0 ± 0.3
5b	ACAAA ATGTGTTTTT TAAAA	7483-7492	Single		
5b	TGTGT TTTTTTAAAT AGTTC	7489-7498	Single		
6	TGACA TAGAACTATT TAAAA	7505-7496	Single		
7	AACTA TAGTTTAAAC TTGTA	7518-7527	Single	7519	2.6 ± 1.4
8	CGTAC AAGTTTAAAC TATAG	7529-7520	Single	7528	1.5 ± 0.5
9	ATGGC AAGCAGGAAA CGTAC	7544-7535	Triple	7542	6.8 ± 0.7
10	CATGC GTGCCAAATC CCTGT	7552-7561	Quadruple	7556	1.9 ± 0.5
11	AATCC CTGTTTTCTT GACCT	7563-7572	Triple	7572	3.1 ± 2.7
12	ACTGC TTGCCAACCA TTCCA	7585-7594	Quadruple	7592	5.9 ± 1.6
13	TTCCA TTGTTTTTTA CACTG	7600-7609	Perfect	7608	1.9 ± 0.7
14a	GTGCA GTGTAAAAAA CAATG	7612-7603	Single		
14b	GCACT ATGTGCAACT ACTGA	7619-7628	Triple		
15	TCACT ATGTACATTG TGTCA	7640-7649	Double	7648	1.5 ± 0.5
16	ACATT GTGTACATA AAATA	7649-7658	Double		
16	ACACA ATGTACATAG TGATT	7638-7647	Double		
17	GCATA GTGATTTATT TTATA	7669-7660	Single	7662	1.4 ± 0.6
18	CCAAC GCCTTACATA CCGCT	7681-7690	Quadruple	7685	4.4 ± 2.1
18	GGTTG CGGAATGTAT GGCGA	7690-7681	Quadruple		
19	AACAG CCGTATGTAA GGCGT	7693-7684	Triple		
20	AATCC GTGTATAAAA ACCGA	7702-7711	Single	7704	1.9 ± 0.0
20	AGGCA CATATTTTTG GCTTG	7704-7713	Triple		
20	TAGGC ACATATTTTT GGCTT	7703-7712	Double		
21	TTGGC TTGTTTTAAC TAACC	7716-7725	Single	7722	1.5 ± 0.5
22a	TAGGT TAGTTAAAAA AAGCC	7727-7718	Single		
22b	CCTAA TTGCATATTT GGCCA	7734-7743	Single	7739	2.6 ± 1.2
23	GGCAT AAGGTTTAAA CTTCCT	7749-7758	Single	7754	1.6 ± 0.2
24a	CTTAG AAGTTTAAAC CTTAT	7761-7752	Single		
24b	GGCCT TAGAAGTTTA AACCT	7764-7755	Single		
24b	TCCGG TTGATTTACT GTGGG	7770-7779	Single	7774	10.8 ± 0.2
25	CTAGG GTGTCATTTA GTTGG	7782-7773	Double		
25	CACCC TAGTTACATC ATGAA	7785-7794	Double	7792	7.3 ± 2.0
26	AGTTC ATGTATGAAC TAGGG	7796-7787	Double		
27	AAGGT TAGTCATACA TTGTT	7812-7821	Double	7817	19.8 ± 8.0
28a	ATACA TTGTTCATTT GTAAA	7822-7831	Single	7827	13.1 ± 1.1
28b	TTTCAT TTGTAAAACT GCACA	7830-7839	Single		
29	GTTGC TTGTAATGT GTAAC	7882-7873	Single	7878	8.1 ± 1.4
29	CAAGC AACTTATATA ATAAT	7885-7894	Single	7890	0.8 ± 0.2
30	TAGTT TAGTATTTT ATATA	7903-7894	Perfect	7899	0.7 ± 0.3
31	TATAC ATGAATTATT GTAGT	15-6	Perfect	11	1.4 ± 0.8
31	AATTC ATGTATAAAA CTAAG	14-23	Perfect	18	1.1 ± 0.1
32	GCCCT TAGTTTTATA CATGA	26-17	Perfect		
33	GGCGT AACCGAAATC GGGTTG	35-44	Quadruple	35	3.5 ± 0.8
34	CCGGT TAGTATAAAA GCAGA	62-71	Perfect	64	1.2 ± 0.5

<sup>a</sup> Position corresponds to the PA arrowhead/where the sequence begins. 4 nt flanking sequences are included.

<sup>b</sup> Reverse orientation.

genome, eight perfect match forward and four perfect match reverse PA1 binding sites were observed by AC. A large number of single, double, triple, and, more rarely, quadruple mismatch sites that bind in both forward and reverse fashion were also observed in this sequence.

### 3.2.2 Quantitative footprinting studies of PA1

DNase I footprinting was used to evaluate all PA binding sites on this portion of the HPV16 genome. Integration positions were chosen based on location with respect to the footprint and for optimum signal-to-noise ratios [21]. In addition to the footprint itself, signal-to-noise ratios are influenced by the distance of the position from the dye and the slight sequence preference of DNase I [44,45]. HPV16 sequences that are not amenable to cleavage by the enzyme have been reported [21] and are present in this region as well (7831-7840 and 7857-7866). It is also important to note that since DNase I is larger than PA1, steric encumbrance occurs and the footprint is always larger than that covered by the PA, so integration sites can lie outside the PA1 binding site.

The integration sites for DNase I footprinting are indicated with solid boxes in Fig. 2. Twenty-nine nucleotide positions were integrated to obtain  $K_d$ 's (Table 1). In most cases, a single  $K_d$  corresponds to each identified AC site, even if there are multiple possible sequence possibilities for an AC pattern. Note that the  $K_d$  for the PA1-EDTA conjugate as measured by footprinting is indistinguishable to that obtained for unmodified polyamide PA1 [21].

Table 1 summarizes the  $K_d$  values obtained by DNase I. All of the  $K_d$ 's for perfect match sites range from 0.7 nM to 1.9 nM, similar to those obtained previously for such sites on a different part of the HPV16 genome [21]. Most  $K_d$ 's for single mismatch sites fall between 0.8 and 2.6 nM, demonstrating good tolerance for such sites (see below). There are three higher  $K_d$ 's: 8.1 nM at 7878, 10.8 nM  $K_d$  quantitated at 7774, and 13.1 nM at 7827. The 7770-7779 site is either a reverse single mismatch or a forward double mismatch site. The  $K_d$  quantitated at 7827 could be the result of the binding affinity being determined at the mismatch nucleotide, where local binding may not be as avid. In this case, this was the location of the most favorable signal-to-noise ratio in the footprint.

$K_d$ 's for the double mismatch sites range from those typical of perfect and single mismatch sites (1.5 nM to 3.4 nM) to those for three mismatches, with  $K_d$ 's between 7 and 20 nM, the latter with a high relative error. Triple mismatch sites displayed slightly higher  $K_d$ 's than typical of most with fewer mismatches (3.1 to 6.8 nM). The same pattern holds for the quadruple mismatch binding sites (1.9 nM to 5.9 nM). Given the emphasis on perfect matches and the general intolerance of mismatches in the smaller, 6-8-ring hairpin polyamides [18,46,51], we were somewhat surprised to observe such high affinities at sites with a large number of mismatches. However, we note that Sugiyama reports 18 bp specificity with a tandem hairpin in which three hairpins are linked together to form one large PA, though it is for a human telomeric repeat sequence and the actual measured specificity is low ( $K_d(\text{mismatch})/K_d(\text{match})$  is 2.6) [52].

### 3.2.3 Effect of a chiral g turn on PA binding orientation

Previous reports indicated that for some shorter hairpin PAs (i.e. 6-ring) and longer PAs recognizing 10 bp, the introduction of a chiral center in the gamma turn of the hairpin PA resulted in a strong preference for the forward binding orientation [35,53,54]. Sugiyama found discrimination between A/T and T/A base pairs with chiral *b*-hydroxy-*g*-aminobutyric acid/*b*-alanine pairs [55]. To examine how the introduction of a chiral group in the gamma turn of PA1, a chiral version of PA1 was prepared with chiral *g*-turn (R)-2,4-diaminobutyric acid, giving PA2; the EDTA conjugate of PA2 was also prepared: PA2-EDTA (or amino-PA1). Binding sites on the

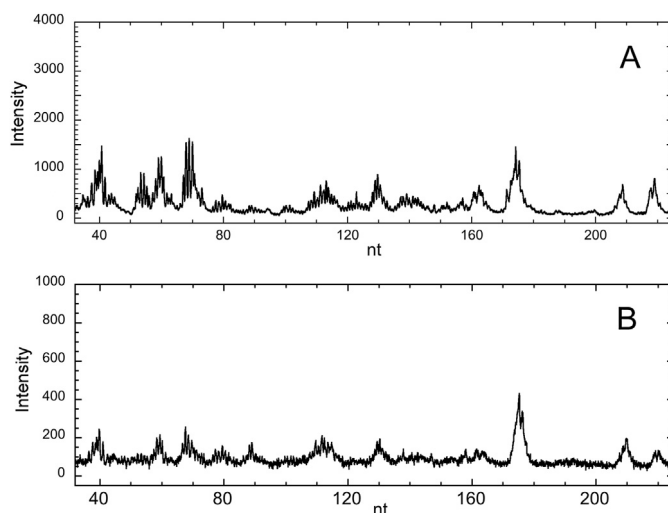


Fig. 3. Effect of chiral turn on affinity cleavage patterns. Partial electropherograms of the 7348-7798 portion of HPV16 when treated with A) PA1-EDTA; B) amino-PA1-EDTA. 1 nM DNA duplex, 50 nM PA. See experimental for additional details.

450 bp fragment (7348-7798) were evaluated using affinity cleavage at 50 mM conjugate. As shown in Fig. 3, there is no appreciable difference between the binding sites for PA1-EDTA and amino-PA1-EDTA, or PA2-EDTA. Thus in the context of this larger PA, the introduction of an (R)-amino-gamma turn does not influence binding orientation of the PA family.

## 3.3 Studies of PA25-HPV16 LCR interactions

### 3.3.1 Analysis of PA25 affinity cleavage patterns

In spite of the fact that there are no perfect match sites for PA25 in this 678 region, 31 distinct AC patterns were observed for PA25 and are numbered in Fig. 4. Extended AC patterns 1, 2, 4, 6, 8, 17, 18, 22, 25, 28 have multiple peaks (high intensity surrounded by those with lower intensity) in their patterns for a total of 43 discernable PA25 binding sites.

PA25 was initially designed to recognize the sequence  $5'$ - $W_2GW_5GW_4-3'$ , the longest PA studied in this fashion to date. In attempting to reconcile AC patterns with probable binding sites, the first question is: How is the PA25 site anchored? Unlike monimidazole derivative PA1, PA25 has two Im residues and therefore the theoretical ability to recognize two G's on either side of a 5 residue/nt spacer ( $5'$ - $GN_5G-3'$ ). Is the  $GW_5G$  site, or at least a  $GN_5G$  site, required for PA25 binding? To address this, all  $5'$ - $GN_5G-3'$  patterns were mapped on the fragment (Fig. S1). There are 14 on the top strand and 27 on the bottom, with a number on the latter overlapping.

What is more striking is that 16 AC patterns, roughly half of the total number of AC patterns for PA25 (1, 5, 7, 12, 13, 16, 17abc, 18a, 20, 25ab, 26, 27, 31) exist in locations that are not near a  $5'$ - $GN_5G-3'$  sequence at all. Conversely, there are two  $5'$ - $GN_5G-3'$  at 7620-7640 that could be rationalized as multiple mismatch sites (triple at 7638 and quadruple at 7632) but for which no AC patterns are observed. It is possible that low signal-to-noise ratios common for positions over 200 bp from the dye prevent us from detecting these sites. Regardless of this possibility, it is clear that the  $5'$ - $GN_5G-3'$  motif is not required for PA25 binding.

This begs the question of which, if either, G the first (N-terminal) or second (internal) is more important in anchoring a site/rationalizing an AC pattern. To address this, the sequences at the observed AC patterns were analyzed (Tables S2-S4). The goal was

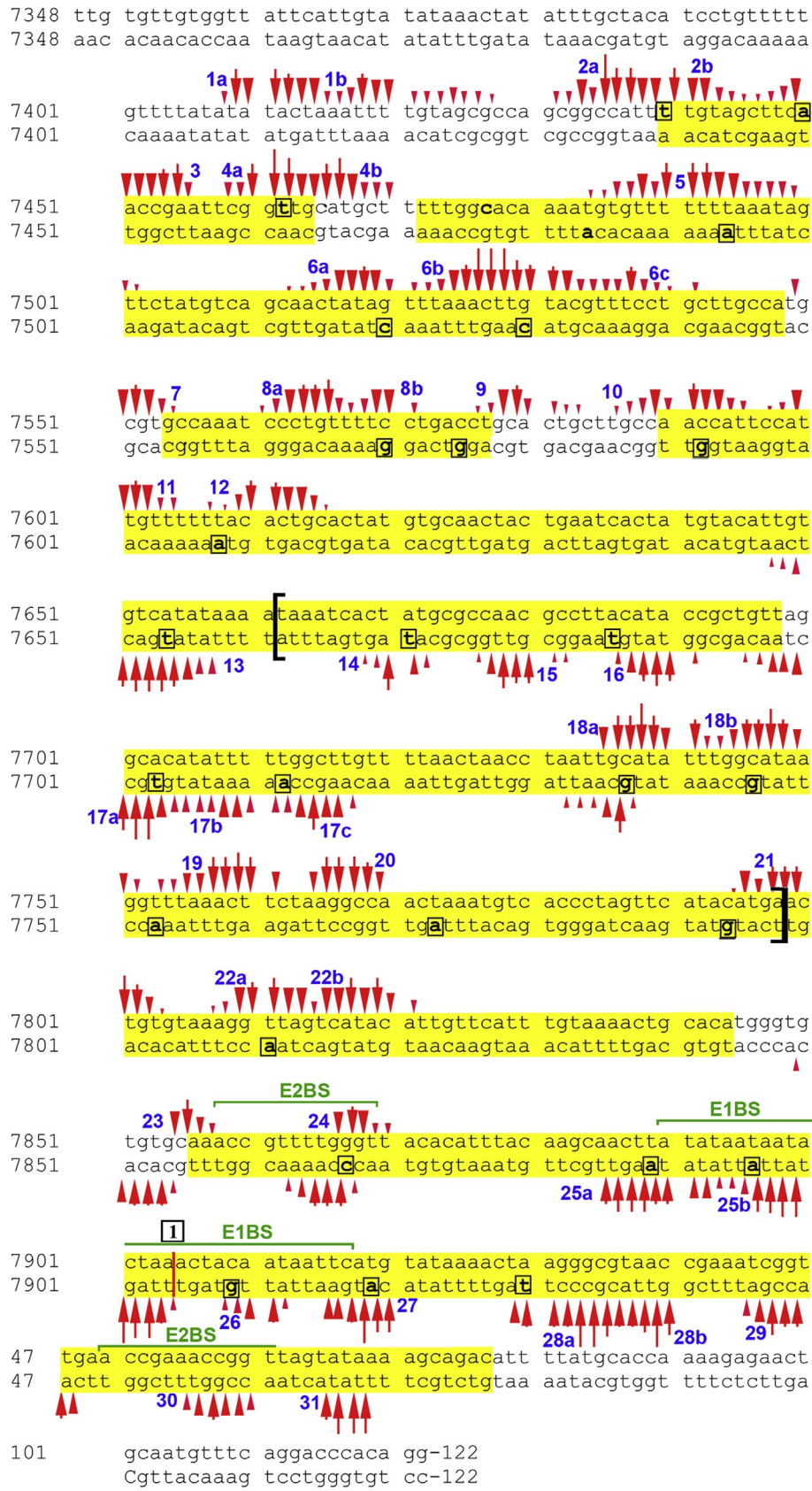


Fig. 4. Sequence map of PA25 affinity cleavage (AC) and footprinting in the 7348-122 LCR region of HPV16.



to identify a binding site with the smallest number of mismatches, forward preferred if the number of mismatches is minimized, and then catalog whether the sequence has a first G, second G, both or neither. Two thirds of these sequences are either triple, quadruple, or quintuple mismatches. Over 80% of these sequences were anchored with one G or the other with no preference for one over the other with no obvious correlation with the number of mismatches. Only 6 sites had either zero or both G's. The A/T rich character of the genome is reflected in this observation.

Thus it appears that neither G is critical to the PA25 recognition sequence, and if the fragment were even more A/T rich, we predict that both PAs would populate those sites as much as any other part of the sequence.

### 3.3.2 Quantitative footprinting studies with PA25

Unlike for PA1 where the binding sequences were far more straightforward, the data do not permit us to establish exact sequences for PA25 binding. For this reason, regions supported by DNase I footprints are presented *in lieu* of specific sequences (Fig. 4). Almost all of the sequence accessible via CE exhibits PA25 footprinting behavior at 10 nM PA (DNA is 1 nM). These results are consistent with the size of DNase I, the relatively large size of PA25, and the high tolerance for mismatches. Discrepancies between footprinting and AC patterns cannot be interpreted due to signal-to-noise issues in those regions.

The 27  $K_{ds}$  obtained from the sites of integration appear in Table S1 and are summarized in Fig. 5. 18 of the  $K_{ds}$  are between 0.7 and 1.9 nM. There are a few slightly weaker  $K_{ds}$  but all are less than 3-fold above this range and thus not significantly different. All of these values compare well to the  $K_d$  of  $1.2 \pm 0.5$  nM obtained for two overlapping perfect match sites integrated at position 3662 (outside the LCR; Gaofei He, unpublished results). Given that a large majority of sites are likely to feature triple or quadruple mismatch sequences, this indicates a high tolerance for mismatches. This is discussed in comparison to PA1 below (Section 4.2).

## 4. Discussion

### 4.1 Hairpin polyamide orientation of DNA binding

There are numerous examples of modifications to smaller hairpin PAs (6-10 rings) that influence binding orientation. For

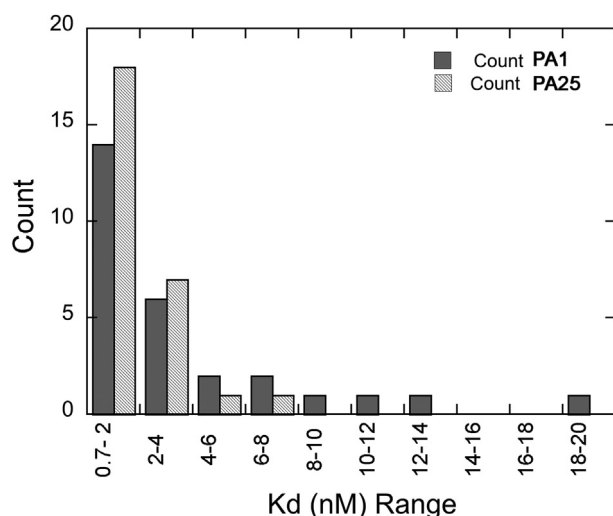


Fig. 5. Distribution of  $K_{ds}$  obtained from the 7348-122 portion of the LCR of HPV16 for PA1 and PA25. Table S1 summarizes the integration positions and  $K_{ds}$  with errors for PA25.

Table 2  
Influence of substitutions in hairpin PA on orientation of binding.<sup>a</sup>

Case#	Polyamide	# rings	R vs. S	Target sequence	Fwd/Rev <sup>b</sup>
1 <sup>c</sup>		6		WWWCW	16
					4
2 <sup>d</sup>		6	R	WWWCW	>1000
			S		5
3 <sup>e</sup>		8	R	WCWCCWW	>1000
			S		0.05
4 <sup>f</sup>				WCCWCCW	2:500
					2
5 <sup>g,h</sup>		8		WCGWC	Fwd
					Rev
			R		Fwd

<sup>a</sup> Open circles represent pyrrole, closed circles represent imidazole, **b** represents **b**-Ala, half circle is **g**-turn. Dp  $N,N'$ -dimethylaminopropylamine.

<sup>b</sup> Refers to ratio of  $K_a$ 's for binding in the forward orientation to the  $K_a$  for binding in the reverse orientation.

<sup>c</sup> [56].

<sup>d</sup> [35].

<sup>e</sup> [57].

<sup>f</sup> [53].

<sup>g</sup> [58].

<sup>h</sup> Preference determined by melting temperature experiments.

example, the 8-ring ImPyPy-**g**-PyPyPy**b**Dp binds in the forward direction with a  $K_d$  16-fold stronger than that of a reverse sequence. Neutralizing the C terminal charge via acetylation decreases this preference 4-fold (Table 2, Entry 1; [56]). More dramatic are chiral modifications to the turn of the hairpin. When a chiral  $\alpha$ -amino substitution is introduced into the **g**-aminobutyric acid (**g**-turn) in the R configuration of a 6-ring and an 8-ring hairpin PA, preferences for the forward orientation over the reverse are over 1000-fold (Table 2, Entries 2,3; [35,57]).

The study of Rucker et al. [53] with 10-ring PAs and **b** alanine substitution illustrates the complexity that can emerge. ImImPyPyIm-**g**-PyPyPyPyPy**b**-Dp binds in a decidedly forward orientation to the extent that it binds a predicted reverse site in a forward fashion, which corresponds to a double mismatch (Table 2, Entry 4). Reverse binding is too weak to quantitate but represents a forward preference of at least 500-fold. Substituting **b** adjacent to the N-terminal imidazoles (ImIm-**b**-PyIm-**g**-PyPy-**b**-PyPy-**b**-Dp) results in a PA that prefers to bind what is a predicted reverse site in the forward fashion, with a double mismatch but with a  $K_d$  similar to that predicted for the reverse orientation (i.e. no thermodynamic discrimination between forward and reverse orientations).

In another case, the substitution of one pyrrole for **b**-Ala in the polyamide with a **b**-substituted **g**-turn results in a preference for the reverse orientation of binding. This was based on thermal melting data (Table 2, Entry 5) [58].  $\alpha$ -substitution (R) at the **g**-turn resulted in a forward binding orientation preference.

Engineering sequences to test forward and reverse binding is much more difficult for a large PA with a demonstrated tolerance for mismatches [21] than for the PAs just discussed. Here we rely on

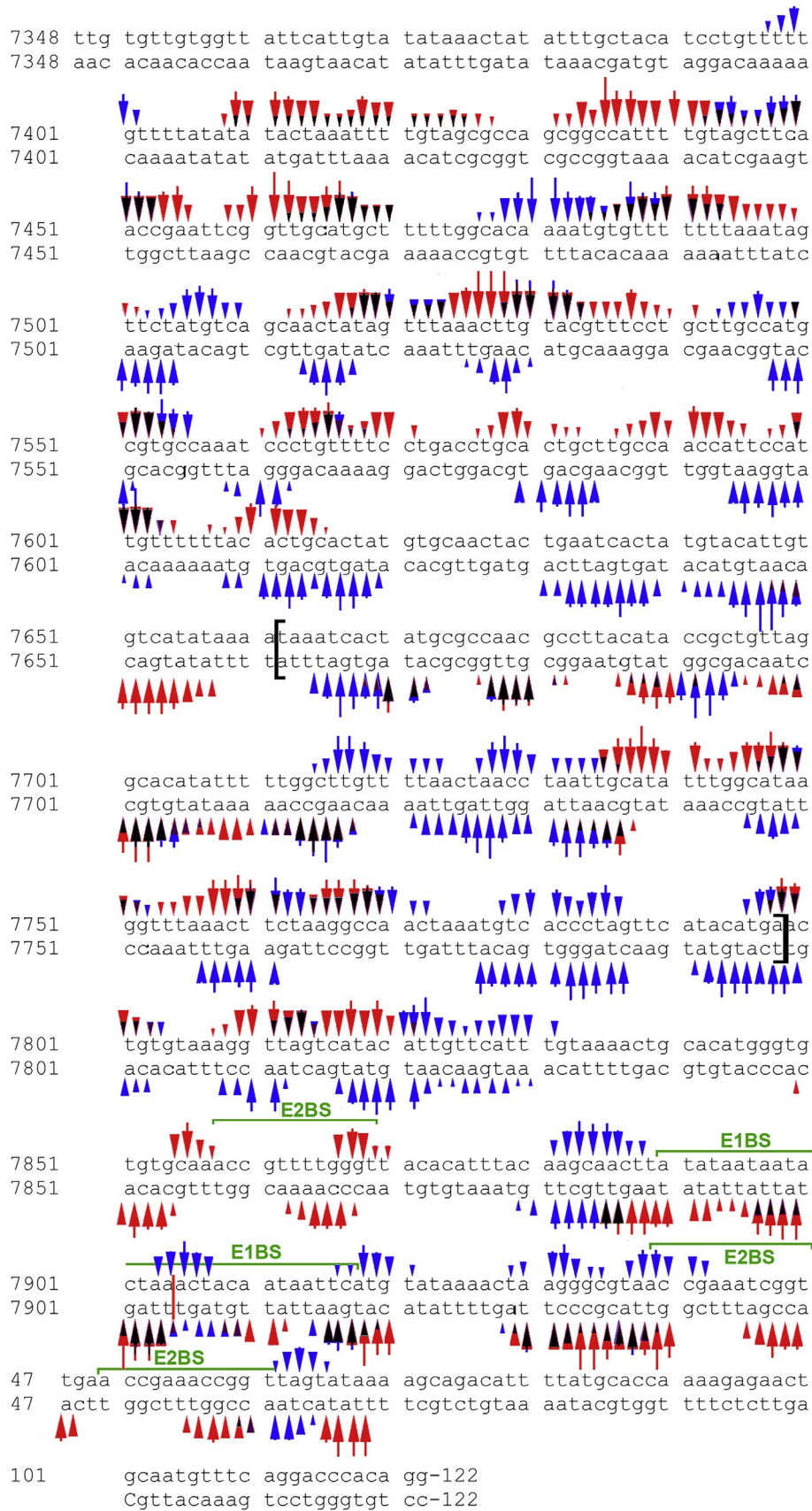


Fig. 6. Overlay of PA1 (blue) and PA25 (red) AC patterns in the 7348-122 LCR Region of HPV16. Black arrows indicate overlap.

the affinity cleavage technique and a PA2-EDTA, a version of PA1 that is modified with an amine in a chiral (R)-g-turn, (R)-g(NH<sub>2</sub>). While such modifications have been shown to influence binding orientation in smaller PAs, (R)-g(NH<sub>2</sub>) has no appreciable effect on AC patterns as exhibited by PA2-EDTA (vs. PA1-EDTA as shown in Fig. 3). This result is entirely consistent with the high tolerance of large PAs for mismatches.

## 42. Comparison of PA1 and PA25 binding in the LCR

### 42.1. Affinity cleavage

Given that both PA1 and PA25 are active against HPV16 and were originally designed to bind at or near the E1/E2 protein-DNA binding sites for an HPV sequence, a comparison of binding behaviors is instructive. Since the size difference between these PAs results in footprints that also differ in size, AC patterns are more readily compared than footprints themselves (Fig. 6). In making comparisons, we note that signal-to-noise ratios were generally better for PA1 than PA25; this had an impact on interpreting data far from the dye (the 5' end of the DNA fragment). Due in part to this fact, we were better able to obtain PA1 data further out from the dye (longer fragments) than for PA25. This resulted in some locations where data for one strand (top or bottom, depending on the label) was more accessible for PA1 than PA25.

With that in mind, some observations emerge: First, the number of AC patterns (i.e. binding events) is similar for PA1 and PA25: 34 for the former (5 compounded, that is, having multiple maxima) and 31 for the latter (10 compounded). This means that the general specificity is not dramatically different, even though PA25 is 40% longer than PA1 (see below). Second, while the footprints are extensive, the fact that there are numerous but distinct AC patterns for both indicate there are discrete binding sites/orientations in both cases. In other words, while the high tolerance for mismatches means that exact sequence binding rules do not exist, neither PA binds indiscriminately. Third, while there is some overlap of AC intensities (black, Fig. 6), there are many places where the AC patterns are distinct from one another (red for PA25 and blue for PA1, Fig. 6). This indicates that the PAs are not targeting the same local regions (e.g. 10-30 bps) and therefore are not similarly targeting any DNA structure driven by local sequence-dependent structures (e.g. minor groove width, roll, propeller twist, and helical twist; [59]). Finally, while more-active PA25 may target approximately the same number of sites in the LCR as PA1, the greater length of PA25 means that it more completely blankets or enshrouds the viral DNA in a way that PA1 cannot. This correlates with activity and may provide important insights into antiviral behavior.

### 42.2. Binding constants

A comparison of the distribution of K<sub>d</sub>s for the two PAs appears in Fig. 5. While most K<sub>d</sub>s for both PAs are between 0.7 and 4 nM, there is clearly a greater proportion of tighter K<sub>d</sub>s measured for PA25 than for PA1. This means that taken as aggregate data in this region of the HPV16 genome, PA25 tolerates a greater number of mismatches than PA1. That is, the average K<sub>d</sub> is lower (1.7 nM for PA25; 4.5 nM for PA1). This is reasonable given that the number of potential hydrogen bonds with bases is also greater (24 vs. 17) for PA25 than PA1. Thus it appears that when it comes to hairpin PA length and sequence specificity, PA1 length is probably at or possibly past the maximum in terms of benefit to specificity. Longer PA25 tolerates mismatches better than PA1 and is therefore less specific. These properties are at least partly correlated with antiviral behavior, in that PA25 is a better, broad-spectrum antiviral compound, with superior anti-HPV18 activity than PA1 [13].

## 43. Mechanistic implications

The above discussion leads to the question of whether or not it matters where either PA binds or if there is some larger effect that is responsible for the antiviral behavior. Both compounds have low IC<sub>50</sub>'s toward HPV16 [13] and both have high binding-site density in the LCR region. It is expected that this pattern would extend to other parts of the genome. Recently, we reported that PA25 induces the DNA damage response (DDR) in HPV-infected cells bearing the unintegrated viral genome (episome) [14], presumably triggered by a change the shape of the episome brought on by PA binding. The high number of binding sites for both PA1 and PA25 is consistent with this suggestion, as is the relatively rigid nature of the PA molecules, which would strain a supercoiled structure. It is less clear is if there are specific binding events, rather than a global effect, that are responsible for inducing the DDR. Extending from that is whether or not PA1 and PA25 share a common antiviral mechanism at the molecular level; more specifically, if precisely the same binding events are responsible for exactly the same elements of the DDR being induced by each compound, or if there is a combination of binding events that can induce viral DNA degradation. This question continues to drive our efforts to elucidate the details of the interactions of antiviral PAs with their DNA targets.

## Acknowledgments

We thank NIH for AI083803 awarded to JKB and NanoVir, LLC for financial support. JKB is part owner of NanoVir, LLC. We thank NanoVir, LLC for the loan of equipment and the use of active antiviral compounds. We thank the DNA Core at the University of Missouri for DNA fragment analysis. We thank the Danforth Plant Science Center for HRMS [National Science Foundation (NSF) Grant DBI 0922879]. We thank Prof. Ben Bythell for mass spectrometry data. The Agilent 600 MHz NMR spectrometer was obtained using funds from the NSF (#0959360).

## References

- [1] M.E. Parks, E.E. Baird, P.B. Dervan, Recognition of 5'-(A,T)GG(A,T)2-3' sequences in the minor groove of DNA by hairpin polyamides, *J. Amer. Chem. Soc.* 118 (1996) 6153-6159.
- [2] J.M. Turner, E.E. Baird, P.B. Dervan, Recognition of seven base pair sequences in the minor groove of DNA by ten-ring pyrrole-imidazole polyamide hairpins, *J. Amer. Chem. Soc.* 119 (1997) 7636-7644.
- [3] S. White, E.E. Baird, P.B. Dervan, On the pairing rules for recognition in the minor groove of DNA by pyrrole-imidazole polyamides, *Chem. Biol.* 4 (1997) 569-578.
- [4] Z.-F. Tao, T. Fujiwara, I. Saito, H. Sugiyama, Rational design of sequence-specific DNA alkylating agents based on duocarmycin A and pyrrole-imidazole hairpin polyamides, *J. Am. Chem. Soc.* 121 (1999) 4961-4967.
- [5] T. Vajjayanthi, T. Bando, K. Hashiya, G.N. Pandian, H. Sugiyama, Design of a new fluorescent probe: pyrrole/imidazole hairpin polyamides with pyrene conjugation at their g-turn, *Bioorg. Med. Chem.* 21 (2013) 852-855.
- [6] D. Harris, M. Stewart, A. Sielaff, K. Mulder, T. Brown, H. Mackay, M. Lee, Solution phase synthesis of imidazole- and pyrrole-containing hairpin polyamides, *Heterocycl. Commun.* 13 (2007) 17-24.
- [7] J.A. Henry, N.M. Le, B. Nguyen, C.M. Howard, S.L. Bailey, S.M. Horick, K.L. Buchmueller, M. Kotecha, D. Hochhauser, J.A. Hartley, W.D. Wilson, M. Lee, Targeting the inverted CCAAT box 2 in the topoisomerase II $\alpha$  promoter by JH-37, an imidazole-pyrrole polyamide hairpin: design, synthesis, molecular biology, and biophysical studies, *Biochemistry* 43 (2004) 12249-12257.
- [8] A. Sielaff, A. Cooper, H. Mackay, T. Brown, C. O'Hare, J. Kluzza, M. Kotecha, M. Le, D. Hochhauser, J.A. Hartley, M. Lee, Binding of f-PIP and JH-37 to the inverted CCAAT box-2 of the topoisomerase II $\alpha$  promoter, in: Abstracts of Papers, 233rd ACS National Meeting, Chicago, IL, United States, March 25-29, 2007, MEDI, 2007, p. 134.



- [9] X. Xiao, P. Yu, H.-S. Lim, D. Sikder, T. Kodadek, A cell-permeable synthetic transcription factor mimics, *Angew. Chem. Int. Ed. Engl.* 46 (2007) 2865–2868. S2865/2861–S2865/2810.
- [10] X. Xiao, P. Yu, H.-S. Lim, D. Sikder, T. Kodadek, Design and synthesis of a cell-permeable synthetic transcription factor mimic, *J. Comb. Chem.* 9 (2007) 592–600.
- [11] N.G. Nickols, J.O. Szablowski, A.E. Hargrove, B.C. Li, J.A. Raskatov, P.B. Dervan, Activity of a Py-Im polyamide targeted to the Estrogen response element, *Mol. Cancer Ther.* 12 (2013) 675–684.
- [12] J.M. Gottesfeld, C. Melander, R.K. Suto, H. Raviol, K. Luger, P.B. Dervan, Sequence-specific recognition of DNA in the nucleosome by pyrrole-imidazole polyamides, *J. Mol. Biol.* 309 (2001) 615–629.
- [13] T.G. Edwards, K.J. Koeller, U. Slomczynska, K. Fok, M. Helmus, J.K. Bashkin, C. Fisher, HPV episome levels are potentially decreased by pyrrole-imidazole polyamides, *Antivir. Res.* 91 (2011) 177–186.
- [14] T.G. Edwards, T.J. Vidmar, K. Koeller, J.K. Bashkin, C. Fisher, Dna damage repair genes controlling human papillomavirus (HPV) episome levels under conditions of stability and extreme instability, *PLoS One* 8 (2013) e75406.
- [15] T.G. Edwards, M.J. Helmus, K. Koeller, J.K. Bashkin, C. Fisher, Human papillomavirus episome stability is reduced by aphidicolin and controlled by DNA damage response pathways, *J. Virol.* 87 (2013) 3979–3989.
- [16] C. Fisher, Recent insights into the control of human papillomavirus (HPV) genome stability, loss, and degradation, *J. Clin. Med.* 4 (2015) 204–230.
- [17] T.F. Martínez, J.W. Phillips, K.K. Karanja, P. Polaczek, C.-M. Wang, B.C. Li, J.L. Campbell, P.B. Dervan, Replication stress by PyIm polyamides induces a non-canonical ATR-dependent checkpoint response, *Nucleic Acids Res.* 42 (2014) 11546–11559.
- [18] J.M. Turner, S.E. Swalley, E.E. Baird, P.B. Dervan, Aliphatic/aromatic amino acid pairings for polyamide recognition in the minor groove of DNA, *J. Amer. Chem. Soc.* 120 (1998) 6219–6226.
- [19] C.A. Hawkins, E.E. Baird, P.B. Dervan, D.E. Wemmer, Analysis of hairpin polyamide complexes having DNA binding sites in close proximity, *J. Amer. Chem. Soc.* 124 (2002) 12689–12696.
- [20] 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, DNA binding polyamides and the importance of DNA recognition in their use as gene-specific and antiviral agents, *Med. Chem.* 4 (2014) 338–344.
- [21] G. He, E. Vasilieva, G.D. Harris Jr., K.J. Koeller, J.K. Bashkin, C.M. Dupureur, Binding studies of a large antiviral polyamide to a natural HPV sequence, *Biochimie* 102 (2014) 83–91.
- [22] E.E. Baird, P.B. Dervan, Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids, *J. Amer. Chem. Soc.* 118 (1996) 6141–6146.
- [23] C.M. Dupureur, J.K. Bashkin, K. Aston, K.J. Koeller, K.R. Gaston, G. He, Fluorescence assay of polyamide-DNA interactions, *Anal. Biochem.* 423 (2012) 178–183.
- [24] J. Bashkin, C. Fisher, K.J. Koeller, Inventors, NanoVir LLC., Polyamides for treating human papilloma virus, US Patent 2007/US6133 (2007).
- [25] D.M. Chenoweth, D.A. Harki, J.W. Phillips, C. Dose, P.B. Dervan, Cyclic pyrrole-imidazole polyamides targeted to the androgen response element, *J. Amer. Chem. Soc.* 131 (2009) 7182–7188.
- [26] J.K. Bashkin, K. Aston, J.P. Ramos, K.J. Koeller, R. Nanjunda, G. He, C.M. Dupureur, W. David Wilson, Promoter scanning of the human COX-2 gene with 8-ring polyamides: unexpected weakening of polyamide-DNA binding and selectivity by replacing an internal N-Me-pyrrole with beta-alanine, *Biochimie* 95 (2013) 271–279.
- [27] A.M. Maxam, W. Gilbert, Sequencing end-labeled DNA with base-specific chemical cleavages, *Methods Enzymol.* 65 (1980) 499–560.
- [28] T.D. Schaal, W.G. Mallet, D.L. McMinn, N.V. Nguyen, M.M. Sopko, S. John, B.S. Parekh, Inhibition of human papilloma virus E2 DNA binding protein by covalently linked polyamides, *Nucleic Acids Res.* 31 (2003) 1282–1291.
- [29] H. Romanczuk, F. Thierry, P.M. Howley, Mutational analysis of cis elements involved in E2 modulation of human papillomavirus type 16 P97 and type 18 P105 promoters, *J. Virol.* 64 (1990) 2849–2859.
- [30] K. Muzikar, J. Meier, D. Gubler, J. Raskatov, P. Dervan, Expanding the repertoire of natural product-inspired ring pairs for molecular recognition of DNA, *Org. Lett.* 13 (2011) 5612–5615, <http://dx.doi.org/10.1021/ol202285y>.
- [31] P.B. Dervan, B.S. Edelson, Recognition of the DNA minor groove by pyrrole-imidazole polyamides, *Curr. Opin. Struct. Biol.* 13 (2003) 284–299.
- [32] C.C. Wang, U. Ellervik, P.B. Dervan, Expanding the recognition of the minor groove of DNA by incorporation of beta-alanine in hairpin polyamides, *Bioorg. Med. Chem.* 9 (2001) 653–657.
- [33] A.R. Urbach, P.B. Dervan, Toward rules for 1:1 polyamide:DNA recognition, *Proc. Natl. Acad. Sci. U. S. A.* 98 (2001) 4343–4348.
- [34] P.B. Dervan, Molecular recognition of DNA by small molecules, *Bioorg. Med. Chem.* 9 (2001) 2215–2235.
- [35] D.M. Herman, E.E. Baird, P.B. Dervan, Stereochemical control of the DNA binding affinity, sequence specificity, and orientation preference of chiral hairpin polyamides, *J. Amer. Chem. Soc.* 120 (1998) 1382–1391.
- [36] M.E. Parks, E.E. Baird, P.B. Dervan, Optimization of the hairpin polyamide design for recognition of the minor groove of DNA, *J. Amer. Chem. Soc.* 118 (1996) 6147–6152.
- [37] D.O. Wilson, P. Johnson, B.R. McCord, Nonradiochemical DNase I footprinting by capillary electrophoresis, *Electrophoresis* 22 (2001) 1979–1986.
- [38] M. Zianni, K. Tessanne, M. Merighi, R. Laguna, F. Tabita, Identification of the DNA bases of a DNase I footprint by the use of dye primer sequencing on an automated capillary DNA analysis instrument, *J. Biomol. Tech.* 17 (2006) 103–113.
- [39] S. Sivapragasam, A. Pande, A. Grove, A recommended workflow for DNase I footprinting using a capillary electrophoresis genetic analyzer, *Anal. Biochem.* 481 (2015) 1–3.
- [40] S. White, E.E. Baird, P.B. Dervan, Effects of the A/T/T/A degeneracy of pyrrole-imidazole polyamide recognition in the minor groove of DNA, *Biochemistry* 35 (1996) 12532–12537.
- [41] J.S. Taylor, P.G. Schultz, P.B. Dervan, DNA affinity cleaving: sequence specific cleavage of DNA by Distamycin-EDTA • Fe(II) and EDTA-distamycin Fe(II), *Tetrahedron* 40 (1984) 457–465.
- [42] T.D. Tullius, J.A. Greenbaum, Mapping nucleic acid structure by hydroxyl radical cleavage, *Curr. Opin. Chem. Biol.* 9 (2005) 127–134.
- [43] P. Schultz, J. Taylor, P. Dervan, Design and synthesis of a sequence-specific DNA cleaving molecule. (Distamycin-EDTA)iron(II), *J. Amer. Chem. Soc.* 104 (1982) 6861–6863.
- [44] H. Koohy, T.A. Down, T.J. Hubbard, Chromatin accessibility data sets show bias due to sequence specificity of the DNase I enzyme, *PLoS One* 8 (2013) e69853.
- [45] B. Heddi, J. Abi-Ghanem, M. Lavigne, B. Hartmann, Sequence-dependent DNA flexibility mediates DNase I cleavage, *J. Mol. Biol.* 395 (2010) 123–133.
- [46] Y.W. Han, G. Kashiwazaki, H. Morinaga, T. Matsumoto, K. Hashiya, T. Bando, Y. Harada, H. Sugiyama, Effect of single pyrrole replacement with beta-alanine on DNA binding affinity and sequence specificity of hairpin pyrrole/imidazole polyamides targeting 5'-GCGC-3', *Bioorg. Med. Chem.* 21 (2013) 5436–5441.
- [47] Y.N.V. Gopal, M.W. Van Dyke, Combinatorial determination of sequence specificity for nanomolar DNA-binding hairpin polyamides, *Biochemistry* 42 (2003) 6891–6903.
- [48] C.C. O'Hare, D. Mack, M. Tandon, S.K. Sharma, J.W. Lown, M.L. Kopka, R.E. Dickerson, J.A. Hartley, DNA sequence recognition in the minor groove by crosslinked polyamides: the effect of N-terminal head group and linker length on binding affinity and specificity, *Proc. Natl. Acad. Sci. U. S. A.* 99 (2002) 72–77.
- [49] D.S. Pilch, N. Poklar, E.E. Baird, P.B. Dervan, K.J. Breslauer, The thermodynamics of polyamide-DNA recognition: hairpin polyamide binding in the minor groove of duplex DNA, *Biochemistry* 38 (1999) 2143–2151.
- [50] S. Swalley, E.E. Baird, P.B. Dervan, Discrimination of 5'-GGGG-3', 5'-GCGC-3', and 5'-GGCC-3' sequences in the minor groove of DNA by eight-ring hairpin polyamides, *J. Amer. Chem. Soc.* 119 (1997) 6953–6961.
- [51] S.E. Swalley, E.E. Baird, P.B. Dervan, Effects of g-turn and b-tail amino acids on sequence-specific recognition of DNA by hairpin polyamides, *J. Amer. Chem. Soc.* 121 (1999) 1113–1120.
- [52] Y. Kawamoto, A. Sasaki, K. Hashiya, S. Ide, T. Bando, K. Maeshima, H. Sugiyama, Tandem trimer pyrrole-imidazole polyamide probes targeting 18 base pairs in human telomere sequences, *Chem. Sci.* 6 (2015) 2307–2312.
- [53] V.C. Rucker, C. Melander, P.B. Dervan, Influence of beta-alanine on hairpin polyamide orientation in the DNA minor groove, *Helvetica Chim. Acta* 86 (2003) 1839–1851.
- [54] P.M. Reddy, J.W. Toporowski, A.L. Kahane, T.C. Bruce, Recognition of a 10 base pair sequence of DNA and stereochemical control of the binding affinity of chiral hairpin polyamide-Hoechst 33258 conjugates, *Bioorg. Med. Chem. Lett.* 15 (2005) 5531–5536.
- [55] W. Zhang, S.K. Jiang, Y.L. Wu, C.X. Guo, H.F. Zhang, H. Sugiyama, X.L. Chen, Discrimination between T/A and A/T base pairs of pyrrole-imidazole polyamides substituted with chiral beta-hydroxy-gamma-aminobutyric acid/beta-alanine pairs, *ChemBioChem* 13 (2012) 47–50.
- [56] S. White, E.E. Baird, P.B. Dervan, Orientation preferences of pyrrole-imidazole polyamides in the minor groove of DNA, *J. Am. Chem. Soc.* 119 (1997) 8756–8765.
- [57] S.M. Tsai, M.E. Farkas, C.J. Chou, J.M. Gottesfeld, P.B. Dervan, Unanticipated differences between alpha- and gamma-diaminobutyric acid-linked hairpin polyamide-alkylator conjugates, *Nucleic Acids Res.* 35 (2007) 307–316.
- [58] J.L. Meier, A.S. Yu, I. Korf, D.J. Segal, P.B. Dervan, Guiding the design of synthetic DNA-binding molecules with massively parallel sequencing, *J. Amer. Chem. Soc.* 134 (2012) 17814–17822.
- [59] T. Zhou, L. Yang, Y. Lu, I. Dror, A.C. Dantas Machado, T. Ghane, R. Di Felice, R. Rohs, DNASHape: a method for the high-throughput prediction of DNA structural features on a genomic scale, *Nucleic Acids Res.* 41 (2013) W56–W62.