Genetic and Structural Analysis of Merkel Cell Polyomavirus Large T Antigen from Diverse Biological Samples

Camila Freze Baez^a Nuria Cirauqui Diaz^b Silvia Maria Baeta Cavalcanti^a
Rafael Brandão Varella^a

^aDepartment of Microbiology and Parasitology, Universidade Federal Fluminense, and ^bDepartment of Pharmaceutical Biotechnology, Universidade Federal do Rio de Janeiro, Niteroi, Brazil

Key Words

Merkel cell polyomavirus · Large T antigen · Single nucleotide polymorphism · Structural biology · Molecular modeling · Molecular dynamics

Abstract

Merkel cell polyomavirus (MCPyV) large T antigen (LT-ag) is frequently found truncated in Merkel cell carcinomas (MCC) and it is considered a major tumor-specific signature. Nonetheless, the biological role of LT-ag nontruncated mutations is largely unknown. In this study, MCPyV LT-ag second exon from 11 non-MCC oral samples and NCBI sequences derived from different anatomical sites were studied from the genetic and structural standpoint. As expected, the LT-ag mutation profile was influenced by the geographical origin of the sample, although nonsynonymous mutations were more frequent in lesional tissues. Our *in silico* study suggests that the mutations found would not significantly affect protein functions, regardless of sample category. This work presents a thorough investigation of the structural and functional properties of LT-ag nontruncated mutations in MCPyV. Our results sustain the geographical influence of the MCPyV genetic profile, but do not discard genetic tissue specificities. Further investigation involving other genetic segments in healthy and lesional tissues are necessary to improve our knowledge on MCPyV pathogenesis.

© 2014 S. Karger AG, Basel

Introduction

In 2008, Feng et al. [1], using digital transcriptome subtraction, discovered a new polyomavirus DNA clonally integrated in about 80% of Merkel cell carcinoma (MCC) cases, and therefore named the virus Merkel cell polyomavirus (MCPyV). MCC is a rare and aggressive neuroectodermal malignancy arising in sun-exposed skin of elderly and immunosuppressed patients [2, 3]. Nevertheless, MCPyV DNA was also found in several non-MCC sites, from MCC patients and healthy individuals, especially in skin and oral samples [4–7], although the role of MCPyV in non-MCC sites is unknown.

Large T antigen (LT-ag) polyomaviruses are multifunctional proteins responsible for gene expression regulation and viral DNA replication, with potential oncogenic activity, as seen for the prototype LT-ag from SV40 and MCPyV [8–10]. Sequencing of the integrated MCPyV genome from several MCC and cell lines revealed various point mutations in the LT-ag N-terminal DnaJ domain that preserves the LXCXE motif, and a stop codon before the original binding domain (OBD). These truncated forms of MCPyV LT-ag are unable to bind to the viral ORI and therefore to initiate bidirectional replication, suggesting these mutations disrupt LT-ag ability to replicate the integrated viral genome [10–12].

Although the consequences of LT-ag truncation for MCPyV oncogenicity have been broadly documented, the effects of other nonsynonymous mutations for MCPyV biology remain obscure and scarcely explored. Therefore, the objective of this study is to analyze the genetic profile of MCPyV LT-ag complete second exon sequences, and to investigate the possible influence of nonsynonymous mutations in protein function through in silico analysis.

Material and Methods

DNA Sequencing and Analysis

The whole second exon of LT-ag (861–3,080 bp) of eleven oral MCPyV-positive samples (8 of saliva and 3 of oral tissue biopsies) from renal transplant recipients and nontransplanted individuals were sequenced using the BigDye™ terminator chemistry kit (Applied Biosystems, Foster City, Calif., USA) according to previously established conditions [5]. They were analyzed along with the whole second exon of all nontruncated sequences available at the NCBI databank to date, with the aid of Molecular Evolutionary and Genetics Analysis (MEGA 5.2) software. Cloned sequences were not used due to potential bias in the number and distribution of mutations.

Overall, thirty-three LT-ag complete second exons were retrieved, comprised of 16 from skin (7 from healthy skin of non-MCC patients; 3 from healthy skin of MCC patients; 3 from non-MCC skin lesions and 3 from MCC lesions), 3 from urine samples, 4 from lung cancer tissue, 3 from nasal swabs, 3 from intestines and 4 from peripheral blood mononuclear cells (PBMC; table 1). The sequence MCC350 (JN038583) was used as a standard truncated LT-ag, and the HF consensus genome (JF813003) as a nontruncated prototype sequence. This study was approved by the Institutional Committee on Research Ethics (protocol 1138/2005).

Structural Analysis of LT-ag

The MCPyV LT-ag sequences were extracted from the UniProtKB database (UniProtKB accession No. E2IPT4). However, the numbering scheme within this article also considers the first exon (78 amino acids). Homologs of the protein with a solved 3D structure were researched using the Basic Local Alignment Search Tool (BLAST) and the HHPred server (<http://toolkit.tuebingen.mpg.de/hhpred>). For the helicase domain, a homology model was built with the program Modeller (<http://salilab.org/modeller/>) using the structure of the SV40 helicase domain as a template (PDB 1SVL). For the retinoblastoma protein binding site (pRb) domain, a homologue with a solved 3D structure was not found, and therefore a model was built using the QUARK structure prediction automated server (<http://zhanglab.cmb.med.umich.edu/QUARK>). For the molecular dynamics study, the OBD Merkel structure with PDB ID 3QFQ was used as a wild-type structure. For the molecular dynamics study, all structures of mutant proteins were built with the Modeller program.

Energy minimization and molecular dynamics simulations were performed with the GROMACS package version 4.5.4 [13], using the amber99sb force field [14]. MD simulations were carried out on 20 cores of 2.4 GHz 32 GB of our Linux cluster. The TIP3P

water model [15] was used to solvate the protein in an octahedral water box. All simulations were performed in periodic boundary conditions. Bond stretches involving hydrogen atoms were constrained using an LINCS algorithm [16]. The electrostatic interactions were treated by a combination of the particle mesh Ewald method and a switch function. Simulations of solvent molecules and counterions were performed at 310 K and 1 atm for 100 ps, with the protein atoms restrained with harmonic potentials. The equations of motion were integrated using the Verlet Leapfrog algorithm [17]. Systems were minimized for 2,000 steps of steepest descent and then heated from 50 to 310 K using 6 blocks of calculation in which the temperature was gradually increased. Each block lasted 20 ps, totaling 120 ps of the heating process for each system. The equilibrated systems were simulated using a time step of 0.002 ps. All systems were simulated for 20 ns.

Results

In regard to the oral MCPyV samples obtained in this study, the profile of synonymous mutations found in LT-ag were similar between transplanted versus nontransplanted individuals and saliva versus oral biopsy (fig. 1). Only two nonsynonymous mutations, S133T (1,024 nt) and R204T (1,237 nt), were found in saliva samples from transplanted individuals (TxS 1 and TxS 2). In contrast, the mutation pattern between MCPyV LT-ag Brazilian sequences and those retrieved from the NCBI databank from diverse biological samples revealed few similarities (fig. 1). Also, Brazilian and Asian sequences apparently formed homogeneous polymorphic profiles each, which did not occur in other sequences.

With the exception of the pRb region, nonsynonymous mutations were found in all LT-ag domains and inter-domain areas (table 1), and its frequency was significantly higher in samples from lesional tissues (MCC, psoriasis, Kaposi sarcoma, lung cancer, squamous cell carcinoma) than in healthy tissue ($p = 0.0018$; data not shown). LT-ag sequences of healthy skin from MCC patients are of particular interest given the high number of nonsynonymous mutations found, equivalent to that of the lesional samples.

In order to gain some insight into the possible roles of these mutations in the protein function, a molecular dynamics analysis of the predicted protein structure was performed. The pRb domain is predicted to be intrinsically disordered. Not surprisingly, a close homologue with a solved 3D structure was not found, either by a sequence search or by protein threading methods, and therefore an ab initio model was built using the QUARK structure prediction automated server. According to the model, all of the mutated residues are located in loops at

Table 1. MCPyV LT-ag second exon amino acid substitutions according to sample type (n = 44)

Sample origin	n	Nonsynonymous mutations, n	Amino acid substitutions (nucleotide position)						References
			← pRb (861–1,252 nt)	pRb (1,253–1,281 nt)	pRb ↔ OBD (1,282–1,565 nt)	OBD (1,566–1,838 nt)	OBD ↔ helicase (1,839–1,946 nt)	helicase (1,947–3,017 nt)	
<i>Skin</i>									
Healthy non-MCC ¹	7	3	–	–	P243S (1,354 nt)	K351R (1,679 nt) A337T (1,636 nt)	–	F448S (1,970 nt); M531T (2,219 nt)	Foulongne et al. [27], 2009 Foulongne et al. [28], 2012
Healthy MCC ²	3	8	S192R (1,198 nt)	–	S226P (1,303 nt) Y261C (1,409 nt)	–	N423D (1,894 nt)	S597R (2,419 nt); V681A (2,669 nt) C749R (2,872 nt)	Schowalter et al. [23], 2010 Foulongne et al. [27], 2009
MCC tumor ³	3	6	H112Y (961 nt); T114S (967 nt)	–	–	L319P (1,583 nt)	–	F714S (2,768 nt); M728T (2,810 nt); L736F (2,833 nt)	Hashida et al. [25], 2014
Kaposi sarcoma	1	2	–	–	–	–	–	–	
Psoriasis	1	3	–	–	Y261C (1,409 nt)	–	L426P (1,904 nt); N423D (1,894 nt)	M531T (2,219 nt); V681A (2,669 nt)	
SCC	1	4	A153T (1,085 nt)	–	–	–	–	Y586H (2,383 nt); I798V (3,019 nt)	
Lung cancer	4	9	H112Y (961 nt); A153T (1,085 nt); H168R (1,130 nt); R200G (1,225 nt)	–	Y261C (1,409 nt)	N386D (1,783 nt)	Y429H (1,912 nt); F431S (1,919 nt)	T527S (2,419 nt)	Hashida et al. [25], 2014
Urine swab	3	1	G114R (964 nt)	–	–	–	–	–	Laude et al. [5], 2010
PBMC	4	0	–	–	–	–	–	–	Laude et al. [5], 2010; Feng et al. [1], 2008
Nasal swab	3	1	–	–	–	–	–	T753I (2,285 nt)	Laude et al. [5], 2010
Intestine	3	3	P121S (988 nt)	–	F296S (1,514 nt)	I352V (1,681 nt)	–	–	Feng et al., unpublished data
Saliva ⁴	8	2	S133T (1,025 nt); R204T (1,238 nt)	–	–	–	–	–	–
Oral biopsy ⁴	3	0	–	–	–	–	–	–	–

¹ Healthy skin from non-MCC patients. ² Healthy skin area from MCC patients. ³ Nontruncated MCPyV LT-ag. ⁴ Samples obtained from the current study.

Fig. 1. Genomic position of synonymous and nonsynonymous mutations found in the second exon of MCPyV LT-ag sequences, according to the sample category and geographical area. Open circles represent synonymous mutations, filled circles are nonsyn-

onymous mutations, black lines indicate sequences obtained from this study and gray lines show NCBI sequences. Tx = Transplanted individual; nTX = nontransplanted individual; S = saliva; OB = oral biopsy.

the protein surface, distant from the LXCXE motif, and also far-off the only predicted alpha helix (Aa 80–90). Four out of ten mutations in this region involve either a proline or a glycine residue (table 1), and would therefore modify the flexibility of the pRb domain. The same high percentage of mutations involving glycine or proline residues was found on the loop regions between domains (table 1).

The structure of the OBD domain is already known (PDB 3QFQ). Five mutations were found in this domain (table 1). Residue A337 is located inside the protein in a hydrophobic core. The A337T mutant stability was studied by molecular dynamics simulations. Its structure and dynamics were not different from that of the wild-type domain. Concerning K351R, due to the similarities between lysine and arginine residues, it is plausible that there is no effect of this mutation on protein function.

However, this family of proteins can undergo acetylation at lysine residues, a mechanism important for T-ag stability and DNA replication [18, 19]. The mutation N386D found in lung cancer tissue could be expected to impact on protein function. Located at the monomer-monomer interface (by comparison with SV40), the mutation of this residue could affect hexamerization [20].

At the helicase domain, 13 nonsynonymous mutations were found. Due to either the residue location in the protein structure or the similar properties of wild-type and mutant residues, some mutations can be hypothesized to not influence protein function: T527S, M531T, V681A, M728T, T753I and I798V. Four mutations, namely F448S, S597R, F714S and L736F, as with the wild-type domain, were submitted to 20 ns of molecular dynamics simulations and no significant differences in protein stability or structure were observed. A swinging of the N-terminal

Fig. 2. p53 interaction site at the Merkel helicase domain and close view of C749 and the residues within interacting distance (4 Å). The p53 protein is shown in dark gray and the Merkel helicase domain in light grey.

region was observed in all protein models, produced by a conformational change at residues 525–527. Finally, some samples of healthy skin in patients with MCC lesions presented a cysteine-arginine substitution in residue 749. This mutation is located at the p53 binding interface (by comparison to SV40) [21]. As can be seen in figure 2, an arginine residue at position 749 could make a hydrogen bond with Arg438 in p53, increasing the interaction energy between these proteins.

Discussion

In this study, the analysis of the second exon of the MCPyV LT-ag from non-MCC oral samples revealed no stop codon or any high impact mutations. This result was expected as truncated LT-ag has been described as a tumor-specific signature found in MCC [1, 5, 10, 12] and in other malignancies [22]. Nevertheless, the position of the synonymous mutations detected in all oral samples suggests a pattern and possibly a phylogenetic proximity, regardless the immunosuppressive status and the sample category. Similar results have been found by other authors when studying different biological samples, where genetic polymorphisms were described in MCPyV DNA from urine, nasal swab, PBMC [5] and malignant blood cells [22] from different subjects or the same subject [23], suggesting a strain-specific signature.

Looking at figure 1, it appears that the MCPyV LT-ag polymorphic profile has a geographical influence, as observed by others [24, 25]. However, the role of intra-host selective forces in shaping MCPyV DNA remains to be determined, as well as the possibility that certain strains might influence the course of the infection [25]. It should be considered that our apparent mutation profile may also have been influenced by methodological differences between studies, especially the sampling process, number of samples and the characteristics of the participants. However, our scarce knowledge regarding MCPyV phylogeny and pathogenesis limits further conclusions. Yet, it seems that the MCPyV LT-ag polymorphic profile has a geographical conservation.

We could hypothesize that the higher number of non-synonymous mutations in MCPyV LT-ag from lesional tissues is the consequence of viral adaptation and perhaps escape from host immunity. However, since all Asian (Japanese) samples were collected from lesional tissues, it is difficult to attribute the frequency of mutations to any pathological event.

The *in silico* study suggests that the mutations found in the second exon of LT-ag would not significantly affect protein functions, regardless the clinical status of the host and the sample type. Also, no apparent pattern of mutations was related to any specific lesion. Stable DNA viruses, like polyomaviruses, are not subjected to strong adaptive selection in the way that high impact mutations

are frequently eliminated from the population in favor to neutral/low-impact mutations [24, 26]. Therefore, it is tempting to speculate that MCPyV-associated pathologies are 'all-or-nothing' processes, relying on rare LT-ag truncation events to take place, although the role of other viral genes such as VP1, ST-ag and the NCCR region in MCPyV pathogenesis deserves investigation.

To the best of our knowledge, this work presents the most detailed investigation of the structural and functional properties of LT-ag nontruncated mutations in MCPyV se-

quences from diverse biological samples. As the information on MCPyV LT-ag structure (mainly the pRb domain), interaction partners and MCPyV pathogenesis increases, the data presented here will gain new light, and the function of nonsynonymous mutations will be better understood.

Acknowledgments

The authors acknowledge the financial support of FAPERJ.

References

- Feng H, Shuda M, Chang Y, Moore PS: Clonal Integration of a polyomavirus in human Merkel cell carcinoma. *Science* 2008;319:1096–1100.
- Penn I, First MR: Merkel's cell carcinoma in organ recipients: report of 41 cases. *Transplantation* 1999;68:1717–1721.
- Harwood CA, McGregor JM, Swale VJ, Proby CM, Leigh IM, Newton R: High frequency and diversity of cutaneous appendageal tumors in organ transplant recipients. *J Am Acad Dermatol* 2003;48:401–408.
- Foulongne V, Kluger N, Dereure O, Mercier G, Molès JP, Guillot B, Segondy M: Merkel cell polyomavirus in cutaneous swabs. *Emerg Infect Dis* 2010;16:685–687.
- Laude HC, Jonchère B, Maubec E, Carlotti A, Marinho E, Couturaud B, Peter M, Sastre-Garau X, Avril MF, Dupin N, Rozenberg F: Distinct Merkel cell polyomavirus molecular features in tumour and non tumour specimens from patients with Merkel cell carcinoma. *PLoS Pathog* 2010;6:e1001076.
- Loyo M, Guerrero-Preston R, Brait M, Hoque MO, Chuang A, Kim MS, Sharma R, Liégeois NJ, Koch WM, Califano JA, Westra WH, Sidransky D: Quantitative detection of Merkel cell virus in human tissues and possible mode of transmission. *Int J Cancer* 2010;126:2991–2996.
- Mertz KD, Pfaltz M, Junt T, Schmid M, Figueras MTF, Pfaltz K, Barghorn A, Kempf W: Merkel cell polyomavirus is present in common warts and carcinoma in situ of the skin. *Hum Pathol* 2010;41:1369–1379.
- Cheng J, Rozenblatt-Rosen O, Paulson KG, Nghiem P, DeCaprio JA: Merkel cell polyomavirus large T antigen has growth-promoting and inhibitory activities. *J Virol* 2013;87:6118–6126.
- Topalis D, Andrei G, Snoeck R: The large tumor antigen: a 'Swiss Army knife' protein possessing the functions required for the polyomavirus life cycle. *Antiviral Res* 2013;97:122–136.
- Shuda M, Feng H, Kwun HJ, Rosen ST, Gjoerup O, Moore PS, Chang Y: T antigen mutations are a human tumor-specific signature for Merkel cell polyomavirus. *Proc Natl Acad Sci USA* 2008;105:16272–16277.
- Duncavage EJ, Magrini V, Becker N, Armstrong JR, Demeter RT, Wylie T, Abel HJ, Pfeifer JD: Hybrid capture and next-generation sequencing identify viral integration sites from formalin-fixed, paraffin-embedded tissue. *J Mol Diagn* 2011;13:325–333.
- Schmitt M, Wieland U, Kreuter A, Pawlita M: C-terminal deletions of Merkel cell polyomavirus large T-antigen, a highly specific surrogate marker for virally induced malignancy. *Int J Cancer* 2012;131:2863–2868.
- Hess B, Kutzner C, van der Spoel D, Lindahl E: GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 2008;4:435–447.
- Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C: Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* 2006;65:712–725.
- Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML: Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983;79:926–935.
- Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG: A smooth particle mesh Ewald method. *J Chem Phys* 1995;103:8577–8593.
- Snyman JA: A new and dynamic method for unconstrained minimization. *Appl Math Modelling* 1982;6:449–462.
- Xie AY, Bermudez VP, Folk WR: Stimulation of DNA replication from the polyomavirus origin by PCAF and GCN5 acetyltransferases: acetylation of large T antigen. *Mol Cell Biol* 2002;22:7907–7918.
- Shimazu T, Komatsu Y, Nakayama KI, Fukazawa H, Horinouchi S, Yoshida M: Regulation of SV40 large T-antigen stability by reversible acetylation. *Oncogene* 2006;25:7391–7400.
- Meinke G, Phelan P, Moine S, Bochkareva E, Bochkarev A, Bullock PA, Bohm A: The crystal structure of the SV40 T-antigen origin binding domain in complex with DNA. *PLoS Biol* 2007;5:e23.
- Liljestrom W, Klein MG, Zhang R, Joachimiak A, Chen XS: Crystal structure of SV40 large T-antigen bound to p53: interplay between a viral oncoprotein and a cellular tumor suppressor. *Genes Dev* 2006;20:2373–2382.
- Pantulu ND, Pallasch CP, Kurz AK, Kassem A, Frenzel L, Sodenkamp S, Kvasnicka HM, Wendtner CM, Zur Hausen A: Detection of a novel truncating Merkel cell polyomavirus large T antigen deletion in chronic lymphocytic leukemia cells. *Blood* 2010;116:5280–5284.
- Schowalter RM, Pastrana DV, Pumphrey KA, Moyer AJ, Buck CB: Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin. *Cell Host Microbe* 2010;7:509–515.
- Martel-Jantin C, Filippone C, Cassar O, Peter M, Tomasic G, Vielh P, Brière J, Petrella T, Aubriot-Lorton MH, Mortier L, Jouvion G, Sastre-Garau X, Robert C, Gessain A: Genetic variability and integration of Merkel cell polyomavirus in Merkel cell carcinoma. *Virology* 2012;426:134–142.
- Hashida Y, Imajoh M, Kamioka M, Taniguchi A, Kuroda N, Hayashi K, Nakajima H, Sano S, Daibata M: Phylogenetic analysis of Merkel cell polyomavirus based on full-length LT and VP1 gene sequences derived from neoplastic tumours in Japanese patients. *J Gen Virol* 2014;95:135–141.
- Ciccozzi M, Babakir-Mina M, Lo Presti A, Farchi F, Zehender G, Ebranati E, Perno CF, Ciotti M: Genetic variability of the small T antigen of the novel KI, WU and MC polyomaviruses. *Arch Virol* 2010;155:1433–1438.
- Unpublished data (GenBank accession numbers: FM955590, FM955592, FM955589).
- Foulongne V, Sauvage V, Hebert C, Dereure O, Cheval J, Gouilh MA, Pariente K, Segondy M, Burgière A, Maniguerra JC, Caro V, Eliot M: Human skin microbiota: high diversity of DNA viruses identified on the human skin throughput sequencing. *PoS One* 2012;7:e38499.