

1 Identification of putative G-quadruplex forming sequences in 2 three manatee papillomaviruses

3
4 Maryam Zahin¹, William L. Dean¹, Shin-je Ghim¹, Joongho Joh^{1,2}, Robert D. Gray¹, Sujita
5 Khanal^{1,3}, Gregory D. Bossart^{4,5}, Antonio A. Mignucci-Giannoni⁶, Eric C. Rouchka^{7,9}, Alfred B.
6 Jenson¹, Jonathan B. Chaires¹ and Julia H. Chariker^{8,9*}

7
8 ¹James Graham Brown Cancer Center, University of Louisville, Louisville, Kentucky, USA.

9
10 ²Department of Medicine, University of Louisville, Louisville, Kentucky, USA.

11
12 ³Department of Biochemistry and Molecular Genetics, University of Louisville, Louisville,
13 Kentucky, USA.

14
15 ⁴Georgia Aquarium, Atlanta, Georgia, USA.

16
17 ⁵Division of Comparative Pathology, Department of Pathology, Miller School of Medicine,
18 University of Miami, Miami, Florida, USA.

19
20 ⁶Puerto Rico Manatee Conservation Center, Inter American University of Puerto Rico, Bayamon,
21 Puerto Rico

22
23 ⁷Department of Computer Engineering and Computer Science, University of Louisville, Duthie
24 Center for Engineering, Louisville, Kentucky, USA.

25
26 ⁸Department of Psychological and Brain Sciences, University of Louisville, Louisville,
27 Kentucky, USA.

28
29 ⁹KBRIN Bioinformatics Core, 522 East Gray Street, University of Louisville, Louisville,
30 Kentucky, USA.

31
32
33 *** Corresponding Author:**

34 **E-mail:** julia.chariker@louisville.edu (JHC)

35 **Running title:** G-quadruplex structures in manatee papillomaviruses

36 **Abstract**

37 The Florida manatee (*Trichechus manatus latirostris*) is considered a threatened aquatic mammal
38 in United States coastal waters. Over the past decade, the appearance of papillomavirus-induced
39 lesions and viral papillomatosis in manatees has been a concern for those involved in the
40 management and rehabilitation of this species. To date, three manatee papillomaviruses (PVs)
41 have been identified in Florida manatees, one forming cutaneous lesions (TmPV1) and two
42 forming genital lesions (TmPV3 and TmPV4). In this study, we identified DNA sequences with
43 the potential to form G-quadruplex structures in all three PVs. G-quadruplex structures (G4) are
44 guanine-rich nucleic acid sequences capable of forming secondary structures in DNA and RNA.
45 In humans, G4 are known to regulate molecular processes such as transcription and translation.
46 Although G4 have been identified in several viral genomes, including human PVs, no attempt
47 has been made to identify G4 in animal PVs. We found that sequences capable of forming G4
48 were present on both DNA strands and across coding and non-coding regions on all PVs. The
49 vast majority of the identified sequences would allow the formation of non-canonical structures
50 with only two G-tetrads. The formation of one such structure was supported through biophysical
51 analysis. Computational analysis demonstrated enrichment of G4 sequences on the reverse strand
52 in the E2/E4 region on all manatee PVs and on the forward strand in the E2/E4 region on one
53 genital PV. Several G4 sequences occurred at similar regional locations on all PVs, most notably
54 on the reverse strand in the E2 region. In other cases, G4 were identified at similar regional
55 locations only on PVs forming genital lesions. On all PVs, G4 sequences were located near
56 putative E2 binding sites in the non-coding region. Together, these findings suggest that G4 are
57 likely regulatory elements in manatee PVs.

58

59 **Author summary**

60 G-quadruplex structures (G4) are found in the DNA and RNA of many species and are known to
61 regulate the expression of genes and the synthesis of proteins, among other important molecular
62 processes. Recently, these structures have been identified in several viruses, including the
63 human papillomavirus (PV). As regulatory structures, G4 are of great interest to researchers as
64 drug targets for viral control. In this paper, we identify the first G4 sequences in three PVs
65 infecting a non-human animal, the Florida manatee. Through computational and biophysical
66 analysis, we find that a greater variety of sequence patterns may underlie the formation of these
67 structures than previously identified. The sequences are found in all protein coding regions of the
68 virus and near sites for viral replication in non-coding regions. Furthermore, the distribution of
69 these sequences across the PV genomes supports the notion that sequences are conserved across
70 PV types, suggesting they are under selective pressure. This paper extends previous research on
71 G4 in human PVs with additional evidence for their role as regulators. The G4 sequences we
72 identified also provide potential regulatory targets for researchers interested in controlling this
73 virus in the Florida manatee, a threatened aquatic mammal.

74

75 **Introduction**

76 G-quadruplex structures (G4) are four-stranded, inter- and intramolecular structures
77 formed from guanine-rich DNA and RNA sequences. The sequences fold to form stacks of G-
78 tetrads, planar structures composed of four guanine bases held together by Hoogsteen hydrogen
79 bonds, Fig 1 (1). The stacked G-tetrads are connected by loops which vary in size and sequence
80 composition, affecting the stability of the structure.

81 G4 are known to be involved in a series of key biological functions. In humans, G4 are
82 found in telomeric repeats and serve to prevent degradation and genomic instability (2). Their
83 formation in this region is also known to decrease telomerase activity which is selectively
84 expressed in a vast majority of cancers (3). G4 located in the promoter region of genes act as
85 transcriptional regulators (4) while those found in intronic and exonic regions play a role in
86 alternative splicing (5, 6). In RNA, G4 identified in 3' and 5' untranslated regions are known to
87 regulate protein synthesis (7, 8).

88 G4 function as regulators through at least a couple of different mechanisms (1). For
89 example, G4 formation can inhibit transcription by blocking the activity of RNA polymerase.
90 Alternatively, G4 can bind with other regulatory elements that either activate or repress
91 transcription. A variety of different proteins are now known to bind with G4 in DNA and RNA
92 (9). In RNA, this includes proteins involved in splicing as well as protein synthesis (10).

93 A regulatory role for G4 in prokaryotic cells has been well-established (11). This has led
94 to an interest in examining the role G4 may play in organisms such as viruses (12, 13). Although
95 one might presume that viruses have evolved analogous regulatory mechanisms, research on the
96 role of G4 in viral genomes has been limited to human immunodeficiency virus 1 (HIV-1) (14),
97 human papillomaviruses (HPV) (15), and Epstein-Barr virus (EBV) (16). In HIV-1, ligand

98 stabilization of a G4 located in the *nef* gene reduced gene expression and repressed HIV-1
99 infectivity in an antiviral assay (8). In EBV, destabilization of a G4, located in virally encoded
100 mRNA, reduced translation (11). Both findings are in line with research on the regulatory role of
101 G4 in prokaryotes. In HPV, DNA sequences capable of forming G4 have been identified in four
102 regions (NCR, L2, E1, and E4) of eight HPV types, and their ability to form in the laboratory has
103 been established. However, it remains unclear how they might affect PV replication and
104 transcription (10, 15).

105 PVs cause a number of benign and malignant tumors in humans and animals. There are
106 approximately 100 human PV types and at least 112 non-human PV types found in 54 different
107 species (17). The specific location of tumor formation (cutaneous, oral, genital, or anal) depends
108 on the type of PV. The existence of G4 in non-human animals would provide further support for
109 their potential functional relevance in PVs and would also provide a valuable comparison to the
110 pattern of distribution seen in human PVs. In the current paper, we identify and characterize G4
111 sequences in three PVs infecting the Florida manatee.

112 The study of G4 in manatee PVs has ecological as well as biological significance. The
113 Florida manatee is an aquatic mammal living in the coastal waters of Florida that has been
114 classified as an endangered species since 1967. Its population declined for a variety of reasons,
115 not the least of which was that its gentle, slow-moving nature made it vulnerable to injury from
116 boat propellers. Efforts at restoring the population have been successful to the extent that the
117 species was downlisted to threatened status in 2016. However, in the midst of these efforts,
118 animals undergoing rehabilitation frequently showed signs of high sensitivity to environmental
119 stress, one sign being the development of cutaneous or mucosotropic genital papillomatous
120 lesions. Some animals, both in captivity and in the wild, showed antibody titers indicating the

121 presence or exposure to *Trichechus manatus* PV 1 (TmPV1), a virus first characterized in our
122 laboratory (18). More recently, genital lesions appeared in a single Florida manatee used as a
123 surrogate animal for manatee rehabilitation at the Puerto Rico Center (PRMCC), and DNA
124 sequencing, also performed in our laboratory, indicated the presence of two new PVs, *Trichechus*
125 *manatus* PV 3 (TmPV3) and 4 (TmPV4) (19, 20). These are the first known genital
126 mucosotropic PVs in a manatee, presenting a potential health threat to this species should the
127 virus spread in wild populations, if not already present there.

128 Similar to HPVs, manatee PV genomes are comprised of double stranded DNA,
129 approximately 8 Kb in length that encodes a maximum of seven genes. Five genes encode non-
130 structural or early proteins E1, E2, E4, E6 and E7, and two encode structural or late proteins L1
131 and L2, with all coding regions located on the forward DNA strand. A non-coding region holds
132 the origin of replication and at least a couple of promotor sites. Much of what is known about the
133 function of these sites comes from molecular biological research on human PVs (21). E1 and E2
134 proteins form a complex that initiates viral replication at the origin, resulting in amplification of
135 the virus. E2 also functions as a negative regulator of E6 and E7, two coding regions that
136 stimulate cell growth and function as oncogenes in human PVs. The late proteins L1 and L2 code
137 for the major and minor capsid proteins encapsidating viral DNA with E4 having a possible role
138 in facilitating virion release.

139 During our initial sequencing of TmPV4, a glycine rich GGA repeat sequence identified
140 in the E2 region created an obstacle to sequencing due to the formation of a secondary structure,
141 necessitating the use of power-read sequencing analysis to complete the genome (19, 20). We
142 reasoned this was likely to be a G4, given that a GGA repeat would be a sequence pattern likely
143 to form one of these structures. In fact, in 2001 Matsugami and colleagues reported the folding of

144 a GGA repeat into an intramolecular parallel G-quadruplex in the laboratory (22). This is
145 significant in that a G4 in the E2 region could play a vital role in altering the regulatory functions
146 of the virus due to the role E2 has in regulating the expression of oncoproteins E6 and E7.
147 Moreover, integration of E2 into a human cervical host cell chromosome is considered by
148 epidemiologists to be an important event leading to the development of cervical cancer. From a
149 functional standpoint, blockage of E2 transcription by integration could potentially be equivalent
150 to blockage of E2 by a G-quadruplex structure.

151 In this paper, we identify sequences with the potential to form G4 on both DNA strands
152 in each coding and non-coding region on all three manatee PV genomes. In contrast to the
153 findings for G4 in HPV, we find that the majority of sequences identified were capable of
154 forming G4 structures with two rather than three G-tetrads, and we provide laboratory support
155 for the formation of a secondary structure from one such sequence. We find several G4 in similar
156 locations on all three PVs as well as several G4 in similar locations unique to the two PVs
157 forming genital lesions. G4 were also located near putative E2 binding sites in non-coding
158 regions on all PVs. Although G4 were found in all coding and non-coding regions, G4 were
159 significantly enriched in the E2/E4 region on all three genomes, suggesting that G4 are
160 evolutionarily preserved in this region.

161 **Results**

162 **G4 sequence distribution**

163 The number of putative G4 sequences, broken down by DNA strand, genomic region, and
164 TmPV genome, is displayed in Table 1. As described in the Materials and methods section,
165 longer sequences supported the development of more than one G4 at a time. As a result, in
166 several regions, the number of G4 possible was slightly higher than the number of sequences

167 identified, and these values are displayed alongside the number of G4 sequences in Table 1.
 168 TmPV4 had the highest number of sequences identified, with 20 on the forward DNA strand and
 169 17 on the reverse strand. Somewhat fewer sequences were identified on TmPV3 (13 forward, 11
 170 reverse) and TmPV1 (14 forward, 15 reverse).

171

172 **Table 1. The number of putative G4 sequences identified and the number of structures**
 173 **possible across different regions on forward and reverse DNA strands for each TmPV**
 174 **genome.**

Region	Forward DNA Strand			Reverse DNA Strand		
	Number Sequences (Number Possible Structures)			Number Sequences (Number Possible Structures)		
	TmPV1	TmPV3	TmPV4	TmPV1	TmPV3	TmPV4
E6	-	1 (1)	-	-	-	-
E7	2 (2)	1 (1)	-	-	-	-
E1	5 (5)	3 (3)	4 (5)	-	-	-
E2	1 (1)	1 (1)	2 (2)	1 (1)	-	-
E1/E2	1 (1)	1 (1)	1 (1)	-	-	-
E2/E4	1 (1)	-	7 (12)	3 (5)	3 (6)	6 (9)
Total Early Region	10 (10)	7 (7)	14 (20)	4 (6)	3 (6)	6 (9)
L2	2 (4)	2 (4)	3 (3)	5 (6)	5 (6)	7 (7)
L1	2 (2)	3 (3)	2 (2)	4 (5)	3 (3)	3 (3)
Total Late Region	4 (6)	5 (7)	5 (5)	9 (11)	8 (9)	10 (10)
NCR	-	1 (1)	1 (3)	2 (2)	-	1 (1)
Total Genome	14 (16)	13 (15)	20 (28)	15 (19)	11 (15)	17 (20)

175

176

177 All identified G4 sequences, with one exception, are capable of forming G4 with only
 178 two G-tetrads. The exception to this pattern was found in the L2 region of TmPV1 where a
 179 sequence capable of forming a three G-tetrad structure on the forward DNA strand was
 180 identified. This sequence was embedded in a much longer sequence also capable of forming a

181 two G-tetrad structure. The individual sequences along with sequence locations and sequence
182 descriptors for putative G4 identified in each TmPV genome are available in S1, S2, and S3
183 Tables.

184 **G4 enriched in E2/E4 region on all TmPVs**

185 For all TmPVs, the number of nucleotides covered by G4 sequences was greater than
186 expected in the E2/E4 region when compared to a random distribution of G4 across the genome.
187 This occurred primarily on the reverse DNA strand (E2: TmPV1, $p = 0.05$; TmPV3, $p = 0.053$;
188 TmPV4, $p = 0.015$; E4: TmPV1, $p = 0.005$; TmPV3, $p = 0.008$; TmPV4, $p = 0.001$). However,
189 TmPV4 also showed enrichment on the forward DNA strand (E2: $p = 0.013$; E4: $p = 0.009$). The
190 number of observed G4 nucleotides, the number of random simulations with G4 nucleotides
191 greater than or equal to the observed G4 nucleotides, and the associated significance values are
192 available in S4 Table for each DNA strand in each genomic region on each TmPV.

193 **Co-occurring G4 locations across TmPVs**

194 To identify similar patterns in the distribution of G4 sequences across the three manatee
195 PV genomes, the coding/non-coding regions were aligned at their starting locations, and G4
196 sequences with one or more nucleotides at the same distance from the beginning of the region
197 were identified as occurring at the same location within a region. There were several regions
198 with G4 sequences in the same location on all genomes. In Fig 2, on the forward DNA strand
199 (left), G4 sequences were found in the same location in E2, L2, and L1. On the reverse DNA
200 strand (Fig 2, right), G4 sequences were found in the same location in E2 and L1. There were
201 also regions with G4 sequences occurring in the same location on TmPV3 and TmPV4 but not
202 TmPV1. On the forward strand this pattern was found in E1 and NCR, and on the reverse strand,
203 this pattern was found in E4.

204 **G4 located near putative E2 binding sites**

205 On each PV, G4 sequences were identified in the non-coding region (NCR) where the
206 origin of replication is located. Given the known role of G4 in replication (23, 24), these regions
207 were searched for E2 binding sites to determine whether the G4 sequences might be positioned
208 close to the origin of replication. E2 binding sites were identified using the consensus sequence
209 ACCgNNNNcGGT, allowing some variation in the fourth and ninth nucleotide positions
210 (lowercase g and c) with most variation occurring from nucleotide positions 5 through 8 (25).
211 Table 2 lists all sequences identified with the pattern ACCNNNNNNGGT. Nine of the 21
212 sequences identified have the more conservative consensus sequence of ACCGNNNNCGGT.
213 The locations of consensus sequences identified in the NCR are displayed in Fig 3 along with the
214 location of G4 in that region. On each PV genome, one or more G4 are located within 100 nt of a
215 putative E2 binding site.

216

217

218

219

220

221

222

223

224

225

226 **Table 2. Putative E2 binding site sequences and locations on three manatee PVs along with**
 227 **the location and distance of the nearest putative G4 sequence.**

PV	Region	Sequence	Genomic Position	Genomic Position of Nearest Upstream G4 (Distance)	Genomic Position of Nearest Downstream G4 (Distance)
TmPV1	NCR	ACC CATGG CGGT	7203	6851 (-352)	7215 (+12)
TmPV1	NCR	ACCG CCTT CGGT*	7276	7215 (-61)	7355 (+79)
TmPV1	NCR	ACCG TCTC GGGT	7383	7355 (-28)	-
TmPV1	NCR	ACC TAATGA GGT	7436	7355 (-81)	-
TmPV1	NCR	ACCG GATTA GGT	7585	7355 (-230)	-
TmPV3	E1	ACCG ATGTT GGT	2344	1195 (-1149)	2542 (+198)
TmPV3	E4	ACCG CTGGA GGT	3576	3316 (-260)	3792 (+216)
TmPV3	NCR	ACCG TAAC CGGT*	7033	6906 (-127)	7383 (+350)
TmPV3	NCR	ACCG TTTGT GGT	7327	6906 (-421)	7383 (+56)
TmPV3	NCR	ACCG TTCC CGGT*	7426	7383 (-43)	-
TmPV3	NCR	ACCG GGAG CGGT*	7466	7383 (-83)	-
TmPV3	NCR	ACCG TTAGG GGT	7561	7383 (-178)	-
TmPV4	E6	ACCG GGTG CGGT*	44	-	1170 (+1126)
TmPV4	E6	ACC AATAT CGGT	320	-	1170 (+850)
TmPV4	E1	ACCG CTATT GGT	2439	2064 (-375)	2634 (+195)
TmPV4	E4	ACC CAGGC GGGT	3809	3777 (-32)	3840 (+31)
TmPV4	L2	ACCG GTTCC CGGT*	4392	4270 (-122)	4395 (+3)
TmPV4	L2	ACC CCAGT GGT	4472	4456 (-16)	4548 (+76)
TmPV4	NCR	ACCG CCAG CGGT*	7315	7205 (-110)	7538 (+223)
TmPV4	NCR	ACCG GGTG CGGT*	7699	7681 (-18)	-
TmPV4	NCR	ACCG GGA ACGGT*	7740	7681 (-59)	-

228 *Conservative search sequence ACCGNNNCGGT; Variable nucleotide positions are
 229 highlighted in red bold.

230

231 **TmPV4 laboratory analysis**

232 To study the possible formation of G4 secondary structures, analytical ultracentrifugation
 233 (AUC) was performed on three oligonucleotides selected from the TmPV4 E2 region, TmPV4-1
 234 (Mw=11,453), TmPV4-2 (Mw=22,045), and TmPV4-3 (Mw=3,511). TmPV4-1 sedimented as a
 235 mixture of several molecular species as monomers (total accounted for 90% monomers, Fig 4A).
 236 TmPV4-2 was also a mixture of structures, about half monomeric and the rest aggregated (57%

237 monomers, Fig 4B). In contrast, TmPV4-3 sedimented as a 90% dimer (Fig 4C), indicating the
238 possible formation of a two-stranded G-quadruplex structure. The circular dichroism (CD)
239 spectrum is consistent with such a structure as illustrated in Fig 5 and 6.

240 **Discussion**

241 The current study represents the first G4 sequences identified in PVs infecting a non-
242 human animal. Prior to this, G4 sequences had been identified solely in PVs infecting humans
243 (15). The G4 sequences identified in the current study consisted, almost exclusively, of
244 sequences capable of forming structures comprised of two G-tetrads, whereas the authors of the
245 recent study of G4 in human PVs searched only for canonical structures comprised of three G-
246 tetrads. As reported in the study, these three G-tetrad sequences were identified in only eight of
247 all human PV types listed in the NCBI Entrez Gene database. However, two G-tetrad sequences
248 are found throughout the human genome (26) and are known to form secondary structures (27).
249 In at least one case, these two-G-tetrad structures have been found to be more stable than three
250 G-tetrad structures (28). In a second case, a similar structure with different loop interactions and
251 capping structures was identified (29), supporting the notion that sequences capable of forming
252 G4 are more variable than once thought. This suggests that the search for G4 in PVs should be
253 extended to include two-tetrad structures.

254 Our biophysical analysis confirmed that one such sequence in TmPV4, a GGA₄ motif
255 located on the forward strand of the E2/E4 region, formed a secondary structure in the
256 laboratory. GGA repeat sequences are common across eukaryotic genomes (30), and the
257 biological relevance of GGA repeats is well-known (31-35). One GGA repeat sequence has been
258 found to form an intramolecular G4 in the laboratory (22, 36). A second GGA repeat sequence

259 located in the *c-myb* promoter forms a G4 that may repress *c-myb* activity through interaction
260 with the *myc*-associated zinc finger protein (37).

261 In the current study, sequences capable of forming structures with two or more G-tetrads
262 are located more broadly across coding regions than is seen in human PVs when searching for
263 three G-tetrad structures. In the human viruses, three G-tetrad sequences were identified only in
264 the E1 region on the forward DNA strand (15), whereas in manatee PVs we find two G-tetrad
265 structures across all coding regions. Interestingly, the pattern we find on the reverse DNA strand
266 for manatees is somewhat similar to that seen in human PVs; no sequences capable of forming
267 G4 structures are found in E6, E7, or E1. In manatee PVs, all subsequent regions contain two G-
268 tetrad sequences. In human PVs, three G-tetrad sequences are found in E2/E4, L2, and the non-
269 coding region.

270 The distribution of G4 sequences within coding and non-coding regions across the
271 viruses suggests that G4 may have a variety of regulatory roles. G4 sequences located on the
272 forward DNA strand that are transcribed to mRNA could regulate splicing and translation of
273 early and late genes by binding splicing factors or other related proteins or by serving to block
274 the machinery necessary for each of these processes. In one other virus, Epstein-Barr virus, G4
275 have been found to inhibit translation of mRNA (16) making this an important avenue to explore
276 in future studies. Similarly, on the reverse DNA strand, the formation of G4 could serve to
277 inhibit transcription through blocking polymerase or binding proteins that enhance or inhibit
278 transcription. In all cases, G4 serve as valuable potential drug targets for viral control.

279 On all three manatee PVs, G4 were identified at the same location on the forward DNA
280 strand in the L1 and L2 coding regions, perhaps indicating a conserved function for these
281 sequences in the formation of the capsid proteins. New evidence suggests that G4 have a role in

282 immune evasion through antigenic variation and viral silencing (13). In *Neisseria gonorrhoeae*, a
283 G4 structure is essential to variation in the surface protein pilin, allowing the bacteria to evade
284 detection by the host immune system (38). The proposed mechanism involves transcription of an
285 sRNA from the G4 site that base-pairs with the complementary location on the opposite DNA
286 strand, separating the two strands and allowing the formation of the G4. The G4 secondary
287 structures are known to be mutagenic (39), and in HIV-1 and HIV-2 mutated capsids are known
288 to affect the ability of dendritic cells to detect the virus (40).

289 G4 sequences were also identified in non-coding areas on all three manatee PVs and were
290 located near putative E2 binding sites, indicating a potential role in replication and/or
291 transcription initiation. This would not be surprising given that G4 have been associated with
292 origins of replication in mouse and humans (24, 41). In a study of two vertebrate replicators, G4
293 were required for initiation of replication and determination of the replication start site (23).
294 On the two manatee PVs producing genital lesions, three putative E2 binding sites were located
295 in a pattern similar to that of genital human PVs (42). One E2 binding site was located on the 5'
296 end of the non-coding region and two adjacent E2 binding sites were located on the 3' end
297 separated by 6 and 28 bases. Interestingly, a G4 sequence is located just upstream of the two
298 adjacent 3' E2 binding sites in an area that should be near the origin of replication. The location
299 of these G4 within the PV promoter region may also be indicative of a role in transcriptional
300 regulation. G4 are found in many promoter sites in humans and are known to regulate
301 transcription (43). In HPV16, two adjacent E2 binding sites in this area were found necessary for
302 negative transcriptional regulation (44).

303 Interestingly, G4 sequences were enriched on the reverse DNA strand in the E2/E4
304 regions, suggesting an evolutionary advantage for G4 in this region. This is a significant location

305 given that E2 is a negative regulator of E6/E7, two potential oncogenes. However, the formation
306 of G4 on the template DNA strand would appear to serve simply as a block to the polymerase,
307 inhibiting expression of the early genes, and it is not clear how this would provide an
308 evolutionary advantage to the virus. Alternatively, G4 are known to have a mutagenic effect
309 during replication (45), and G4 sequences in this area may provide some evolutionarily
310 advantageous disruption of the sequence in this region.

311 **Conclusion**

312 This study provides further confirmation of the existence of G4 in PVs. It extends
313 previous work in human PVs by demonstrating the existence of non-canonical sequences, more
314 broadly located across the genome, that are capable of forming G4 in non-human PVs.
315 Distribution patterns that are indicative of G4 sequence conservation in specific coding regions
316 support the notion that these structures regulate activities similar to those of G4 in other species.
317 As regulatory structures, G4 offer potential drug targets for researchers interested in controlling
318 disease processes (4). Given the threatened status of the Florida manatee and the concerns of
319 scientists working to protect the health of this species, this research also provides an important
320 first step in exploring the biological significance of these structures in this gentle aquatic
321 mammal.

322 **Materials and methods**

323 **Bioinformatics**

324 The complete genomes for *Trichechus manatus* PVs (TmPV) 1, 3, and 4 were obtained
325 from the NCBI nucleotide division of GenBank (46). Information for accessing individual
326 sequences can be found in Table 3.

327

328 **Table 3. GenBank Sequence Information for TmPV Genomes.**

Accession no.	Description	Size
NC_006563.1	TmPV1	7722 bp
KP205502.1	TmPV3	7622 bp
KP205503.2	TmPV4	7771 bp

329

330 Putative G4 forming sequences were identified along each genome using the Quadparser
331 algorithm (47). For each genome, two separate analyses were performed to identify sequences
332 capable of forming unimolecular structures with two and three G-tetrads. In each case, loop
333 lengths were restricted to one to seven bases. The regular expression
334 $(G\{2,\}[ATGC]\{1,7\})\{3,\}G\{2,\}$ was used to identify structures with two G-tetrads, and the
335 regular expression $(G\{3,\}[ATGC]\{1,7\})\{3,\}G\{3,\}$ was used to search for structures with three
336 G-tetrads. Quadparser was instructed to search for G4 sequences on the forward DNA strand by
337 searching for runs of guanine bases. However, G4 sequences on the reverse DNA strand were
338 identified by searching for runs of cytosine, guanine's complement.

339 Sequences identified by Quadparser generally vary in the number of guanine tracts. A G4
340 requires four guanine tracts to form. Therefore, sequences with five or more guanine tracks can
341 form G4 at different locations in the sequence, and sequences with eight or more guanine tracts
342 support the development of more than one G4 at a time. To highlight these variations,
343 Quadparser provides a sequence descriptor in the form $x:y:z$ (Fig 7), indicating the number of
344 guanine tracts in the sequence (x), the number of locations at which a G4 could form (y), and the
345 number of G4 possible at a given time within the sequence (z).

346 To establish values for the number of G4 sequences expected in each genomic region at
347 random, Bedtools v2.16.2 (48) was used to randomly distribute the G4 sequences along each PV
348 genome 1,000 times without overlap of the sequences. A Perl program was written to count the

349 number of nucleotides covered by G4 sequence in each coding/non-coding region on each DNA
350 strand for each of the random distributions of G4 sequences and for the observed G4 sequences
351 identified by Quadparser. Significance values for G4 sequence coverage in each region were
352 estimated as $P = (r + 1)/(n + 1)$, where r is the number of random simulations in which the
353 coverage of G4 sequences in a region was greater than or equal to the observed coverage of G4
354 sequences in that region and n is the number of random simulations (49, 50). Significance levels
355 were calculated for G4 sequence coverage on each DNA strand in each region of the PV
356 genomes.

357 Putative E2 binding sites were identified using the consensus sequence 5'-
358 ACCgNNNNcGGT-3' derived from a comprehensive study of human PVs (25). The consensus
359 sequence includes some variation in the fourth and ninth nucleotide positions with most variation
360 occurring from nucleotide positions 5 through 8. To cast the widest net in searching for E2
361 binding sites in manatee PVs, positions 4 through 9 were allowed to vary over all nucleotides in
362 the regular expression designed to search for these sequences.

363 **Oligonucleotides**

364 Oligodeoxynucleotides TmPV4-1, TmPV4-2 and TmPV4-3 (sequences are given in
365 Table 4) were obtained as desalted, lyophilized solids from Eurofins MWG Operon (Huntsville,
366 AL). Each was reconstituted in MQ H₂O to give ~1 mM stock solutions based on the
367 manufacturer's yield. Concentrations were estimated from the absorbance at 260 nm of suitable
368 dilutions into K⁺-free tBAP buffer (10 mM tetrabutyl ammonium phosphate, 1 mM EDTA, pH
369 7.0) in conjunction with extinction coefficients supplied by the manufacturer. For NMR
370 experiments, the oligonucleotide was dialyzed vs. 10 mM LiPO₄, 50 mM KCl, pH 7.0, prior to
371 measurement.

372 **Table 4. Oligodeoxynucleotides used in this study.**

Oligonucleotide	Sequence (length)	ϵ (mM ⁻¹ cm ⁻¹)	MW
TmPV4-1	GGACGAGGAGGAGGACGAGGAGGAC GAGGAGGGAGC (36 nt)	397.5	11453.4
TmPV4-2	GGAGCAGGAGAAGGAGGAGGAGGAG GACGAGGAGGACGAGGAGGAGGACG AGGAGGACGAGGAGGGAGC (69 nt)	774.7	22045.3
TmPV4-3	GGAGGAGGAGG (11 nt)	125.1	3511.3

373

374 **Analytical ultracentrifugation (AUC) method**

375 Sedimentation velocity experiments were carried out in a Beckman Coulter ProteomeLab
376 XL-A analytical ultracentrifuge (Beckman Coulter Inc., Brea, CA) at 20°C and 50,000 rpm in
377 standard 2 sector cells. Buffer density was determined on a Mettler/Parar Calculating Density
378 Meter DMA 55A at 20.0 °C and viscosity was measured using an Anton Parr AMVn Automated
379 Microviscometer at 20°C. Data were analyzed with the program Sedfit (free software:
380 www.analyticalultracentrifugation.com) using the continuous c(s) distribution model. A value of
381 0.55 ml/g was used for the DNA oligonucleotides as described (51).

382 **Circular dichroism spectra**

383 Ultraviolet (UV) and circular dichroism (CD) spectra were measured at 25 °C in a
384 stoppered 1-cm cuvette with a Jasco J-810 spectropolarimeter equipped with a programmable
385 Peltier thermostatted cuvette holder and magnetic stirrer. Instrumental parameters were: 1.0 nm
386 bandwidth, 2 s integration time, 200 nm/min scan rate, four scans averaged. CD data were
387 corrected by subtracting a buffer blank and then normalized using the relationship $\epsilon_L - \epsilon_R = \Delta\epsilon$
388 $=\theta/(32980 \cdot c \cdot l)$, where θ is the observed ellipticity in millidegrees, c is the DNA strand
389 concentration in mol·L⁻¹, and l is the path length in cm (52).

390

391 **Circular dichroism melts**

392 Thermal denaturation studies of oligonucleotide TmPV4-3 in 10 mM LiPO₄, 50 mM KCl,
393 pH 7.0, were carried out essentially as previously described (52). CD spectra were recorded from
394 320 nm to 220 nm over the temperature range of 4°C to 98 °C at intervals of 1 °C using the
395 instrumental parameters described above. Thermal denaturation was carried out with a
396 temperature ramp of 4 °C/min, ±0.05 °C equilibration tolerance and 60 s delay after
397 equilibration. The resulting temperature/wavelength data matrices were analyzed by singular
398 value decomposition (SVD) as described (53) to obtain melting temperatures (T_m) and enthalpy
399 values (ΔH). Two melts of the same sample were carried out on subsequent days to assess
400 reversibility of the thermal denaturation process.

401 **Acknowledgements**

402 The original sampling of TmPV3 and TmPV4 was carried out under USFWS permit
403 #MA231088.

404 **References**

- 405 1. Bochman ML, Paeschke K, Zakian VA. DNA secondary structures: stability and function of
406 G-quadruplex structures. *Nat Rev Genet.* 2012;13(11):770-80.
- 407 2. Yang DZ, Okamoto K. Structural insights into G-quadruplexes: towards new anticancer
408 drugs. *Future Med Chem.* 2010;2(4):619-46.
- 409 3. Sun D, Thompson B, Cathers BE, Salazar M, Kerwin SM, Trent JO, et al. Inhibition of
410 human telomerase by a G-quadruplex-interactive compound. *J Med Chem.*
411 1997;40(14):2113-6.
- 412 4. Balasubramanian S, Hurley LH, Neidle S. Targeting G-quadruplexes in gene promoters: a
413 novel anticancer strategy? *Nat Rev Drug Discov.* 2011;10(4):261-75.

- 414 5. Fisette JF, Montagna DR, Mihailescu MR, Wolfe MS. A G-rich element forms a G-
415 quadruplex and regulates BACE1 mRNA alternative splicing. *J Neurochem*.
416 2012;121(5):763-73.
- 417 6. Ribeiro MM, Teixeira GS, Martins L, Marques MR, de Souza AP, Line SR. G-quadruplex
418 formation enhances splicing efficiency of PAX9 intron 1. *Hum Genet*. 2015;134(1):37-44.
- 419 7. Beaudoin JD, Perreault JP. Exploring mRNA 3'-UTR G-quadruplexes: evidence of roles in
420 both alternative polyadenylation and mRNA shortening. *Nucleic Acids Res*.
421 2013;41(11):5898-911.
- 422 8. Kumari S, Bugaut A, Huppert JL, Balasubramanian S. An RNA G-quadruplex in the 5' UTR
423 of the NRAS proto-oncogene modulates translation. *Nat Chem Biol*. 2007;3(4):218-21.
- 424 9. Brazda V, Haronikova L, Liao JC, Fojta M. DNA and RNA quadruplex-binding proteins. *Int*
425 *J Mol Sci*. 2014;15(10):17493-517.
- 426 10. Millevoi S, Moine H, Vagner S. G-quadruplexes in RNA biology. *Wiley Interdiscip Rev*
427 *RNA*. 2012;3(4):495-507.
- 428 11. Wieland M, Hartig JS. Investigation of mRNA quadruplex formation in *Escherichia coli*. *Nat*
429 *Protoc*. 2009;4(11):1632-40.
- 430 12. Metifiot M, Amrane S, Litvak S, Andreola ML. G-quadruplexes in viruses: function and
431 potential therapeutic applications. *Nucleic Acids Res*. 2014;42(20):12352-66.
- 432 13. Harris LM, Merrick CJ. G-quadruplexes in pathogens: a common route to virulence control?
433 *PLoS Pathog*. 2015;11(2):e1004562.
- 434 14. Perrone R, Nadai M, Poe JA, Frasson I, Palumbo M, Palu G, et al. Formation of a unique
435 cluster of G-quadruplex structures in the HIV-1 Nef coding region: implications for antiviral
436 activity. *PLoS One*. 2013;8(8):e73121.

- 437 15. Tluckova K, Marusic M, Tothova P, Bauer L, Sket P, Plavec J, et al. Human papillomavirus
438 G-quadruplexes. *Biochemistry*. 2013;52(41):7207-16.
- 439 16. Murat P, Zhong J, Lekieffre L, Cowieson NP, Clancy JL, Preiss T, et al. G-quadruplexes
440 regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. *Nat Chem Biol*.
441 2014;10(5):358-64.
- 442 17. Rector A, Van Ranst M. Animal papillomaviruses. *Virology*. 2013;445(1):213-23.
- 443 18. Rector A, Bossart GD, Ghim SJ, Sundberg JP, Jenson AB, Van Ranst M. Characterization of
444 a novel close-to-root papillomavirus from a Florida manatee by using multiply primed
445 rolling-circle amplification: *Trichechus manatus latirostris* papillomavirus type 1. *J Virol*.
446 2004;78(22):12698-702.
- 447 19. Ghim SJ, Joh J, Mignucci-Giannoni AA, Rivera-Guzmán AL, Falcón-Matos L, Alsina-
448 Guerrero MM, et al. Genital papillomatosis associated with two novel mucosotropic
449 papillomaviruses from a Florida manatee (*Trichechus manatus latirostris*). *Aquatic*
450 *Mammals*. 2014;40(2):195-200.
- 451 20. Zahin M, Ghim SJ, Khanal S, Bossart GD, Jenson AB, Joh J. Molecular characterization of
452 novel mucosotropic papillomaviruses from a Florida manatee (*Trichechus manatus*
453 *latirostris*). *J Gen Virol*. 2015.
- 454 21. Johansson C, Schwartz S. Regulation of human papillomavirus gene expression by splicing
455 and polyadenylation. *Nature Rev. Microbiol*. 2013;11(4):239-51.
- 456 22. Matsugami A, Ouhashi K, Kanagawa M, Liu H, Kanagawa S, Uesugi S, et al. New
457 quadruplex structure of GGA triplet repeat DNA--an intramolecular quadruplex composed of
458 a G:G:G:G tetrad and G(:A):G(:A):G(:A):G heptad, and its dimerization. *Nucleic Acids Res*.
459 *Suppl*. 2001(1):271-2.

- 460 23. Valton AL, Hassan-Zadeh V, Lema I, Boggetto N, Alberti P, Saintome C, et al. G4 motifs
461 affect origin positioning and efficiency in two vertebrate replicators. *EMBO J.*
462 2014;33(7):732-46.
- 463 24. Besnard E, Babled A, Lapasset L, Milhavet O, Parrinello H, Dantec C, et al. Unraveling cell
464 type-specific and reprogrammable human replication origin signatures associated with G-
465 quadruplex consensus motifs. *Nat Struct Mol Biol.* 2012;19(8):837-44.
- 466 25. Rogers A, Waltke M, Angeletti PC. Evolutionary variation of papillomavirus E2 protein and
467 E2 binding sites. *Virology J.* 2011;8:379.
- 468 26. Qin M, Chen Z, Luo Q, Wen Y, Zhang N, Jiang H, et al. Two-quartet G-quadruplexes
469 formed by DNA sequences containing four contiguous GG runs. *J Phys Chem B.*
470 2015;119(9):3706-13.
- 471 27. Chambers VS, Marsico G, Boutell JM, Di Antonio M, Smith GP, Balasubramanian S. High-
472 throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat*
473 *Biotechnol.* 2015;33(8):877-81.
- 474 28. Lim KW, Amrane S, Bouaziz S, Xu W, Mu Y, Patel DJ, et al. Structure of the human
475 telomere in K⁺ solution: a stable basket-type G-quadruplex with only two G-tetrad layers. *J.*
476 *Am. Chem. Soc.* 2009;131(12):4301.
- 477 29. Zhang Z, Dai J, Veliath E, Jones RA, Yang D. Structure of a two-G-tetrad intramolecular G-
478 quadruplex formed by a variant human telomeric sequence in K⁺ solution: insights into the
479 interconversion of human telomeric G-quadruplex structures. *Nucleic Acids Res.*
480 2010;38(3):1009-21.
- 481 30. Beckmann JS, Weber JL. Survey of human and rat microsatellites. *Genomics.*
482 1992;12(4):627-31.

- 483 31. Aoki T, Koch KS, Leffert HL. Attenuation of gene expression by a trinucleotide repeat-rich
484 tract from the terminal exon of the rat hepatic polymeric immunoglobulin receptor gene. *J.*
485 *Mol. Biol.* 1997;267(2):229-36.
- 486 32. Derry JM, Wiedemann P, Blair P, Wang Y, Kerns JA, Lemahieu V, et al. The mouse
487 homolog of the Wiskott–Aldrich syndrome protein (WASP) gene is highly conserved and
488 maps near the scurfy (sf) mutation on the X chromosome. *Genomics.* 1995;29(2):471-7.
- 489 33. Heller M, Flemington E, Kieff E, Deininger P. Repeat arrays in cellular DNA related to the
490 Epstein-Barr virus IR3 repeat. *Mol Cell Biol.* 1985;5(3):457-65.
- 491 34. Hirsch M, Gaugler L, Deagostini-Bazin H, Bally-Cuif L, Goridis C. Identification of positive
492 and negative regulatory elements governing cell-type-specific expression of the neural cell
493 adhesion molecule gene. *Mol Cell Biol.* 1990;10(5):1959-68.
- 494 35. Koch KS, Gleiberman AS, Aoki T, Leffert HL, Feren A, Jones AL, et al. Discordant
495 expression and variable numbers of neighboring GGA-and GAA-rich triplet repeats in the
496 3'untranslated regions of two groups of messenger RNAs encoded by the rat polymeric
497 immunoglobulin receptor gene. *Nucleic Acids Res.* 1995;23(7):1098-112.
- 498 36. Matsugami A, Okuizumi T, Uesugi S, Katahira M. Intramolecular higher order packing of
499 parallel quadruplexes comprising a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad of
500 GGA triplet repeat DNA. *J Biol Chem.* 2003;278(30):28147-53.
- 501 37. Palumbo SL, Memmott RM, Uribe DJ, Krotova-Khan Y, Hurley LH, Ebbinghaus SW. A
502 novel G-quadruplex-forming GGA repeat region in the c-myb promoter is a critical regulator
503 of promoter activity. *Nucleic Acids Res.* 2008;36(6):1755-69.
- 504 38. Cahoon LA, Seifert HS. Transcription of a cis-acting, noncoding, small RNA is required for
505 pilin antigenic variation in *Neisseria gonorrhoeae*. *PLoS Pathog.* 2013;9(1):e1003074.

- 506 39. Lemmens B, Van Schendel R, Tijsterman M. Mutagenic consequences of a single G-
507 quadruplex demonstrate mitotic inheritance of DNA replication fork barriers. *Nat Commun.*
508 2015;6.
- 509 40. Lahaye X, Satoh T, Gentili M, Cerboni S, Conrad C, Hurbain I, et al. The capsids of HIV-1
510 and HIV-2 determine immune detection of the viral cDNA by the innate sensor cGAS in
511 dendritic cells. *Immunity.* 2013;39(6):1132-42.
- 512 41. Cayrou C, Coulombe P, Puy A, Rialle S, Kaplan N, Segal E, et al. New insights into
513 replication origin characteristics in metazoans. *Cell Cycle.* 2012;11(4):658-67.
- 514 42. Lu JZ, Sun YN, Rose RC, Bonne W, McCance DJ. Two E2 binding sites (E2BS) alone or
515 one E2BS plus an A/T-rich region are minimal requirements for the replication of the human
516 papillomavirus type 11 origin. *J Virol.* 1993;67(12):7131-9.
- 517 43. Huppert JL, Balasubramanian S. G-quadruplexes in promoters throughout the human
518 genome. *Nucleic Acids Res.* 2007;35(2):406-13.
- 519 44. Tan SH, Gloss B, Bernard HU. During negative regulation of the human papillomavirus-16
520 E6 promoter, the viral E2 protein can displace Sp1 from a proximal promoter element.
521 *Nucleic Acids Res.* 1992;20(2):251-6.
- 522 45. Kruisselbrink E, Guryev V, Brouwer K, Pontier DB, Cuppen E, Tijsterman M. Mutagenic
523 capacity of endogenous G4 DNA underlies genome instability in FANCD1-defective *C.*
524 *elegans.* *Curr Biol.* 2008;18(12):900-5.
- 525 46. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al.
526 GenBank. *Nucleic Acids Res.* 2013;41(D1):D36-D42.
- 527 47. Huppert JL, Balasubramanian S. Prevalence of quadruplexes in the human genome. *Nucleic*
528 *Acids Res.* 2005;33(9):2908-16.

- 529 48. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features.
530 Bioinformatics. 2010;26(6):841-2.
- 531 49. Davison AC, Hinkley DV. Bootstrap methods and their application: Cambridge university
532 press; 1997.
- 533 50. North BV, Curtis D, Sham PC. A note on the calculation of empirical P values from Monte
534 Carlo procedures. Am J Hum Genet. 2002;71(2):439-41.
- 535 51. Chaires JB, Dean WL, Le HT, Trent JO. Hydrodynamic Models of G-Quadruplex Structures.
536 Methods Enzymol. 2015;562:287-304.
- 537 52. Gray RD, Buscaglia R, Chaires JB. Populated intermediates in the thermal unfolding of the
538 human telomeric quadruplex. J Am Chem Soc. 2012;134(40):16834-44.
- 539 53. Gray RD, Chaires JB. Analysis of multidimensional G-quadruplex melting curves. Curr
540 Protoc Nucleic Acid Chem. 2011;Chapter 17:Unit17 4.
- 541 54. Chariker JH, Miller DM, Rouchka EC. Computational Analysis of G-Quadruplex Forming
542 Sequences across Chromosomes Reveals High Density Patterns Near the Terminal Ends.
543 PLoS One. 2016;11(10):e0165101.
544

545 **Figure captions**

546 **Fig 1. Secondary intramolecular G4 structures (top) and corresponding DNA sequences**
547 **(bottom) with varying numbers of G-tetrads.** This figure has been modified from (54).

548 **Fig 2. Coding and non-coding regions aligned across TmPV1 (outer), TmPV3 (middle), and**
549 **TmPV4 (inner) to illustrate G4 sequences occurring at the same location within regions on**
550 **the forward (left) and reverse (right) DNA strands.** Blue arrows indicate G4 sequences at the
551 same location on TmPV3 and TmPV4. Red arrows indicate G4 sequences at the same location
552 on all three genomes.

553 **Fig 3. Location of G4 sequences in relation to putative E2 binding sites in the NCR region**
554 **of each manatee papillomavirus.** Only locations where the more conservative sequence
555 ACCGNNNCGGT is found are displayed.

556 **Fig 4. Analytical centrifugation (AUC) of TmPV4 oligos (A, B, and C).**

557 **Fig 5. CD-Spectrum analysis of TmPV4 oligos (A, B, and C).** Analysis was performed in
558 tBAP/1 mM EDTA/50 mM KCl/pH 7.0.

559 **Fig 6. Melting curve (SVD) analysis of temperature-dependent CD spectra of TmPV4-3.**

560 Analysis was performed in 10 mM LiPO₄/1 mM EDTA/50 mM KCl/pH 7.0. Optimized
561 parameters were $\Delta H1 = -56.6$, $Tm1 = 81.42$, $\Delta H2 = -55.05$, $Tm2 = 72.98$, $\Delta G1 = -9.8$, $\Delta G2 = -$
562 8.43 , $\Delta G_{total} = -18.23$.

563 **Fig 7. A description of the sequence codes provided by Quadparser for each putative G4**
564 **sequence.** This figure has been modified from (54).

565

566

567

568 **Supporting information captions**

569 **S1 Table. Sequences, locations, and descriptors for putative G4 identified on TmPV1.** Note
570 that all sequences are identified on a reference genome. Thus, G4 sequences on the reverse DNA
571 strand are identified by searching for C-tracts.

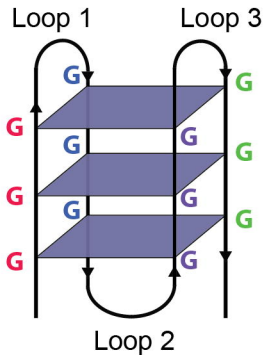
572 **S2 Table. Sequences, locations, and descriptors for putative G4 identified on TmPV3.** Note
573 that all sequences are identified on a reference genome. Thus, G4 sequences on the reverse DNA
574 strand are identified by searching for C-tracts.

575 **S3 Table. Sequences, locations, and descriptors for putative G4 identified on TmPV4.** Note
576 that all sequences are identified on a reference genome. Thus, G4 sequences on the reverse DNA
577 strand are identified by searching for C-tracts.

578 **S4 Table. The number of observed G4 nucleotides (NT), the number of random simulations**
579 **with G4 nucleotides greater than or equal to the observed G4 nucleotides, and the**
580 **associated significance values for each DNA strand in each genomic region on each TmPV.**

581

G4 (3 G-tetrads)



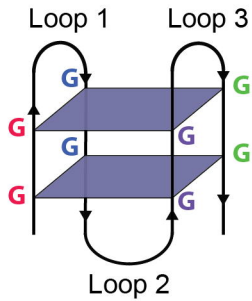
GGG TGAAAAGGG AGCAGCTGGG ATTGGG

↑
Loop 1

↑
Loop 2

↑
Loop 3

G4 (2 G-tetrads)



GGTATGGCGAATGGTGG

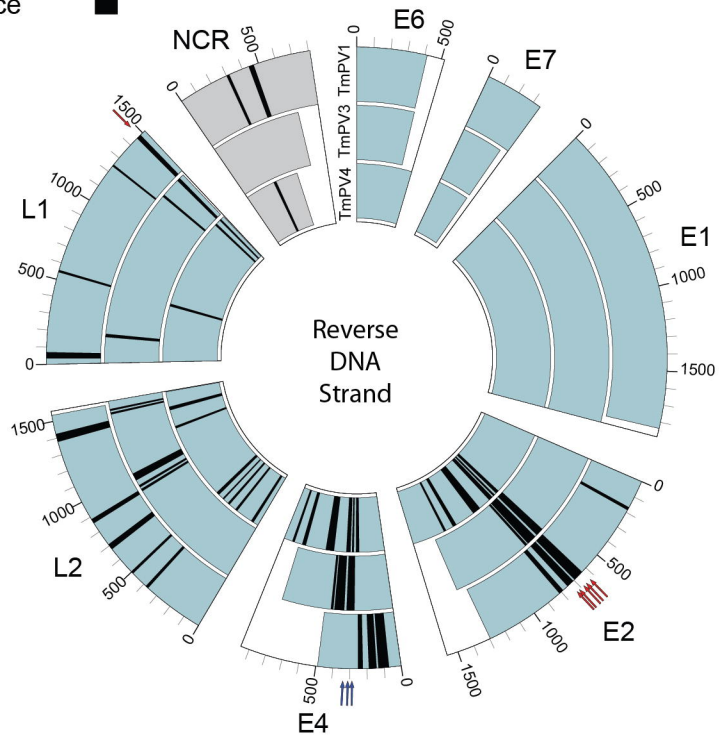
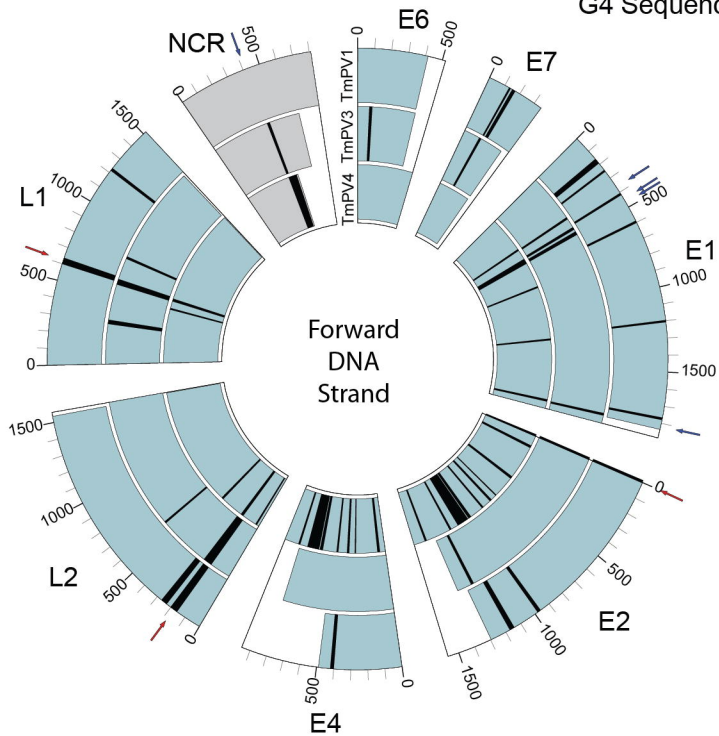
↑
Loop 1

↑
Loop 2

↑
Loop 3

Trichechus manatus latirostris Papillomavirus, Types 1, 3, 4

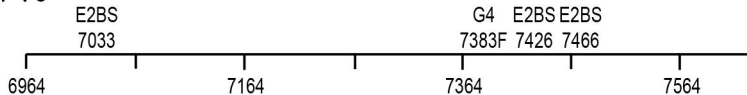
Coding Region
Non-coding Region
G4 Sequence



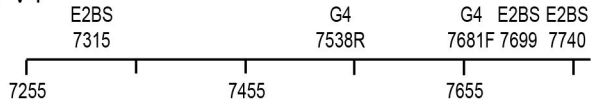
TmPV1

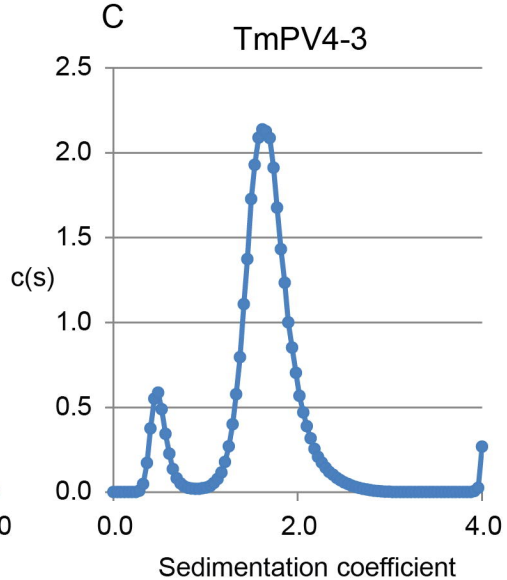
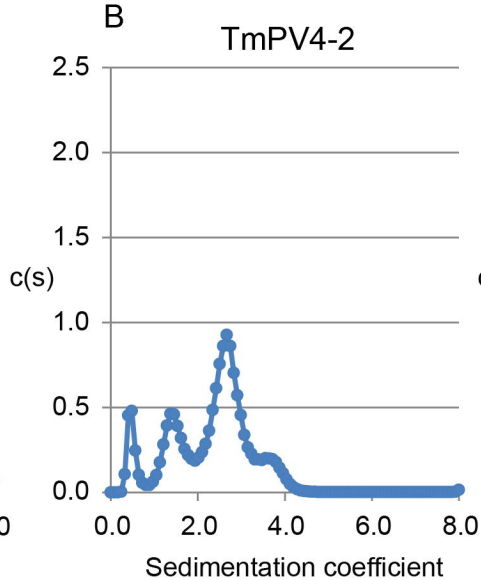
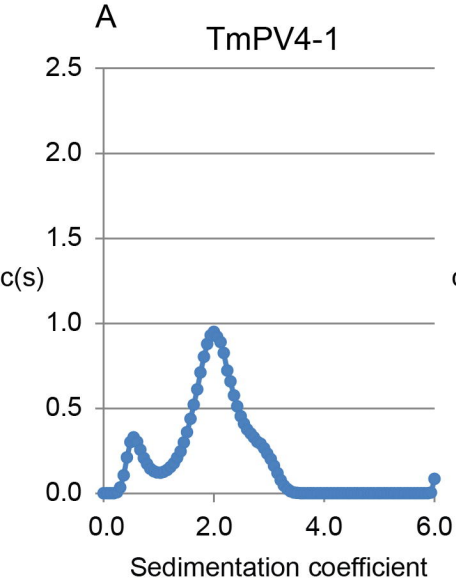


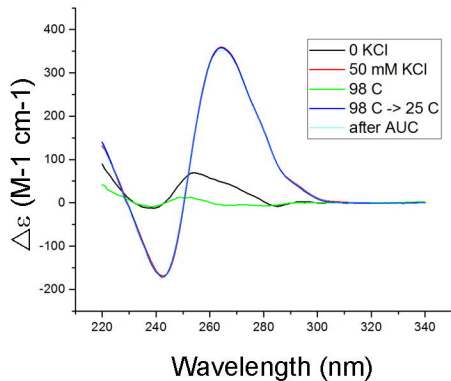
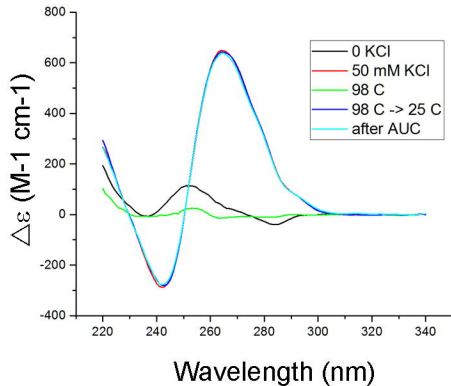
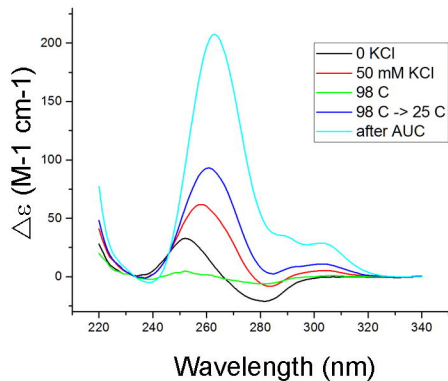
TmPV3

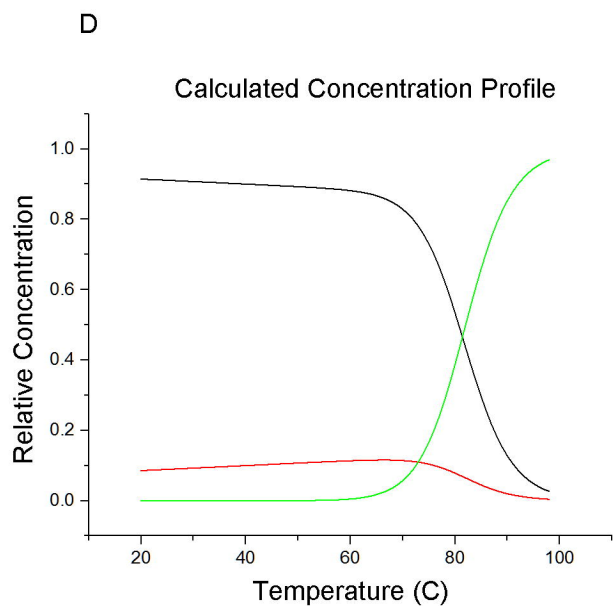
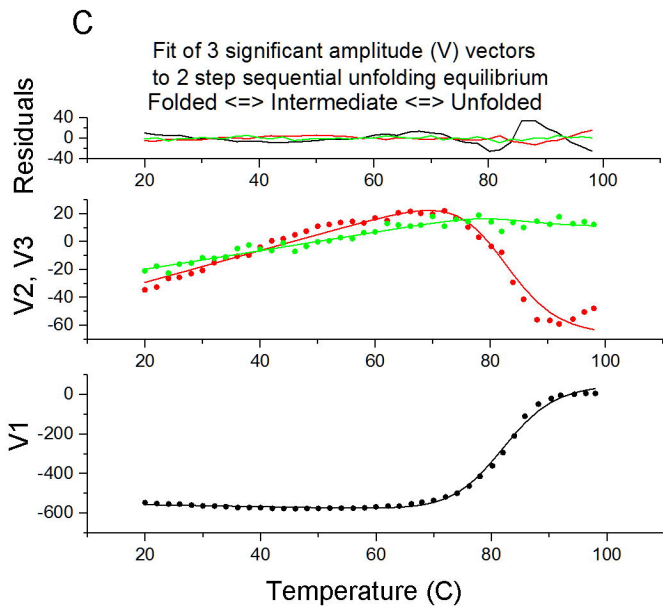
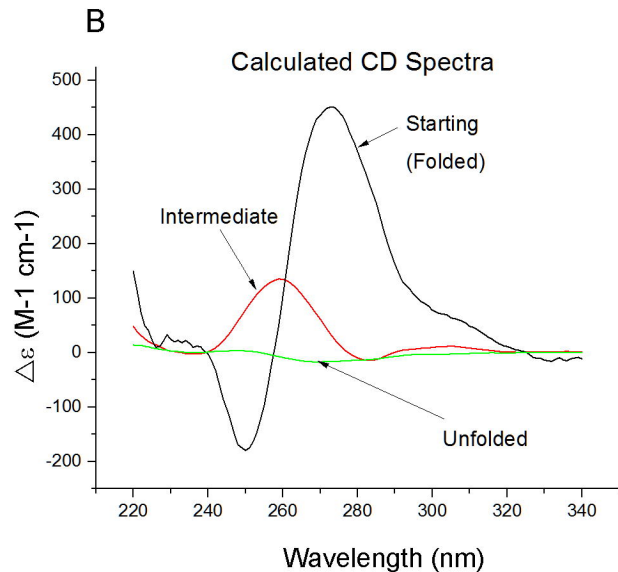
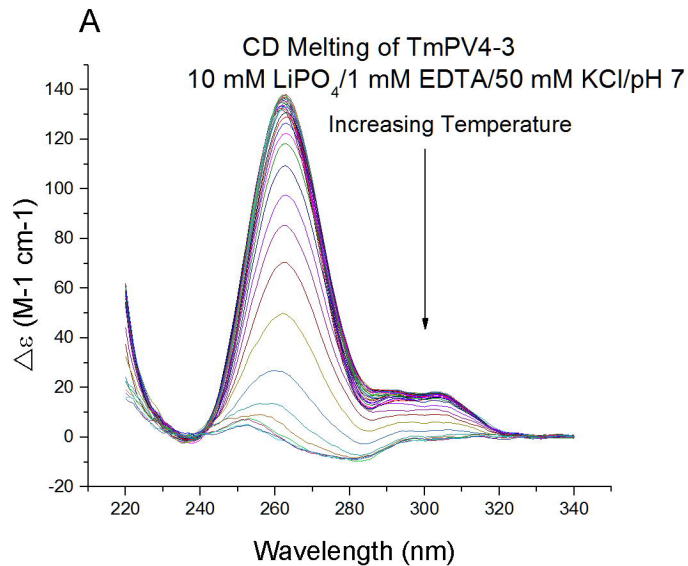


TmPV4





A TmPV4-1**B** TmPV4-2**C** TmPV4-3



X. Number of guanine tracts

4:1:1 $G_{2+}, N_{1-7}, G_{2+}, N_{1-7}, G_{2+}, N_{1-7}, G_{2+}$

Y. Number of locations for G4 formation

5:2:1 $G_{2+}, N_{1-7}, G_{2+}, N_{1-7}, G_{2+}, N_{1-7}, G_{2+}, N_{1-7}, G_{2+}$

Z. Number of possible simultaneous G4 structures

8:5:2 $G_{2+}, N_{1-7}, G_{2+}, N_{1-7}, G_{2+}, N_{1-7}, G_{2+}, N_{1-7}, G_{2+}, N_{1-7}, G_{2+}, N_{1-7}, G_{2+}$