Functional characterization of novel mutations in UL54 of ganciclovir resistant HCMV strain using structural analysis

Jambulingam Malathi[1], Vetrivel Umashankar[2], Ravichandran Sathyabaarathi[2], Sivashanmugan Muthukumaran[2], Murali Ishwarya[1], Hajib Narahari Madhavan[1]*

[1]Department of Microbiology, Vision Research Foundation, Sankara Nethralaya, Old no, 18, College Road, Chennai - 600 006, Tamilnadu, India; [2]Centre for Bioinformatics, Vision Research Foundation, Sankara Nethralaya, Old no, 18, College Road, Chennai - 600 006, Tamilnadu, India; Hajib Narahari Madhavan - Email: drhnm@snmail.org; Phone: 91-44-282716161; Fax: 91 -44-28254180; *Corresponding author

Abstract:
This study reports the probable impact of the coupled mutations observed in our clinical isolate of HCMV UL54 polymerase, through structural bioinformatics approaches. The reported variant was found to be resistant to Ganciclovir (GCV) as per the clinical records. The presence of Glutamine deletion at 639 (Glu639) and a mis sense mutation of Serine 655 Leucine (Ser655Leu) in UL54 were identified by DNA sequencing and were predicted to lie in the DNA polymerase type-II domain. Docking simulation studies of the phosphorylated forms of Ganciclovir (GCV), Cidofovir (CDV) and Foscarnet (PFA) with the reported mutants showed significant variation in terms of binding affinity and inhibitory constant (Ki) in comparison to wild type UL54. The findings of this study revealed that the observed coupled mutation could potentially induce allosteric effects in the binding pockets of UL54 and thereby alter the drug binding affinity. In specific, it was observed that this coupled mutation could confer changes in the binding affinity of GCV and PFA by altering the binding energies and inhibitory constants to -0.88Kcal/mol and 226.71mM, -5.81Kcal/mol and 54.83μM, respectively, in comparison to Wild Type. On the other hand, CDV showed increased susceptibility for the reported mutant with a binding energy of -6.16Kcal/mol and inhibitory constant of 30.47μM.

Keywords: HCMV, UL54, GCV, Modeling, Docking, Drug resistance

Background:
Increased rate of morbidity and mortality among the immunocompromised patients are mainly due to Human Cytomegalovirus (HCMV; species Human herpesvirus 5) infection and specifically causes sight threatening problems as per the frequent reports [1]. DNA polymerase of HCMV (strain AD 169), a 1242 amino acids long protein coded by 3,729-bp gene UL54, is mainly targeted to control the severity of HCMV [2] by the phosphorylated forms of Ganciclovir (GCV), Cidofovir (CDV) and Foscarnet (PFA) [3]. Many Clinical studies have documented increased frequency of HCMV resistance towards GCV, CDV and PFA in relation to UL54 mutations [4]. The Crystal structure of UL54 is solved only for C-terminal peptide region which is in complex with UL44 [5]. The complete crystal structure of UL54 is yet to be resolved.

Therefore, this study attempts to explain the probable structural factors conferring inhibitor resistance due to the novel Glutamine (Gln) deletion at 639th position coupled with the most frequently reported missense mutation, Serine(Ser) to Leucine(Leu) at 655th position [6] in the DNA polymerase type – II domain, through sequencing and structural bioinformatics approaches. Polymerase Chain Reaction (PCR) amplified form of this viral strain DNA was produced using primers which were custom-designed in the laboratory. The viral strain isolated patient was non-responsive to GCV treatment as per the clinical records.

Since the minimum inhibitory concentration of GCV could not be established in the laboratory, this study was formulated to predict the phenotypic effects of the observed coupled mutation by modeling the 3 Dimensional (3D) structure of UL54, and its role in binding of the currently practiced drugs. Hence, we attempted to predict the i*n silico* structural models for the wild type (WT) UL54 and for the coupled mutant (MT). To get the deeper insight into the mechanisms behind the drug resistance, molecular docking studies were performed for the WT and MT with GCV, PFA and CDV.

Methodology:
Clinical Specimen:
Five ml of peripheral blood sample was collected in EDTA anticoagulant from a renal transplanted patient who was receiving valganciclovir 400 mg twice a day for three months. The specimen was transported in ice to L & T Microbiology Research Center. Two hundred micro liters of DNA was subjected to DNA extraction using QIAGEN blood extraction kit (Hilden, Germany). The DNA was extracted according to the manufacturer's instructions. HCMV copy numbers were estimated using the kit and protocol as described earlier. The DNA extract of the blood sample was screened for the presence of mutations conferring GCV resistance by PCR –DNA sequencing. Partial UL97 region was amplified using primers and protocols as described earlier [7, 8]. For UL54 DNA region, the amplification primers were designed using Primer Premier 5 software. Out of the 3729bp coding region of the UL54 gene, a 1900 nucleotide region harboring extensive mutations was targeted. The 1900bp was amplified as four overlapping regions and each of the four regions were amplified by

nested PCRs to obtain better sensitivity, with the standardized primers with optimal annealing temperatures **(Table 1 see Supplementary material).**

Specificity of the primers:
The specificity of the primer sets were determined against DNA extracted from *Staphylococcus aureus* ATCC 6538, *Bacillus subtilis* lab isolate, *Escherichia coli* ATCC 4157, *Pseudomonas aeruginosa* ATCC 9742, *Klebsiella pneumoniae* lab isolate, *Mycobacterium tuberculosis* H37 RV, *Proteus vulgaris* lab isolate, *Haemophilus influenza* ATCC16211, *Acinetobacter* ATCC 9956, *Salmonella typhi* lab isolate, *Enterobacter aerogenes* lab isolate, *Citrobacter koseri* lab isolate, *Candida albicans* ATCC90028, Culture infiltrate of Marmoset cell line infected EBV B958, *Varicella* Zoster virus (OKA vaccine strain), Herpes Simplex Virus 1 - ATCC 733-VR and human DNA. DNA extracted from HCMV AD 169 (ATCC VR 538) was used as a positive control. The DNA extract from peripheral blood of a renal transplant recipient was amplified with primers specific for all four regions with DNA extracts of HCMV AD 169 as positive control and deionized autoclaved water was used as negative control. The amplified product was visualized in a 2% agarose gel electrophoresis along with the molecular weight marker. All universal precautions were taken for setting up the PCR reaction.

The PCR amplified products of UL97 and UL54 were further purified by gel elution using QIAGEN gel elution kit. The eluted products were cycle sequenced with the reverse primers using ABI big Dye terminator. The cycle sequenced products were further purified and denatured with formamide and were loaded in ABI sequencer. The sequenced data was BLAST analyzed.

Structural bioinformatics approaches:
The amino acid sequence of HCMV UL54 (Accession No: P08546) retrieved from SWISSPROT database (http://www.uniprot.org/), and found to be 1242 amino acids in length. As per the SWISSPROT entry, the complete crystal data of UL54 is yet to be determined, and only partial structure information is available for the last 20 residues of the C-terminal, which interacts with UL44. Therefore, in this study we attempted to predict the complete structure of UL54 WT and MT with the reported mutations, through structural bioinformatics approaches.

Fold Recognition based method was implemented using ITASSER server [9] to model the WT protein. The generated structure was energy minimized for 500 steps of conjugate gradient using OPLS force field of Gromacs 3.3.4 in order to refine the structure to its near native conformation [10]. Since MT sequence shared 98% sequence identity to WT, homology modeling was implemented to generate the MT model with WT model as a template using MODELLER 9v7 [11]. The modeled structures of WT and MT were validated by analyzing Ramachandran plot generated by PROCHECK [12].

Docking of ligands with UL54:
Triphosphate form of Ganciclovir [PubChem: 506602], Foscarnet [PubChem: 3415] and diphosphate form of Cidofovir [PubChem: 163311] were the set of inhibitors studied in this work. The structural coordinates of these ligand molecules were retrieved from NCBI-Pubchem Compound database [13] and the ligands were further geometry optimized using gromos96 force field via PRODRG server [14].

Automated docking of ligands to their macromolecular receptors was performed using AutoDock4 molecular docking suite [15]. Kollman United atom charges [16] and polar hydrogens were added to the receptor molecules (WT & MT) and Gasteiger charges [17] were added to the ligand molecules (GCV, CDV, PFA) for optimization of the docking process. In this docking simulation, semi-flexible docking protocols [18] were implemented, wherein, the target protein UL54 was kept rigid, and the ligands being docked were kept flexible in order to investigate an arbitrary number of torsional degrees of freedom. Preliminary docking studies were performed to identify potential binding sites of inhibitors by generating a grid box which could potentially cover the entire surface of

the macromolecule. The grid box size was set to 126x126x126Å and was kept constant for all the ligands. The spacing between the grid points was set to 0.836 Å. The docking parameters were set to the default values provided by the software, except for the maximum number of energy evaluations before the Lamarckian Genetic Algorithm (LGA) terminates, which was set to 50000 generations to improve precision. Graphical User Interface program of "Auto Dock 4.0 Tools" was used to prepare, run, and analyze the docking simulations. Finally, the plausible protein-ligand hydrogen bonding interactions within 3Å distances in the docked complexes were analyzed using DS visualizer tool [19].

Discussion:
HCMV copy number in the peripheral blood of patient was estimated as 83,478 copies/ml. The primers designed for UL54 were specific in amplifying only HCMV DNA. The UL97 amplified product on sequencing with reverse primers did not show any mutations. All the four sets of primers targeted against the partial gene sequence of UL54 gave an amplified product with the specimen DNA **(Figure 1)**. The amplified product of the fourth region on sequence analysis (done with reverse primers) showed a deletion of Glu at 639th position **(Figure 2)**. The patient expired and was non-responsive to GCV treatment. The protein sequence of the same sample contained Leu at the 655th position instead of Ser **(Figure 2)**. (Gen Bank Accession number: BankIt1396010 Strain HQ340161).

Hence, the present study was attempted to understand the complexity behind the phenotypic changes due to structural variation via structural bioinformatics approaches. Since the crystal coordinate information of UL54 is not available for the entire structure and crystallized only for the short peptide of C-terminal region, the structure is elucidated. Hence, UL54 -WT structure was generated using ITASSER, a top ranked protein structure prediction server as per CASP [20], which utilizes fold recognition strategy to model the proteins. Moreover, in this, ITASSER server will make use of the folds conferred by crystal structure of C-termini region of UL54 present in PDB for modeling the entire UL54 protein. From the results, the top scoring model with C-Score of -1.59 and TM-Pred value 0.52±0.15 was chosen as a significant model. In order to remove the steric clashes and to optimize the stereo chemical properties, the chosen model was further optimized by energy minimization through 500 steps of conjugate gradient algorithm using Gromacs 3.3.4. The backbone conformation of the refined model was validated using Ramachandran plot obtained through PROCHECK. The generated model was found to be highly plausible, as none of the residues were found to span the disallowed region of Ramachandran plot.

To further validate the plausibility of the structure generated, the crystal structure of Herpes Simplex Virus type 1 DNA polymerase (HSV-1) [21], which is reported to be a probable remote homologue of HCMV-UL54 [22] sharing 29.7 % of sequence identity as predicted using pairwise alignment program of EMBOSS (http://www.ebi.ac.uk/Tools/emboss/align/), was superposed to WT using PyMOL [23] and the Root Mean Square Deviation (RMSD) between the backbone was found to be 2.34 Å. Hence, this was suggestive of fold level homology of the generated model **(Figure 3)**.

The structure of MT UL54 with the deletion of Gln at 639th position coupled with the missense mutation, Ser to Leu at 655th position, shared 98% of sequence identity with the generated WT model. Hence, the WT model was chosen as a template to model the 3D structure for MT based on homology modeling approach using MODELLER 9v7. The predicted MT structure was further validated by superposing its backbone using PyMOL with that of WT structure and the Root Mean Square Deviation (RMSD) was found to be 0.2 Å, and was suggestive of fold level similarity and structure plausibility. Furthermore, the backbone conformation of the model generated was validated using PROCHECK. The generated model as found to be highly plausible **(Figure 4)**, since none of the residues spanned the disallowed region of Ramachandran plot.
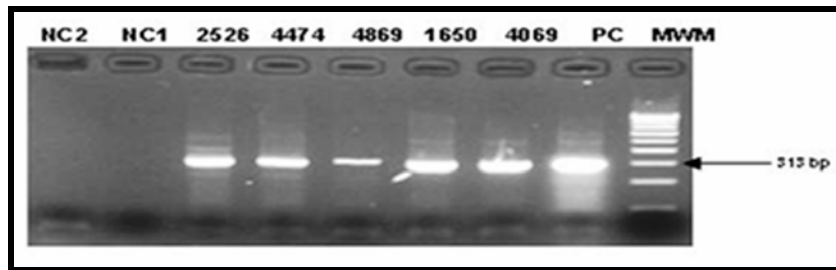
Figure 1: The amplified products of region IV of the specimens along with others specimens detected positive to HCMV. Agarose Gel Electrophoretogram of PCR amplified product targeting Human cytomegalovirus UL54 gene of Region IV. Round II applied on Clinical Specimens. Some of the specimens positive for the region IV along with sample 2256 is shown.
Lanes
NC2: Negative Control - II round
NC1: Negative Control - I round
PC: Positive Control DNA extract from Human cytomegalovirus AD 169
MWM: Molecular weight marker 100 bp Ladder.
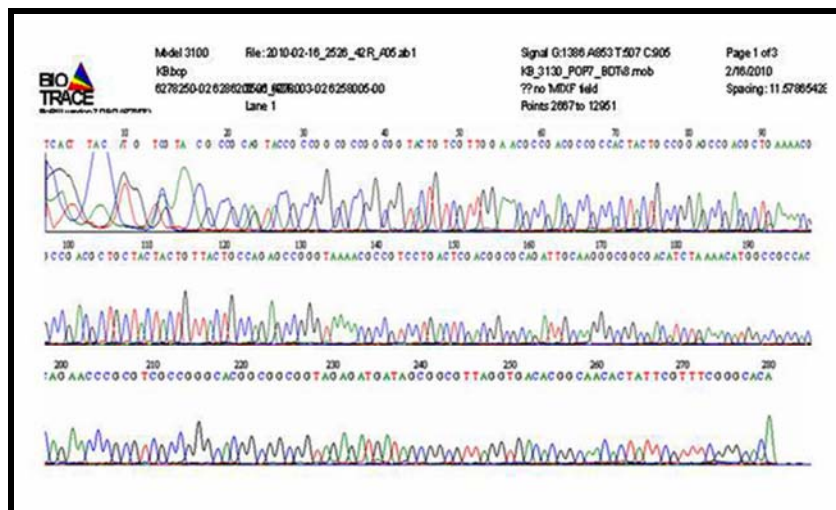Specimens positive for UL54 partial region: 2526, 4474, 4869, 1650, 4069

Figure 2: Electrophoretogram of the DNA sequenced product. DNA sequencing performed with the PCR amplified product of UL54 II round product region IV of a clinical specimen of 2526.
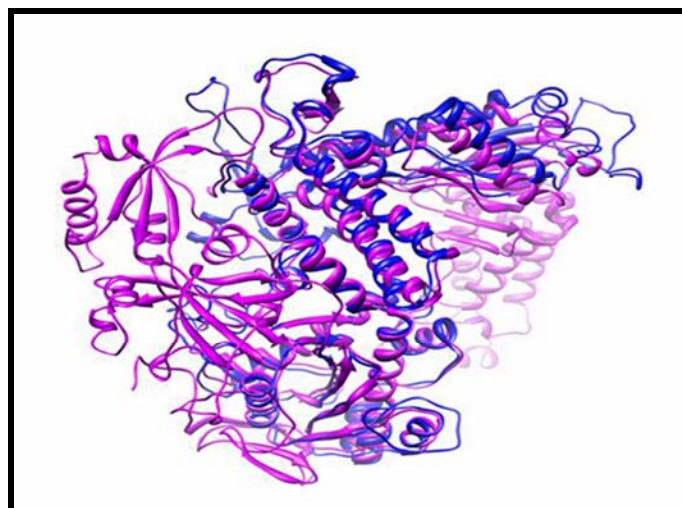

Figure 3: Superposition of crystal structure of HSV-1(magenta) with WT (blue) shows backbone RMSD of 2.343Å.
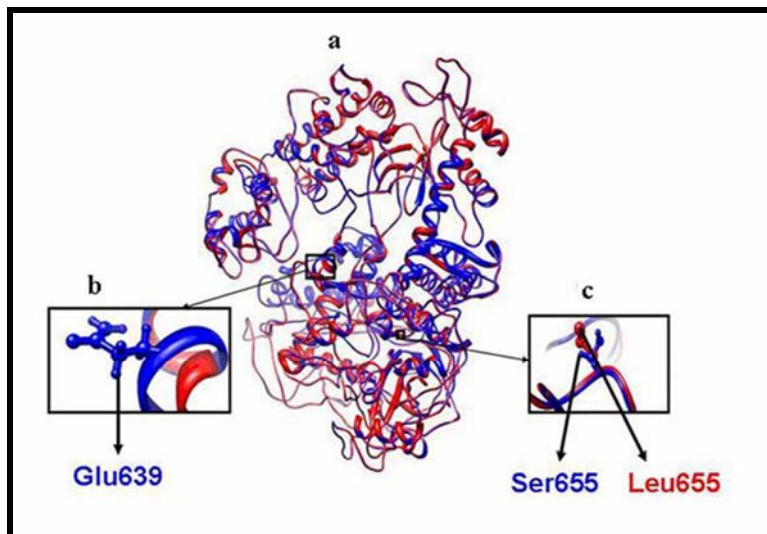
Figure 4: (a) Superposition of WT (blue) with MT(red) shows backbone RMSD of 0.2Å. **(b)** Zoom view of Glu639 deletion. **(c)** Zoom view of Missense Mutation Ser 655 Leu.
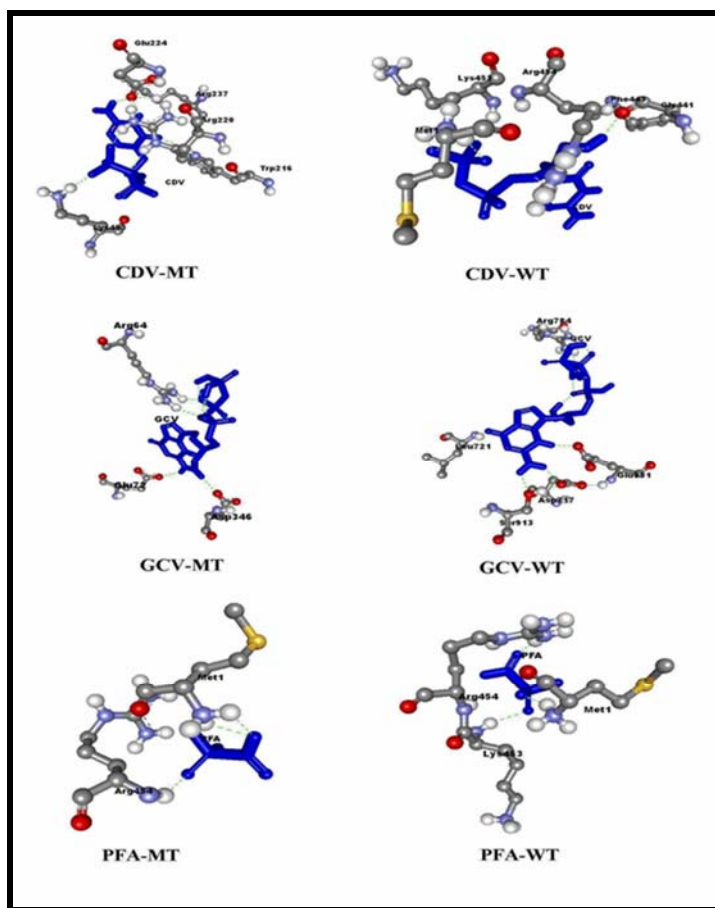

Figure 5: Docking analysis of WT and MT with GCV, PFA and CDV .CDV-MT and CDV-WT represents the hydrogen bonding interaction of diphosphorylated form of Cidofovir (blue) with MT and WT, respectively. GCV-MT and GCV-WT represents the hydrogen bonding interaction of triphosphorylate form of Ganciclovir (blue) with MT and WT, respectively. PFA-MT and PFA-WT represents the hydrogen bonding interaction of PFA (blue) with MT and WT, respectively. Dotted lines (green) represent the hydrogen bonding interaction.

Comparison of GCV inhibitory effects in WT and MT of UL54:
Triphosphorylated form of GCV was used for the docking purposes. In case of WT, GCV formed hydrogen bonding interactions with Asp717, Arg784, Leu721, Ser913 and Glu951 of UL54 WT (**Figure 5**). The functional motif of UL54**,** DNA polymerase family B signature**,** found to span the 908 to 916th residual position was identified using ScanProsite (http://expasy.org/tools/scanprosite/). Among the observed hydrogen bonding residues, Ser913 was found to interact with GCV, which forms a part of the above mentioned signature. This suggests that this residue may play an important role in polymerization inhibitory activity of GCV. The binding free energy and Ki of WT with GCV was found to be -1.67Kcal/mol and of 59.67mM, respectively. In case of docking of MT with GCV, the drug was found to be hydrogen bonded with Arg64, Glu72 and Asp346, and these residues were not part of the DNA polymerase family B signature motif. Furthermore, the stability of the MT-GCV complex was less in comparison to WT-GCV complex with the binding energy of -0.88 Kcal /mol (**Table 2 see Supplementary material**). Moreover, the Ki for MT-GCV showed more than 3 fold increment with 226.71mM value in comparison to WT-GCV. These results indicate that the acquired mutation is an adaptive evolution of resistance strain, wherein it has reduced the affinity of GCV with MT by conferring structural variations (**Figure 5**).

Comparison of GCV inhibitory effects in WT and MT of UL54:
Triphosphorylated form of GCV was used for the docking purposes. In case of WT, GCV formed hydrogen bonding interactions with Asp717, Arg784, Leu721, Ser913 and Glu951 of UL54 WT (**Figure 5**). The functional motif of UL54, DNA polymerase family B signature was found to span from 908 to 916th residual position was identified using ScanProsite (http://expasy.org/tools/scanprosite/). Among the observed hydrogen bonding residues, Ser913 was found to interact with GCV, which forms a part of the above mentioned signature. This suggests that this residue may play an important role in polymerization inhibitory activity of GCV. The binding free energy and Ki of WT with GCV was found to be -1.67Kcal/mol and of 59.67mM, respectively. In case of docking of MT with GCV, the drug was found to be hydrogen bonded with Arg64, Glu72 and Asp346, and these residues were not the part of DNA polymerase family B signature motif. Furthermore, the stability of the MT-GCV complex was less in comparison to WT-GCV complex with the binding energy of -0.88 Kcal /mol (**Table 2**). Moreover, the Ki for MT-GCV showed more than 3 fold increment with 226.71mM value in comparison to WT-GCV. These results indicate that the acquired mutation is an adaptive evolution of resistance strain, wherein it has reduced the affinity of GCV with MT by conferring structural variations (**Figure 5**).

Increased binding affinity of CDV to UL54 mutant:
Since cidofovir is a nucleoside analogue used for anti-HCMV treatment, the diphosphorylated form of Cidofovir was used for docking purposes in this study. CDV-WT docked complex showed hydrogen bonding with Met1, Gly441, Phe443, Lys453 and Arg 454. Whereas**,** in case of MT-CDV complex, except for Lys453 of DNA polymerase type-II domain, other hydrogen bonding interactions as observed in WT were lost. Alternatively, CDV formed stable hydrogen bonding interactions with Trp216, Arg220, Glu224 and Arg237 of MT. This pattern of MT-CDV complex formation was found to be more stable with a binding energy of -6.16 Kcal/mol in comparison to WT-CDV wherein, it was -5.62 Kcal/mol. Moreover, the Ki value for MT-CDV was found to be 30.47 μM, which was two fold lesser than WT-CDV with 75.83μM (Table 2) (Fig. 5).

Docking analysis of PFA:
WT- PFA complex showed hydrogen bonding interactions with Met1, Lys453 and Arg454 with binding energy of -6.04 Kcal/mol and a Ki value of 37.3μM, whereas, in case of MT-PFA complex, the binding energy and

Ki value were found to be increased to -5.81 Kcal/mol and 54.83 μM, respectively. In both the cases, Met1 and Arg454 binding residues were conserved and Lys453 hydrogen bonding was lost in MT-PFA. This is suggestive of MT resistance towards PFA (**Table 2**) (**Figure 5**).

Conclusion:
In the present study, the structural bioinformatics analysis reveals the probable mechanisms of drug interactions conferred by GCV, CDV and PFA with UL54-WT as well as with the MT. Lys453 of UL54 was predicted to play an important role in conferring stability of the complex formation as it was found to exhibit interaction in cases of WT-PFA, WT-CDV and MT-CDV. Furthermore, the observed variant was found to be more susceptible to CDV in comparison to other drugs. Hence, this study provides deeper insights on mutation and drug resistance relationship in UL54 of HCMV.

References:
[1] MA Jacobson & J Mills *et al*. *Ann Intern Med* **108**: 585 (1988) [PMID: 2831765]
[2] AM Fillet *et al*. *Antimicrob Agents Chemother* **48**: 1865(2004) [PMID: 15105145]
[3] F Baldanti & G Gerna. *J Antimicrob Chemother* **52:** 324 (2003) [PMID: 12888590]
[4] F Picard-Jean *et al*. *The Biochemical journal* **407**: 331(2007) [PMID: 17672827]
[5] BA Appleton *et al*. *J Biol Chem* **8**: 5224 (2006) [PMID: 16371349]
[6] EF Bowen *et al*. *J Med Virol* **58**: 402 (1999) [PMID: 10421408]
[7] S Chou *et al*. *J Infect Dis* **188**: 32 (2003) [PMID: 12825168]
[8] J Castor *et al*. *J Clin Microbiol* **45**: 2681(2007) [PMID: 17537934]
[9] A Roy *et al*. *Nat Protoc* **5**: 725 (2010) [PMID: 20360767]
[10] CM Summa & M Levitt. Proc Natl Acad Sci USA **104**: 3177(2007) [PMID: 17360625]
[11] A Fiser & A Sali. *Methods Enzymol* **374**: 461 (2003) [PMID: 14696385]
[12] RA Laskowski *et al*. *J Biomol NMR* **8**: 477(1996) [PMID: 9008363]
[13] Y Wang *et al*. *Nucleic acids research* **37**: W623 (2009) [PMID: 19498078]
[14] AW Schuttelkopf & DM. van Aalten. *Acta Crystallogr D Biol Crystallogr* **60**: 1355 (2004) [PMID: 15272157]
[15] GM Morris *et al*. *Journal of Computational Chemistry* **19**: 1639 (1998) (http://mgl.scripps.edu/people/sanner/html/papers/1998.Morris-J.Comput.Chem.19.1639.pdf)
[16] UC Singh *et al*. *Journal of Computational Chemistry* **5**: 129 (1984) (http://onlinelibrary.wiley.com/doi/10.1002/jcc.540050204/abstract)
[17] J Gasteiger *et al*. *Tetrahedron* **36**: 3219(1980)
[18] U Gowthaman *et al*. *BMC Bioinformatics* **9**: S14 (2008) [PMID: 19091013]
[19] TT Talley *et al*. *Proc Natl Acad Sci USA* **105**: 7606 (2008) [PMID: 18477694]
[20] Y Zhang, *BMC Bioinformatics* **9**:40 (2008) [PMID: 18215316]
[21] S Liu *et al*. *J Biol Chem* **281**:18193 (2006) [PMID: 16638752]
[22] J Ghuman *et al*. *J Mol Biol* **353**: 38(2005) [PMID: 16169013]
[23] http://www.uniprot.org
[24] http://www.ebi.ac.uk/Tools/emboss/align/
[25] http://expasy.org/tools/scanprosite/

Supplementary material:

Table 1: List of primers designed to amplify the partial sequence of UL54 gene with the corresponding annealing temperature and the amplified product size.

Region	Round	Primers	Annealing Temperature	Amplicon Size
I	1	FP: 5'GTGAGGTCATTGCCAGCGT 3' RP: 5'AATGTCATCGGACTTCTCGG 3'	64ºC	457bp
	2	FP: 5' CCGATTTGCAGTGTGTGTCT 3' RP: 5' CAGTCGTAACGATTCACGGA 3'	65ºC	170bp
II	1	FP: 5' GGATCAGGGGATCCCAAAC 3' RP: 5'AGGTACTCGAGACGCGTGAG 3'	60ºC	268bp
	2	FP: 5' TGCTGGGCTTTATGCTTTTC 3' RP: 5'CGCGTGAGGATGTACTTCAA 3'	60ºC	90bp
III	1	FP: 5'GTATTGGTGCGCGATCTGTT 3' RP: 5'CTACTGTTACTGCCGGAGCC 3'	65ºC	369bp
	2	FP: 5'GTGCCCGAAACGAATAGTGT 3' RP: 5'CTACTGTTACTGCCGGAGCC 3'	67ºC	183bp
IV	1	FP: 5'TCATCTTTGACGGACAGCAG 3' RP: 5'CACTTGTTGAGCAGTTCCGA 3'	60ºC	694bp
	2	FP: 5'GTGCCCGAAACGAATAGTGT 3'	60ºC	313bp

Table 2: Molecular interactions of WT and MT with Cidofovir, Ganciclovir and Foscarnet

Ligand	Binding Energy Kcal/mol		Inhibitory Constant		Hydrogen Bonding Residues	
	Wild	Mutant	Wild	Mutant	Wild	Mutant
Cidofovir	-5.62	-6.16	75.83μM	30.47 μM	Met 1 Gly441 Phe443 Lys 453 Arg 454	Trp216 Arg 220 Glu224 Arg237 Lys 453
Ganciclovir	-1.67	-0.88	59.67mM	226.71mM	Asp 717 Leu 721 Arg 784 Ser913 Glu 951	Arg 64 Glu72 Asp 346
Foscarnet	-6.04	-5.81	37.3μM	54.83 μM	Met 1 Lys 453 Arg 454	Met 1 Arg 454