doi:10.1016/j.jmb.2011.03.051 J. Mol. Biol. (2011) xx, xxx–xxx

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Asymmetric Assembly of Merkel Cell Polyomavirus Large T-Antigen Origin Binding Domains at the Viral Origin

Celia J. Harrison[1], Gretchen Meinke[1], Hyun Jin Kwun[2], Henry Rogalin[3], Paul J. Phelan[1], Peter A. Bullock[1], Yuan Chang[2], Patrick S. Moore[2] and Andrew Bohm[1]*

[1]*Department of Biochemistry, Sackler School of Graduate Biomedical Sciences, Tufts University School of Medicine, 136 Harrison Avenue, Boston, MA 02111, USA*
[2]*Molecular Virology Program, University of Pittsburgh Cancer Institute, University of Pittsburgh, 5117 Center Avenue, Pittsburgh, PA 15213, USA*
[3]*Integrated Studies Program, Sackler School of Graduate Biomedical Sciences, Tufts University School of Medicine, 136 Harrison Avenue, Boston, MA 02111, USA*

Received 24 January 2011; received in revised form 17 March 2011; accepted 24 March 2011

Edited by K. Morikawa

Keywords:
viral replication;
polyomavirus;
T-antigen;
protein–nucleic acid cocrystals;
Merkel cell polyomavirus

The double-stranded DNA polyomavirus Merkel cell polyomavirus (MCV) causes Merkel cell carcinoma, an aggressive but rare human skin cancer that most often affects immunosuppressed and elderly persons. As in other polyomaviruses, the large T-antigen of MCV recognizes the viral origin of replication by binding repeating G(A/G)GGC pentamers. The spacing, number, orientation, and necessity of repeats for viral replication differ, however, from other family members such as SV40 and murine polyomavirus. We report here the 2.9 Å crystal structure of the MCV large T-antigen origin binding domain (OBD) in complex with a DNA fragment from the MCV origin of replication. Consistent with replication data showing that three of the G(A/G)GGC-like binding sites near the center of the origin are required for replication, the crystal structure contains three copies of the OBD. This stoichiometry was verified using isothermal titration calorimetry. The affinity for G(A/G)GGC-containing double-stranded DNA was found to be ~740 nM, approximately 8-fold weaker than the equivalent domain in SV40 for the analogous region of the SV40 origin. The difference in affinity is partially attributable to DNA-binding residue Lys331 (Arg154 in SV40). In contrast to SV40, a small protein–protein interface is observed between MCV OBDs when bound to the central region of the origin. This protein–protein interface is reminiscent of that seen in bovine papilloma virus E1 protein. Mutational analysis indicates, however, that this interface contributes little to DNA binding energy.

**Corresponding author*. E-mail address: andrew.bohm@tufts.edu.

Abbreviations used: MCV, Merkel cell polyomavirus; OBD, origin binding domain; JCV, JC virus; dsDNA, double-stranded DNA; LT-ag, large T-antigen; EP, early palindrome; BKV, BK virus; NSLS, National Synchrotron Light Source; ITC, isothermal titration calorimetry; PDB, Protein Data Bank; SSM, secondary-structure matching; BPV, bovine papillomavirus; PEG, polyethylene glycol.

0022-2836/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.

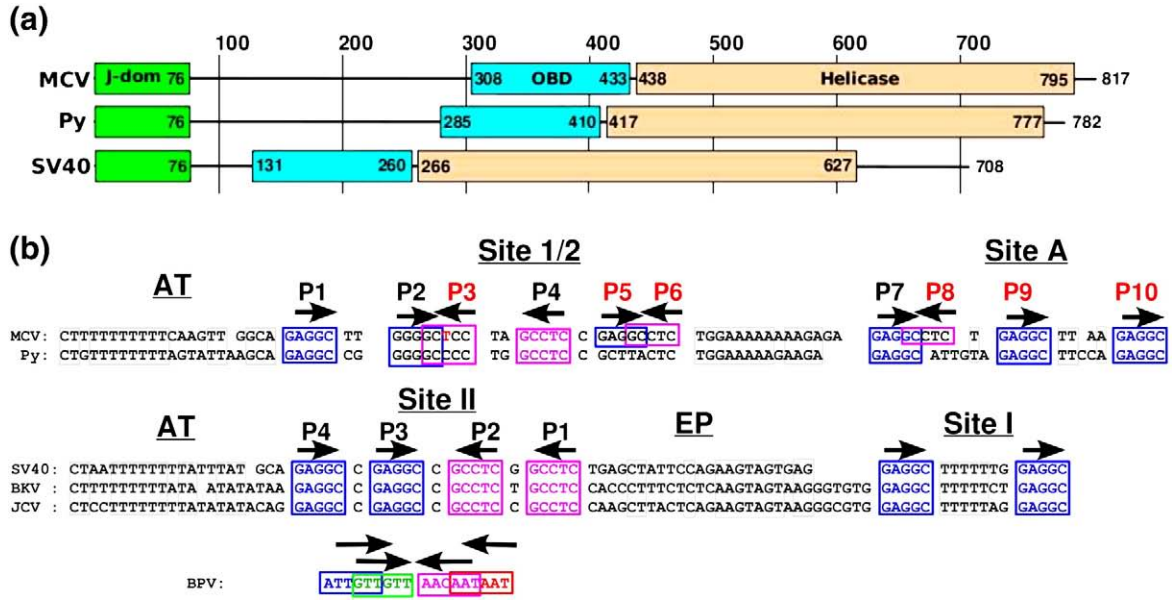

Fig. 1. Domain map of large T-antigens and alignment of origin sequences from Merkel cell virus, murine polyomavirus, SV40, BKV, and JCV. (a) Domains of known/predicted structure in large T-antigens (LT-ag) from SV40, MCV, and murine polyomavirus. J-domains, green; OBDs, cyan; helicase domains, orange. (b) The origin sequences from MCV, murine polyomavirus (Py), SV40, BKV, and JCV have been aligned and colored to emphasize the conservation of GAGGC/GCCTC repeats. The regions labeled Site I are not required for *in vitro* replication and are not strictly part of the origin. However, P7 in the corresponding Site A of MCV is required for replication. These regions are shown to highlight the similarities and differences among the viruses. GAGGC repeats are boxed in blue, and the inverse complement GCCTC is shown in pink. The base that causes the P3 repeat to be imperfect is shown in red. The overlapped repeats of a BPV are shown at the bottom for comparison. The AT-rich regions and the EP regions are indicated. MCV pentameric sequences are numbered P1–P8 from the AT-rich side to the EP side, whereas in SV40, the pentamers are numbered P1–P4 from the EP side to the AT-rich side. P3 is either GGAGC with a 2-bp overlap with P2, analogous to Py,[7] or GAGCC with a 3-bp overlap with P2, as described previously.[3]

Introduction

Merkel cell polyomavirus (MCV) is a recently discovered polyomavirus that has been detected in 80% of samples of human neuroectodermal tumors arising from mechanoreceptor Merkel cells.[1–3] Polyomaviruses have long served as model systems for understanding replication, viral pathogenesis, cellular transformation, and tumorigenesis. The recent discovery of MCV and unanswered questions regarding the related and ubiquitous JC virus (JCV), a causative agent of progressive multifocal leukoencephalopathy in immunosuppressed individuals,[4] highlight the need for a deeper understanding of human DNA polyomaviruses.

Tumorigenic polyomaviruses, such as the monkey virus SV40 and MCV, have double-stranded DNA (dsDNA) genomes that efficiently encode viral capsid and T-antigen proteins.[1,5,6] T-antigen proteins are generated by variable splicing of an early mRNA to yield multiple mature forms.[2] The largest form of T-antigen (LT-ag) contains three well-characterized functional domains: an N-terminal J-domain, an origin binding domain (OBD), and a C-terminal helicase domain (Fig. 1a). The linker region between the J-domain and the OBD varies significantly in length from one virus to another and contains regions generally thought to be critical for cell transformation and tumorigenicity, such as the retinoblastoma protein binding domain.[8] However, the J-domain and the linker are not needed for replication *in vitro*.[9]

In SV40, the most extensively studied model system, the central region of the viral origin (termed Site II) contains four copies of the pentameric DNA sequence GAGGC. These pentamers are named P1 through P4, and are arranged as an inverted repeat with two copies on either side of the dyad. Flanking Site II is the AT-rich region and the early palindrome (EP) region. This architecture allows full-length large T-antigen molecules to assemble to form a dodecameric structure wherein two head-to-head hexameric LT-ag rings encircle the DNA.[10–12] By recognizing the repeating pentameric sequences within the origin, the OBD plays an essential role in positioning these hexameric rings. Having at least two head-to-head OBD binding sites at the origin flanked by easily melted AT-rich regions seems logical. It remains unclear, however, why the origins of polyomaviruses tend to have inverted repeats

Please cite this article as: Harrison, C. J. *et al.*, Asymmetric Assembly of Merkel Cell Polyomavirus Large T-Antigen Origin Binding Domains at the Viral Origin, *J. Mol. Biol.* (2011), doi:10.1016/j.jmb.2011.03.051

within this critical Site II region, or why alterations in the spacing or number of these repeats impair replication.[13] To address this issue, we examined the OBD interactions at the origin of MCV.

In SV40, and in the closely related JCV and BK virus (BKV), there is a one-base spacer between each of the four DNA pentamers (Fig. 1b). In MCV and murine polyomavirus, the analogous region of the viral origin (termed Site 1/2 or simply referred to as ORI in these viruses) contains G(A/G)GGC-like sequences with a different spacing.[3] The sequences approximate an inverted repeat, but the centermost repeats are overlapped, and there is a two-base spacer between the inner set and the outer set of G(A/G)GGC-like pentamers. Moreover, the P3 repeat in MCV (but not in Py) differs from the consensus, with changes that would be predicted to interfere with (but not abrogate) binding, based on work with other viruses.[14] Interestingly, MCV (but not Py or SV40) contains two additional perfect GAGGC pentamers (P5 and P6) in this central: Site 1/2 region of the origin.[3] These are on opposite strands and are overlapped by 2 bp. Mutation analyses indicate that P3, P5, and P6 are not required for efficient replication from the MCV origin.[3] These data are surprising because in SV40 and Py, all four G(A/G)GGC pentamers are thought to be bound, and available data suggest that all are required for replication.[15,16] Thus, while SV40, Py, and more distantly related papillomaviruses all have two sets of 2-fold symmetric DNA binding sites, MCV differs in having an asymmetric arrangement of G(A/G)GGC pentamers. Although the MCV and SV40 LT-ag OBD sequences are ~50% identical and both bind the same pentameric DNA sequence, replication is sensitive to alterations in spacing between the G(A/G)GCC repeats.[13] Thus, the LT-ag molecules are tailored to fit their own particular origin sequence, and the SV40 or MCV origins are not replicated by heterologous viral LT-ag proteins.[2]

Polyomaviruses contain still more GAGGC repeats outside of the ORI [e.g., the early enhancer region in MCV (also known as Site A in Py) and Site I in SV40, as shown in Fig. 1b]. Some of these additional pentamers appear to bind DNA more tightly than those at the origin and help control transcription of viral genes.[7,17] In the MCV origin, P7 (part of an overlapping P7/P8 head-to-head pair) is also required for LT-ag-initiated DNA replication.[3] Mutations at this site in a naturally occurring, tumor-derived viral strain abolish origin replication. A 71-bp minimal core origin spanning from the MCV poly(T) tract to P7/P8 is required for MCV LT-ag-initiated replication.[3]

Crystal structures have shown that the OBDs of SV40 LT-ag bind the major groove of the DNA and interact with all 5 bp within the GAGGC repeat.[18,19] The spacing of adjacent repeats within SV40 Site II results in a protein–DNA complex in which SV40 LT-ag OBDs are noninteracting when all four pentamers are bound. In MCV, however, the first four repeats are more closely spaced, and modeling based on the SV40 structures suggests that the OBDs of MCV LT-ag would either interact with one another when bound to the origin or compete with one another for a subset of potential binding sites. To better understand how the central repeats at the MCV origin are utilized, we present here the crystal structure of the MCV OBD in complex with a 26-nt-long oligonucleotide duplex from the MCV Site 1/2 origin. We also report the results from a calorimetric study of MCV that confirm and extend our structural results.

Results and Discussion

Crystallization and structure solution

The protein used for structure determination contained the MCV OBD (residues 308–433) and an N-terminal hexahistidine tag, which was not removed prior to crystallization and is disordered in the crystal structure. Crystals of the OBD lacking the His-tag were grown under the same conditions but were of poorer quality than those with the His-tag. The oligonucleotide chosen for structural studies has the sequence of the MCV Site 1/2 origin that spans the first through fourth pentameric sequences (P1–P4), starting from the AT-rich side (nucleotides 5382–14; GenBank accession no. EU375804) (Fig. 2). The oligo was designed with 5′ overhangs to promote the formation of pseudo-continuous DNA helices in the crystal. The crystals reported here belong to space group $P2_12_12$ and contain the dsDNA and three OBDs in the asymmetric unit. Diffraction data were collected at National Synchrotron Light Source (NSLS) beamline X-29, and the crystal structure was solved and refined to 2.9 Å resolution, as described in Materials and Methods.

Stoichiometry of the protein–DNA complex

The structure revealed that only three MCV OBDs bound to the Site 1/2 containing DNA (Fig. 2a). The DNA oligo is an imperfect palindrome that contains two head-to-tail-oriented GAGGC-like sequences on each strand. Sites P1, P2, and P4 are occupied by the OBD, but P3, which overlaps with P2 on the opposite strand, is not. Following murine polyomavirus conventions (which differ from those used with SV40), the G(A/G)GGC sequences corresponding to P1 and the P2 site are on the "top strand" (Fig. 2b).

Since the DNA sequence is a pseudo-palindrome and the resolution is not exceptionally high, special care was taken to ensure that the pentamers were

Please cite this article as: Harrison, C. J. *et al.*, Asymmetric Assembly of Merkel Cell Polyomavirus Large T-Antigen Origin Binding Domains at the Viral Origin, *J. Mol. Biol.* (2011), doi:10.1016/j.jmb.2011.03.051

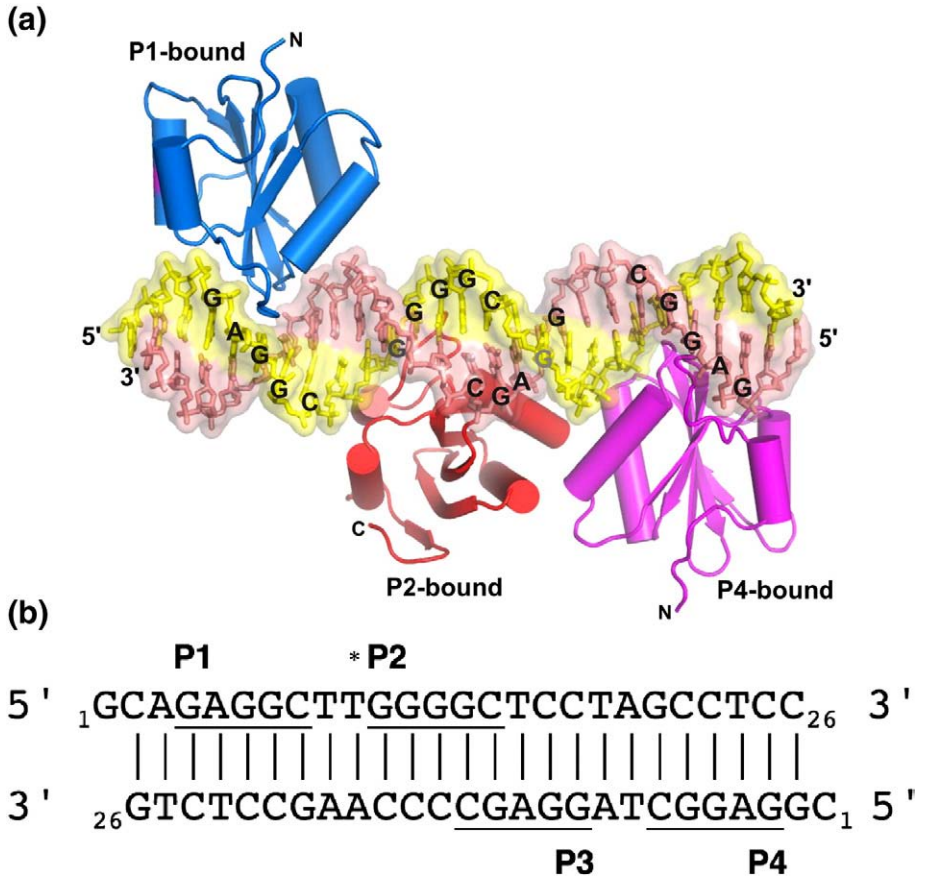

Fig. 2. Overall architecture of the Site 1/2 MCV origin P1–P4 pentamers bound by MCV OBDs, as determined by X-ray crystallography and the oligonucleotide sequence used for crystallization. (a) As indicated, three MCV OBDs are bound to sites P1, P2, and P4. The top DNA strand (yellow) corresponds to the origin sequence shown in Fig. 1b, such that the AT-rich region would be located to the left of the DNA shown in this figure and the EP would be found to the right. Despite the closeness of the P1 and P2 sites in the DNA sequence, the OBDs bound at P1 and P2 are on opposite sides of the helix, a consequence of the spacing in the linear DNA. The amino-terminus and the carboxy-terminus are indicated with N and C, respectively. (b) The 26-mer used for cocrystals with the MCV LT-ag OBD is shown; it contains the P1–P4 pentamers of Site 1/2 from the MCV origin. The position of T10, used for bromouridine substitution, is indicated with an asterisk.

correctly assigned. The assignment was based on the location of a break in the simulated annealing composite omit map electron density between the ends of the DNA oligos (Supplementary Fig. 1). This assignment was confirmed using data from a derivatized oligonucleotide with bromouridine substitution for one of the two thymidines (position T10 in Fig. 2b) between the P1 pentamer and the P2 pentamer. A 3.5σ peak was observed in the difference Fourier map very near the C5 carbon in the pyrimidine ring corresponding to T10, but not at the position that would be expected if the oligo were traced incorrectly (i.e., if P4 were really P1; data not shown).

To confirm the crystallographic results, we used isothermal titration calorimetry (ITC) to investigate the solution-state stoichiometry of binding and to determine the affinity of the MCV OBD for the pentamers of the Site 1/2 origin. We tested two oligos containing the sites P1 through P4. The first oligo was the 26-mer used in the crystal structure (Fig. 2b). The second oligo was a 32-mer with blunt ends (described in Materials and Methods). In both cases, an analysis of the binding isotherms indicated three monomers of the MCV OBD bound to the DNA oligos. The data fitted well to a single binding site model, indicating that the affinities of all three OBDs for the DNA are quite similar. At 25 °C, the K_d of MCV OBD is ~740 nM, with a stoichiometry of 2.79 ± 0.04 MCV OBDs per oligo (Fig. 3, Table 1).

The binding data thus support the crystal structure, which revealed that the P2 site, but not the P3 site, was bound in the crystal. Consistent with these results, earlier work has shown that the P3 pentamer is not necessary for replication.[3] In that study, single mutations unique to the P3 site that do not overlap with the opposite strand P2 have no effect on replication. In contrast, single substitution mutations

Please cite this article as: Harrison, C. J. *et al.*, Asymmetric Assembly of Merkel Cell Polyomavirus Large T-Antigen Origin Binding Domains at the Viral Origin, *J. Mol. Biol.* (2011), doi:10.1016/j.jmb.2011.03.051

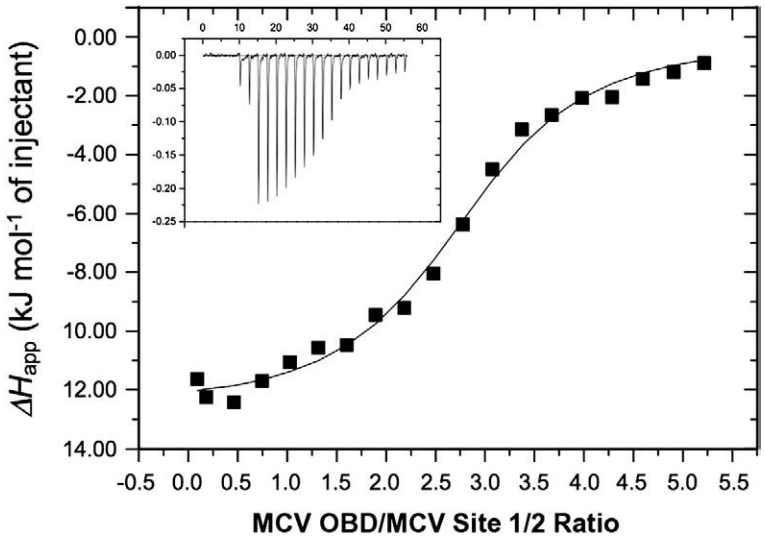

Fig. 3. ITC of MCV origin Site 1/2 with MCV LT-ag OBD reveals three of four sites bound in solution. Titration of MCV LT-ag OBD into a 26-bp duplex DNA containing the P1–P4 sites from the MCV origin Site 1/2 at a concentration of ~4.8 μM in 10 mM sodium phosphate buffer (pH 7.0) and 50 mM sodium chloride at 25 °C. The protein concentration in the syringe was ~182 μM. The actual calorimetric trace is shown in the inset, where the *y*-axis represents power (μcal/s) and the *x*-axis represents time (min). The stoichiometry and association constants are determined from a curve-fitting of the integrated calorimetric trace. We determined a stoichiometry of 2.79:1 for MCV OBDs to duplex DNA, and a dissociation constant (K_d) of 740 nM.

in P1, P2, and P4 impair replication. The earlier study also examined the effect of mutating P5, P6, P7, and P8 outside the central region. Alterations of P5, P6, and P8 had no significant effects on replication efficiency. In contrast, P7 is critical for replication.[3]

Structure of the MCV OBD

The MCV OBD is a well-folded domain of ~120 amino acids (residues 309–428). The structure is a five-stranded anti-parallel β-sheet flanked on both sides by α-helices, with a topological fold of α1β1α2β2β3α3β4α4β5 (αβαββαβαβ). The β-strands are arranged β5–β2–β3–β1–β4 from one side of the β-sheet to the other; α-helices α1 and α4 lie on one face of the β-sheet, while α2 and α3 lie on the other face (Fig. 2a). The fold is similar to that seen in the SV40 OBD,[20,21] papillomavirus E1 initiator protein,[22] RNA binding domains of U1A[23] and PTB,[24] adenoassociated virus Rep protein (replication initiator protein),[25] and the tomato leaf yellow curl virus Rep.[26] These proteins all contain two helices that lie on one face of a four-stranded or five-stranded β-sheet. In MCV, these helices are α2 and α3. As discussed below, α3 interacts with the DNA, and α2 participates in the intermolecular interaction between P2-bound monomers and P4-bound monomers. Among the proteins mentioned above, SV40 OBD is the most similar in structure [r.m.s.d. of ~1.0 Å for ~108 equivalent C^{α} positions; Protein Data Bank (PDB) ID: 2NTC] and has the greatest sequence homology (Fig. 4). For the ~120 residues that comprise the MCV and SV40 OBDs, ~50% of the residues are identical, and 25% of the residues are substituted nonconservatively.

In the three crystallographically independent views of the MCV OBD, the last 10 amino acids are disordered and cannot be modeled. In structures of the SV40 OBD, equivalent residues form a short α-helix that varies in conformation among the different SV40 structures. These residues likely serve as a flexible hinge between the OBD and the helicase domain. This linker region is predicted to be four amino acids shorter in MCV than in SV40. Thus, we expect that the loss of this final α-helix in MCV allows the full-length protein at least as much flexibility as SV40 in this region.

Table 1. ITC

Protein	DNA	K_d (nM)	*n*
MCV	MCV–P1–P2–P3–P4	741	2.79
K331A-MCV	MCV–P1–P2–P3–P4	4630	2.68
K331R-MCV	MCV–P1–P2–P3–P4	546	2.98
F391A-MCV	MCV–P1–P2–P3–P4	990	2.86
F391R-MCV	MCV–P1–P2–P3–P4	2123	2.92
MCV	MCV–P4–P5–P6	205	2.09
MCV	SV40 Site II P1–P2–P3–P4	595	4.09
SV40	MCV–P1–P2–P3–P4	214	2.58
SV40	SV40 Site II P1–P2–P3–P4	93	3.57

Specific DNA interactions

A combination of hydrogen bonds, ionic interactions, and van der Waals interactions is seen at the interface between the DNA and the OBDs (Fig. 5). Owing to the resolution of the structure, no water-mediated hydrogen bonds between the protein and DNA were modeled, although these have been seen in SV40 LT-ag OBD–DNA cocrystal structures. All interactions occur either in the major groove or with the phosphate backbone; there are no minor groove interactions. The DNA is in B-form, and there is a 5° bend between P1 monomers and P2 monomers.

Please cite this article as: Harrison, C. J. *et al.*, Asymmetric Assembly of Merkel Cell Polyomavirus Large T-Antigen Origin Binding Domains at the Viral Origin, *J. Mol. Biol.* (2011), doi:10.1016/j.jmb.2011.03.051

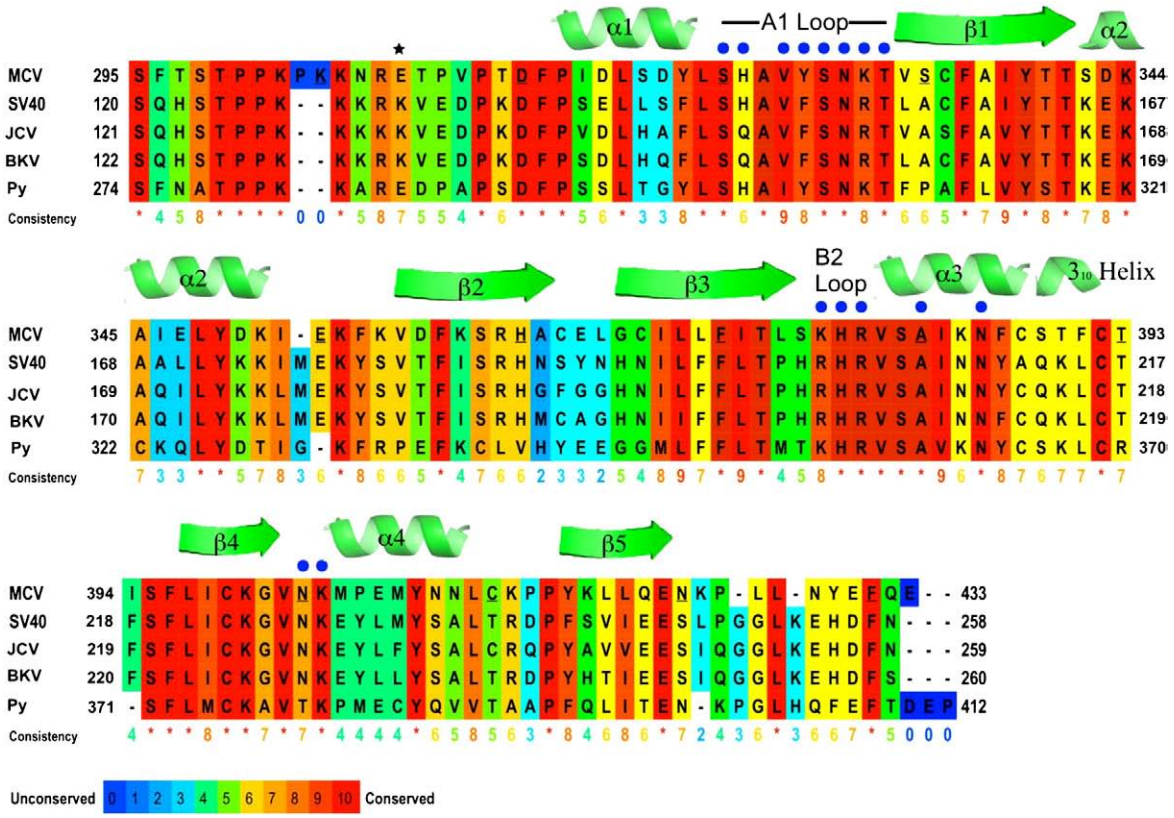

Fig. 4. Protein sequence alignment for viral OBDs. Viral OBDs from MCV, murine polyomavirus (Py), SV40, BKV, and JCV are aligned. Conservation is indicated by color (given by key), where blue is the least conserved and red is the most conserved. The nuclear localization sequences located before the N-terminus of the OBDs are included, but the region of the MCV OBD used in this study was amino acid residues 308–433. Secondary structure elements are indicated for α-helices, β-strands, and a 3_{10}-helix. Blue dots represent residues that interact with the DNA. The star represents amino acid residue 308, the start of the MCV OBD. This figure is based on the output of PRALINE (http://www.ibi.vu.nl/programs/pralinewww/).

Most of the bending occurs at T10 and G11. By comparison, in the SV40 LT-ag OBD–DNA cocrystal (PDB ID: 2NTC), there is a 12° bend between the P1 site and the P3 site; in the SV40 OBD structure with four LT-ag OBDs on four GAGGC pentamers, the DNA is essentially straight (PDB ID: 2ITL).

Four regions of the MCV OBD bind the DNA (Fig. 5). Some protein–phosphate backbone interactions occur outside of the GAGGC pentamers, but specific contacts with DNA bases occur only within the pentameric sequences. The first two DNA interaction motifs involve the A1 and B2 loops.[27] As in the SV40 structures, these loops contribute most of the atoms that interact with the DNA pentamers. The A1 loop (OBD residues 325–333) is located between the first α-helix and the first β-strand. The B2 loop (residues 376–381) is located between the third β-strand and the third α-helix (Fig. 5). The A1 loop extends from the core of the OBD and traverses the major groove from one phosphate backbone to the other. The B2 loop sits at the edge of the major groove, roughly following the phosphate backbone of the 'bottom' strand.

Additional DNA interactions are borne by the third α-helix and by a turn between the fourth β-strand and the fourth α-helix. The third α-helix (residues 382–390) contributes two side chains Ser382 and Asn386, which interact with the DNA phosphate backbone of the 'bottom' strand in some of the OBDs (discussed further below). A final pair of interactions is made by Asn403 and Lys404, which lie at the turn between the fourth β-strand and the fourth α-helix. These residues interact with the phosphate backbone of the 'top' strand.

The three MCV LT-ag OBDs in this structure are similar in their interactions with the DNA, and a pairwise comparison results in an r.m.s.d. of ~0.5 Å for all C^{α} positions. The OBD bound to P2 makes a few contacts that are slightly different from those in the P1- and P4-bound monomers. Given the limited resolution of the crystal structure, it is difficult to determine if the relatively subtle structural differences we observe are real. As discussed below, our binding experiments clearly suggest that the affinities for all three sites are similar. Thus, the structural

Please cite this article as: Harrison, C. J. *et al.*, Asymmetric Assembly of Merkel Cell Polyomavirus Large T-Antigen Origin Binding Domains at the Viral Origin, *J. Mol. Biol.* (2011), doi:10.1016/j.jmb.2011.03.051

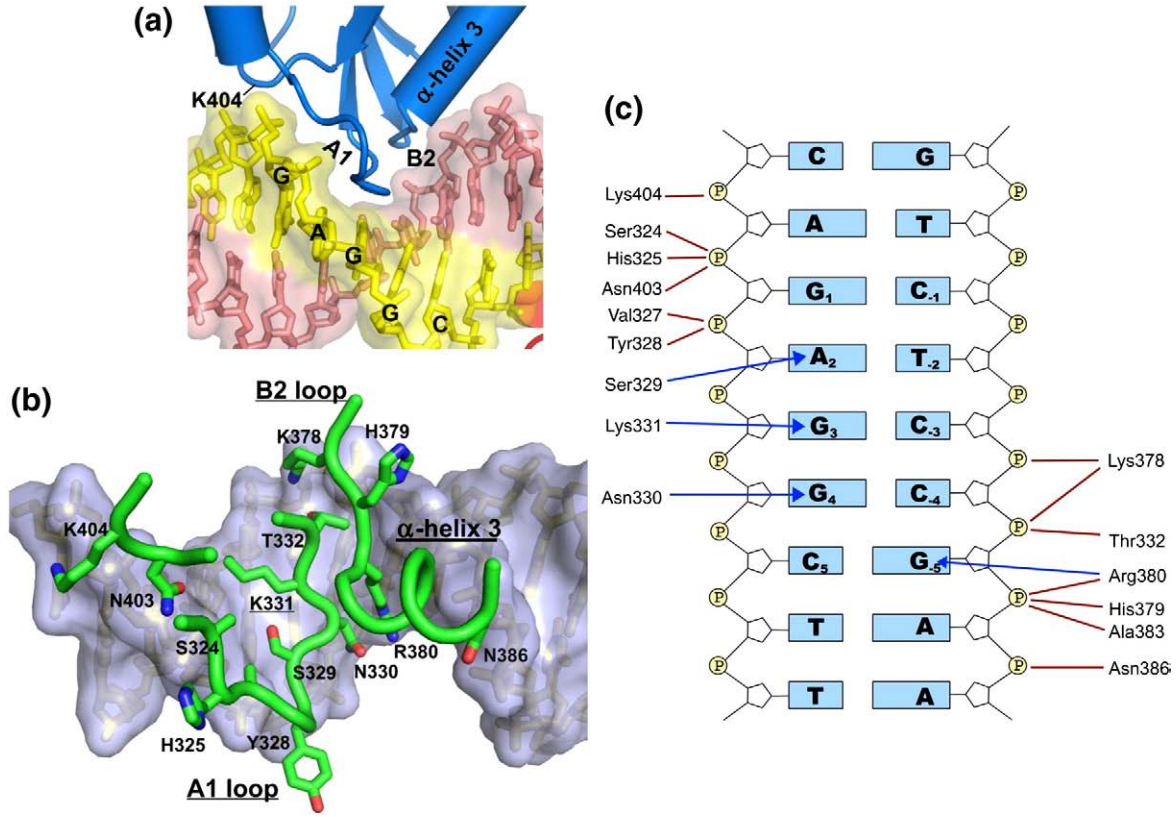

Fig. 5. DNA–protein interactions in the MCV origin. (a) A close-up of protein–DNA interactions in the P1-bound monomer is shown in the same orientation as in Fig. 2a. The A1 and B2 DNA-binding loops of the MCV OBD are indicated. (b) The view looking down into the major groove of the P1 pentamer, showing the side chains of the P1-bound OBD. (c) Consensus representation of the interactions between the MCV OBD and the GAGGC pentameric sequence. Blue lines represent hydrogen bonds to the bases, and red lines represent hydrogen bonds or van der Waals contacts with the phosphoribose backbone.

differences appear to have little effect on the strength of DNA binding.

Given the resolution of the crystallographic data and the slight variations observed in the three MCV OBDs on the DNA, it is difficult to draw absolute conclusions regarding some of the specific intermolecular interactions. For example, the backbone nitrogen of His325 makes a hydrogen bond with a phosphate backbone oxygen in the OBDs bound to P1 and P2, but not P4. The DNA interactions, as judged by the consensus of the three crystallographically unique copies of the OBD in our structure, are summarized in Fig. 5. The interactions are quite similar to those observed in the SV40/DNA complexes, with a few notable differences. Tyr328 is a phenylalanine in SV40 (SV40 Phe151), but it makes similar interactions with the DNA in both structures. Lys378 is an arginine in SV40 (Arg202); this SV40 side chain contacts a backbone sugar, but the analogous residue in MCV contacts the phosphate backbone. The most important difference between MCV and SV40 is at Lys331. In SV40, the residue at this position is an arginine (Arg154). The side chain of Arg154 lies at the base of the major groove and makes base-specific DNA contacts with the first and third nucleotides of the GAGGC pentamer. In MCV, Lys331 interacts only either with the base of the first G or with the third G of the pentamer, but not with both as Arg154 does in SV40.

Mutagenesis of Lys331

SV40 Arg154 has been implicated in several T-antigen functions. In addition to sequence-specific DNA binding, it also plays a role in single-stranded DNA binding,[28] RPA binding,[29] and nonspecific dsDNA binding.[30] We mutated the analogous MCV residue Lys331 to alanine and arginine, and analyzed the affinity of the mutants by ITC. When Lys331 was mutated to alanine, the affinity of MCV OBD for the Site 1/2 oligo used in other ITC experiments decreased markedly, whereas the stoichiometry was unchanged (Supplementary Fig. 2a, Table 1). Compared to the wild-type affinity ($K_d \sim 740$ nM) under the same conditions, the Lys331>Ala mutant binds approximately ~6-fold

Please cite this article as: Harrison, C. J. *et al.*, Asymmetric Assembly of Merkel Cell Polyomavirus Large T-Antigen Origin Binding Domains at the Viral Origin, *J. Mol. Biol.* (2011), doi:10.1016/j.jmb.2011.03.051

less well ($K_d \sim 4.6$ μM). When Lys331 was mutated to arginine, the affinity of the MCV OBD mutant became about one-third stronger than that of wild-type (Lys331 > Arg; $K_d \sim 550$ nM), although it was still weaker than the SV40 OBD for Site II (Supplementary Fig. 2b; Table 1).

Modeling an OBD at P3

In the origins of MCV and Py, sites P1 and P4 have the sequence GAGGC, but P2 and P3 are variations of this sequence (Fig. 2b). P2 is GGGGC in both viruses. Historically, the consensus binding site has been given as G(A/G)GGC. However, recent fluorescence-based binding studies with SV40 LT-ag OBD have demonstrated that changing the second base from A to G has a moderately negative effect on DNA affinity.[14] There is some ambiguity regarding the specific sequence that constitutes P3 in MCV, as there are two possibilities. P3 is either GGAGC with a 2-bp overlap with P2, as it is in Py,[7] or GAGCC with a 3-bp overlap with P2, as described previously.[3] Studies of the SV40 OBD have highlighted the importance of guanosine in the third and fourth positions,[14] suggesting that the P3 sequence is less than ideal in either case.

An interesting question is whether an OBD at P3 could theoretically bind to the MCV Site 1/2 region of the origin. This question is not merely academic, since in Py, the sequence of P3 is the same as P2. To investigate this, we created a molecular model in which a fourth OBD "binds" at the P3 site by superimposing the coordinates of the DNA sequence of the P4 pentamer onto the P3 pentamer, and by allowing the P4-bound OBD to be shifted accordingly. In this model, the P2-bound and P3-bound OBDs have approximately 2-fold symmetry, but collisions occur within the major groove of the DNA between the two copies of Arg380 and Asp330 (Supplementary Fig. 3a). These two residues are conserved in Py and SV40. In our crystal structure and in the SV40 structures, these residues always make the same contacts with the DNA. Rearrangement of these side chains to alleviate clash would result in a loss of at least two hydrogen bonds to the DNA for each monomer. Thus, it is unlikely that they rearrange without significant effects on DNA binding.

To test the theory that the overlapped P2 and P3 pentamers cannot be simultaneously bound, as well as to investigate binding to the P5 and P6 pentamers, we designed an oligo containing P4, P5, and P6 for use with ITC. As shown in Fig. 1b, the P5 and P6 pentamers have the same spacing and orientation as P2 and P3, but whereas P3 has a sequence that is predicted to make it somewhat defective for OBD binding,[14] P5 and P6 both have the same perfect pentamer sequence. Thus, the oligo used in this ITC experiment has three perfect GAGGC pentamers. The ITC revealed that only two of these three sites are bound ($K_d \sim 200$ nm; $n = 2.0$) (Supplementary Fig. 3b; Table 1). These data indicate that, in MCV, only four of the six potential OBD binding sites in Site 1/2 can be occupied at any time. These are P1, P2, P4, and either P5 or P6 (but not both of the latter). The sequence at P7/P8 is the same as that at P5/P6. Thus, presumably, these also cannot be simultaneously occupied.

The modeling and ITC results reported here suggest that although all four of these binding sites have the canonical G(A/G)GGC sequence, only three of the four can be bound at any one time. The outer two sites should always be bound, but a steric clash involving two conserved protein side chains that bind the DNA will prevent simultaneous occupancy of the inner two sites. The well-studied murine polyomavirus origin has the same spacing of binding sites as the first four sites in MCV. This issue has been somewhat addressed in two earlier studies. DNase I footprinting suggested that the entire Site 1/2 region was protected,[31] but the size of the nuclease may have prevented digestion of the incompletely occupied P2 and P3 pentamers. A second study examined the effect of guanine methylation of specific bases within the Py origin and found that alteration of specific bases within the inner two sites had an effect different from those of the outer pentamers.[7] This suggests that both inner sites in Py are bound, and the orientation of the OBDs when bound to these sites is different from that observed in this complex and in related SV40 DNA–protein complexes. Alternatively, mutually exclusive binding of the middle two sites may have complicated the data interpretation. Clearly, further investigation is required to fully understand OBD interactions at the Py origin.

Protein–protein interactions

The P2 and P4 pentamers lie on the same face of the DNA in a head-to-head orientation, and a roughly 2-fold symmetric protein–protein interface is created by the MCV OBDs bound to these sites (Fig. 2a; Supplementary Fig. 4). Secondary structure element α-helix α2 (amino acids Glu347, Lys351, and Lys354) and the second half of α-helix α3 (Phe387, Thr390, and Phe391) contribute side chains to this interface. At this resolution, it is difficult to say if water is fully excluded, but approximately 400 Å^2 of surface area is buried. The interface is primarily hydrophobic, with two phenylalanine side chains (Phe391 on both molecules) in close proximity to the aliphatic portions of the two Thr390 side chains. Two Phe387 residues contribute to the general hydrophobic environment around the two Phe391 and Thr390 side chains.

Despite the 2-fold appearance of the protein–protein interaction site, the interactions are not

Please cite this article as: Harrison, C. J. *et al.*, Asymmetric Assembly of Merkel Cell Polyomavirus Large T-Antigen Origin Binding Domains at the Viral Origin, *J. Mol. Biol.* (2011), doi:10.1016/j.jmb.2011.03.051

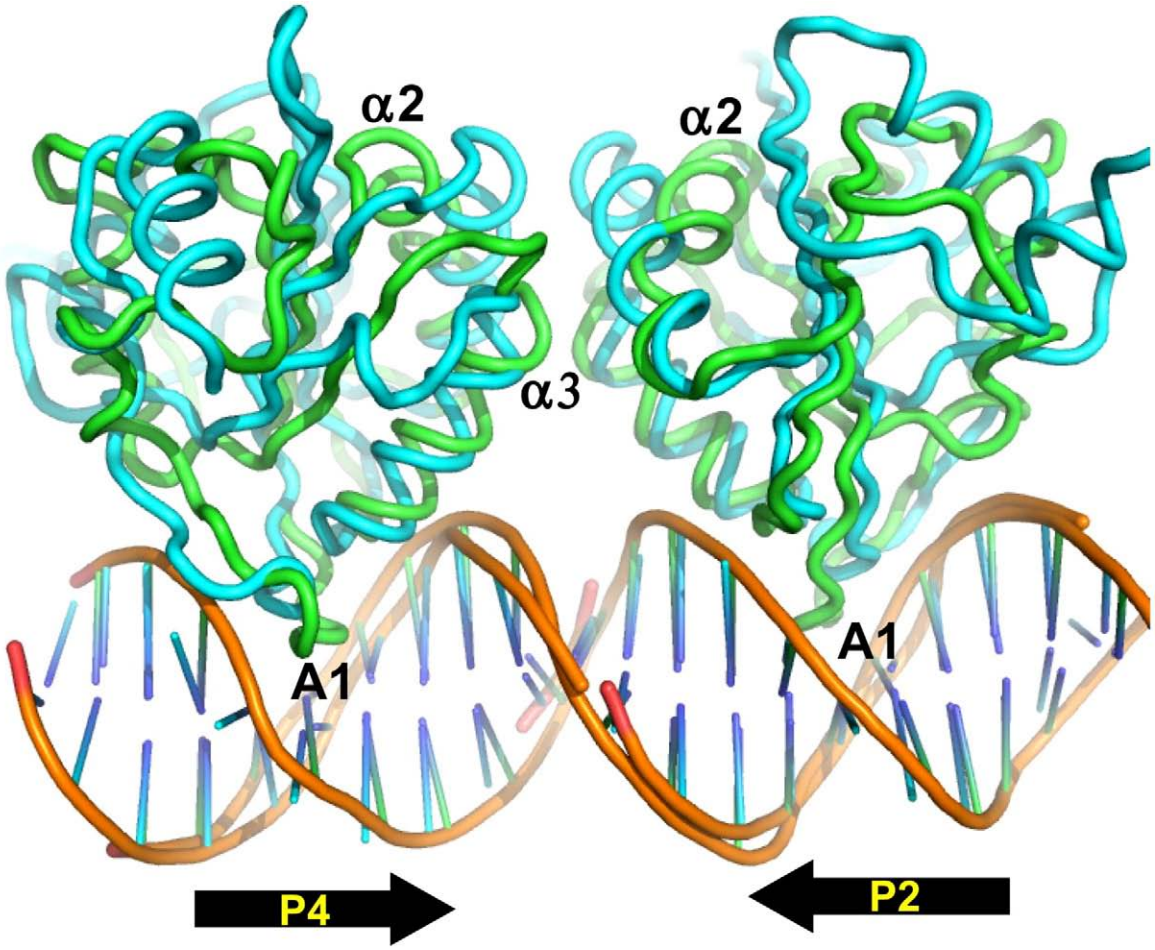

Fig. 6. Superimposition of MCV LT-ag and BPV E1 OBDs. The superimposition of the MCV LT-ag and BPV E1 OBDs reveals a common protein–protein interaction interface and conserved spatial relationship to the cognate DNA. Both of the MCV OBDs bound to P2 and P4 were aligned by SSM algorithms to both of the E1 initiator proteins bound to the E1-1 and E1-3 sites, as found in PDB ID 1KSY; the DNA associated with the E1 initiator proteins was not forced to be superimposed on the DNA in the MCV structure.

perfectly symmetric. For example, Lys351 on the MCV OBD bound to P2 is in a good position to make a water-mediated contact with Glu347 of the P4-bound MCV OBD. However, the cognate Lys351 of the P4-bound OBD is disordered and cannot be modeled in the electron density. A few side chains near, but not immediately at, the protein–protein interface remain disordered (Lys351 of P4, and Thr393 of both P2-bound and P4-bound OBDs).

This sort of protein–protein interface, which results from the P2 and P4 pentamers being oriented head-to-head on the same face of the DNA, has not been seen in other polyomaviral OBD costructures. It is not yet known what structural and functional consequences this has for assembly of the helicase hexamers or double hexamers. Surprisingly, the MCV OBDs bound to P2 and P4, when aligned by secondary-structure matching (SSM) algorithms to the distantly related bovine papillomavirus (BPV) E1 initiator protein (an origin-binding protein), revealed a conserved protein–protein interaction and spatial relationship to the DNA (Fig. 6). The r.m.s.d. of C^{α} positions by SSM superimposition of both of the MCV OBDs bound to P2 and P4, and both of the E1 initiator proteins bound to the E1-1 and E1-3 sites, as found in PDB ID: 1KSY, is ~3.4 Å. The DNA associated with the E1 initiator proteins was not forced to be superimposed on the DNA in the MCV, so that the good alignment of bases seen in Fig. 6 is a consequence of the conserved spatial relationships between the protein and the DNA in MCV and BPV. Several side chains of MCV and BPV E1 at the protein–protein interface appear to be similar in conformation and/or conserved. For example, Phe391 of MCV LT-ag, which is at the center of the 2-fold symmetry that relates the two OBDs, is Met253 in BPV's E1. Thr390 of MCV corresponds to Asn252 of E1, and MCV Lys177 lies close to Lys213 of E1. The secondary structure elements of α-helix

Please cite this article as: Harrison, C. J. *et al.*, Asymmetric Assembly of Merkel Cell Polyomavirus Large T-Antigen Origin Binding Domains at the Viral Origin, *J. Mol. Biol.* (2011), doi:10.1016/j.jmb.2011.03.051

α2 and the second half of α-helix α3, which bear these amino acid residues, thus form a conserved feature of protein–protein interactions in some viral origins.

To probe the functional importance of the OBD–OBD interface, we undertook a calorimetric analysis. Phe391, which is at the center of the protein–protein interface, was mutated to both alanine and arginine to investigate the energetics of the interface. When Phe391 was mutated to alanine, no significant effect on protein–DNA affinity or stoichiometry could be measured by ITC ($K_d \sim 990$ nM; $n = 2.86$) (Supplementary Fig. 5a; Table 1). When Phe391 was mutated to arginine, a slight decrease in affinity was found, with normal stoichiometry ($K_d \sim 2.1$ μM; $n = 2.92$) (Supplementary Fig. 5b; Table 1). One might expect that the mutation of Phe391 to arginine would affect the affinity of OBDs bound to P2 and P4, but not to P1, which would suggest that a two-site model could fit the calorimetric data better. The data were fitted both ways, and the single-site model had a lower value for chi square/degrees of freedom compared to the two-site model. Thus, the MCV interface appears to contribute little to binding energy. For BPV E1 proteins, the dimerization interface was confirmed by mutagenesis; the interaction is not orientationally rigid but rather is a flexible hinge point.[32]

Viral specificity

In vitro origin replication assays performed previously demonstrate that despite similarities in sequence between the replication origin of MCV and the replication origin of SV40, neither MCV LT-ag nor SV40 LT-ag initiates replication from each other's viral origin.[2] It is not known whether assembly on the origins fails or whether the functional defect is farther downstream in the replication process. To better understand the differences between the MCV replication system and the SV40 replication system, we performed ITC experiments in which we examined the DNA binding properties of SV40 and MCV OBDs that were titrated with origin-like oligonucleotides from the alternate virus. We found that the SV40 LT-ag OBD binds to the MCV origin containing the P1–P4 pentamers with weaker affinity than it binds to an oligo containing the SV40 Site II pentamers [$K_d \sim 210$ nM ($n = 2.58$) and $K_d \sim 93$ nM ($n = 3.57$), respectively] (Fig. 7a and b, Table 1). Previously, a K_i constant of 31 nM was reported for SV40 OBDs and Site II-containing oligos, using fluorescence anisotropy.[33] Consistent with our conclusion that the P2 and P3 pentamers of the MCV origin cannot be bound simultaneously, only ~3 SV40 OBDs bound to the four potential pentamers at the MCV origin. The deviations in the value of n from the expected integer values are discussed in Supplementary Materials.

The reverse experiment, in which the MCV OBD was titrated into SV40 Site II DNA containing four pentamers, revealed that the MCV OBD binds SV40 Site II ($K_d \sim 595$ nM, stoichiometry $n = 4.09$; data not shown; Table 1) and MCV Site 1/2 with similar affinities. Furthermore, all four pentamers of the SV40 Site II oligo were bound by the MCV OBD. These reciprocal origin binding experiments demonstrate that the differing (and lower) affinity of MCV for its origins is more likely a function of its protein sequence, and not the geometry of the SV40 or MCV origin. Furthermore, the SV40 OBD binding to the MCV origin recapitulates the stoichiometry for the MCV OBD, confirming our hypothesis that the P2 and P3 pentamers of the MCV Site 1/2 cannot be simultaneously bound. These *in vitro* data indicate that the MCV and SV40 OBDs can interchangeably bind each other's viral origin, suggesting that quaternary interactions with complete origins and full-length T-antigen molecules are likely to be responsible for the specificity.

Conclusion

We have determined the 2.9 Å costructure of the MCV OBD in complex with DNA corresponding to the first four GAGGC-like sequences (P1–P4) of Site 1/2 within the MCV origin of replication. We show that the OBD of MCV binds the pentanucleotides at positions P1, P2, and P4. The murine polyomavirus origin has an architecture similar to that of the four binding sites at its origin, and we believe it likely that it will exhibit similar OBD binding properties. Although there are some structural similarities with the E1 initiator protein costructures, the asymmetry observed in the DNA–protein interaction has not been seen in earlier structures of the analogous regions of the SV40 and papillomavirus origin in complex with the equivalent domain. In those structures, all four sites were bound. We have also addressed binding to two additional binding sites (P5 and P6) within the MCV origin. These are unique to MCV and are dispensable for replication. We have shown that these sites can also bind the OBD, but they cannot be simultaneously occupied.

Although the atomic details of the MCV OBD interaction with DNA are quite similar to those seen in SV40, the number and arrangement of essential sites seem to represent a significant difference. Even though all four GAGGC pentamers at the SV40 origin are required for efficient replication, and all four can be simultaneously occupied by the OBD, only two pentamer sequences are required for double hexamer assembly of full-length T-antigen.[16,34] Moreover, DNA footprinting studies with SV40 have indicated that in spite of the high degree of symmetry and the absence of steric impediments, three of the four

Please cite this article as: Harrison, C. J. *et al.*, Asymmetric Assembly of Merkel Cell Polyomavirus Large T-Antigen Origin Binding Domains at the Viral Origin, *J. Mol. Biol.* (2011), doi:10.1016/j.jmb.2011.03.051

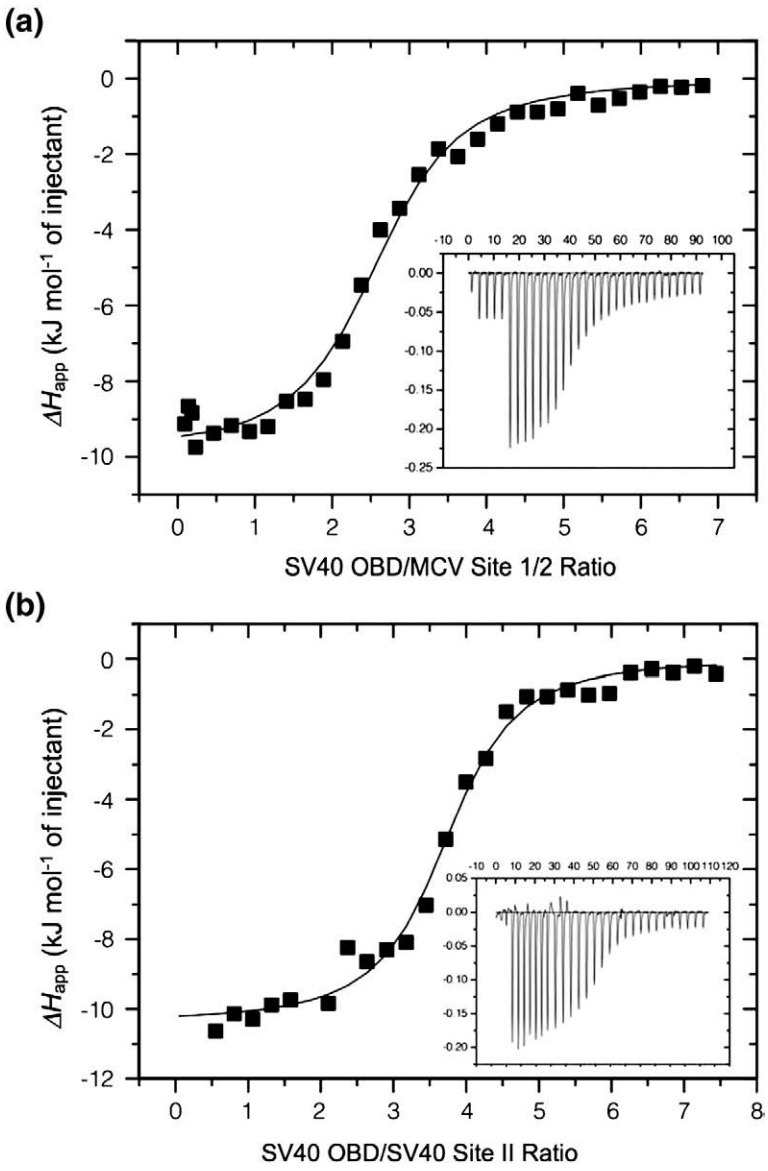

Fig. 7. ITC of SV40. The calorimetric trace is shown in the inset, where the *y*-axis represents power (μcal/s) and the *x*-axis represents time (min). The stoichiometry and association constants are determined from curve-fitting of the integrated calorimetric trace. All experiments were performed in 10 mM sodium phosphate buffer (pH 7.0) and 50 mM sodium chloride at 25 °C. (a) Titration of SV40 LT-ag OBD with a 32-bp duplex DNA containing the P1–P4 sites from the MCV origin Site 1/2. The SV40 LT-ag OBD was used at a concentration of 65 μM, and the MCV Site 1/2 oligo was used at a concentration of 2.0 μM. A stoichiometry of 2.58:1 for SV40 LT-ag OBDs to duplex DNA, with a dissociation constant K_d of 214 nM, was determined. (b) Titration of SV40 LT-ag OBD with a 33-bp duplex DNA containing the P1–P4 pentamers from the SV40 Site II. The SV40 LT-ag OBD was used at a concentration of 55 μM, and the SV40 Site II oligo was used at a concentration of 1.6 μM. A stoichiometry of 3.57:1 for SV40 LT-ag OBDs to duplex DNA, with a dissociation constant K_d of 93 nM, was determined.

sites (P1 through P3) are preferentially occupied.[16] The footprinting studies suggest that, under some conditions or during some stages of replication, only three of the four SV40 pentamers are occupied.

It is unclear from a mechanistic standpoint why some polyomaviruses have more GAGGC-like repeats than others, or why the spacing between GAGGC sites is often critical for efficient replication. The results presented here provide insight into some mechanistic avenues that do *not* seem to play a part in LT-ag function. First, in this structure, intermolecular OBD interactions do not appear to distort the DNA in Site 1/2. Thus, even when the OBD binding sites are in very close proximity, no large perturbations in DNA structure are necessarily observed. Second, the intermolecular interaction between MCV OBDs on the DNA is not necessarily important for DNA binding. Mutation of the interacting residues in MCV did not significantly alter the affinity for the DNA. Interestingly, interactions among OBDs, when bound to their origin DNA, are present in some viruses (MCV, papillomaviruses, and most likely Py) and absent in others (SV40 and most likely JCV and BKV). Thus, although the interactions do seem to be structurally conserved between MCV and papillomavirus E1, they are unlikely to be of general mechanistic importance for the assembly of viral initiators. Finally, we conclude that the specificity of viral T-antigens for their own origin does not appear to reside within the OBDs. Presumably, other parts of the large T-antigen are responsible for specificity.

The data presented here are generally consistent with the prevailing notion that sequence-specific

Please cite this article as: Harrison, C. J. *et al.*, Asymmetric Assembly of Merkel Cell Polyomavirus Large T-Antigen Origin Binding Domains at the Viral Origin, *J. Mol. Biol.* (2011), doi:10.1016/j.jmb.2011.03.051

interactions between the OBD and the repeating elements within the origin help nucleate the formation of the LT-ag double hexamer at the site of replication initiation.[12] The work with MCV has shown that the molecular details associated with the assembly of the LT-ag complex vary in significant ways from one virus to another. Additional work involving intact LT-ag is clearly warranted, as are studies aimed at understanding the role of the GAGGC sequence found at P7. This sequence lies outside of the central region of the origin, and its essential nature with respect to MCV replication is not explained by the prevailing models of LT-ag assembly.

Materials and Methods

Protein overexpression and purification

Briefly, the DNA sequence encoding the MCV OBD residues 308–433 was cloned into pET15b and overexpressed in BL21(DE3) *Escherichia coli* cells in a standard IPTG-dependent method. Cells were lysed, and the N-terminally His-tagged protein was purified by nickel-affinity chromatography and cation-exchange chromatography. Purified protein was concentrated to 15 mg/mL and stored at −80 °C in 25 mM Tris (pH 7.5), 100 mM NaCl, and 5% (vol/vol) glycerol.

Crystallization, data collection, and structure solution

Oligonucleotides were purchased from IDT DNA on the 1.0 μm scale and purified by high-pressure ion-exchange chromatography, as previously described.[18] Complexes of the MCV origin-containing oligonucleotide (given in Fig. 2b) and the OBD were made by mixing 100 nmol of protein and 28 nmol of oligo (containing four binding sites) and concentrating to a final volume of 100 μl, resulting in a complex that was about 1 mM in concentration with respect to the protein, with a 1.12 molar excess of binding sites. High-throughput crystallization methods were used with commercially available crystallization screens.

The most productive crystal form grew in 100 mM Tris (pH 7.1), 200 mM lithium acetate, and 19–21% polyethylene glycol (PEG) 3350 (wt/vol) at 4 °C. Crystals grew as orthogonal rods or plates, and belonged to space group $P2_12_12$. Data were collected on crystals that had been slowly dehydrated by moving hanging drop coverslips from reservoirs containing about 19–20% PEG 3350 (wt/vol) to reservoirs containing 27–28% PEG 3350 (wt/vol), as well as salt and buffer. The desiccated crystals required no cryoprotectant, nor were they stable to the addition of any cryoprotectant, and were stream frozen prior to data collection at NSLS/Brookhaven on beamline X-29. Data were collected, integrated, and reduced with Mosflm.[35] The data extend to 2.9 Å resolution. Extensive data pruning was required to remove data in the resolution ranges normally associated with ice rings (Supplementary Table 1).

Molecular replacement methods (Phaser[36]) were used to solve the crystal structure. Initial search models were created by modeling the MCV OBD on the SV40 OBD, swapping side chains, and including the dsDNA pentamer GAGGC from PDB ID 2NTC. Phaser readily found three OBD–DNA models for which translation and rotation functions were meaningful. The models were built in Coot;[37] CNS[38] and PHENIX[39] were used for refinement. The model was refined to 2.9 Å resolution with all appropriate geometrical and crystallographic constraints. Axial bending of the DNA was analyzed with the program Curves+.[40] The protein–protein interface and the protein–DNA interactions were analyzed using PISA† and PDBSUM‡.

ITC data were collected with either the VP-ITC Calorimeter or the ITC-200 Calorimeter (Microcal, Northampton, MA), and the data were analyzed with Origin software provided by the manufacturer. Prior to the experiments, the dsDNA oligos and proteins were buffer-exchanged into 10 mM sodium phosphate buffer (pH 7.0) and 50 mM NaCl, using PD-10 columns (GE Healthcare). Protein and DNA concentrations were determined spectrophotometrically, using calculated extinction coefficients from the ProtParam Web server§ and the IDTDNA Web site‖, respectively.

PDB accession codes

Atomic coordinates and structure factors for the MCV OBD–DNA cocrystal reported here have been deposited with the Research Collaboratory for Structural Bioinformatics PDB. The accession code is 3QFQ.

Supplementary materials related to this article can be found online at doi:10.1016/j.jmb.2011.03.051

Acknowledgements

We wish to thank Howard Robinson of the NSLS for data collection. We also thank the following for access and assistance during the ITC experiments: Knut Langsetmo, David Jeruzalmi, Michael Durney, Rachel Nager, Gillian Henry, and Jim Baleja. We also thank Brian Schaffhausen for critical reading of the manuscript. This work was supported by National Institutes of Health grant R21A1082496 to A.B. P.A.B. was supported by National Institutes of Health grant R01GM055397. Y.C. and P.S.M. were supported by National Institutes of Health grants CA136363 and CA120726 and were also supported as American Cancer Society professors.

† http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html

‡ http://www.ebi.ac.uk/pdbsum/

§ http://ca.expasy.org/tools/protparam.html

‖ http://biophysics.idtdna.com/UVSpectrum.html

References

1. Feng, H., Shuda, M., Chang, Y. & Moore, P. S. (2008). Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science*, **319**, 1096–1100.
2. Shuda, M., Feng, H. C., Kwun, H. J., Rosen, S. T., Gjoerup, O., Moore, P. S. & Chang, Y. (2008). T antigen mutations are a human tumor-specific signature for Merkel cell polyomavirus. *Proc. Natl Acad. Sci. USA*, **105**, 16272–16277.
3. Kwun, H. J., Guastafierro, A., Shuda, M., Meinke, G., Bohm, A., Moore, P. S. & Chang, Y. (2009). The minimum replication origin of Merkel cell polyomavirus has a unique large T-antigen loading architecture and requires small T-antigen expression for optimal replication. *J. Virol.* **83**, 12118–12128.
4. White, M. K. & Khalili, K. (2011). Pathogenesis of progressive multifocal leukoencephalopathy—revisited. *J. Infect. Dis.* **203**, 578–586.
5. Fanning, E. & Knippers, R. (1992). Structure and function of simian virus 40 large tumor antigen. *Annu. Rev. Biochem.* **61**, 55–85.
6. Simmons, D. T. (2000). SV40 large T antigen functions in DNA replication and transformation. *Adv. Virus Res.* **55**, 75–134.
7. Cowie, A. & Kamen, R. (1986). Guanine nucleotide contacts within viral DNA sequences bound by polyomavirus large T antigen. *J. Virol.* **57**, 505–514.
8. Pipas, J. M. (2009). SV40: Cell transformation and tumorigenesis. *Virology*, **384**, 294–303.
9. Weisshart, K., Bradley, M. K., Weiner, B. M., Schneider, C., Moarefi, I., Fanning, E. & Arthur, A. K. (1996). An N-terminal deletion mutant of simian virus 40 (SV40) large T antigen oligomerizes incorrectly on SV40 DNA but retains the ability to bind to DNA polymerase alpha and replicate SV40 DNA *in vitro*. *J. Virol.* **70**, 3509–3516.
10. Mastrangelo, I. A., Hough, P. V. C., Wall, J. S., Dodson, M., Dean, F. B. & Hurwitz, J. (1989). ATP-dependent assembly of double hexamers of SV40 T antigen at the viral origin of DNA replication. *Nature*, **338**, 658–662.
11. Valle, M., Gruss, C., Halmer, L., Carazo, J. M. & Donate, L. E. (2000). Large T-antigen double hexamers imaged at the simian virus 40 origin of replication. *Mol. Cell. Biol.* **20**, 34–41.
12. Gai, D., Chang, Y. P. & Chen, X. S. (2010). Origin DNA melting and unwinding in DNA replication. *Curr. Opin. Struct. Biol.* **20**, 756–762.
13. Parsons, R. & Tegtmeyer, P. (1992). Spacing is crucial for coordination of domain functions within the simian virus 40 core origin of replication. *J. Virol.* **66**, 1933–1942.
14. Fradet-Turcotte, A., Vincent, C., Joubert, S., Bullock, P. A. & Archambault, J. (2007). Quantitative analysis of the binding of simian virus 40 large T antigen to DNA. *J. Virol.* **81**, 9162–9174.
15. Dean, F. B., Borowiec, J. A., Ishimi, Y., Deb, S., Tegtmeyer, P. & Hurwitz, J. (1987). Simian virus 40 large tumor antigen requires three core replication origin domains for DNA unwinding and replication *in vitro*. *Proc. Natl Acad. Sci. USA*, **84**, 8267–8271.
16. Joo, W. S., Kim, H. Y., Purviance, J. D., Sreekumar, K. R. & Bullock, P. A. (1998). Assembly of T-antigen double hexamers on the simian virus 40 core origin requires only a subset of the available binding sites. *Mol. Cell. Biol.* **18**, 2677–2687.
17. Muller, D., Ugi, I., Ballas, K., Reiser, P., Henning, R. & Montenarh, M. (1987). The AT-rich sequence of the SV40 control region influences the binding of SV40 T-antigen to binding site-2 and site-3. *Virology*, **161**, 81–90.
18. Meinke, G., Phelan, P. J., Moine, S., Bochkareva, E., Bochkarev, A., Bullock, P. A. & Bohm, A. (2007). The crystal structure of the SV40 T-antigen origin binding domain in complex with DNA. *PloS Biol.* **5**, e23.
19. Bochkareva, E., Martynowski, D., Seitova, A. & Bochkarev, A. (2006). Structure of the origin-binding domain of simian virus 40 large T antigen bound to DNA. *EMBO J.* **25**, 5961–5969.
20. Luo, X., Sanford, D. G., Bullock, P. A. & Bachovchin, W. W. (1996). Solution structure of the origin DNA-binding domain of SV40 T-antigen. *Nat. Struct. Biol.* **3**, 1034–1039.
21. Meinke, G., Bullock, P. A. & Bohm, A. (2006). Crystal structure of the simian virus 40 large T-antigen origin-binding domain. *J. Virol.* **80**, 4304–4312.
22. Enemark, E. J., Chen, G., Vaughn, D. E., Stenlund, A. & Joshua-Tor, L. (2000). Crystal structure of the DNA binding domain of the replication initiation protein E1 from papillomavirus. *Mol. Cell*, **6**, 149–158.
23. Oubridge, C., Ito, N., Evans, P., Teo, C. & Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, **372**, 432–438.
24. Conte, M., Grüne, T., Ghuman, J., Kelly, G., Ladas, A., Matthews, S. & Curry, S. (2000). Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold. *EMBO J.* **19**, 3132–3141.
25. Hickman, A. B., Ronning, D. R., Kotin, R. M. & Dyda, F. (2002). Structural unity among viral origin binding proteins: crystal structure of the nuclease domain of adeno-associated virus Rep. *Mol. Cell*, **10**, 327–337.
26. Campos-Olivas, R., Louis, J. M., Clerot, D., Gronenborn, B. & Gronenborn, A. M. (2002). The structure of a replication initiator unites diverse aspects of nucleic acid metabolism. *Proc. Natl Acad. Sci. USA*, **99**, 10310–10315.
27. Simmons, D. T., Loeber, G. & Tegtmeyer, P. (1990). Four major sequence elements of simian virus 40 large T antigen coordinate its specific and nonspecific DNA binding. *J. Virol.* **64**, 1973–1983.
28. Meinke, G., Phelan, P. J., Fradet-Turcotte, A., Bohm, A., Archambault, J. & Bullock, P. A. (2011). Structure-based analysis of the interaction between the simian virus 40 T-antigen origin binding domain and single-stranded DNA. *J. Virol.* **85**, 818–827.
29. Jiang, X., Klimovich, V., Arunkumar, A. I., Hysinger, E. B., Wang, Y., Ott, R. D. *et al.* (2006). Structural mechanism of RPA loading on DNA during activation of a simple pre-replication complex. *EMBO J.* **25**, 5516–5526.
30. Lin, H. J., Upson, R. H. & Simmons, D. T. (1992). Nonspecific DNA binding activity of simian virus 40 large T antigen: evidence for the cooperation of two regions for full activity. *J. Virol.* **66**, 5443–5452.

Please cite this article as: Harrison, C. J. *et al.*, Asymmetric Assembly of Merkel Cell Polyomavirus Large T-Antigen Origin Binding Domains at the Viral Origin, *J. Mol. Biol.* (2011), doi:10.1016/j.jmb.2011.03.051

31. Lorimer, H. E., Wang, E. H. & Prives, C. (1991). The DNA-binding properties of polyomavirus large T antigen are altered by ATP and other nucleotides. *J. Virol.* **65**, 687–699.
32. Enemark, E. J., Stenlund, A. & Joshua-Tor, L. (2002). Crystal structures of two intermediates in the assembly of the papillomavirus replication initiation complex. *EMBO J.* **21**, 1487–1496.
33. Titolo, S., Welchner, E., White, P. W. & Archambault, J. (2003). Characterization of the DNA-binding properties of the origin-binding domain of simian virus 40 large T antigen by fluorescence anisotropy. *J. Virol.* **77**, 5512–5518.
34. Dean, F. B., Bullock, P., Murakami, Y., Wobbe, C. R., Weissbach, L. & Hurwitz, J. (1987). Simian virus 40 (SV40) DNA replication: SV40 large T antigen unwinds DNA containing the SV40 origin of replication. *Proc. Natl Acad. Sci. USA*, **84**, 16–20.
35. Leslie, A. (1992). *Jt. CCP4 + ESF-EAMCB Newsl. Protein Crystallogr.* **26**.
36. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674.
37. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D*, **60**, 2126–2132.
38. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr. Sect. D*, **54**, 905–921.
39. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N. *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D*, **66**, 213–221.
40. Lavery, R., Moakher, M., Maddocks, J. H., Petkeviciute, D. & Zakrzewska, K. (2009). Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **37**, 5917–5929.

Please cite this article as: Harrison, C. J. *et al.*, Asymmetric Assembly of Merkel Cell Polyomavirus Large T-Antigen Origin Binding Domains at the Viral Origin, *J. Mol. Biol.* (2011), doi:10.1016/j.jmb.2011.03.051