 doi:10.20944/preprints202012.0671.v1

Article

Modeling and molecular dynamics of the 3D structure of the HPV16 E7 protein and its variants

Ciresthel Bello-Rios[1], Sarita Montaño[2,*], Olga Lilia Garibay-Cerdenares[1,3], Lilian Esmeralda Araujo-Arcos[1], Marco Antonio Leyva-Vázquez [1], Berenice Illades-Aguiar [1,*]

[1] Molecular Biomedicine Laboratory, Faculty of Chemical-Biological Sciences, Autonomous University of Guerrero, Chilpancingo, Mexico. CP 39087; ciresthelbello@uagro.mx (C.B.R.), esme_1018@hotmail.com (L.E.A.A.), leyvamarco13@gmail.com (M.A.L.V.), b.illadesaguiar@gmail.com (B.I.A.).

[2] Laboratorio de Bioinformática y simulación molecular, Facultad de Ciencias Químico Biológicas, Universidad Autónoma de Sinaloa, Culiacán Sinaloa. CP 80030; mmontano@uas.edu.mx (S.M.)

[3] CONACyT-Autonomous University of Guerrero, Chilpancingo, Mexico. CP 39087; olgaribayce@conacyt.mx (O.L.G.C.).

* Correspondence: mmontano@uas.edu.mx (S.M.) and b.illadesaguiar@gmail.com (B.I.A.), Tel. +52 747-499-15-68

Abstract: The oncogenic potential of high-risk HPVs is focused on producing the E6 and E7 oncoproteins responsible for disrupting the control of the cell cycle. Epidemiological studies propose the presence of the N29S and H51N variants of the HPV16 E7 protein as a significant association with cervical cancer. It has been suggested that changes in the amino acid sequence of E7 variants may affect the oncoprotein 3D structure; however, this remains unknown. Analysis of the structural differences of the HPV16 E7 protein and its variants (N29S and H51N) was performed through homology modeling and structural refinement by molecular dynamics simulation. We propose for the first time a 3D structure of the E7 reference protein and two of its variants (N29S and H51N) and conclude that the mutations induced by the variants in N29S and H51N have a significant influence on the 3D structure of the E7 protein of HPV16, which could be related to the oncogenic capacity of this protein.

Keywords: HPV, E7, variants, molecular dynamics simulation

1. Introduction

Cervical cancer (CC) is a serious public health problem worldwide [1]. Human papillomavirus (HPV) infection is the main cause of CC [2]. However, not all infections cause tumor development due to the oncogenic potential of different types of HPV [2-5]. For instance, high-risk HPV types 16 and 18 are reported with the greatest frequency in cervical cancer [4, 6, 7]. The oncogenic potential of high-risk HPV is due to E6 and E7 oncoprotein production, which initiates a series of alterations associated with cell transformation through the inactivation of p53 and pRB, respectively [4, 8, 9]. The E7 protein plays a central role in the life cycle of the virus and tumor transformation [10-13]. The E7 protein has the ability to transform due to its interaction with multiple targets involved in different cellular processes, such as apoptosis, angiogenesis, cell immortality, resistance to cell death, and Mesenchymal Epithelial Transition (EMT) [14-18].

The E7 protein of HPV16 is composed of 98 amino acids and is divided into three conserved regions, CR1, CR2, and CR3 [19-21]. CR1 of E7 consists of residues M1-L15 in the N-terminal region. CR2 (Q16- I38) contains the LXCXE motif, which establishes high-affinity interactions with the retinoblastoma protein (pRB) [21] [22, 23]. In addition, E7 contains a recognition motif for casein kinase II (CKII), which phosphorylates serine residues 31 and 32. These phosphorylation events confer structural stability to the protein [23-25]. Finally, CR3 (D39-P98), located in the C-terminus, contains two CXXC motifs that form a zinc finger structure involved not only in protein stability but also in its dimerization [19, 20, 26-29].

 doi:10.20944/preprints202012.0671.v1

The three-dimensional structure of the E7 protein allows it to associate with cell cycle regulatory proteins [23-25, 30]. The N-terminus confers conformational plasticity to promote interactions with other proteins. These transition properties and the plasticity of E7 play key roles in its ability to interact with different components of cellular processes and thus initiate malignant transformation [2, 19, 28, 31, 32].

Mutations in the E7 gene of HPV16 are related to the pathogenicity of the virus; these intratypical variations may or may not generate changes in the sequence of amino acids. At least 23 variants have been reported to induce amino acid changes in the protein. These changes are epidemiologically related to the oncogenicity of the virus in different populations worldwide [33-36]. Studies have shown a significant association between the presence of the G647 (N29S) and A712 (H51N) variants of the E7 gene and CC [33-35]. These studies suggest that mutations in the E7 gene affect its 3D structure and modify its interaction with target proteins to induce their oncogenic potential [33-35, 37]. Nevertheless, these amino acid changes are poorly understood. The aim of this study was to elucidate the structural changes in the E7 reference protein and its variants by employing structural analysis *in silico* and molecular dynamics (MD) simulations.

2. Results and Discussion

2.1. Modeling and structural alignment of the E7 reference and its variants

Multiple alignments of the sequences were performed to observe the variants' residue changes with respect to the E7 reference. The changes were in the N29S variant (blue) were located inside the CR2 region, and those of the H51N variant (orange) were in the CR3 region (Figure 1A). The 2D structures of the E7 reference and variants were predicted and were identified via alignment (green line); the results show that the secondary structure was not changed between the reference and the variants.

To predict the 3D model of the E7 reference, we used the crystal structure of the C-terminal domain of the E7 HPV45 protein (PDB ID: 2EWL) as a template in the I-TASSER server. The structural alignment between 2EWL and the 3D model of the E7 reference is presented in pink and green, respectively (Figure 1B). After obtaining the E7 model, we generated the variants with their respective mutations N29S (orange) and H51N (blue) in PyMOL (Figure 1C). The E7 reference contains a Zn finger. To place the Zn finger of the CR3 region in E7, we first deprotonated the cysteine residue with the CYS_D patch, and subsequently, the ZN_C structural patch was added to the Zn finger at C 58, 61, 91, and 94. These proteins have phosphorylation sites, and we used the SP2 structure patch to place the phosphorylation sites of S31 and 32 (in the reference and H51N) and S29, 31, and 32 (in N29S). Figure 1D shows the phosphorylation and Zn fingers.

2.2. N29S presents the most stable trajectory compared to the E7 reference and H51N.

We performed an MD simulation of 200 ns to understand the effect of the mutations on the structural motion and stability of the E7 reference and the N29S and H51N variants. Changes in protein stability were evaluated with the RMSD (Figure 2A), and the E7 reference (green) reached equilibrium after 80 ns. N29S (orange) reached equilibrium after 20 ns and remained stable throughout the trajectory simulation. In contrast, H51N (blue) reaches equilibrium after 20 ns and remained stable until 100 ns, but in the last 100 ns of the trajectory simulation, we observed an equilibrium period from 100 to 150 ns and then an oscillating pattern of approximately 2 Å. This behavior of the variant may be due to the nature of the residue change. To understand how the mutations affected protein backbone motion, the root mean square fluctuation (RMSF) was calculated. RMSF is used to explore the protein flexibility of the Cα of a protein throughout the trajectory of the MD simulation. Higher RMSF values indicate greater flexibility during the MD simulation. Interestingly, the principal peak of fluctuation was located between residues N29 and H51 (shaded in pink and violet) in all the structures. However, this fluctuation was highest for the H51N variant, and the fluctuation corresponding to N29S was minor with respect to the E7 reference (Figure 2B). The principal peak of fluctuation corresponds to the CR2 and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 December 2020 doi:10.20944/preprints202012.0671.v1

CR3 regions of these proteins, where structural differences were observed. The radius of gyration (Rg) is the parameter that defines the balance of the conformations of the total system in terms of compaction with respect to the folding and unfolding of the proteins throughout the simulation. The Rg between the E7 reference (green) and N29S (blue) showed similar compaction during dynamics, while the H51N (blue) variant exhibited an expansion in the final 50 ns of the trajectory simulation. This behavior agrees with the RMSD and RMSF values, showing that H51N has the highest peak of fluctuation among the proteins. The N29S variant shows a more compact structure throughout the MD simulation than the E7 reference and the H51N variant (Figure 3C). This behavior on the trajectory of N29S may suggest the importance of the extra phosphorylation site on S29.

To validate the quality of the 3D structure of the E7 reference and variants, we generated a Ramachandran plot. The results indicated that, for the E7 reference, 85.3% of the amino acids were in the favored region, 7.1% were in the allowed region, and 3.6% were not allowed (Figure 3A). The N29S variant showed that 88.1% of the residues are in the favored region, while 10.7% are in the permitted region, and 1.2% are in the disallowed region (Figure 3B). With respect to the H51N variant, 90.8% of its amino acids are found in the favored region, 8.0% in the allowed region, and 1.1% in the not allowed region (Figure 3C). These data show that the refined models are reliable. To observe the structural changes between reference E7 and the variants, an alignment of the average structure of the reference E7 and the N29S and H51N variants was performed. The alignment showed that there were differences in the 3D structure between the proteins (Figure 3D), and changes were observed in the three regions of the proteins; however, the main changes were found in the CR1 and CR2 regions (Figure S1), which contain the main motifs that enable E7 to interact with its molecular targets. These changes may explain the different oncogenic potential of the E7 variants.

2.3. H51N loses β-sheets throughout the trajectory simulation

The E7 protein has a disordered N-terminus that includes the CR1 and CR2 domains, which are involved in most of the protein interactions with its molecular targets [37, 38]. To understand the differences between the E7 reference and its variants, we explored the changes in the secondary structure of the protein throughout the trajectory of the MD simulation. The most significant changes in the 2D structures are in the E7 reference since it loses the α helix and the β-sheets (Figure 4A). However, most of the secondary structures of N29S are preserved (Figure 4B). The H51N variant conserves the α-helix but loses the β-sheets after 140 ns (Figure 4C). The loss of the β-sheet structures may explain the oscillating behavior observed in the final 50 ns, as indicated by the RMSD values and the compaction of the RG. This outcome reveals that the mutation makes the N29S variant more structurally stable than the E7 reference and H51N. These findings agree with the trajectory analysis, where N29S had the most stable RMSD, the lowest RMSF, and major compaction, compared to the others.

2.4. Surface electrostatic potential of E7 and its variants

The LxCxE motif of the CR2 region of E7 is crucial for the interaction with pRB, which is its main molecular target. Studies have shown that mutations in the E7 LxCxE motif drastically influence pRB degradation [19, 39]. Our results demonstrated an essential structural change induced by mutations. To determine whether the LxCxE motif of E7 was different between the systems, an analysis of the electrostatic potential was performed using the adaptive Poisson-Boltzmann solver (APBS) plugin in PyMOL to map the electrostatic potential of the surface of the E7 reference and its variants (Figure 5). The charge differences of the variants with respect to the E7 reference were noticeable. The E7 reference has a negative charge on the LxCxE motif (inside the green circle), while in the N29S variant, it has not only a prominent negative charge but also an electrostatically positive region and a neutral region. In the H51N variant, the LxCxE motif appears to be maintained by a combination of hydrophobic interactions and electrostatic

 doi:10.20944/preprints202012.0671.v1

complementation, as it contains negative, neutral, and positive regions. These differences may be related to the affinity of their interactions with target proteins and explain the differences in the oncogenic potential between them.

2.5. The CR3 region allows homodimer formation of the E7 protein.

It has been reported that the E7 protein can form homodimers, allowing interactions with more than one protein at the same time and thus enabling the formation of protein complexes [23, 27, 29]. Studies have reported that the amino acids in CR3 of E7 are involved in homodimer formation [13, 22, 23, 29]. Our results show that E7-E7 homodimer formation involves binding in the CR3 regions, leaving the N-terminal regions free for interactions with other proteins (Figure 6). This finding agrees with those experimentally reported.

2.6. Principal component analysis

PCA was performed to support the results of the MD simulation and to understand the structure and conformational changes of the E7 reference, N29S, and H51N by calculating the atomic fluctuation covariance matrix (Figure 7A). From this result, it was observed that the first 20 main components represent 72 to 75% of the total movement (72.5, 73.9, and 74.3% for the E7 reference and the N29S and H51N, respectively). This analysis suggests that the properties of the movements described by the first PCs are different in the three protein systems (Figure 7B). The projection of PC1, and PC2 for each of the proteins (E7 reference, N29S, and H51N) shows that the H51N variant presents more significant displacement since it covers a wide range of space with respect to the E7 reference and the N29S variant. The N29S variant has the most restricted movement, making it the most stable of the three protein systems (Figure 7C, D, and E). This result shows differential distribution between the protein systems, suggesting that the conformational mobility is changed because of the mutation of a single residue (Figure 7D and E).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 December 2020 doi:10.20944/preprints202012.0671.v1

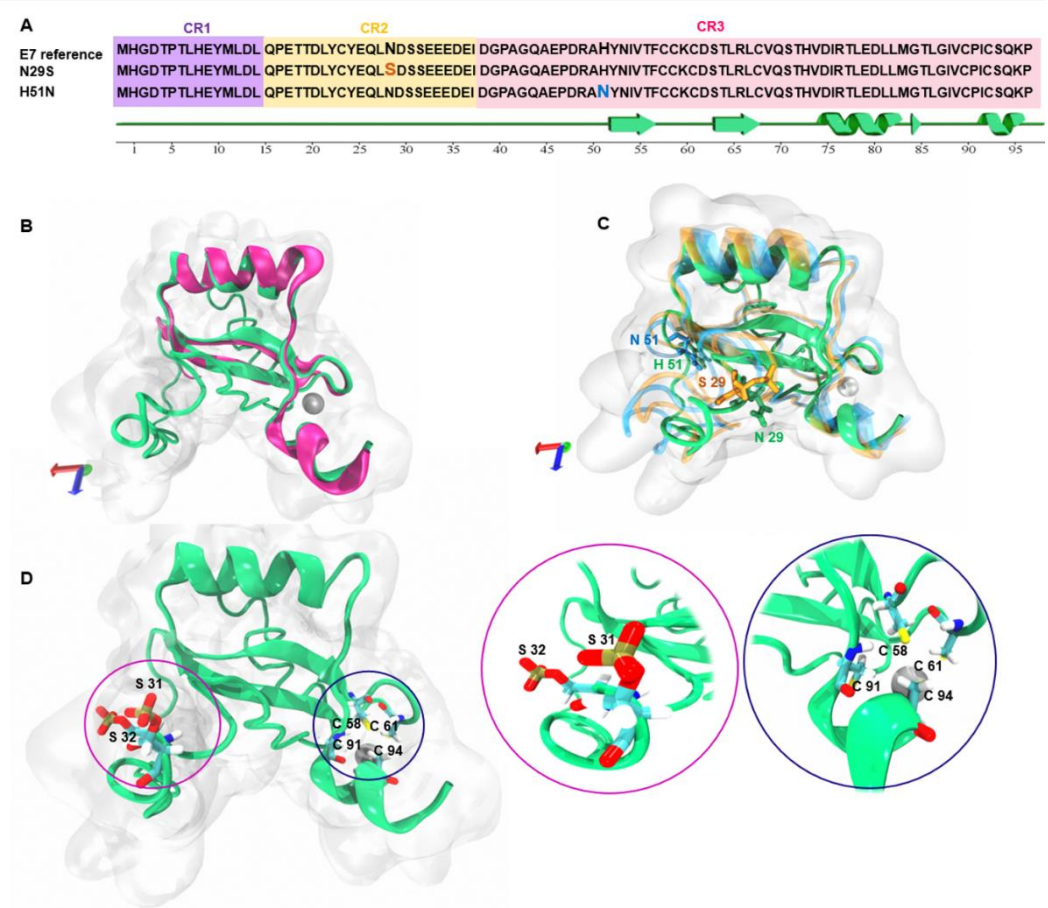

Figure 1. Amino acid sequence alignment and 3D model of E7 and variants. A) The amino acid sequence of HPV16 E7 (P03129) is aligned with the sequences of the N29S and H51N variants. Regions highlighted in orange and blue indicate the amino acid change corresponding to each variant. B) 3D structure of the E7 reference (green) and HPV45 crystal template (PDB ID: 2EWL) in pink. C) 3D structural alignment of the E7 reference (green), N29S (orange), and H51N (blue) variants. D) Phosphorylation sites at serines 31 and 32 in the N-terminal region are in the pink circle, and the blue circle shows the presence of Zn finger in cysteines 58, 61, 91, and 94 at the C-terminal, the zoom of these regions are next to the figure D.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 December 2020 doi:10.20944/preprints202012.0671.v1

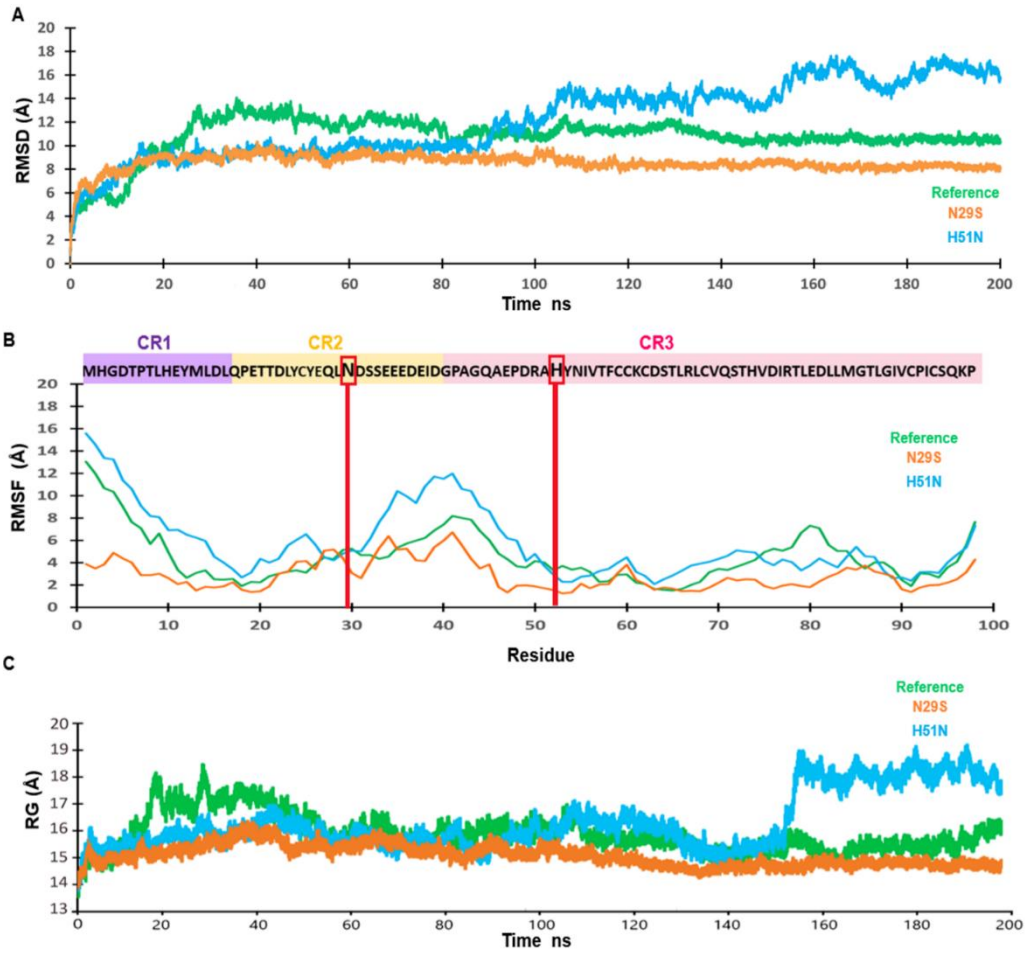

Figure 2. Trajectory analysis of Molecular dynamics simulation of E7 and variants. A) The root mean square deviation (RMSD). The trajectory of the systems is shown after a 200 ns of MD simulation. B) The root mean square fluctuation (RMSF) of Ca atoms after a 200 ns of MD simulation. In the upper panel, the E7 reference amino acid sequence is located in the purple shading are the amino acids from the CR1 region. In yellow CR2 and, in pink CR3, the red lines and boxes indicate the mutations of each variant. C) Radius of gyration (Rg) of the E7 reference (green), N29S (orange), and H51N (blue).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 December 2020 doi:10.20944/preprints202012.0671.v1

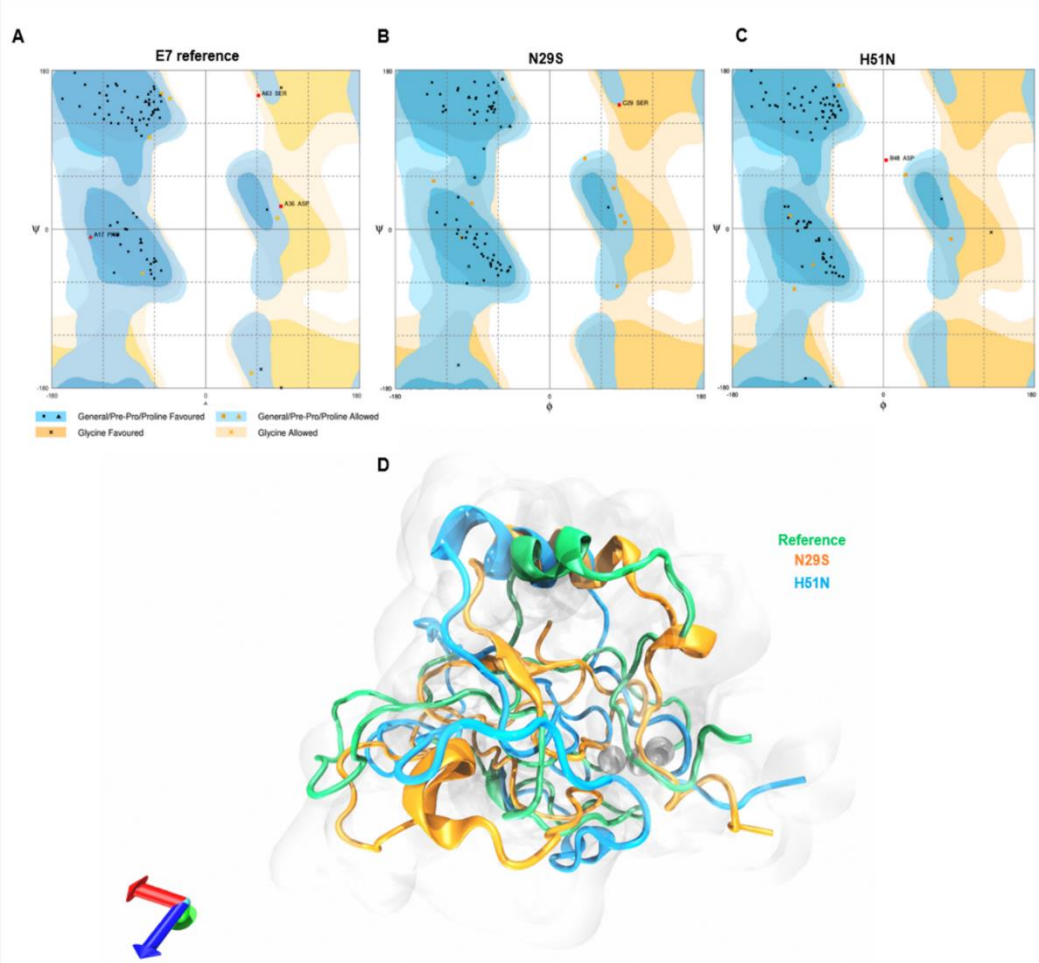

Figure 3. Evaluation and structural alignment of the average structure of E7 reference and variants. A) Ramachandran plot analysis of E7 reference. B) Ramachandran plot analysis N29S. C) Ramachandran plot analysis H51N. D) Structural alignment of the average structures of the E7 reference (green), N29S (orange), and H51N (blue).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 December 2020 doi:10.20944/preprints202012.0671.v1

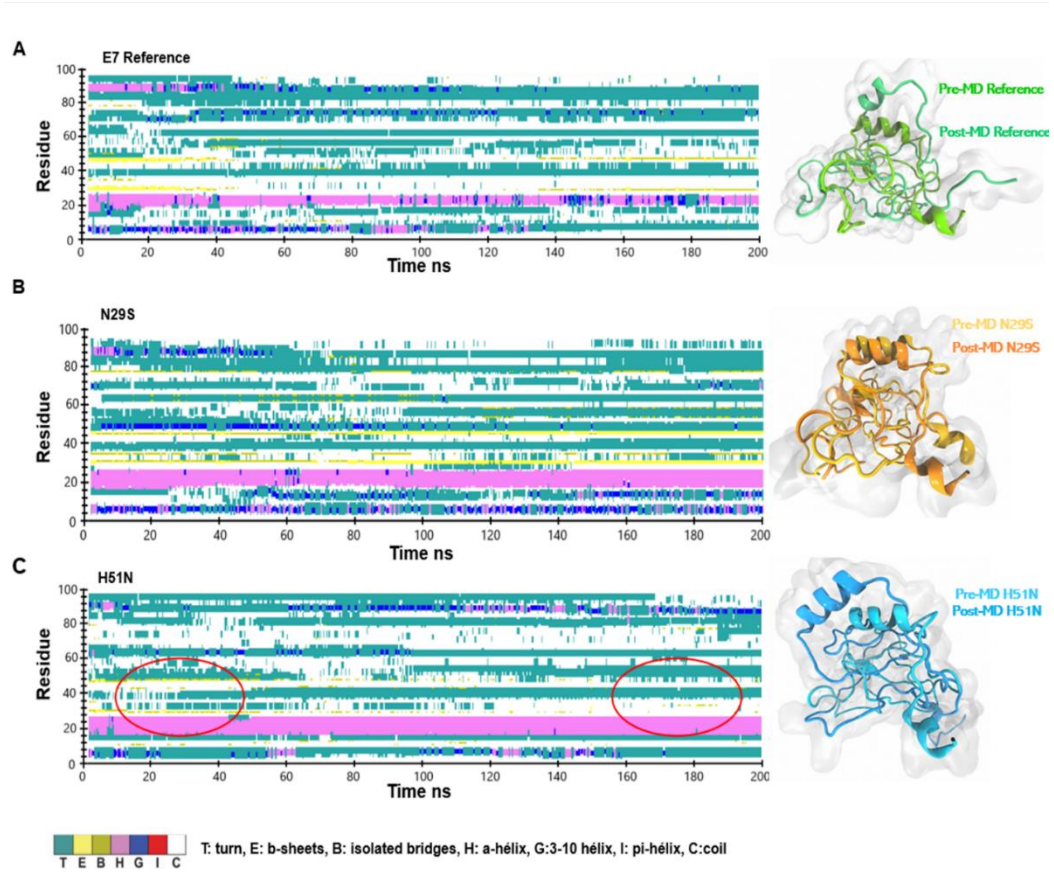

Figure 4. Analysis of the secondary structure through 200 ns of MD simulation of the proteins. A) Secondary structures analysis of the E7 reference. Right: Structural alignment of the initial structure and snapshot corresponding to 200 ns for E7 reference. B) Secondary structure analysis of the N29S variant. Right: Structural alignment of the initial structure and snapshot corresponding to 200 ns for N29S. C) Secondary structure analysis of the H51N variant. Right: Structural alignment of the initial structure and snapshot corresponding to 200 ns for H51N. The principal zones of β-sheets are indicated with circles in red.

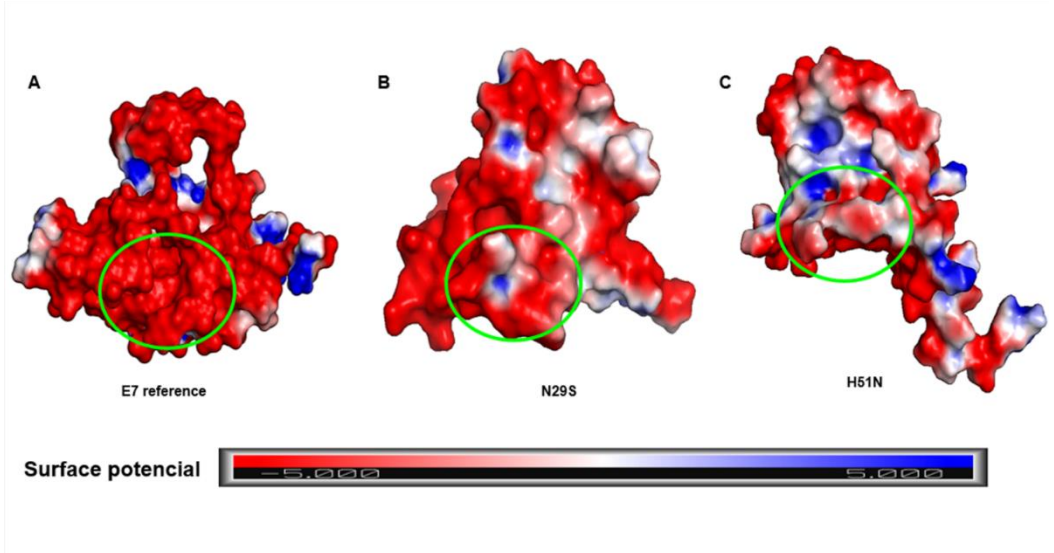

Figure 5. Surface representation of the electrostatic potential of the E7 protein and variants. A) E7 reference. B) N29S variant. C) H51N variant. The surface colors are represented in red (–5) or blue

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 December 2020 doi:10.20944/preprints202012.0671.v1

(+5). The marked regions inside the green circle show the different charge distributions in the LxCxE motif.

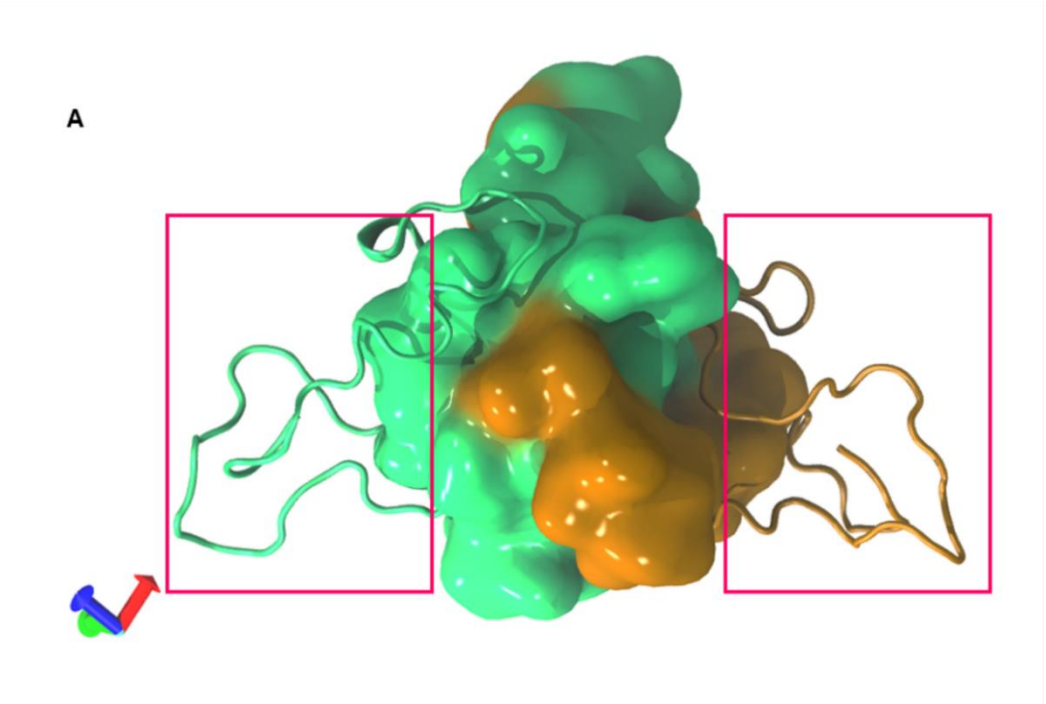

Figure 6. Homodimer prediction of E7. A) Homodimer of the reference E7 average structure model. The CR3 region of each monomer is represented in quicksurf (green and orange). In the boxes, the N-terminal region of each monomer is highlighted.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 25 December 2020 doi:10.20944/preprints202012.0671.v1

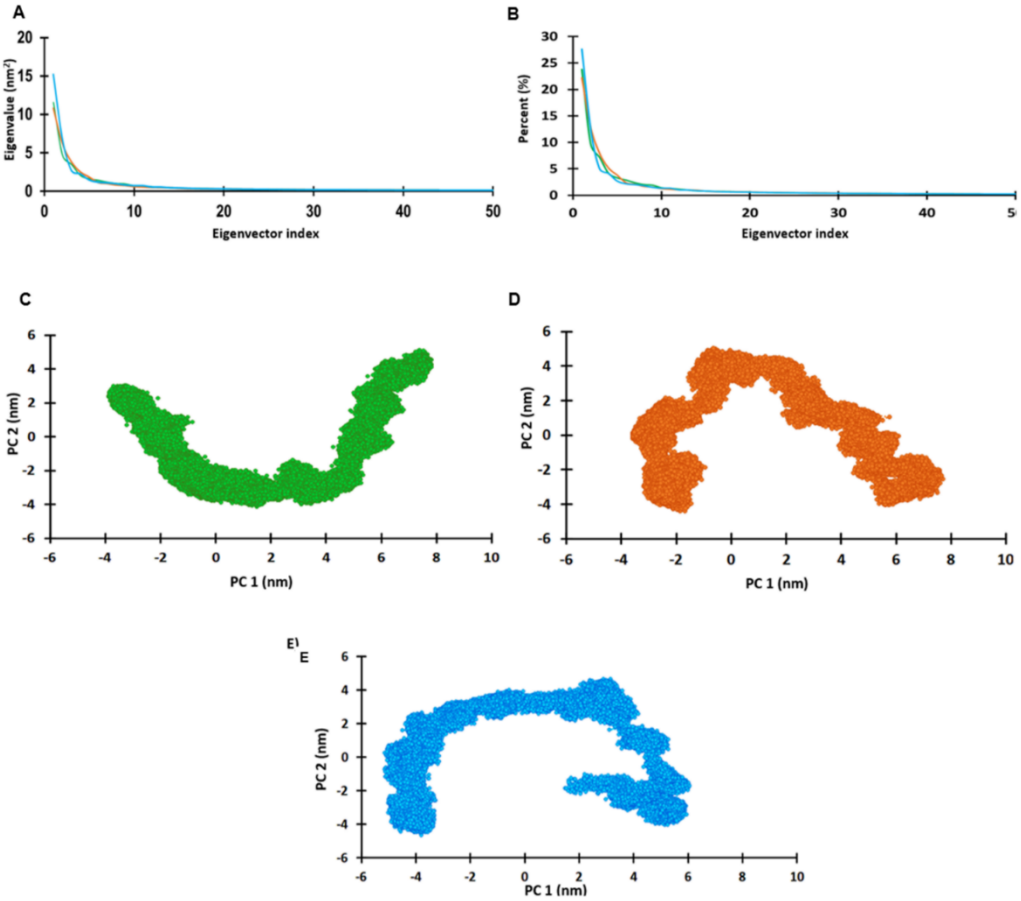

Figure 7. Principal components analysis of the E7 protein and its variants. A) Eigenvectors of the covariance matrix. B) Percentage of each eigenvector vs. eigenvalues, C) Projection of the movement in the phase space between the first and second eigenvectors (PC1 vs PC2) of the reference E7, D) Projection of the movement in the phase space along the first and second eigenvectors (PC1 vs PC2) N29S, E) Projection of the movement in phase space along the first and second eigenvectors (PC1 vs PC2) of H51N. Reference E7 protein is represented in green, N29S orange, and H51N in blue.

3. Materials and Methods

3.1. The 3D structure of E7 and its variants

Multiple alignments of the sequences and the calculations of the percent identity of the amino acid sequences were performed using CLUSTAL X 1.81 [40]. The secondary structure of the E7 protein models and its variants were predicted using PSIPRED server [41], and templates were located in the Protein Data Bank (PDB) (https://www.rcsb.org/). The HPV16 E7 structure was built by the I-TASSER server [42] using the P03129 sequence obtained from the UniProtKB database (http://www.uniprot.org). The homology model was generated using the structure of the C-terminal portion of the E7 oncoprotein of HPV45 as the template (PDB ID: 2EWL) [20] in the I-TASSER server. After selecting the most energetically stable models, we evaluated their quality by generating Ramachandran's plots with the RAMPAGE server (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php). To build the N29S and H51N variants, the 3D structure obtained from the PDB was edited with the PyMOL program (http://www.pymol.org). The alignment of the 3D structures of the models by homology and the refined models were performed with the alignment plugin of the PyMOL program and visualized with VMD software [43].

3.2. Molecular dynamics simulation

For the preparation of the molecular dynamics simulation, a collaboration was carried out with the Laboratory of Molecular Modeling and Bioinformatics of the Faculty of

 doi:10.20944/preprints202012.0671.v1

Biological Chemical Sciences at the Autonomous University of Sinaloa using the Hybrid Cluster "Xiuhcoatl" in CGSTIC-CINVESTAV (http://clusterhibrido.cinvestav.mx). The proteins were prepared with CHARMM/VMD NAMD software [43-45]. The E7 protein contains a Zn finger motif and is phosphorylated; specific structural patches were used in each case using CHARMM. The CHARMM 36 force field was used to create the protein topology file [46]. To neutralize the system, we added 4507 water molecules and 35 Na+ and 12 Cl^- to the E7 reference. To the N29S variant, we added 4578 water molecules and 35 Na+ and 13 Cl^-, and to the H51N variant, 4731 water molecules, and 33 Na+ and 13 Cl^- were added. Each system was solvated in a cubic box of TIP3P water with a minimum distance of 10 Å imposed between the solute atoms and the edge of the box. The system charges were balanced by adding the chloride and sodium ions required for each protein. The system was minimized to 1000 steps followed by an equilibrium step at a constant temperature and pressure using NPT at 300 K for one ns followed by a 200-ns MD simulation for each protein (E7 and its variants). To visualize and analyze the results of the simulation, we used the VMD graphics program [43].

3.3. Trajectory and secondary structure analysis

The stability and conformational changes of the trajectory for E7 and its variants were evaluated by analyzing the root mean square deviation (RMSD), the root mean square fluctuation (RMSF), and the radius of gyration (Rg). The dynamics data were calculated by the Carma program [47] and plotted in the Excel program. The secondary structure calculation was carried out in the VMD program using the timeline plugin through the entire trajectory simulation of 200 ns.

3.4. Calculation of surface electrostatic potential

The electrostatic potential energy was calculated using the adaptive Poison-Boltzmann solver (APBS) package in PyMOL. This plugin solves the Poisson-Boltzmann equation for biomolecules. The potential areas were colored using -5 and +5 kT/e, where k is the Boltzmann constant, T is the temperature, and e is the electron charge. The figures were visualized and generated with the PyMOL program.

3.5. Homodimer prediction

To predict the homodimer formation of the E7 reference, the COTH server was employed (https://zhanglab.ccmb.med.umich.edu/COTH/) [42], and the complex with the highest score was selected.

3.6. Principal component analysis

The goal of protein simulation is to generate enough configurations of the system of interest to identify functionally relevant movements. Principal component analysis (PCA) has been widely used to reduce the complexity of the data obtained from MD simulation trajectories by recovering the collective movement of atoms in the simulated trajectories that are essential for biological processes. PCA was performed with the data extracted from a 200 ns MD simulation. The projected eigenvectors and eigenvalues obtained during the first two principal components (PC1 and PC2) runs were calculated with the Carma program. The dPCA was calculated considering only the Cα of the protein.

4. Conclusion

HPV has been described as the principal cause for the development of cervical cancer, and this potential is due to the production of its oncoproteins. One of the most important is the E7 oncoprotein, which disrupts the cell cycle through direct interactions with regulatory proteins during this cellular process. This protein presents intratypical variants that have been epidemiologically related to the oncogenic potential of this virus, even there has been no report about the effect of these variations in the structure of the protein;

 doi:10.20944/preprints202012.0671.v1

however, these effects may be closely related to the oncogenic difference present in the variants.

Our findings show, for the first time, the effect that an amino acid change in a variant (N29S and H51N) changes the 3D structure of the E7 protein. These changes translate into increased or reduced stability of the 3D structure of the protein. In our systems, the H51N variant showed lower 3D structure stability than the other systems. Interestingly, the N29S mutation confers greater 3D structure stability compared to that of the H51N variant and the E7 reference; this outcome may be related to the fact that N29S is an additional phosphorylation site. Our results show that the stability that we observed in the MD simulation trajectories of N29S can be induced by this additional phosphorylation, which affects the global structure of the protein. Experimental studies have shown that the phosphorylation of the E7 protein is related to an increased half-life and a greater capacity for protein transformation [48, 49]. Thus, phosphorylation may explain why this protein showed greater more stability in the MD simulation trajectory. The mutation of asparagine 29 to serine in the N29S variant was reported to confer an additional phosphorylation site at S29 due to its proximity to the CKII recognition motif, which phosphorylates S31 and S32. It has been reported that S29 phosphorylation is involved in the increase in the number and size of colonies with respect to the E7 reference, which indicates an improvement in the oncogenic potential of this variant [50]. These conformational changes induced by mutations may explain the difference between their oncogenic potential. However, the effect of these structural changes on the interaction with the molecular targets of the E7 protein remains to be elucidated.

5. Supplementary Materials:

Figure S1. Superposition of the initial model with the average structure of the reference E7 protein and variants A) The amino acid sequence of HPV16 E7 (P03129) is aligned with the sequences of the N29S and H51N variants. Regions highlighted in orange and blue indicate the amino acid change corresponding to each variant. B) Initial model of E7 reference (green) and average structure of E7 reference. C) Initial model of N29S (orange) and average structure of N29S. D) Initial model of H51N (blue) and average structure of H51N.

List of abbreviations:

3D: Three Dimensional
CC: Cervical Cancer
HPV16: Human Papillomavirus Type 16
PDB: Protein Data Bank
MD: molecular dynamics.
HPV: Human Papillomavirus
Dihedralangles PCA: dPCA.

Funding: This work was partially supported by grants from Basic Science 2016 (288612) from CONACyT.

Acknowledgments: Bello-Rios C was a recipient of a CONACyT fellowship (No. 628901). Thanks PhD. Justin Lemkul for his generous advice on the use of the structure patch. The author wants to express their gratitude to LANCAD for the supercomputer time support. The MDS was performed in the Laboratory of Bioinformatics at FCQB-UAS, and in the Hybrid Cluster Xiuhcoatl, http://clusterhibrido.cinvestav.mx) at CINVESTAV-IPN, México.

Author Contributions: This work was carried out with the collaboration of all the authors. BIA, SM, OLGC, CBR were responsible for the research design. CBR and SM Carried out theoretical studies, results analysis and wrote the manuscript; and together with MALV, BIA, AALE and OLGC analyzed the data and wrote the manuscript. SM and CBR wrote the final version of the manuscript. All authors had read and approved the final manuscript.

Conflicts of Interest:"The authors declare no conflict of interest"

 doi:10.20944/preprints202012.0671.v1

Disclosure statement: The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript

References

1. Jemal, A.; Siegel, R.; Xu, J.; Ward, E., Cancer Statistics, 2010. *CA: A Cancer Journal for Clinicians* **2011,** 60, (5), 277-300.
2. Kudela, E.; Holubekova, V.; Farkasova, A.; Danko, J., Determination of malignant potential of cervical intraepithelial neoplasia. *Tumor Biology* **2015,** 37, (2), 1521-1525.
3. Burd, E. M., Human papillomavirus and cervical cancer. *Clinical microbiology reviews* **2003,** 16, (1), 1-17.
4. Woodman, C. B. J.; Collins, S. I.; Young, L. S., The natural history of cervical HPV infection: unresolved issues. *Nature Reviews Cancer* **2007,** 7, (1), 11-22.
5. Yugawa, T.; Kiyono, T., Molecular mechanisms of cervical carcinogenesis by high-risk human papillomaviruses: novel functions of E6 and E7 oncoproteins. *Reviews in Medical Virology* **2009,** 19, (2), 97-113.
6. Baseman, J. G.; Koutsky, L. A., The epidemiology of human papillomavirus infections. *Journal of Clinical Virology* **2005,** 32, 16-24.
7. Schiffman, M.; Doorbar, J.; Wentzensen, N.; de Sanjosé, S.; Fakhry, C.; Monk, B. J.; Stanley, M. A.; Franceschi, S., Carcinogenic human papillomavirus infection. *Nature Reviews Disease Primers* **2016,** 2, (1), 16086.
8. Shanmugasundaram, S.; You, J., Targeting Persistent Human Papillomavirus Infection. *Viruses* **2017,** 9, (8), 229.
9. Hoppe-Seyler, K.; Bossler, F.; Braun, J. A.; Herrmann, A. L.; Hoppe-Seyler, F., The HPV E6/E7 Oncogenes: Key Factors for Viral Carcinogenesis and Therapeutic Targets. *Trends in Microbiology* **2018,** 26, (2), 158-168.
10. Thomas, J. T.; Hubert, W. G.; Ruesch, M. N.; Laimins, L. A., Human papillomavirus type 31 oncoproteins E6 and E7 are required for the maintenance of episomes during the viral life cycle in normal human keratinocytes. *Proceedings of the National Academy of Sciences of the United States of America* **1999,** 96, (15), 8449-8454.
11. Flores, E. R.; Allen-Hoffmann, B. L.; Lee, D.; Lambert, P. F., The Human Papillomavirus Type 16 E7 Oncogene Is Required for the Productive Stage of the Viral Life Cycle. *Journal of Virology* **2000,** 74, (14), 6622.
12. Oh, S. T.; Longworth, M. S.; Laimins, L. A., Roles of the E6 and E7 proteins in the life cycle of low-risk human papillomavirus type 11. *Journal of virology* **2004,** 78, (5), 2620-2626.
13. McLaughlin-Drubin, M. E.; Münger, K., The human papillomavirus E7 oncoprotein. *Virology* **2009,** 384, (2), 335-344.
14. Liu, X.; Disbrow, G. L.; Yuan, H.; Tomaić, V.; Schlegel, R., Myc and Human Papillomavirus Type 16 E7 Genes Cooperate To Immortalize Human Keratinocytes. *Journal of Virology* **2007,** 81, (22), 12689.
15. Hellner, K.; Mar, J.; Fang, F.; Quackenbush, J.; Münger, K., HPV16 E7 oncogene expression in normal human epithelial cells causes molecular changes indicative of an epithelial to mesenchymal transition. *Virology* **2009,** 391, (1), 57-63.
16. Zhou, X.; Münger, K., Expression of the human papillomavirus type 16 E7 oncoprotein induces an autophagy-related process and sensitizes normal human keratinocytes to cell death in response to growth factor deprivation. *Virology* **2009,** 385, (1), 192-197.
17. Wang, N.; Zhan, T.; Ke, T.; Huang, X.; Ke, D.; Wang, Q.; Li, H., Increased expression of RRM2 by human papillomavirus E7 oncoprotein promotes angiogenesis in cervical cancer. *British journal of cancer* **2014,** 110, (4), 1034-1044.
18. Shankar, S.; Prasad, D.; Sanawar, R.; Das, A. V.; Pillai, M. R., TALEN based HPV-E7 editing triggers necrotic cell death in cervical cancer cells. *Scientific Reports* **2017,** 7, (1), 5500.
19. Phelps, W. C.; Münger, K.; Yee, C. L.; Barnes, J. A.; Howley, P. M., Structure-function analysis of the human papillomavirus type 16 E7 oncoprotein. *Journal of virology* **1992,** 66, (4), 2418-2427.
20. Ohlenschläger, O.; Seiboth, T.; Zengerling, H.; Briese, L.; Marchanka, A.; Ramachandran, R.; Baum, M.; Korbas, M.; Meyer-Klaucke, W.; Dürst, M.; Görlach, M., Solution structure of the partially folded high-risk human papilloma virus 45 oncoprotein E7. *Oncogene* **2006,** 25, (44), 5953-5959.
21. McLaughlin-Drubin, M. E.; Bromberg-White, J. L.; Meyers, C., The role of the human papillomavirus type 18 E7 oncoprotein during the complete viral life cycle. *Virology* **2005,** 338, (1), 61-68.
22. Liu X, C. A., Zhao K, Marmorstein R., Structure of the human Papillomavirus E7 oncoprotein and its mechanism for inactivation of the retinoblastoma tumor suppressor. . *J Biol Chem* **2006,** Jan 6;, 578-86.
23. Jansma, A. L.; Martinez-Yamout, M. A.; Liao, R.; Sun, P.; Dyson, H. J.; Wright, P. E., The high-risk HPV16 E7 oncoprotein mediates interaction between the transcriptional coactivator CBP and the retinoblastoma protein pRb. *Journal of molecular biology* **2014,** 426, (24), 4030-4048.
24. Barbosa, M. S.; Edmonds, C.; Fisher, C.; Schiller, J. T.; Lowy, D. R.; Vousden, K. H., The region of the HPV E7 oncoprotein homologous to adenovirus E1a and Sv40 large T antigen contains separate domains for Rb binding and casein kinase II phosphorylation. *The EMBO journal* **1990,** 9, (1), 153-160.
25. Zhang, B.; Chen, W.; Roman, A., The E7 proteins of low- and high-risk human papillomaviruses share the ability to target the pRB family member p130 for degradation. *Proceedings of the National Academy of Sciences of the United States of America* **2006,** 103, (2), 437.
26. Chellappan, S.; Kraus, V. B.; Kroger, B.; Munger, K.; Howley, P. M.; Phelps, W. C.; Nevins, J. R., Adenovirus E1A, simian virus 40 tumor antigen, and human papillomavirus E7 protein share the capacity to disrupt the interaction between transcription factor E2F and the retinoblastoma gene product. *Proceedings of the National Academy of Sciences* **1992,** 89, (10), 4549.
27. McIntyre, M. C.; Frattini, M. G.; Grossman, S. R.; Laimins, L. A., Human papillomavirus type 18 E7 protein requires intact Cys-X-X-Cys motifs for zinc binding, dimerization, and transformation but not for Rb binding. *Journal of Virology* **1993,** 67, (6), 3142.
28. Balsitis, S.; Dick, F.; Lee, D.; Farrell, L.; Hyde, R. K.; Griep, A. E.; Dyson, N.; Lambert, P. F., Examination of the pRb-Dependent and pRb-Independent Functions of E7 In Vivo. *Journal of Virology* **2005,** 79, (17), 11392.

29. Todorovic, B.; Massimi, P.; Hung, K.; Shaw, G. S.; Banks, L.; Mymryk, J. S., Systematic analysis of the amino acid residues of human papillomavirus type 16 E7 conserved region 3 involved in dimerization and transformation. *Journal of virology* **2011,** 85, (19), 10048-10057.
30. Berezutskaya, E.; Yu, B.; Morozov, A.; Raychaudhuri, P.; Bagchi, S., Differential regulation of the pocket domains of the retinoblastoma family proteins by the HPV16 E7 oncoprotein. *Cell Growth Differentiation* **1997,** 8, (12), 1277-1286.
31. Pang, C. L.; Toh, S. Y.; He, P.; Teissier, S.; Ben Khalifa, Y.; Xue, Y.; Thierry, F., A functional interaction of E7 with B-Myb-MuvB complex promotes acute cooperative transcriptional activation of both S- and M-phase genes. (129 c). *Oncogene* **2014,** 33, (31), 4039-4049.
32. Hatterschide, J.; Bohidar, A. E.; Grace, M.; Nulton, T. J.; Kim, H. W.; Windle, B.; Morgan, I. M.; Munger, K.; White, E. A., PTPN14 degradation by high-risk human papillomavirus E7 limits keratinocyte differentiation and contributes to HPV-mediated oncogenesis. *Proceedings of the National Academy of Sciences* **2019,** 116, (14), 7033.
33. Song, Y. S.; Kee, S. H.; Kim, J. W.; Park, N. H.; Kang, S. B.; Chang, W. H.; Lee, H. P., Major Sequence Variants in E7 Gene of Human Papillomavirus Type 16 from Cervical Cancerous and Noncancerous Lesions of Korean Women. *Gynecologic Oncology* **1997,** 66, (2), 275-281.
34. Pillai, M. R.; Hariharan, R.; Babu, J. M.; Lakshmi, S.; Chiplunkar, S. V.; Patkar, M.; Tongaonkar, H.; Dinshaw, K.; Jayshree, R. S.; Reddy, B. K. M.; Siddiqui, M.; Roychoudury, S.; Saha, B.; Abraham, P.; Gnanamony, M.; Peedicayil, A.; Subhashini, J.; Ram, T. S.; Dey, B.; Sharma, C.; Jain, S. K.; Singh, N., Molecular variants of HPV-16 associated with cervical cancer in Indian population. *International Journal of Cancer* **2009,** 125, (1), 91-103.
35. Park, J. S.; Shin, S.; Kim, E.-C.; Kim, J. E.; Kim, Y. B.; Oh, S.; Roh, E. Y.; Yoon, J. H., Association of human papillomavirus type 16 and its genetic variants with cervical lesion in Korea. *APMIS* **2016,** 124, (11), 950-957.
36. Ramas, V.; Mirazo, S.; Bonilla, S.; Ruchansky, D.; Arbiza, J., Analysis of human papillomavirus 16 E6, E7 genes and Long Control Region in cervical samples from Uruguayan women. *Gene* **2018,** 654, 103-109.
37. Chemes, L. B.; Glavina, J.; Alonso, L. G.; Marino-Buslje, C.; de Prat-Gay, G.; Sánchez, I. E., Sequence Evolution of the Intrinsically Disordered and Globular Domains of a Model Viral Oncoprotein. *PLOS ONE* **2012,** 7, (10), e47661.
38. Noval, M. G.; Gallo, M.; Perrone, S. n.; Salvay, A. G.; Chemes, L. B.; de Prat-Gay, G., Conformational Dissection of a Viral Intrinsically Disordered Domain Involved in Cellular Transformation. *PLOS ONE* **2013,** 8, (9), e72760.
39. Giarrè, M.; Caldeira, S.; Malanchi, I.; Ciccolini, F.; Leão, M. J.; Tommasino, M., Induction of pRb degradation by the human papillomavirus type 16 E7 protein is essential to efficiently overcome p16INK4a-imposed G1 cell cycle Arrest. *Journal of virology* **2001,** 75, (10), 4705-4712.
40. Thompson, J. D.; Gibson, T. J.; Plewniak, F. d. r.; Jeanmougin, F. o.; Higgins, D. G., The CLUSTAL_X Windows Interface: Flexible Strategies for Multiple Sequence Alignment Aided by Quality Analysis Tools. *Nucleic Acids Research* **1997,** 25, (24), 4876-4882.
41. Jones, D. T., Protein secondary structure prediction based on position-specific scoring matrices11Edited by G. Von Heijne. *Journal of Molecular Biology* **1999,** 292, (2), 195-202.
42. Mukherjee, S.; Zhang, Y., Protein-protein complex structure predictions by multimeric threading and template recombination. *Structure (London, England : 1993)* **2011,** 19, (7), 955-966.
43. Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **1996,** 14, (1), 33-38.
44. Jo, S.; Kim, T.; Iyer, V. G.; Im, W., CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* **2008,** 29, (11), 1859-1865.
45. Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B. K.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E., Scalable molecular dynamics on CPU and GPU architectures with NAMD. *The Journal of Chemical Physics* **2020,** 153, (4), 044130.
46. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D., CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* **2016,** 14, (1), 71-73.
47. Glykos, N. M., Software news and updates carma: A molecular dynamics analysis program. *Journal of Computational Chemistry* **2006,** 27, (14), 1765-1768.
48. Valdovinos-Torres, H.; Orozco-Morales, M.; Pedroza-Saavedra, A.; Padilla-Noriega, L.; Esquivel-Guadarrama, F.; Gutierrez-Xicotencatl, L., Different Isoforms of HPV-16 E7 Protein are Present in Cytoplasm and Nucleus. *Open Virol J* **2008,** 2, 15-23.
49. Nogueira, M. O.; Hošeka, T.; Calçadaa, E. O.; Castiglia, F.; Massimi, P.; Banks, L.; Felli, I. C.; Pierattelli, R., Monitoring HPV-16 E7 phosphorylation events. *Virology* **2017,** 503, 70-75.
50. Zine El Abidine, A.; Tomaić, V.; Bel Haj Rhouma, R.; Massimi, P.; Guizani, I.; Boubaker, S.; Ennaifer, E.; Banks, L., A naturally occurring variant of HPV-16 E7 exerts increased transforming activity through acquisition of an additional phospho-acceptor site. *Virology* **2017,** 500, 218-225.