## Electronic Supplementary Information:

# The Critical Residues of Helix 5 for *in vitro* Pentamer Formation and Stability of Papillomavirus Capsid Protein, L1

**Shi Jin[a,b], Dong Pan[a], Xiao Zha[a,d], Xianghui Yu[c], Yuqing Wu[*a], Yongjiang Liu[e], Fei Yin[e], Xiaojiang S. Chen[*b]**

[a]*State Key Laboratory for Supramolecular Structure and Materials, Jilin University, No. 2699, Qianjin Street, Changchun, 130012, China. E-mail: yqwu@jlu.edu.cn.*

[b]*Molecular and Computational Biology/Chemistry/Norris Cancer Center, University of Southern California, Los Angeles, CA 90089, USA. E-mail: xiaojiac@usc.edu.*

[c]*The State Engineering Laboratory of AIDS Vaccine, Jilin University, No. 2699, Qianjin Street, Changchun, 130012, China.*

[d]*Sichuan Tumor Hospital & Institute, Chengdu, 610041, China.*

[e]*Beijing Health Guard Inc, Beijing, 100176, China.*

## 1. Materials and methods

### 1.1 Cloning and deletion constructs

The clones with deletion of N-terminal 4 residues (ΔN4) and C-terminal 29 residues (ΔC29) of HPV16 L1 (16L1ΔN4ΔC29, 16WT) and HPV18 L1 (18L1ΔN4ΔC29, 18WT) in pGEX-6P-1 were used as parent ones for generating the mutants of L464A, G465A, R466A, R466H, K467A, F468A, L469A of HPV16 L1 and R467H of HPV18 L1. In these vectors, L1 genes of the mutants were cloned downstream of glutathione-S-transferase (GST) to express as GST-L1 fusions protein in *E. coli*. The standard molecular cloning methods, including PCR amplification, enzyme digestion, DNA ligation, and transformation were used to obtain the mono-site mutants. All mutant clones were confirmed by the DNA sequencing of the entire coding sequence for this study.

### 1.2 Protein expression and purification

The protein expression and purification of wild-type (WT) L1 and all L1 mono-site mutants were carried out essentially as described previously[1,2]. Briefly, 0.2 mM IPTG was used to induce protein expression overnight at room temperature (RT). The cell was lysed by sonication in buffer L (pH 8.0, 50 mM Tris-HCl, 200 mM NaCl, 1 mM DTT, 1 mM EDTA, 10 mM PMSF). After centrifugation at 16,000 rpm for 40 min, the supernatant was loaded onto a glutathione-sepharose column to bind GST-L1 fusion proteins. Then the column was washed with 10× bed volumes buffer L to remove away the contaminating proteins. After that, the GST-L1 was cleaved on column by using PreScission™ Protease (PPase) (GE Healthcare Life Sciences) in an approximate ratio of 100 μg GST-L1 to 1 NIH unit of enzyme. The digestion was carried out at 4 ℃ overnight, and then, the L1 was eluted from the column by adding 20 mM DTT to prevent aggregation and assembly into virus-like particles (VLPs). After that, the L1 was purified by size-exclusion chromatography (SEC).

### 1.3 Size-exclusion chromatography (SEC)

For SEC analysis, sample was injected directly into the sample loop and onto a HiLoad 16/60 Superdex 200 pg column (GE Healthcare Life Sciences) at a flow rate of 1 ml/min in the buffer W (pH 8.0, 50 mM Tris-HCl, 200 mM NaCl) at 4 ℃. During the purification process, all the reagents were kept at 4 ℃ whenever possible. The protein peaks

in the fast protein liquid chromatography (FPLC) elution profile were detected by an UV monitor at a wavelength of 280 nm. The approximate L1 pentamer and monomer sizes of WT and the mono-site mutants were determined on the basis of a comparison to three standard proteins Ferritin (440 kDa), Albumin (66 kDa) and Ovalbumin (45 kDa).

**1.4 Dynamic light scattering (DLS) measurements**

The proteins solutions containing the purified proteins L1-M were filtered using a 0.22 μm filter (Millipore) to remove aggregates, then 1 ml L1 protein was added into a 4 ml cuvette (1.0×1.0 cm). The particle size analyzer (Malvern Zetasizer Nano ZS90) with the Dispersion Technology Software (DTS) (V6.01) was used for data collection and analysis. The graphs and statistical analysis for the L1-M proteins size in solution were reported.

**1.5 Circular dichroism (CD) spectroscopy**

The CD spectroscopic measurements were performed on a PMS 450 spectropolarimeter (Biologic, France). The buffer W containing target protein L1 (~5 μM) was used for the measurement of far-UV CD spectra, which was recorded in a 0.1 cm path length cuvette at 25 ℃. The CD spectra was obtained by averaging 10 wavelength scan from 190 to 260 nm, in 0.05 nm steps, with a signal averaging time of 1s and a bandwidth of 1 nm. Data was recorded in terms of mean residue ellipticity [θ] (deg · cm2/dmol) during measurement, and then it was calculated and reported for per mol of amide groups present by software Bio-Kine 32 V4.51.

**1.6 The UV cloud point-temperature determination of the thermal stability of proteins.**

The thermal stability of 16WT and the mutants were evaluated with a UV cloud point-temperature (UV CP-Temp) assay, which was carried out by using a UV-2450 spectrophotometer (Shimadzu, Japan) equipped with a DC-0506 LOW-CONSTANTTEMP BATH temperature control system (Hangping, China). The UV CP-Temp determination was performed by observing the ultraviolet absorption signals as the optical density at 350 nm of the solution of HPV L1. All samples were measured at a scan rate of 0.8 ℃/min with the scan temperature ranged from 25 to 70 ℃ to monitor the thermal denaturation and/or aggregation of HPV L1 in solution.

**1.7 The inhibition of HPV16 L1 pentamer formation by a 15-amino acid peptide containing h5.**

The ability of mono-site mutations in the h5 to disrupt L1 pentamer formation suggests a possibility that a peptide sequence containing h5 may disrupt/inhibit the pentamer formation. To substantiate this idea, we designed a peptide of 15 amino acids containing h5, with the sequence of PLGRKFLLQAGLKAK (referred as h5-pep), to test its effect on pentamer formation of HPV16 L1. Meanwhile, a peptide of 15 amino acids (referred as NC-pep) with similar physical properties of h5-pep, KIPKNFQRTVALKLA, was use as negative control. The peptide was synthesized by Shanghai Apeptide Co., Ltd by using the normal solid phase protocol, which was further assayed by Mass spectra and HPLC. Molecular weight (MWt) of h5-pep is 1640.06 as determined by mass spectra, which matches the calculated MWt of the peptide. The purity of the peptide is 99.11% as confirmed by HPLC (Kromasil 100-5C18 column). The expression and purification of 16WT was the same as described in **1.2**, and the process can be summarized as follows. After centrifugation, the supernatant of cell lysates (from 3 liter cell culture) was divided into three aliquots, and each was loaded onto three identical glutathione-sepharose columns, A, B and C, to bind GST-L1 fusion proteins, respectively. After washing with 10× bed volumes buffer L to remove contaminating proteins, 5ml buffer L containing 0.33mM h5-pep and NC-pep was added into column B and C incubated at 4 ℃ for 1h, respectively, while column A was incubated with 5ml buffer L without the peptide as the control at 4 ℃. The GST-L1 was then cleaved on column A, B and C by using the same amount of PPase in an approximate ratio of 100 μg GST-L1 to 1 NIH unit of enzyme. The digestion was carried out at 4 ℃ overnight, and then the L1 proteins were eluted from the three columns, respectively, for further SEC analysis.

## 2. Results

### 2.1 Mono-site mutation of the seven mutants of HPV 16 L1-h5.

The structure of HPV capsid protein L1 reveals a small alpha-helix near the C-terminal end, helix 5 (h5), which may play a role in maintaining the C-terminal fragment conformation by anchoring itself to the core structure. In an effort to understand the importance of h5 in L1 pentamer formation, we then performed the mono-site mutations on the highly conserved $^{464}$LGRKFL$^{469}$ to explore the structural and functional significance of particular amino acids within this primary sequences, and the results were list as follows.

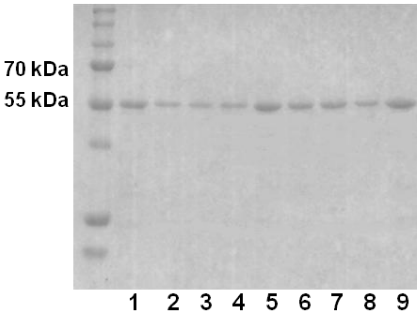


**Fig. S1** SDS-PAGE of 16WT and its mono-site mutants from the SEC. Lane 1: 16WT-P; lane 2: 16WT-M; lane 3: L464A-M; lane 4: G465A-M; lane 5: R466A-M; lane 6: R466H-M; lane 7: K467A-M; lane 8: F468A-M; lane 9: L469A-M. The purity of each L1 monomer was over 98%.

### 2.2 The DLS of the monomers of WT and its mutants.

The results of DLS provided information about the monomer sizes of 16WT and mono-site mutants. The column bar of the hydrodynamic particles diameter distributions for 16WT-M, L464A-M, G465A-M, R466A-M, R466H-M, K467A-M, F468A-M and L469A-M (Fig. S2) indicated the uniform in monomer size and the average size of which was 7.652, 7.736, 7.843, 7.475, 7.561, 7.674, 7.325 and 7.393 nm, respectively, being consistent with the previously reported size of L1 monomer[3].

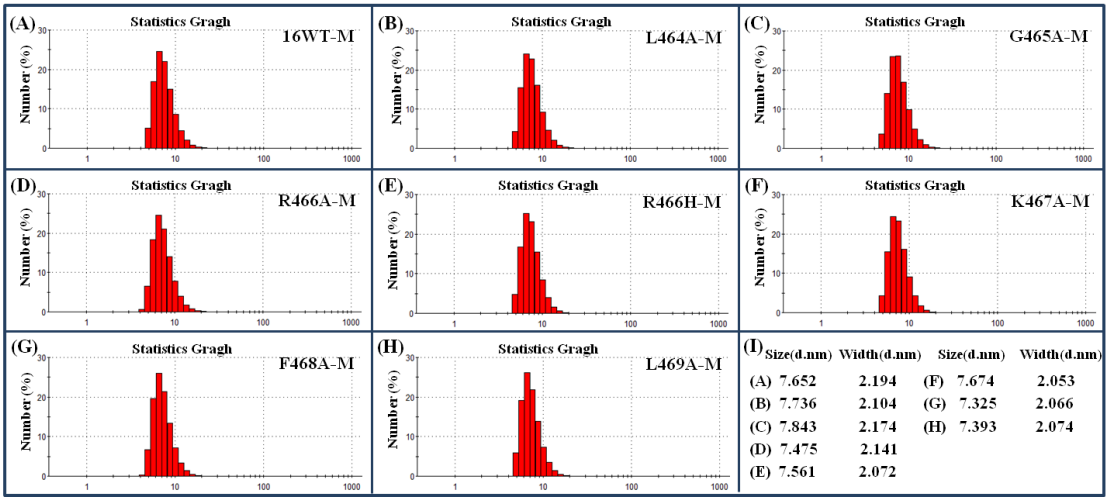


**Fig. S2** The hydrodynamic particles diameter distributions of the monomer of 16WT and its mono-site mutants were determined by dynamic light scattering (DLS) measurements. Comparison the column bar of (A) 16WT-M; (B) L464A-M; (C) G465A-M; (D) R466A-M; (E) (C) R466H-M; (F) K467A-M; (G) F468A-M; (H) L469A-M; (I) The statistic table of the monomer of 16WT and its mono-site mutants.

**2.3 The CD spectra of 16WT and its mutants.**

Far-UV CD spectra were measured to compare the secondary structure contents of the 16WT and six mono-site HPV16L1-h5 mutants (Fig. S3). Compare the far-UV CD spectra of 16WT-P and 16WT-M with the far-UV CD spectra of L464A-M, G465A-M, R466A-M and R466H-M, they exhibited different CD spectra, indicating they had the different the secondary structure, namely, the contents of α-helix, β-sheet, β-turn. The quantitative analysis data of the far-UV CD spectra (Table S1) revealed that the contents of α-helix, β-sheet, β-turn were 13.20%, 36.60%, 22.20% for 16WT-P and 15.30%, 32.10%, 21.40% for 16WT-M, respectively. The former ones were close to the known structural features of the protein as revealed by crystal of small VLPs[1] and L1 pentamers[4]. From the same quantitative analysis method, the contents of α-helix, β-sheet, β-turn for L464A-M were 14.70%, 33.10%, 23.50%, the contents of α-helix, β-sheet, β-turn for G465A-M were 14.30%, 33.80%, 23.10%, the contents of α-helix, β-sheet, β-turn for R466A-M were 13.80%, 35.90%, 22.60%, the contents of α-helix, β-sheet, β-turn for R466H-M were 13.50%, 35.10%, 23.90%. Therefore, these quantitative analysis data on the far-UV CD spectra indicated that the structures of these monomeric proteins were somewhat different from those of 16WT-P and 16WT-M.

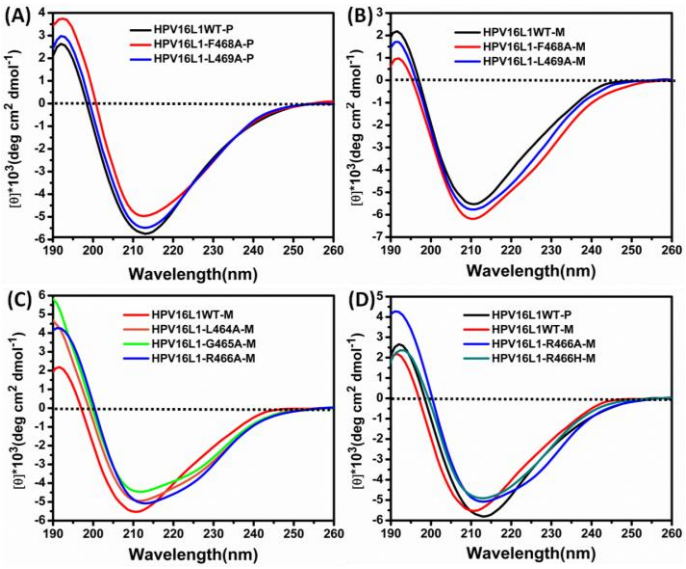


**Fig. S3** CD spectra and their comparisons of 16WT and its six mutants, after purification from SEC in buffer W. (A) Pentamers of 16WT, F468A and L469A; (B) Monomers of 16WT, F468A and L469A; (C) Monomers of 16WT, L464A, G465A and R466A; (D) Comparison of 16WT-P, 16WT-M, R466A-M and R466H-M.

**Table S1** The contents of secondary structure for 16WT and its four mutants.

| | The contents of secondary structures | | | |
|---|---|---|---|---|
| | α-helix | β-sheet | β-turn | Random coil |
| HPV16L1WT-P | 13.20% | 36.60% | 22.20% | 27.60% |
| HPV16L1WT-M | 15.30% | 32.10% | 21.40% | 30.60% |
| HPV16L1-L464A-M | 14.70% | 33.10% | 23.50% | 28.00% |
| HPV16L1-G465A-M | 14.30% | 33.80% | 23.10% | 27.60% |
| HPV16L1-R466A-M | 13.80% | 35.90% | 22.60% | 25.70% |
| HPV16L1-R466H-M | 13.50% | 35.10% | 23.90% | 27.80% |

**Table S2** UV Cloud point temperature (UV CP-Temp) ramping values for 16WT and its six mutants.

| | CP-Temp (℃) | | |
|---|---|---|---|
| | Pentamer | Monomer | |
| HPV16L1WT | 53.2 | — | 38.7 |
| HPV16L1-F468A | 53.2 | — | 38.6 |
| HPV16L1-L469A | 53.1 | — | 38.8 |
| HPV16L1-L464A | — | — | 38.2 |
| HPV16L1-G465A | — | 53.3 | 38.0 |
| HPV16L1-R466A | — | 51.4 | 38.6 |
| HPV16L1-R466H | — | 52.9 | 38.3 |

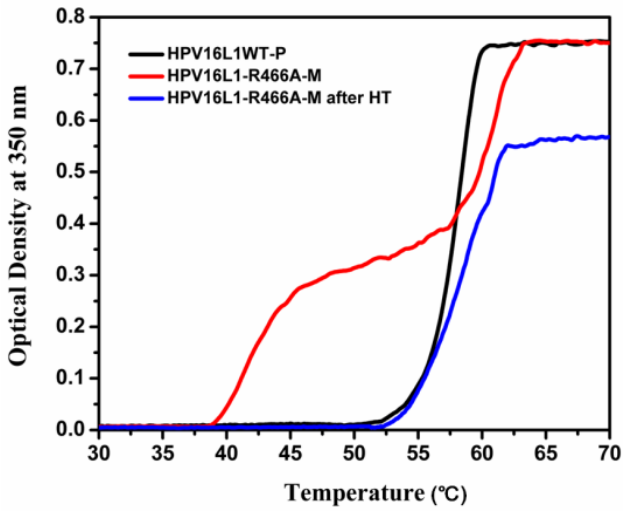


**Fig. S4** The thermal stability of 16WT-P and HPV16 L1 R466A-M before and after the heating treatment (HT), which was performed at 47 ℃ for 10 min, and then the samples were centrifuged at 11,000 rpm for 5 min to remove the precipitates of the less stable from of L1. After that, the more stable protein remained in solution, which was then examined by UV cloud point temperature ramping. The thermal treated R466A-M (blue line) showed a transition curve similar to that of 16WT-P (black line), suggesting that this fraction part of L1 had similar thermal stability with that of 16WT-P.

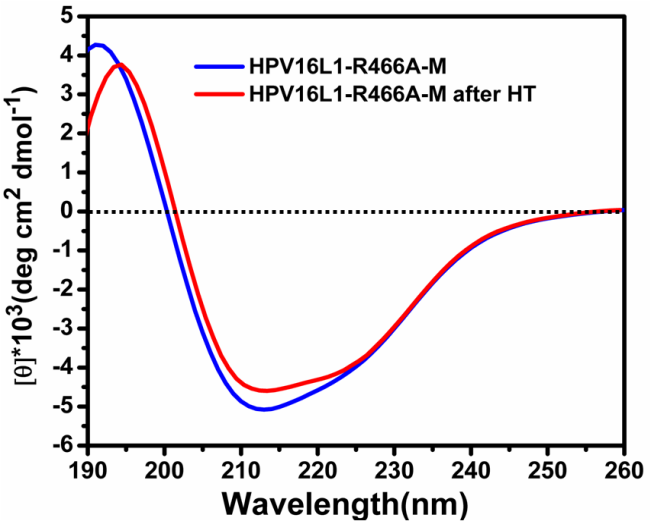


**Fig. S5** CD spectra of HPV16 L1 R466A-M before and after the heating treatment at 47 ℃. The procedure for the thermal treatment was same as that described for Fig. S4 in above.

**2.4 The R467H mutant of HPV18 L1 behaved similarly as R466H of HPV16 L1.**

To probe the significance of Arg residue in h5 for the L1 pentamer formation for a different HPV type, we constructed a corresponding mutation for HPV18 L1 that is another representative of the high-risk viruses most often associated with genital carcinoma. Since Arg466 of HPV16 L1 corresponds to the site of Arg467 in HPV18 L1, the R467H mutant of HPV18 L1 was constructed, expressed and purified using the identical procedure as that for HPV16 L1 R466H. The FPLC elution profile of it showed only the monomeric L1 peak, with an elution position of ~53 kDa (Fig. S6), which is very simililar to the result of HPV16 L1 R466H. The result suggested that the Arg residue in h5 played a critical role in pentamer formation for both HPV16 and HPV18 L1. This absolutely conserved Arg of h5 among the capsid protein L1 of all papillomavirus types suggests its common function in mediating pentamer formation.

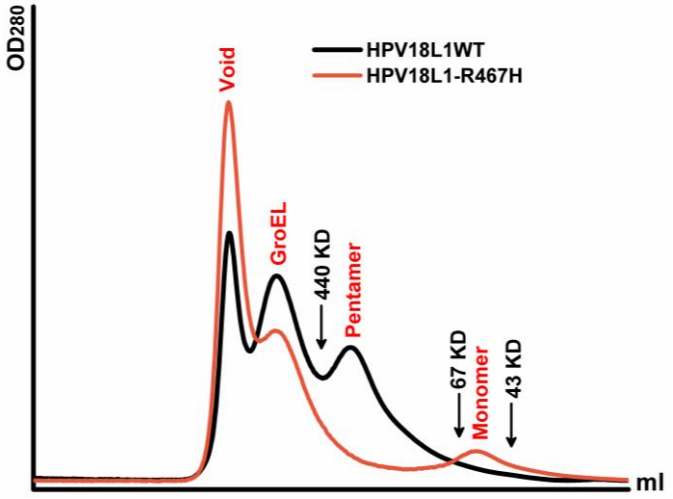


**Fig. S6** The FPLC elution profiles of 18WT and HPV18 L1 R467H mutant. The L1 mutant of HPV18 R467H showed similar result as the mutant of HPV16 R466H, no pentameric L1 peak, but only monomeric L1 peak was observed.

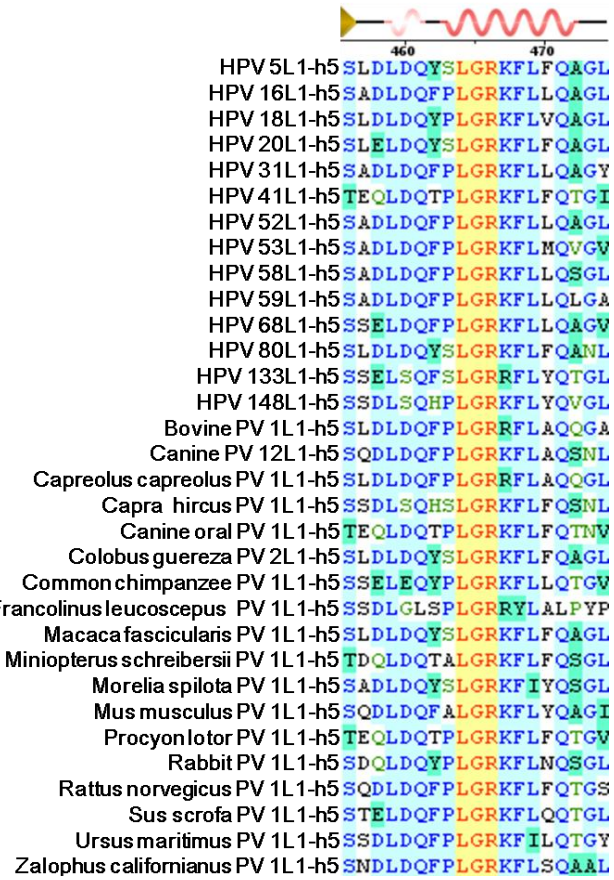


**Fig. S7** The h5 sequence alignment of the L1 proteins in 32 papillomavirus types from human and other species. The highlighted residues indicate absolutely conserved amino acids. These representatives show the high conservation in the h5 sequence.

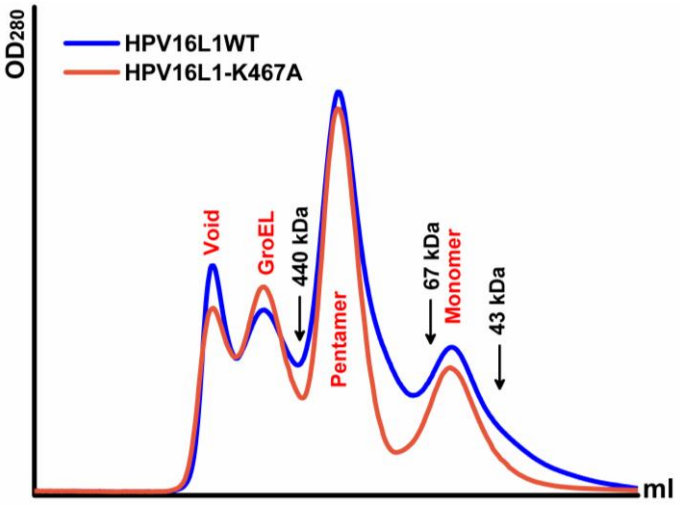


**Fig. S8** The FPLC elution profiles of 16WT and its K467A mutant. They displayed pentameric and monomeric elution peaks.

**2.5 A peptide of 15 amino acids containing h5 was used as an inhibitor of HPV16 L1 pentamer formation.**

After L1 proteins were incubated and eluted from glutathione-sepharose column in the absence and presence of 5ml buffer L containing 0.33mM h5-pep and NC-pep, respectively, they were analyzed further by size-exclusion chromatography (Fig. S9). In the presence of h5-pep (red line), the pentameric fraction peak was significantly reduced, accompanied with the increase of the monomeric L1. This result demonstrated the presence of h5-pep effectively inhibited the L1 pentamer formation. Whereas, adding the NC-pep into the L1 proteins solution had not change the chromatogram profile (olive line), indicating the NC-pep had no effect on L1 pentamer formation, which supported that h5-pep as a competitive and specific peptide for inhibiting L1 pentamer formation.

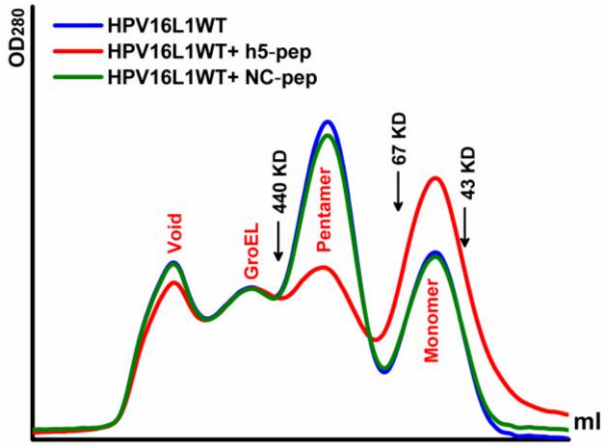


**Fig. S9** The FPLC elution profiles of HPV16 L1 in the absence and presence of h5-pep and NC-Pep, respectively, these two peptides have the similar physical properties.

## References

1. X. S. Chen, R. L. Garcea, I. Goldberg, G. Casini and S. C. Harrison, *Mol. Cell.*, 2000, **5**, 557.
2. X. S. Chen, G. Casini, S. C. Harrison and R. L.Garcea, *J. Mol. Biol.*, 2001, **307**, 173.
3. D. D. Zheng, D. Pan, X. Zha, Y. Q. Wu, C. L. Jiang and X. H. Yu, *Chem. Commun.*, 2013, **49**, 8546.
4. B. Bishop, J. Dasgupta, M. Klein, R. L. Garcea, N. D. Christensen, R. Zhao and X. S. Chen, *J. Biol. Chem.*, 2007, **282**, 31803.