HUMAN CYTOMEGALOVIRUS UL138 PROTEIN: PREDICTION OF PHYSICOCHEMICAL PROPERTIES, PROTEIN STRUCTURE, AND AMINO ACID PHOSPHORYLATION AND MUTATION SITES

Chunyan Qian[1,2], Yuefeng Zhang[1], Hong Yuan[1] and Jiang Zhong[2]

[1]Yuhang Branch of the Second Affiliated Hospital of Zhejiang University, Hangzhou;
[2]School of Life Sciences, Fudan University, Shanghai, PR China

Abstract. Human cytomegalovirus (HCMV) infection is usually asymptomatic in healthy individuals but the virus may be fetal in immunocompromised individuals because of reactivation of latent infection. HCMV *UL138*, a latency-associated determinant, encodes a 169-amino acid protein of unknown function. This study employed online bioinformatics tools to predict UL138 protein physicochemical properties, secondary and tertiary structures, and phosphorylation and mutation sites. UL138 protein was predicted to have a theoretical isoelectric point (pI) of 6.51, an instability index of 53.62, a half-life of 30 hours in mammalian eukaryotic cell, hydropathicity of -0.101 (a hydrophilic protein), 37.3% α-helical structure, 16.6% extended β-strands, 3.0% β-turns and 43.1% random coils. The protein lacked a canonical N-terminal signal sequence or cleavage site, but contained an internal transmembrane domain indicating a group II integral membrane protein with 11 predicted serine phosphorylation sites all located in the ectodomain, which also contains a domain with a tertiary structure highly similar to toll-like receptor 4. Comparison of 38 known UL138 protein sequences deposited in a database with that of reference UL138 ACL51196.1 revealed the presence of 24 mutant amino acid positions, with occurrences ranging from 3 to 16%, all located in the protein ectodomain. In conclusion, the information obtained should assist in furthering knowledge of the role of UL138 protein in HCMV latency state.

Keywords: amino acid mutation, bioinformatics, human cytomegalovirus UL138, physicochemical property, protein structure

Correspondences: Jiang Zhong, School of Life Sciences Fudan University, No. 2005 Songhu Road, Yangpu District, Shanghai 200433, PR China
Tel/Fax: +86 21 51630507; E-mail: jzhong@fudan.edu.cn
and
Hong Yuan, Yuhang Branch of the Second Affiliated Hospital of Zhejiang University, No. 369 Yingbin Road, Yuhang District, Hangzhou City, Zhejiang 311100, PR China
Tel/Fax: +86 571 89369003; E-mail: q448508713@163.com

INTRODUCTION

Human cytomegalovirus (HCMV), a member of *Herpesviridae* family in subfamily *β-herpesvirinae* (also known as human β-herpesvirus 5), is a common human pathogen worldwide (Landolfo *et al*, 2003), infecting approximately 60% of adults in developed world and more than 90% in developing countries (Manicklal *et al*, 2013). Upon infection, HCMV establishes lifelong persistence within the human host, showing a "latency-activation" biological feature (Söderberg-Nauclér and Nelson, 1999) which HCMV infection is usually asymptomatic in healthy individuals (Landolfo *et al*, 2003) but can be fatal in immunocompromised individuals owing to reactivation of latent infection, which results in severe clinical outcomes, including loss of graft function, HCMV pneumonia, HCMV hepatitis and even death (Rawlinson and Scott, 2003; Dioverti and Razonable, 2016; Navarro, 2016). More importantly, until recently there were no effective drugs for treatment of latent HCMV infections but the availability of such drugs as cidofovir (viral DNA polymerase inhibitor), letermovir (viral terminase complex inhibitor) and maribavir (viral DNA polymerase inhibitor) has brought hope that this insidious infection could be successfully treated and controlled (Kotton, 2019).

HCMV contains a double-stranded DNA genome, characterized by presence of two unique regions (unique long UL and unique short US), both flanked by a pair of repeats, UL by sequences ab (terminal) and a'c' (internal) and US by the same a'c' and (terminal) ca (prime indicating inverted orientation) (Sijmons *et al*, 2014). In latent infection, the viral genome is maintained in primary sites (mononuclear cells of myeloid lineage) without production of infectious virions (Taylor-Wiedeman *et al*, 1991; Mendelson *et al*, 1996; Reeves *et al*, 2005). Transcript and protein produced from an open reading frame, UL138, located near UL internal repeat, were detected in individuals with latent infection (Goodrum *et al*, 2007). *UL138* sequence is highly conserved in clinical strains (Qi *et al*, 2009) and has since been identified as a latency-associated determinant (Tey *et al*, 2010). However, the function of UL138 protein remains elusive.

Here, bioinformatics tools were employed to predict UL138 physicochemical properties, protein structure and amino acid mutation sites. The information should assist in furthering our understanding of HCMV UL138 protein function and mechanisms of action.

MATERIALS AND METHODS

Data source

HCMV UL138 (510 bp) encodes a 169-aa protein (MDDLPLNVGLPIIGVMLVLIVAILCYLAYHWHDTFKLVRMFLSYRWLIRCCELYGEYERRFADLSSLGLGAVRRESDRRYRFSERPDEILVRWEEVSSQCSYASSRITDRRAGSSSSSSVHVANQRNSVPPPDMAVTAPLTDVDLLKPVTGSATQFTTVAMVHYHQEYT) (GenBank accession no. ACL51196.1). A total of 38 HCMV UL138 proteins in GenBank database (AAA85877.1, AAO62776.1, AAO62779.1,

AAP12523.1, ABA26298.1, ABA26310.1, ACT81966.1, ACZ81674.1, ADB84733.1, ADB84781.1, ADB84829.1, AFR54948.1, AHB19432.1, AHB19767.1, AHJ82514.1, AHJ82683.1, AHJ82849.1, AHJ83859.1, AKI09856.1, AKI10194.1, AKI10528.1, AKI11530.1, AKI13696.1, AKI14533.1, AKI17873.1, AKI19880.1, AKI20719.1, AKI25106.1, AMJ53490.1, AMJ53989.1, AMJ54491.1, APA45245.1, AQN69908.1, AQN71931.1, AQN72267.1, BA26285.1, KI17036.1, and YP_081578.1) sharing > 90% sequence identity to HCMV UL138 ACL51196.1 were used in analyses.

Prediction of UL138 protein physicochemical properties

Physicochemical properties (amino acid composition, atomic composition, estimated half-life, instability index, grand average of hydropathicity, molecular weight, and theoretical isoelectric point) of UL138 protein were predicted using an online ProtParam tool (https://web.expasy.org/protparam/).

Prediction of UL138 protein secondary and tertiary structures

Secondary structure of UL138 protein was predicted using online service tools PredictProtein (https://www.predictprotein.org/), TMHMM Server version 2.0 (http://www.cbs.dtu.dk/services/TMHMM-2.0/), SMART (http://smart.embl-heidelberg.de/), SOPMA (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) and JPred 4 (http://www.compbio.dundee.ac.uk/jpred/); and UL138 protein tertiary structure using an online tool SWISS-MODEL (https://swissmodel.expasy.org). Protein structure alignment and function prediction were performed using a threading method of a Phyre[2] web portal (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index). Consistency between target and template sequences of >30% is defined as similar and approaching 100% as high goodness of fit.

Prediction of UL138 protein signal sequence, phosphorylation sites and subcellular localization

Signal sequence and phosphorylation sites were predicted using online tool SignalP-5.0 Server (http://www.cbs.dtu.dk/services/SignalP/) and KinasePhos 2.0 (http://KinasePhos2.mbc.nctu.edu.tw/), respectively.

Identification of UL138 protein dominant amino acids and mutational hot spots

An amino acid of UL138 proteins with the highest occurrence at a given location is defined as dominant amino acid, while less frequent amino acids defined as mutant amino acids. UL138 protein consisting of dominant amino acids is considered as the dominant sequence.

RESULTS

Physicochemical properties of UL138 protein

UL138 protein ($C_{858}H_{1341}N_{239}O_{251}S_9$; 19.309 kDa; 169 amino acids) contained 18 basic amino acids (K and R) and 19 acidic amino acids (D and E), with a theoretical isoelectric point (pI) of 6.51, an estimated half-life of 30 hours in mammalian eukaryotic cell *in vitro*, an instability index of 53.62 (indicative of an unstable protein), and an average hydropathicity

of -0.101 (indicative of a hydrophilic protein).

Signal sequence and phosphorylation sites

UL138 protein contained no canonical signal sequence or cleavage site and 11 potential serine sites but no threonine or tyrosine phosphorylation sites (Table 1).

Secondary and tertiary structures of UL138 protein

UL138 protein was predicted to contain 37.3% α-helix, 16.6% extended β-strand, 3.0% β-turn and 43.1% random coil. Despite the putative absence of a cleavable signal sequence, a transmembrane region was predicted to be located at aa 10-32 (using TMHMM Server version 2.0 and SMART) and at aa 15-45 (using Phyre[2]), but no other putative transmembrane sequences, indicating that UL138 protein contains a single transmembrane domain. SWISS-MODEL showed UL138 protein has high similarities with a transmembrane region (aa 624-657) and a tertiary structure domain (aa 624-670) of toll-like receptor 4 (TLR4).

Dominant amino acid sequence and mutation hot spots of UL138 protein

A comparison of 38 UL138 protein sequences deposited in GenBank database with UL138 protein UL138 ACL51196.1 resulted in a dominant sequence containing 24 mutation sites (Table 2). There were three hot spots with a frequency ≥10% (aa 112, 26%, aa 124, 16% and aa 131, 10%), as well as four serine mutants, namely, S86, S113, S124 and S131, and aa 104 is predicted as a phosphorylation site.

DISCUSSION

Although HCMV *UL138* is recognized as one of several viral genes associated with HCMV latency state (Tey *et al*, 2010), its function remains undescribed. Employing several bioinformatics tools, the study predicts UL138 to be hydrophilic, containing approximately one third α-helical content and with pI of 6.51 [properties verified in laboratory assays (unpublished)]. UL138 is polymorphic possessing 24/169 mutant amino acids among 38 known sequences and has 11 potential serine phosphorylation sites. It is worth noting UL138 S104A mutant lacks a possible phosphorylation site, but this could be compensated by a reciprocal P86S or G113S mutation.

The dominant 169-aa sequence contains a single transmembrane sequence with no canonical N-terminal cleavable signal sequence indicating a type II integral membrane protein with the putative phosphorylation sites located in the ectodomain. UL138 protein has been postulated to be first located in the ER membrane, then translocated to the Golgi apparatus and ultimately to the host plasma membrane where it anchors progeny HCMV prior to release via a signal peptide (Petrucelli *et al*, 2009; Gelbmann and Kalejta, 2019). Given the high similarity of UL138 ectodomain tertiary structure with that of TLR4 (Kawai and Akira, 2007), it is hypothesized that cell surface UL138 acts similarly in TLR signaling, and belongs to the innate immune system.

In conclusion, application of

Table 1

Predicted phosphorylation serine positions in human cytomegalovirus UL138 protein

SVM score (protein kinase)	Amino acid position										
	65	66	97	98	101	104	105	116	117	128	152
Highest	0.947 (ATM)	0.967 (ATM)	0.938 (ATM)	0.915 (ATM)	0.939 (ATM)	0.939 (GSK-3)	0.974 (GSK-3)	0.920 (ATM)	0.937 (ATM)	0.958 (Aurora)	0.975 (ATM)
Second highest	-	0.902 (CK1)	-	-	-	0.951 (ATM)	0.967 (ATM)	-	-	-	0.909 (Aurora)

ATM: ataxia telangiectasia mutated kinase; Aurora: aurora-related kinase; CK1: casein kinase 1; GSK-3: glycogen synthase kinase-3; SVM: support vector machine (http://www.cbs.dtu.dk/services/TargetP/)

Table 2

Mutation sites in human cytomegalovirus UL138 protein

Amino acid substitution	Amino acid position																							
	2	5	12	18	24	36	38	64	72	85	86	103	104	107	112	113	122	123	124	131	138	153	161	162
Mutation (dominant/ mutant)	Asp/ Asn	Pro/ Leu	Ile/ Val	Val/ Met	Leu/ Phe	Lys/ Arg	Val/ Met	Leu/ Met	Val/ Ile	Arg/ Gln	Pro/ Ser	Ala/ Val	Ser/ Ala	Ile/ Val	Ala/ Met or Thr or Val	Gly/ Ser	Val/ Ile	Ala/ Thr	Asn/ Ser	Pro/ Ala or Ser or Thr	Ala/ Thr	Ala/ Thr or Val	Met/ Thr	Val/ Ile
Number of occurrences (%) (38 sequences)	3(8)	1(3)	1(3)	1(3)	1(3)	1(3)	1(3)	1(3)	2(5)	1(3)	1(3)	1(3)	3(8)	1(3)	10(26)	1(3)	1(3)	1(3)	6(16)	4(10)	3(8)	2(5)	1(3)	1(3)

Ala: Alanine; Arg: Arginine; Asn: Asparagine; Asp: Aspartic acid; Gln: Glutamine; Gly: Glycine; Ile: Isoleucine; Leu: Leucine; Lys: Lysine; Met: Methionine; Phe: Phenylalanine; Pro: Proline; Ser: Serine; Thr: Threonine; Val: Valine

several bioinformatics tool results not only in physicochemical properties of latency-associated HCMV UL138 protein but its putative location on infected host cell surface with a polymorphic, phosphorylated ectodomain with a TLR4-like property. Our data should provide insights into future elucidation of the function and mechanism of action of HCMV UL138 protein.

ACKNOWLEDGEMENTS

The study was funded by grants from the National Natural Science Foundation of China (grant no. 81601831), the Medical and Health Scientific Research Project of Hangzhou Department of Science and Technology (grant no. 20150633B66), the Key Scientific Research Project of Yuhang Department of Health (grant no. 2015008), the Social Development Project of Zhejiang Provincial Department of Science and Technology (grant no. LGF19H200002), and the Research and Development Project for Novel Technology and Products in Zhejiang Provincial Health Commission (grant no. 2021PY018).

CONFLICTS OF INTEREST DISCLOSURE

The authors declare no conflicts of interest.

REFERENCES

Dioverti MV, Razonable RR. Cytomegalovirus, 2016 [cited 2020 Jan 15]. Available from: URL: https://www.asmscience.org/content/journal/microbiolspec/10.1128/microbiolspec.DMIH2-0022-2015

Gelbmann CB, Kalejta RF. The membrane-spanning peptide and acidic cluster dileucine sorting motif of UL138 are required to downregulate MRP1 drug transporter function in human cytomegalovirus-infected cells. *J Virol* 2019; 93: e00430-19.

Goodrum F, Reeves M, Sinclair J, High K, Shenk T. Human cytomegalovirus sequences expressed in latently infected individuals promote a latent infection *in vitro. Blood* 2007; 110: 937-45.

Kawai T, Akira S. TLR signaling. *Semin Immunol* 2007; 19: 24-32.

Kotton CN. Updates on antiviral drugs for cytomegalovirus prevention and treatment. *Curr Opin Organ Transplant* 2019; 24: 469-75.

Landolfo S, Gariglio M, Gribaudo G, Lembo D. The human cytomegalovirus. *Pharmacol Ther* 2003; 98: 269-97.

Manicklal S, Emery VC, Lazzarotto T, Boppana SB, Gupta RK. The "silent" global burden of congenital cytomegalovirus. *Clin Microbiol Rev* 2013; 26: 86-102.

Mendelson M, Monard S, Sissons P, Sinclair J. Detection of endogenous human cytomegalovirus in CD34+ bone marrow progenitors. *J Gen Virol.* 1996; 77: 3099-102.

Navarro D. Expanding role of cytomegalovirus as a human pathogen. *J Med Virol* 2016; 88: 1103-12.

Petrucelli A, Rak M, Grainger L, Goodrum F. Characterization of a novel Golgi apparatus-localized latency determinant encoded by human cytomegalovirus. *J Virol* 2009; 83: 5615-29.

Qi Y, He R, Ma YP, Sun ZR, Ji YH, Ruan Q. Human cytomegalovirus UL138 open reading frame is highly conserved in clinical strains. *Chin Med Sci J* 2009; 24: 107-11.

Rawlinson W, Scott G. Cytomegalovirus. A common virus causing serious disease. *Aust Fam Physician* 2003; 32: 789-93.

Reeves MB, MacAry PA, Lehner PJ, Sissons JG, Sinclair JH. Latency, chromatin remodeling, and reactivation of human cytomegalovirus in the dendritic cells of healthy carriers. *Proc Natl Acad Sci USA* 2005; 102: 4140-5.

Sijmons S, Van Ranst M, Maes P. Genomic and functional characteristics of human cytomegalovirus revealed by next-generation sequencing. *Viruses* 2014; 6: 1049-72.

Söderberg-Nauclér C, Nelson JY. Human cytomegalovirus latency and reactivation - a delicate balance between the virus and its host's immune system. *Intervirology* 1999; 42: 314-21.

Taylor-Wiedeman J, Sissons JG, Borysiewicz LK, Sinclair JH. Monocytes are a major site of persistence of human cytomegalovirus in peripheral blood mononuclear cells. *J Gen Virol* 1991; 72: 2059-64.

Tey SK, Goodrum F, Khanna R. CD8+ T-cell recognition of human cytomegalovirus latency-associated determinant pUL138. *J Gen Virol* 2010; 91: 2040-8.