A study on heparin in complex with L1 protein of ten high risk Human Papillomavirus: new structural insights based on in silico analysis

Jorge Félix Beltrán Lissabet

Department of Biochemistry, National Center for Scientific Research, Ave 25 and 158, P.O 6412, Cubanacán, Havana City, Cuba.
jorge.beltran@cnic.edu.cu

ABSTRACT

Human papillomavirus (HPV) is one of the most common causes of sexually transmitted diseases in both men and women around the world. It has been suggested that the heparan sulfate constitute the main receptor recognized by HPV, however its role during the interactions with other HPV is not very clear. BC, DE, EF, FG, and HI are the five loops displayed on the surface of the pentamer which can interact with heparin as an analogue for heparan sulfate. In this study, the L1 late protein of ten high risks HPV (HR HPV L1) corresponding to the types 31/33/35/39/45/51/52/56/58/59 were analyzed using in silico methods by looking for insights related to HR HPV L1-loops and heparin interactions. The five loop regions (BC, DE, EF, FG and HI) of the ten HR HPV L1 interact with heparin, where the interaction established between the BC-loop and heparin was found to be present in all the HR HPV L1 analyzed in this study as well as those reported in the scientist literature. Moreover it was found that lysine residues are involves in most of interactions and that the charge-charge and polar interactions are stabilizing the HR HPV L1-heparin interaction. The results obtained in this work with all the in silico methods; suggest that the heparin-binding site in the loops for all the HR HPV plays an important role during HR HPV infections, where BC-loop constitute the most required structure during the HR HPV L1-heparin interactions. The charge-charge and polar interactions are the main forces stabilizing the HR HPV L1-heparin.

KEYWORDS: in silico; heparin; HPV; L1, heparan sulfate

INTRODUCTION

Human papillomavirus (HPV) is one of the most common causes of sexually transmitted diseases in both men and women around the world, with prevalence rates varying with the studied population and geographical localization.(1) HPVs can be divided into low-risk and high-risk types according to the propensity of the lesions to evolve into malignancies.(2) Twelve HPVs (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59) are defined by the World Health Organization (WHO) as being high-risk cancer causing types, with additional types (68, 73) being recognized as ‘possibly’ cancer-causing.(3) The papillomavirus major capsid protein, L1, is a ~55 kD protein with the ability to spontaneously self-assemble into virus-like particles (VLPs). These VLPs present an exterior surface essentially indistinguishable from the native 60nm non-enveloped papillomavirus virion.(4, 5) BC, DE, EF, FG, and HI are the five loops displayed on the surface of the pentamer. These five loops constitute the part of the viral surface that can be recognized as conformational epitopes of HPV.(6) Host cell entry of HPV is initiated by binding of the virus particle to cell surface receptors. It has been suggested that virions bind initially to the basement membrane prior to transfer to the basal keratinocyte cell surface.(7, 8) Early work investigating the cell surface receptors found that HPVs bind to a widely expressed and evolutionary conserved cell surface receptor and that the interaction depends primarily on L1.(8-12) Glycosaminoglycans, especially heparan sulfate, were suggested as initial attachment receptors for HPV VLPs.(8, 13-15) Binding to these negatively charged polysaccharides is usually electrostatic and relatively nonspecific.(16) Despite considerable efforts, (13, 14, 17-21) the interactions between HPV and the heparan sulfate oligosaccharides that initiate infection are poorly understood.(22) To date, most of the work related to the HPV L1 protein and heparan sulfate interactions using heparin as an analogue for heparan sulfate (23) covering structural aspect are very limited.

Moreover, there is a wide controversy in the scientific literature about the role of heparan sulfate as receptor for HPV. For this reason the present study is focused on the discovery of potential binding sites on the loop structures of high risk HPV L1 protein (HR HPV L1) through in silico strategies, in order to provide a better understanding of the establishing of heparan sulfate-HPV L1 interactions.
MATERIALS AND METHODS

The protein sequences of each high risk HPV L1 protein (HR HPV L1) used in this study were downloaded from the NCBI protein sequence database with the accession numbers [GenBank: P17388.1, P06416.1, AGU90581.1, AAY86494.1, P26536.1, ACX32362.1, P36743.1, ACL12325.1 and AGU90696.1], corresponding to HPV types 31/33/39/45/51/52/56/58/59, respectively. Further, the crystal structures and protein sequences of HR HPV L1 of the HPV 16/18/35 were retrieved from the Protein Data Bank (PDB) with the PDB entry ID: [PDB: 1DZL, 2R5I, 2R5J], correspondingly. Subsequently, all these protein sequences were aligned using multiple sequence comparison by log-expectation (MUSCLE) (24, 25) in order to locate the regions corresponding to loops, according to Bishop et al. and Dasgupta et al. (6, 22). MUSCLE utilizes a 3-stage algorithm: (1) generate a progressive alignment; (2) increase the accuracy of the progressive alignment by reconstructing a tree with the Kimura matrix and the clustering method; (3) iterative refinement of progressive alignment.

Homology modeling of the HR HPV L1

Taking into account that all crystal structures of HR HPV L1 reported to date are truncated at the N- and C-terminal (6, 22, 26), the two terminal regions of each sequences retrieved from the NCBI protein sequence database were deleted in order to generate protein models truncated with high quality, based only on templates corresponding to all HPV L1 reported in PDB. Then, the sequences with both terminal regions deleted were submitted into Phyre2 (Protein Homology/analogy Recognition Engine V 2.0) server (27), which is completely automated and determine 3D model of proteins based on multiple templates from the PDB. Also, it applies the ab initio folding simulation for protein modeling of regions where no templates are identified.

Refinement of protein structure models and pentamer building

All HR HPV L1 modelled with Phyre2 were refined by molecular dynamics simulation using the GalaxyRefine web server by selecting the mild and aggressive relaxation refinement mode, which perform an intensive refinement of the whole protein and its loops (28). Subsequently, all the second models refined were selected and submitted to M-ZDOCK (Symmetric Multimer Docking) server, using the symmetry option of five to generate the HPV L1 pentamers (29).

Quality assessment of protein structure models

Structural evaluation and stereoechemical analyses were performed by different evaluation and validation tools before and after refinement. Backbone conformation was evaluated by analyzing the Psi/Phi Ramachandran plot obtained from PROCHECK and RAMPAGE (30, 31). Also the models were evaluated through ERRAT and PROSA web tool (32, 33).

Molecular docking

The molecular docking was performed using the program ClusPro for each monomer refined, including the native crystal structure of the HPV 35. For this process each monomer of HR HPV L1 and the heparin were chosen as receptor and ligand, respectively, by selecting the heparin mode of the program. ClusPro ranks the model solutions by cluster sizes, which were all analyzed looking for interaction among the loop regions of each monomer in complex with heparin (34). All the complexes (monomers bound to heparin) obtained in this study were visualized and analyzed using the Pymol Molecular Graphics software 1.7.0.1. Moreover, Pymol was used to observe the charge distributions of each pentamer and monomer determined with the tools mentioned above.

RESULTS AND DISCUSSIONS

Alignment

The alignment performed with the sequences retrieved from the NCBI protein sequence database and PDB, allowed the discrimination of the five loop regions (BC, DE, EF, FG and HI). Moreover, it was also observed that the N- and C-terminal sequences recovered from PDB align good with each other, which was taken into account as a benchmark in selection of amino acid regions of the sequences retrieved from NCBI for 3D modeling by removing the sequences not aligned in such extremes, since no templates based on HPV L1 are reported for these regions (Figure1). One of the main causes of removing these regions was due the C-terminal amino acids that are reported to bind DNA, which constituted a problem in the initial dockings because of these regions can also bind heparin and thus make difficult the study of the interaction established between HR HPV L1-loops and heparin (data not shown).
Figure 1. Sequence alignment of the L1 protein of twelve high risk Human Papillomavirus showing the portion of the N-and C-terminal region deleted, and the localization of the loops BC, DE, EF, FG, and HI according to Bishop et al and Dasgupta et al.
Homology modeling of the HR HPV L1

The 3D structures of the HR HPV L1 types 31/33/39/45/51/52/56/58/59 were constructed by homology modeling. The final results of the HR HPV L1 modelled in this work showed that in some cases among one and three templates were selected by Phyre2 for proteins modeling as shown in Table 1. The 100% of each model were modelled with a confidence of more than 90%. Further, no region on HPV L1 was modelled by ab initio folding simulation since all templates identified and used for models building by Phyre2 belongs to HPV L1 structures reported in PDB.

<table>
<thead>
<tr>
<th>HR HPV L1 proteins modeled</th>
<th>Code of templates used by Phyre2</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPV 33 L1</td>
<td>1DZL (HPV 16)</td>
</tr>
<tr>
<td>HPV 58 L1</td>
<td>2R5K (HPV 11) and 1DZL (HPV 16)</td>
</tr>
<tr>
<td>HPV 52 L1</td>
<td>2R5K (HPV 11) and 1DZL (HPV 16)</td>
</tr>
<tr>
<td>HPV 31 L1</td>
<td>1DZL (HPV 16)</td>
</tr>
<tr>
<td>HPV 39 L1</td>
<td>2R5I (HPV 18) and 1DZL (HPV 16)</td>
</tr>
<tr>
<td>HPV 45 L1</td>
<td>1DZL (HPV 16) and 2R5I (HPV 18)</td>
</tr>
<tr>
<td>HPV 51 L1</td>
<td>2R5K (HPV 11), 2R5I (HPV 18) and 1DZL (HPV 16)</td>
</tr>
<tr>
<td>HPV 56 L1</td>
<td>2R5K (HPV 11), 2R5I (HPV 18) and 1DZL (HPV 16)</td>
</tr>
</tbody>
</table>

HR HPV L1: High Risk Human Papillomavirus L1 protein

Refinement of protein structure models and quality assessment

All the second models refined by GalaxyRefine showed a good improvement of its quality as shown in Table 2. The stereochemical quality of the homology models of HR HPV L1 types 31/33/39/45/51/52/56/58/59 was analyzed using the PROCHECK and RAMPAGE programs. The Ramachandran plot obtained from RAMPAGE uses a new Cbeta measure and updated of it. Both Ramachandran plot analysis revealed that all these models have good stereochemical properties, even before refinement of it. Similarly, the results obtained by ERRAT and PROSA web tool during the assessment of the overall model quality also showed good scores. Generally, all these results are within the range of values typically found for native proteins.

<table>
<thead>
<tr>
<th>L1 proteins modeled</th>
<th>ERRAT-Overall quality factor</th>
<th>ProSA-web (Z-Score)-Overall model quality</th>
<th>RAMPAGE (Number of residues in favoured region)</th>
<th>PROCHECK (Number of residues in favoured region)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPV 31 L1</td>
<td>67.401/85.349</td>
<td>-6.35/-6.64</td>
<td>94.5%/97.4%</td>
<td>85.5%/93.5%</td>
</tr>
<tr>
<td>HPV 33 L1</td>
<td>66.517/81.860</td>
<td>-6.6/-6.55</td>
<td>94.7%/96.9%</td>
<td>86.8%/93.5%</td>
</tr>
<tr>
<td>HPV 39 L1</td>
<td>66.516/89.302</td>
<td>-6.76/-6.45</td>
<td>92.7%/97.4%</td>
<td>85.9%/92.8%</td>
</tr>
<tr>
<td>HPV 45 L1</td>
<td>72.727/84.944</td>
<td>-6.22/-6.1</td>
<td>93.7%/96.7%</td>
<td>84.1%/91.4%</td>
</tr>
<tr>
<td>HPV 51 L1</td>
<td>64.719/83.900</td>
<td>-6.41/-6.52</td>
<td>94.5%/97.4%</td>
<td>88.6%/92.5%</td>
</tr>
<tr>
<td>HPV 52 L1</td>
<td>67.627/90.389</td>
<td>-6.55/-6.54</td>
<td>95.0%/96.9%</td>
<td>85.9%/92.8%</td>
</tr>
<tr>
<td>HPV 56 L1</td>
<td>74.715/86.256</td>
<td>-6.41/-6.28</td>
<td>96.0%/97.6%</td>
<td>86.4%/93.7%</td>
</tr>
<tr>
<td>HPV 58 L1</td>
<td>65.991/87.298</td>
<td>-6.6/-6.78</td>
<td>95.1%/97.6%</td>
<td>87.5%/94.4%</td>
</tr>
<tr>
<td>HPV 59 L1</td>
<td>61.937/85.126</td>
<td>-5.85/-5.83</td>
<td>94.5%/96.5%</td>
<td>86.3%/92.1%</td>
</tr>
</tbody>
</table>

Before refinement: numerical values in black
After refinement: numerical values in **red**

**Pentamer building and molecular docking**

As a main reference in the discussion of the most results obtained in this work, the work done by Dasgupta and coworkers was taken into account (22).

All pentamers determined in this work show on its surface a variable charge distribution, suggesting different binding mode between pentamers and heparin, taking into account that charge-charge interaction for the formation of this complex have been reported (22). These pentamers have a central hole with a highly negative charge density which varies in size in the different HPV types (Figure 2), similar to those reported by Dasgupta and coworkers in their studies on the HPV 16 and HPV 18 pentamers in complex with heparin (22).

![Figure 2. Charge distributions on the surface of the pentamers. Positive charge (Blue) and Negative charge (Red) visualized with the Pymol Molecular Graphics software 1.7.0.1.](image)

As a previous study it was observed that heparin oligosaccharides bind to the monomers surface mainly through charge-charge and polar interactions (22). In this work, it was observed that heparin binding in the loops were located in monomers regions with a positive charge, which could be due to the negative charge of heparin. Moreover, polar interactions were observed (Figure 3).
Figure 3. Binding sites of the heparin in the HR HPV L1 monomers through charge-charge interactions. Positive charge (Blue) and Negative charge (Red) visualized with the Pymol Molecular Graphics software 1.7.0.1.

For the HR HPV L1 31/33/35/39/45/51/52/56/58/59, all heparin-binding sites are conformational for each monomer, consisting in residues scattered on more than one surface loop that include in some cases a combination of it. The binding modes for the heparin oligosaccharides in HR HPV L1 types are as shown in (Figure 4) and the residues participating in binding the heparin are as follow: HPV 31 (BC-loop: Lys 41, Lys 35, Asn 38), HPV 33 (BC-loop: Lys 46, Ser 32, Lys 34, Lys 40) (FG-loop Ser 269, Lys 247, Arg 244) (DE-loop Asn 106, Arg 126), HPV 35 (BC-loop Lys 40, Lys 35, Ser 38), (DE-loop Ser 120, Asn 119, Lys 115), (FG-loop Ser 269, Ser 267, Thr 264), HPV 39 (BC-loop Lys 32, Lys 45, Gln 41, Asn 36, Arg 39), HPV 45 (BC-loop Lys 46, Arg 32, Gln 42, Lys 41) (DE-loop Thr 118) (FG-loop Tyr 258, Arg 267, Thr 269, Ser 272) (HI-loop Thr 332, Lys 334, Thr 343), HPV 51 (BC-loop Arg 39, Lys 35, Ser 37, Thr 36) (FG-loop Tyr 272, Asn 242, Tyr 270, Arg 264) (DE-loop Thr 109, Ser 112, Asn 111, Asn 116, Arg 113), HPV 52 (BC-loop Lys 42, Lys 34), HPV 56 (BC-loop Asn 37, Lys 35, Lys 39, Asn 41) (FG-loop Tyr 271, Asn 242, Ser 269, Lys 258, Arg 263) (DE-loop Asn 116, Ser 112), HPV 58 (BC-loop Lys 40, Lys 34, Asn 37), HPV 59 (BC-loop Gln 50, Lys 45, Lys 32, Gln 41). It is worth highlighting that these binding sites may be structured between adjacent monomers which are forming pentamers as reported by Dasgupta and coworkers.
It was noted the contribution of the BC-loop regions to heparin binding, which is observed in all the HPV L1 monomers. Also it was observed that most of interactions involve the lysine residues, which can be due to the charge-charge interaction between positive charge of lysine and the negative charge of the sulfate groups of heparin. Although not all loops for each HR HPV types are involved in the interaction with heparin, it is remarkable the presence of the BC-loop in all interactions with heparin which could suggest an important common role of this structure during the HR HPV infections.

It is also important to note that the results determined on the docking analysis of the crystal structure of the L1 protein of HPV 35 are similar to those determined from the 3D structures predicted in this study, regarding the number of loops involved in the interaction with heparin, which can be taking into account as benchmark in the analysis of the interactions established among the HR HPV L1 determined by in silico methods and heparin.

CONCLUSION

The heparin-binding site in the loops for all the HR HPV reported in this work seems to plays an important role during HR HPV infections that involves heparan sulfate as receptor. Among all loops in the HR HPV L1, BC-loop constitute the most required structure during the HR HPV L1-heparin interactions. Lysine residues are mostly interacting with heparin in some loops, indicating an important role of this residue as stabilizer of such interaction. The charge-charge and polar interactions are the main forces stabilizing the HR HPV L1-heparin. The insight reported in this work related to heparin-HPV L1 interactions are very important for the understanding of HPV immunology and drug designing.

REFERENCES


