Università degli Studi di Firenze

Facoltà di Scienze Matematiche Fisiche e Naturali

Corso di Laurea in Chimica

Espressione, purificazione e caratterizzazione iniziale della oncoproteina virale E7 del papilloma virus umano 16 (HPV16)

Expression, purification and initial characterization of the viral oncoprotein E7 from human papillomavirus 16 (HPV16)

RELATORE:

Prof.ssa Isabella Caterina Felli

CORRELATORE:

Prof.ssa Roberta Pierattelli

TESI DI LAUREA DI:

Daniele Mungianu

Anno Accademico 2011/2012

Sommary

1. Abbreviations

HPV = Human PapillomaVirus

IDP = Intrinsically Disordered Protein

CR = Conservative Region

NMR = Nuclear Magnetic Resonance

HSQC = Heteronuclear Single Quantum Resonance

OD = Optical Density

IPTG = Isopropyl β-D-1-thiogalactopyranoside

HEPES = 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

DLS = Dynamic Light Scattering

SDS = Sodium Dodecyl Sulfate

Expression, purification and initial characterization of the viral oncoprotein E7 from human papillomavirus 16 (HPV16)

2. Introduction

Intrinsically Disordered Proteins (IDPs) are proteins characterized by lack of tertiary structure, and this feature gives them functional advantages. IDPs play an important role for biochemical functions, including molecular recognition, signaling and regulation with implications in several human diseases[1]. Intrinsically disordered protein fragments, thanks to the largely extended and exposed backbones, display short aminoacid segments (short linear motives-SLIMs[2]) that may interact with different partners. This strategy of exploiting SLIMs to interact with protein partners has been proposed as a key one extensively used by viruses, which, in virtue of their small genomes, need more economic ways to interfere with the host cell than using folded proteins as they only can code a few ones[2]. Among viral proteins, we decided to focus on one that plays a key role in oncogenesis, E7 from human papilloma virus type 16 (HPV 16).

Human papilloma viruses are DNA viruses, found in many animals as well as in humans. They have double stranded DNA genome encoding eight or nine proteins[3;4]. In humans there are many HPV variants which are the causative agents of benign papillomas/warts and are risk factors for the development of carcinomas of the

genital tract, head and neck, and epidermis. The HPV infection spreads by direct contact, including sexual transmission in the case of genital infections or infection during birth in the case of juvenile laryngeal papillomas[5]. HPVs infect mucosal and cutaneous stratified squamous epithelia and are divided into high risk and low risk viruses based on their pathogenicity.

2.1. The viral oncoprotein E7 from HPV16

HPV16 is known to be the most carcinogenic viral type and E7 plays a key role on infection, interfering in the cell growth and transformation, gene transcription, apoptosis, DNA transcription, and among other process (Figure 1)[6]. E7 is a 98 amino acid acidic protein (12kDa) and shares sequence and functional features with adenovirus E1A protein and Simian Virus type 40 (SV40) T antigen[7]. Based on this similarity, E7 protein can be divided into three conservative regions: CR1 (amino acids 2-15), CR2 (amino acids 16-38) and CR3 (amino acids 39-98)[8]. Even though the structural and functional homology between E7 and E1A is restricted to the N-terminus side, both proteins share the ability to bind Zn^{2+} through the C-terminus. In E7 this peculiarity is due to the presence of two Cys-X-X-Cys motifs. The structure and length of E7 zinc binding domain is unusual and unique and is not shared by any cellular protein with zinc finger DNA-binding domains. Moreover, the zinc-binding domain seems to be responsible for the E7 ability to dimerize/oligomerize.

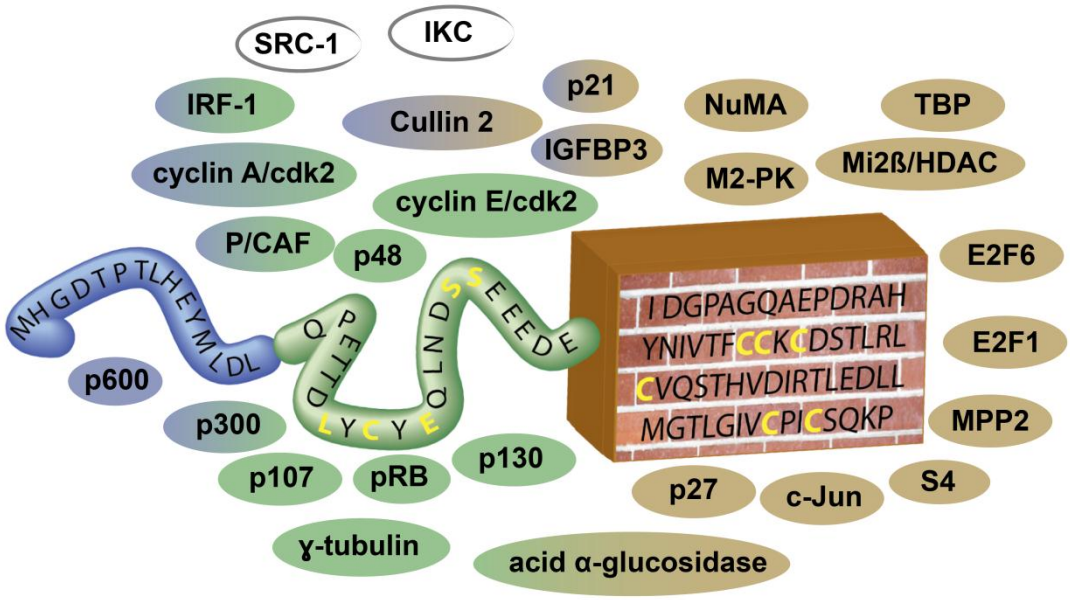

Figure 1 – A schematic cartoon describing E7 and the human proteins with which it interacts interfering in this way with many biological processes, such as cell growth and transformation, gene transcription, apoptosis, DNA transcription, and many other process. All the interactions are shown in the table in the appendix 1

2.2. Interactions and functionalities of E7

The most well characterized interactions of E7 are those established with the members of the retinoblastoma family of "pocket proteins"(a family of proteins that regulate the cell cycle): pRb[9], p107 and p130[10]. They play a crucial role in the cell cycle through the interaction with members of the E2F family of transcription factors.

E7 binds to pRb through a LXCXE motif, which represents the canonical binding site for "pocket proteins". This motif is shared by many cellular proteins that interact with retinoblastoma, and by the other viral oncoproteins able to functionally

inactivate pRb, such as adenovirus E1A and SV40 T-Antigen[10;13-15]. The LXCXE motif belongs to the CR2 region, thus it is included in the unstructured and highly flexible N-terminal moiety of E7. Therefore, this position might represent a way to keep the pRb-binding domain always accessible to pRb. The LXCXE motif is enough to bind with pRb, but several data suggest that other regions of the HPV16 E7 are involved with the pRb association. A peptide containing the LXCXE sequence is 100-fold less efficient than the full length E7 protein in inhibiting pRb in competitive binding inhibition assay[16]. Furthermore, it has been shown that E2F, does not contain an LXCXE motif, and E7 binding sites in pRb pocket domain do not overlap but bind in two different sites. Therefore, the pRb binding domain alone cannot explain either the association or the functional inactivation of pRb. From previous research, it is known that the C-terminus of E7 contains a low-affinity binding site for pRb[16;17]. Furthermore, it seems that this domain harbors binding sites for the "marked box" region of E2F as well[17]. In fact, the CR3 region of E7 is necessary in inducing E2F displacement, whereas the pRb binding motif alone is not capable of promoting the release of transactivation-competent E2F[16;17]. Previous studies have shown that E7 targets two other proteins that are functionally and structurally related to the pRb: p107 (retinoblastoma-like 1,RBL1) and p130 (RBL2)[10;18]. The information available suggests the importance of these interactions, even though they have not yet well characterized[10;18]. In comparison with pRb, p107

and p130 have not been conclusively linked to tumor suppressor activity and their contribution to E7 oncogenic potential is still debated.

2.3. Why study HPV16 E7?

As stared above, we know that HPVs are the causative agents of benign papillomas/warts and risk factors for the development of carcinomas[19]. The cancer due to HPV represents up to 15% of all cancer in women and up to 5% in men. In particular cervical cancer is the second most common cancer among women, accounting for one fifth of all their cancer deaths[20]. Initial detection of pre-cancerous lesions based on regular PAP-testing can significantly reduce cervical cancer death rates. However, not even effective population screening programs nor optimal cure are available in developing countries, where cervical cancer may represent the most prevalent one in women. HPV oncoproteins represent the most interesting candidate for a targeted cancer therapy as they are the primary contributors to the development of malignant transformation. Due to the E7 profound implication in the oncogenesis, the inhibition of the activity of this protein could be an effective approach to block the process of the malignant transformation. E7 has the unique advantage of being constitutively expressed only in HPV-positive cells and of being unrelated to the host proteins. Therefore E7 is a specific target, allowing selective targeting of only the pre-malignant and malignant cells, without any effect on normal cells[6].

The structural characterization of the C-terminal part of the protein, that contains a zing binding motif (two CXXC fragments separated by XX AA), has been achieved for short constructs comprising only the C-terminal part for HPV 1 variant through X-ray and for the HPV45 variant through NMR[17;21]. However the whole protein, also for the 1 and 45 variants of the virus, failed to give crystals and/or was not fully characterized through NMR. Many interesting biophysical studies have been done on E7 form HPV 16, both on the whole protein and on smaller constructs, providing interesting information[22]. Finally the structure of a short polypeptide containing the LXCXE motif bound to pRB is available providing valuable information on how this short fragment binds to pRB[23]. However this polypeptide only has 100 fold lower affinity for pRB respect to the whole protein. This, together with other evidence, thus suggests that the function of E7 actually derives not only from this fragment, the most well characterized one, but from the synergy between the different portions of the protein, responsible for being able to interfere with regulatory mechanisms of the cell. However as of today, despite attempts being done both with NMR and with X-ray, the two major techniques to access atomic resolution information on proteins, no atomic resolution data on E7 are available.

Therefore it would be extremely interesting and helpful to characterize at atomic resolution the structural and dynamic properties of E7 from HPV 16 as well as its interaction with key partners in order to understand the molecular basis of how E7 is

involved in cancer progression and develop some treatment to reduce cervical cancer death rates.

The most well suited technique to access high resolution information on highly dynamic protein is NMR. In order to plan a high resolution NMR study, it is important to express the protein double labeled (^{13}C,^{15}N) to be able to acquire the set of multidimensional NMR experiments necessary to perform sequence specific assignment of the protein, which is the aim of this work.

3. Materials and Methods

3.1 Expression and purification of HPV16 E7

Utilizing the available protocol for expression and purification of HPV16 E7, the protein went all in the inclusion bodies, with only a minimal trace in the soluble fraction when using minimal media with ^{13}C and ^{15}N for isotopic enrichment. The following parameters could be responsible for the delivery of the protein in the inclusion bodies: 1) OD (Optical Density), 2) Amount of IPTG, 3) Temperature of cell growth and 4) Time of cell harvest after induction. In the original method we used the following values: 0.6 OD, 1 mM IPTG, 17°C and overnight expression after induction. The complete protocol is described hereafter.

50 μL of pLysS BL21, a particular strain of *Escherichia coli* used for protein expression, are taken and placed in an eppendorf maintaining the cells in an ice bath. 2 μL of the plasmid solution are inserted directly into the cell stock. The coltures are left in ice for 30 minutes, operating in the cold room at 4°C.Then the sample is placed in a hot bath at 42°C for 1 minute. The samples are returned to ice for 2 minutes. Meanwhile chloramphenicol and ampicillin are added in a agar-agar plate. Then the solution containing the cells is spread over the Petri agar plates. The plates are closed with parafilm and put upside down in the heater at 37°C for overnight growth. An isolated colony is taken and put in 50mL of SOC minimal medium (the

media composition is shown in Appendix 2). The flask is put in the thermoshaker at 37°C for overnight growth. 5mL are taken and inserted in 500 mL of M9 minimal medium (which contains 100mg/L ampicillin, 34 mg/L chloramphenicol and 100 mM ZnCl). The cells are grown at 37°C with constant agitation until the optical density at 600 nm reaches 0.6 OD. The production of recombinant protein is induced by addition of IPTG (isopropyl-β-D-thiogalactoside) to a final concentration of 250 μL. Cells are harvested by centrifugation after 4 h from induction. The pellets are resuspended in lysis buffer (50 mM Tris-HCl, 8M urea, 10 mM DTT, pH 8.5) and disrupted by sonication and centrifuged at 12000 rpm (the composition of M9 and Silantes medium are shown in Appendix 2). The solution, containing the soluble fraction is removed and stocked. The pellets, containing the the insoluble fraction (the insoluble fraction refers to the protein) inside the inclusion bodies, are resuspended, destroyed with the homogeniser and centrifuged at 12000 rpm. The solutions containing the insoluble and the soluble fraction are united, put in one concentrator and centrifuged at 12000 rpm until 500 μL. The supernatants containing the soluble and insoluble protein fraction is loaded onto a metal chelate affinity chromatography column charged with Ni^{2+} and equilibrated with the lysis buffer. After washing out the unbound sample, the E7 protein is eluted with an imidazole gradient of 3 columns volume from 10 to 200 mM. The fraction containing E7 is loaded in a PD-10 desalting column and equilibrated with buffer A (10 mM HEPES,

50 mM KCl, 10 mM DTT). The sample is collected in the new buffer and concentrated until 500 µL using a centrifugal filter device.

All the purification steps are analysed through Bio-Rad Mini-PROTEAN Precast Gel containing 4-20% polyacrylamide and through SDS-PAGE under denaturant conditions, followed by staining with Coomassie brilliant blue. SDS-PAGE differs from the native gel because the presence of SDS, negatively charged, interact with constant ratio with each denaturated protein, thus the separation is by difference of molecular weight.

3.2 Dynamic Light Scattering (DLS)

Purified HPV16 E7 oncoprotein was analyzed by size exclusion chromatography connected to a multiangle light scattering (MALS) equipped with QELS module (quasielastic light scattering) for R_H measurements. 100 µl protein samples were loaded on a Superdex 200HR column (GE Healthcare) equilibrated in 10 mM HEPES, pH 7.5, 150mM KCl, with 10 µM $ZnCl_2$. A constant flow rate of 0.6 ml/min was applied. Elution profiles were detected by an Optilab rEX interferometric refractometer and a Dawn EOS multiangle laser light-scattering system at 690 nm (Wyatt Technology Corp.). Data acquisition and processing were carried out using ASTRA 5.3.4.20 software (Wyatt Technology). Determination of molecular masses and hydrodynamic radii are reported as mean values ± S.D. of duplicate experiments.

3.3 NMR

The ^{1}H detected NMR experiments were carried out at 298 K on 18.8 T Bruker AVANCE 800, operating at 800.13 MHz for ^{1}H, equipped with cryogenically cooled probe.

Protein concentrations of 50-600 μM in 10 mM HEPES buffer, containing 50 mM KCl, 10 mM DTT at pH 8.5 are used for NMR studies. 10% D_2O was added for the lock.

NMR spectra are performed using one-dimensional experiments (with excitation sculpting to suppress the intense water signal[24]) as well as 2D NMR experiments (SFHMQC[25], HSQC[26;27]) . The 1D experiments are set up with following parameters: 64 number of scans (ns), spectral width (sw) of 12 ppm, relaxation time (d1) of 1s and acquisition time (aq) of 300ms. The SFHMQC are set up with 16 scans per increment and using a recycle delay of 0,5s and the other parameters were:

- For the ^{1}H dimension: 12ppm sw, carrier 4,7ppm and 1024 number of points
- For the ^{15}N dimension: 25ppm sw, carrier 117,5ppm and 128 number of points

The HSQC experiments are set up with same parameters of SFHMQC but with a recycle delay of 1s.Spectra were calibrated using DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) and dioxane as reference.

Data were processed with TopSpin 2.0 and were analyzed with program CARA[28] and Sparky[29].

4. Results and discussion

4.1 Expression protocol to obtain ^{13}C, ^{15}N labeled HPV16 E7.

One of the key steps to proceed with the high resolution characterization of HPV16 E7 consists in optimizing the protocol for protein expression and purification in order to obtain a suitable quantity of isotopically enriched protein (in 500µL samples of 0.1 - 1 mM). With the available protocol, the protein was mainly delivered in inclusion bodies and thus very difficult to recover. Several conditions were screened to prevent the delivery of the protein in the inclusion bodies, still maintaining high levels of protein expression and thus high yield in the preparation of the sample.

The following conditions were screened:

- OD - 0.6, 1.0, 1.5. OD refers to the optical density of the *Escherichia coli* culture which gives an indication of the amount of the cells.
- Temperature - 17°C, 25°C, 37°C. Temperature refers to the heat of the culture during the protein expression.

- IPTG - 100μM, 500μM, 1mM. IPTG refers to the concentration of isopropyl-β-D-thiogalactoside that induce the expression of the HPV16 E7.
- Time of cell harvest after induction - 4h, 8h, 16,h.

Combining and discarding some of the them, we obtained several different trial conditions shown in the Table 1.

The plasmid was transformed into pLys BL21 *Escherichia coli* strain, the cells were grown overnight in agar-agar Petri dish at 37°C. An isolated colony was preinoculated in 50 mL of SOC media and the flask was put into a thermoshaker at 37°C for overnight growth. The expression culture was inoculated into 2 mL, for each condition, of two different media: M9 and Silantes (both containing 100mg/L ampicillin, 34 mg/L chloramphenicol and 100 mM ZnCl). The cells were grown until they reached the desired OD. The cells were induced with different amounts of IPTG and they were grown with different harvest times. The cells were harvested by centrifugation at 12000 rpm for 15 minutes.

The cell pellets were resuspended into lysis buffer (50mM NaH_2PO_4, 300 mM KCl, 10 mM DTT, 10 mM imidazole, pH 8.0) and they were sonicated.

Controls ^{15}N Only for M9	0.6	1mM	25	4h	A
			17	8h	B
	1.5		25	4h	C
IPTG, Temperature and Time tests		100µM	17	4h	D
			25		E
			37		F
			17	8h	G
			25		H
			37		I
		500µM	17	4h	J
			25		K
			37		L
			17	8h	M
			25		N
			37		O
		1mM	17	4h	P
			25		Q
			37		R
OD tests	1.0	1mM	17	8h	S
	1.5	1mM			T

Table 1 – Trial conditions used for test expression with Silantes and M9 media

The cell pellets were resuspended into lysis buffer (50mM NaH_2PO_4, 300 mM KCl, 10 mM DTT, 10 mM imidazole, pH 8.0) and they were sonicated. The supernatants containing the soluble fraction of the protein were loaded into native gel and compared against a control (expressed only in M9) labeled with ^{15}N and Invitrogen Marker AnykD, a marker containing IgM Hexamer (1236kDa), IgM Pentamer (1048kDa), Apoferritin band 1 (720kDa), Apoferritin band 2 (480kDa), B-phycoerythin (242kDa), Lactate Dehydrogenase (146kDa), Bovine serum albumin (66kDa) and Soybean trypsin inhibitor (20kDa). The cell pellets were resuspended into the lysis buffer and were loaded into another native gel and compared against the insoluble fraction of control and Invitrogen Marker AnykD.

Based on the gels we note that expression in soluble fraction of HPV16 E7 is better with 0.6 OD, between 100-500 μM IPTG, high temperature and 4h growth after induction (Column F and column L in Figure 2).

The experiment is repeated in 50 mL of both media (M9 and Silantes) to choose the conditions for the final preparation following these experimental conditions: 0.6 OD, 250 μM IPTG, 32°C and 4h expression after induction. In the end we obtain two samples in 500 μL: from M9 1.5 mM and from Silantes 750 μM, whereas concentrantions refer to E7. The pellets of HPV16 E7, coming from M9 expression, are resuspended into 25 mL of the lysis buffer (50 mM Tris-HCl, 8M urea, 10 mM

DTT, pH 8.5) and destroyed with the homogeniser. The solution is centrifuged at 12000 rpm until 500 µL.

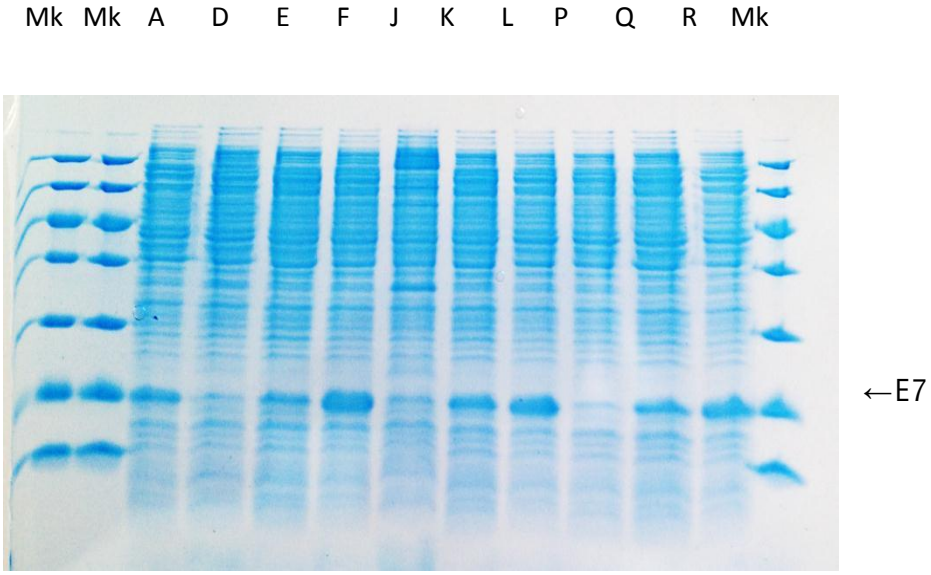

Figure 2 – Native gels of E7 from the soluble fraction expressed in M9: Native gels showing the analysis of test expression of HPV16 E7 in soluble fraction. The letters described the test condition in Table 1 for reference. In each gel also were loaded the marker (Mk) and the control (protein labeled with only ^{15}N) column A) were also loaded in the gel.

The supernatants are loaded onto a metal chelate affinity chromatography column charged with Ni^{2+} and equilibrated with the lysis buffer. After washing out the unbound sample, the E7 protein is eluted with an imidazole gradient of 3 columns volume from 10 to 200 mM. The fraction containing E7 is charged in a PD-10 desalting column and equilibrated with buffer A (10 mM HEPES, 50 mM KCl, 10 mM DTT). The samples are collected in the buffer and centrifuged into a concentrator at 12000 rpm until 500 µL. Comparing the HSQC (^{1}H, ^{15}N) NMR spectra of HPV16 E7 coming from the soluble fraction and from inclusion bodies (M9) we obtained the same results. This means that HPV16 E7 can be properly recovered also from inclusion bodies. From the optimized protocol of the HPV16 E7

expression in M9 we obtain a yield of approximately 40 g/L from the soluble fraction and 20 g/L from the insoluble fraction.

4.2. pH dependence

A second important aspect for the complete characterization of HPV16 E7 consists in improving the quality of the 2D NMR spectra, in particular the SNR (Signal-to-Noise Ratio) of the set of isolated, well dispersed cross peaks that are likely to derive for the CR3 region. Originally, after the purification, we put the sample in buffer A (10 mM HEPES, 50 mM KCl, 10 mM DTT, pH 7.5). The pH may have a high influence on the quality of the NMR spectra, so the following pH values are tested:

- 7.5 - 8.0 - 8.5 - 9.0 - 9.4 - 10.0 - 7.5 - 6.7

The direct addiction of NaOH/HCl may cause large local changes in pH and thus be dangerous for the protein. Therefore we exchange buffers by dialysis. We use a 500 µL dialysis device, put it in 0.5 L of the new buffer, and leave it shaking for 4h by magnetic agitation. The spectra at pH 7.5 are used as control to compare the improvements/worsening after each change. The spectra 2D (^{1}H,^{15}N) correlations are shown in Fig.3. The HPV16 E7 shows a pronounced pH dependence of the observed signals. Going to higher pH we lose the signal of the unfolded part of the HPV16 E7, due to the fast exchange of amide protons with the solvent in this region, improving the SNR of the folded part. At pH 10.0 almost all signals are lost. Going lower with

the pH (6.5) it is not possible to appreciate any meaningful changes respect to pH 7.5, our initial conditions. The best conditions, to see the CR3 region, are obtained at pH 8.5 while to see the IDP part are at pH 7.5.

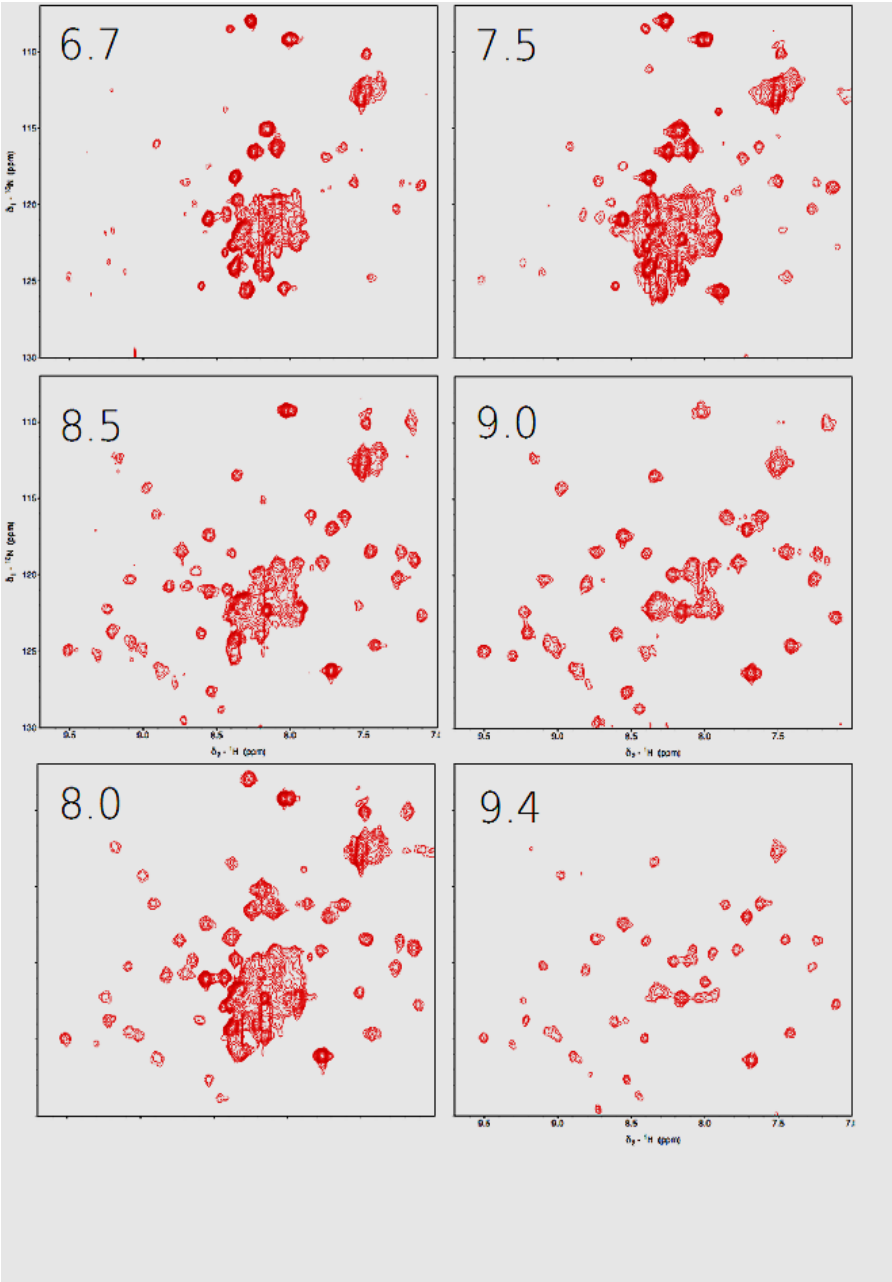

Figure 3 - Series of 2D ^{1}H-^{15}N SFHMQC spectra recorded at different pH. The set of weak well dispersed signals is most intense at pH 8.5. The best results for the set of strong signals clustered in the central crowded region are obtained at pH 7.5.

4.3 Existence of an oligomer

From the literature[22] it is known that HPV16 E7 has extended conformational equilibria in solution and when Zn^{2+} is removed by a chelating agent (as EDTA), HPV16 E7 self-assembles into spherical particles with an average molecular weight of 790 kDa called E7SOs (for E7 Spherical Oligomers).

We have recognized a similar behaviour of the protein to form a soluble molecular aggregate in solution (in presence of 10 mM DTT and without adding any reagents) analysing the protein with the native gels. On the gel (as shown in Fig.4) it is possible to see in the middle the band belonging to the protein and in the top a tiny spot, probably due to the oligomer (that doesn't run due to the huge molecular weight). In the SDS-PAGE (Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis) it is not possible to appreciate this because the action of SDS (a denaturing detergent) disrupts this oligomer as we have tested making a gradient from 1 to 25 mM of SDS on a native gel as shown in Fig.4.

After this information obtained from the gel, we have tried to load the sample into the Dynamic Light Scattering (DLS) to ascertain the presence of this molecular aggregate and, eventually which proportions of the free protein and of the oligomer we have in solution. Figure 5 shows that it is possible to appreciate the presence of a molecular aggregate of molecular weight about 2400 kDa and a peak due to the

protein of almost 44 kDa (so in this case it seems to indicate the presence of a tetramer of E7). Comparing the peaks due to the two species it seems that oligomer is in lower concentration compared to the free protein in solution.

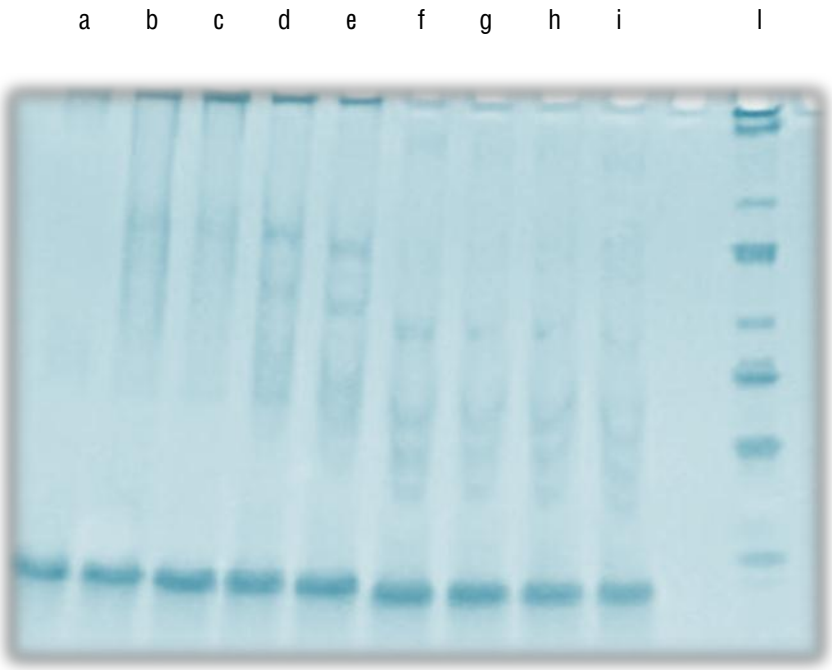

Figure 4 - Native gel of SDS gradient. The samples are placed in the following order: a) E7, b) E7+1mM SDS, c) E7+2mM SDS, d) E7+3mM SDS, e) E7+4mM SDS, f) E7+6mM SDS, g) E7+8mM SDS, h) E7+10mM SDS, i) E7+25mM SDS, l) Marker .

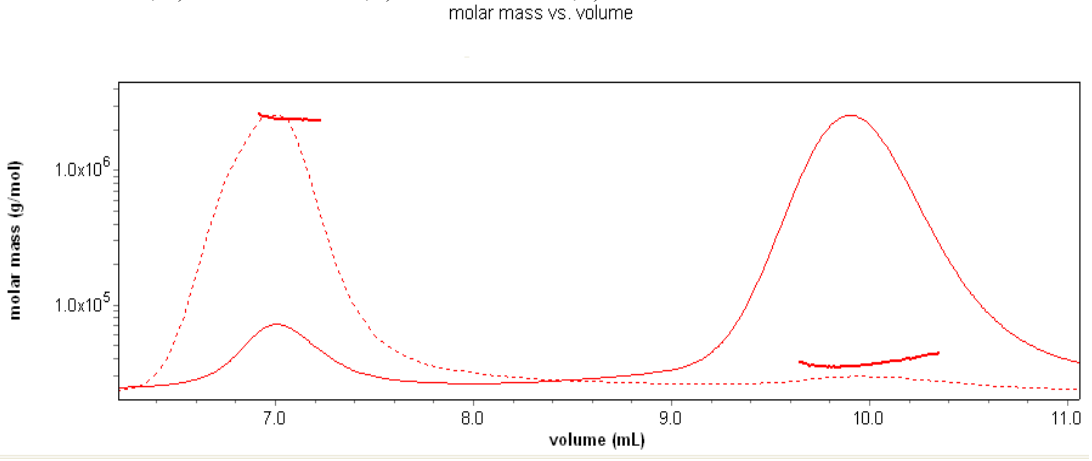

Figure 5 – Dynamic Light Scattering of E7. The report of DLS of viral oncoportein E7 show the presence of an oligomer (peak on the right) with a molecular mass about 2400 kDa and the tetramer (peak on the left) of almost 44 kDa

4.4 Concentration dependence

The existence of an oligomer could cause a line broadening of the NMR signals, depending on the exchange rates between various species present in solution and on their molecular size. For this reason we checked the effect of increasing the concentration of the protein on the intensity of the signals. We have done this test making a concentration gradient from 67 μM to 580 μM of HPV16 E7. We have selected few cross peaks coming from residues of the CR3 part and of the IDP part for the analysis of the protein (Fig.6a). We made the samples at different concentrations from a stock solution at 450 μM diluting and concentrating appropriately. From the analysis of the 2D ^{1}H-^{15}N SFHMQC spectra (Fig.6b) we noted that the intensity of the peaks, of the folded part grew linearly with the concentration but with a lower slope compared to the IDP part. This behaviour may be due to the formation of an oligomer that, exchanging with the free protein in solution, causes a line broadening of the NMR signals of the folded part (Fig.6b) thus lowering the apparent intensity of the peaks.

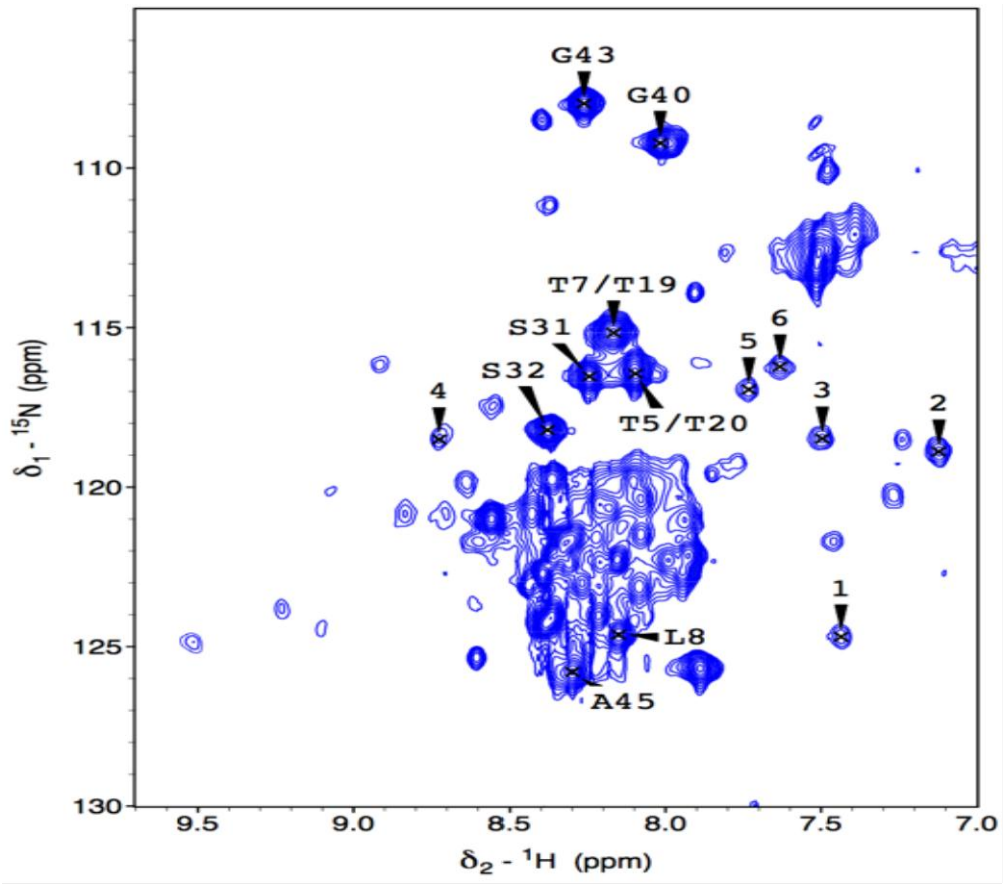

Figure 6a - 2D NMR acquired with SFHMQC (^{1}H,^{15}N) experiment. We have marked with arrows the peaks chosen for the comparison of the intensity at different concentrations.

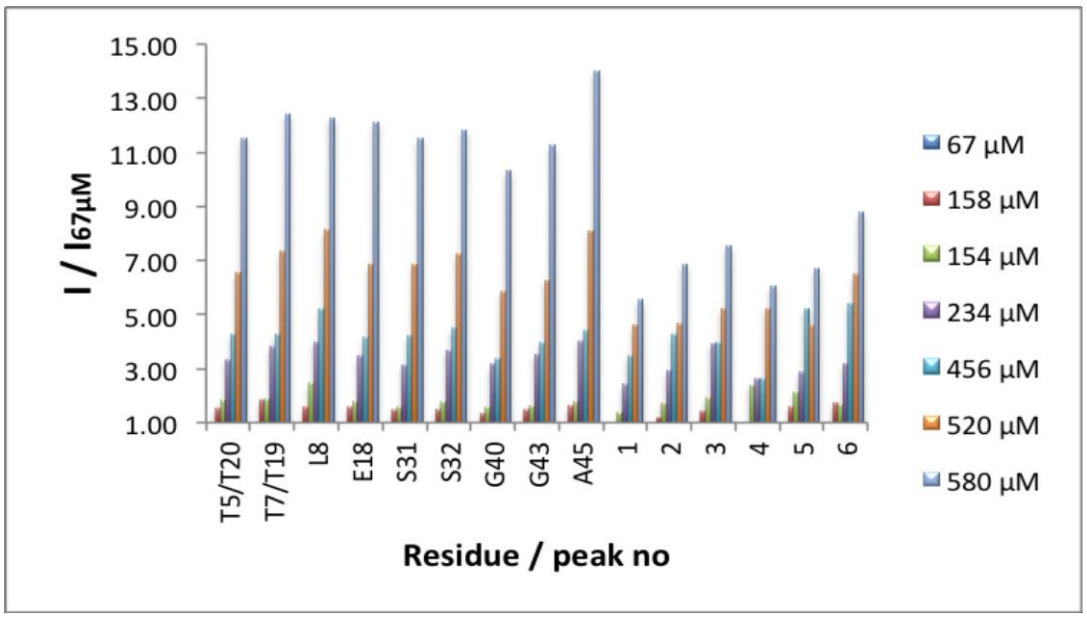

Figure 6b – Intensity of selected signals with increasing the concentration. Intensity of signals coming from the intrinsically disordered CR1 and CR2 region grows linearly with the increasing concentration. On the other hand, the intensity of the peaks coming from the structured CR3 region increase less with increasing concentration, which could be consequence of formation of bigger oligomers.

5. Conclusion

The protocol for expression and purification of E7 in ^{13}C-^{15}N minimal media was optimized and now enable us to obtain a very good yield of ^{13}C-^{15}N HPV16 E7 for high resolution NMR investigation of the protein. The pH and concentration dependence NMR studies revealed very heterogeneous structural and dynamic properties, well beyond the accepted, very schematic view of an assembly of a fully disordered module linked to a well structured one, presenting several extent of disorder which may be important for the function of the protein. In addition this NMR study provides an important platform for the investigation of the interaction of this protein, which seems to behave as a hub protein, with may partners. Concluding it is amazing how E7 from HPV16, with less than 100 amino acid, is able to interact with several key protein partners interfering with key cell regulatory processes. The large structural and dynamic heterogeneity monitored here through NMR may be an important aspect in this scenario and the current study opens the way to the atomic resolution characterization of these interactions that may provide important information for the understanding of the key features that promote progression of cells to malignant and possibly for the design of new drugs.

6. References

(1) Felli, I. C.; Pierattelli, R.; Tompa, P. Intrinsically disordered proteins; Bertini, I., McGreevy, K. S., Parigi, G., eds. Wiley-Blackwell: 2012; pp 137-152.

(2) Davey, N. E., Trave, G., and Gibson, T. J. How viruses hijack cell regulation. Trends Biochem.Sci. 36(3), 159-169. 2011.

(3) Munger, K. *Front Biosci.* 2002, *7*, 641-649.

(4) Munger, K.; Baldwin, A.; Edwards, K. M.; Hayakawa, H.; Nguyen, C. L.; Owens, M.; Grace, M.; Huh, K. *J.Virol.* 2004, *78*, 11451-11460.

(5) Uversky, V. N.; Roman, A.; Oldfield, C. J.; Dunker, A. K. *Journal of Proteome Research* 2006, *5*, 1829-1842.

(6) Donà, M. G. E7 Oncoprotein of Human Papillomavirus: Functions and Strategies of Inactivation for the Treatment of HPV-Associated Cancer; In *Oncogene Proteins: New Research*; Malloy, A. H., Carson, E. C., eds. 2008; pp 19-63.

(7) Phelps, W. C.; Yee, C. L.; Munger, K.; Howley, P. M. *Cell* 1988, *53*, 539-547.

(8) Gage, J. R.; Meyers,C.; Wettstein, F. O. The E7 proteins of the nononcogenic human papillomavirus type 6b (HPV-6b) and of the oncogenic HPV-16 differ in retinoblastoma protein binding and other properties. *J. Virol. 1990, 64, 723-730*.

(9) Dyson, N.; Howley, P. M.; Münger, K.; Harlow, E. *Science* 1989, *243*, 934-937.

(10) Dyson, N.; Guida, P.; Münger, K.; Harlow, E. *J Virol* 1992, *66*, 6893-6902.

(11) Harbour, J. W.; Dean, D. C. *Current opinion in cell biology* 2000, *12*, 685-689.

(12) Godefroy, N.; Lemaire, C.; Mignotte, B.; Vayssiere, J. L. *Apoptosis* 2006, *11*, 659-661.

(13) Dyson, N.; Guida, P.; McCall, C.; Harlow, E. *Journal of virology* 1992, *66*, 4606-4611.

(14) DeCaprio, J. A.; Ludlow, J. W.; Figge, J.; Shew, J. Y.; Huang, C. M.; Lee, W. H. *Cell* 1988, *54*, 275-283.

(15) Chen, S.; Paucha, E. *Journal of virology* 1990, *64*, 3350-3357.

(16) Patrick, D. R.; Oliff, A.; Heimbrook.D.C. *The Journal of Biological Chemistry* 1994, *269*, 6842-6850.

(17) Liu, X.; Clements, A.; Zhao, K.; Marmorstein, R. *J.Biol.Chem.* 2006, *281*, 578-586.

(18) Davies, R.; Hicks, R.; Crook, T.; Morris, J.; Vousden, K. *J Virol* 1993, *67*, 2521-2528.

(19) zur Hausen, H. *Proceedings of the Association of American Physicians* 1999, *111*, 581-587.

(20) Parkin, D. M.; Bray, F.; Ferlay, J.; Pisano, E. *CA Cancer J Clin* 2005, *55*, 74-108.

(21) Ohlenschlager, O.; Seiboth, T.; Zengerling, H.; Briese, L.; Marchanka, A.; Ramachandran, R.; Baum, M.; Korbas, M.; Meyer-Klaucke, W.; Dürst, M.; Görlach, M. *Oncogene* 2006, *25*, 5953-5959.

(22) Chemes, L. B.; Sanchez, I. E.; Alonso, L. G.; de Prat-Gay, G. Intrinsic disorder in the human papillomavirus E7 protein; In *Flexible Viruses: Structural Disorder in Viral Proteins*; Uversky, V., Longhi, S., eds. John Wiley & Sons, Inc.: 2012; pp 313-346.

(23) Xiao, B.; Spencer, J.; Clements, N.; Ali-Khan, N.; Mittnacht, S.; Broceno, C.; Burghammer, M.; Perrakis, A.; Marmorstein, R.; Gamblin, S. J. *Proc.Natl.Acad.Sci.* 2003, *100*, 859-865.

(24) Hwang, T.-J.; Shaka, A. J. *J Magn Reson A* 1995, *112*, 275-279.

(25) Schanda, P.; Kupce, E.; Brutscher, B. *J.Biomol.NMR* 2005, *33*, 199-211.

(26) Kay, L. E.; Keifer, P.; Saarinen, T. *J.Am.Chem.Soc.* 1992, *114*, 10663-10665.

(27) Schleucher, J.; Schwendinger, M.; Sattler, M.; Schmidt, P.; Schedletzky, O.; Glaser, S. J.; Sørensen, O. W.; Griesinger, C. *J.Biomol.NMR* 1994, *4*, 301-306.

(28) keller, R. L. J. The Computer Aided Resonance Assignment Tutorial. (1.3). 2004. Cantina Verlag.

(29) Goddard, T. D. and Kneller, D. G. SPARKY 3, University of Califofnia, San Francisco. 2000.

(30) Huh, K.; Zhou, X.; Hayakawa, H.; Cho, J. Y.; Libermann, T. A.; Jin, J.; Harper, J. W.; Münger, K. *J.Virol.* 2007, *81*, 9737-9747.

(31) Zerfass, K.; Schulze, A.; Spitkovsky, D.; Friedman, V.; Henglein, B.; Jansen-Dürr, P. *J Virol* 1995, *69*, 6389-6399.

(32) Nguyen, C. L.; Münger, K. *Virology* 2008, *380*, 21-25.

(33) He, W.; Staples, D.; Smith, C.; Fisher, Ch. *J Virol* 2003, *77*, 10566-10574.

(34) Funk, J. O.; Waga, S.; Harry, J. B.; Espling, E.; Stillman, B.; Galloway, D. A. *Genes Dev.* 2013, *11*, 2090-2100.

(35) Jones, D. L.; Alani, R. M.; Münger, K. *Genes Dev.* 1997, *11*, 2101-2111.

(36) Zerfass-Thome, K.; Zwerschke, W.; Mannhardt, B.; Tindle, R.; Botz, J. W.; Jansen-Dürr, P. *Oncogene* 2013, *13*, 2323-2350.

(37) Huh, K. W.; DeMasi, J.; Ogawa, H.; Nakatani, Y.; Howley, P. M.; Münger, K. *PNAS* 2005, *102*, 11492-11497.

(38) Barnard, P.; McMillan, N. A. J. *Virology* 1999, *259*, 305-313.

(39) Barnard, P.; Payne, E.; McMillan, N. A. J. *Virology* 2000, *277*, 411-419.

(40) Mannhardt, B.; Weinzimer, S. A.; Wagner, M.; Fiedler, M.; Cohen, P.; Jansen-Dürr, P.; Zwerschke, W. *Mol Cell Biol* 2000, *20*, 6483-6495.

(41) Massimi, P.; Pim, D.; Storey, A.; Banks, L. *Oncogene* 1996, *12*, 2325-2330.

(42) Massimi, P.; Pim, D.; Banks, L. *J Gen Virol* 1997, *78*, 2607-2613.

(43) Antinore, M. J.; Birrer, M. J.; Patel, D.; Nader, L.; McCance, D. J. *EMBO J* 1996, *15*, 1950-1960.

(44) Hwang, S. G.; Lee, D.; Kim, J.; Seo, T.; Choe, J. *J Biol.Chem.* 2002, *277*, 2923-2930.

(45) McLaughlin-Drubin, M. E.; Huh, K. W.; Münger, K. *J Virol* 2008, *82*, 8695-8705.

(46) Lüscher-Firzlaff, J. M.; Westendorf, J. M.; Zwicker, J.; Burkhardt, H.; Henriksson, M.; Müller, M.; Pirollet, F.; Lüscher, B. *Oncogene* 1999, *18*, 5620-5630.

(47) Zwerschke, W.; Mazurek, S.; Massimi, P.; Banks, L.; Eigenbrodt, E.; Jansen-Dürr, P. *PNAS* 1999, *96*, 1291-1296.

(48) Bernat, A.; Avvakumov, N.; Mymryk, J. S.; Banks, L. *Oncogene* 2003, *22*, 7871-7881.

(49) Huang, S. M.; McCance, D. J. *J Virol* 2002, *76*, 8710-8721.

(50) Brehm, A.; Nielsen, S. J.; Miska, E. A.; McCance, D. J.; Reid, J. L.; Bannister, A. J.; Kouzarides, T. *EMBO J* 1999, *18*, 2449-2458.

(51) Spitkovsky, D.; Hehner, S. P.; Hofmann, T. G.; Möller, A.; Schmitz, M. L. *J Biol.Chem.* 2002, *277*, 25576-25582.

(52) Baldwin, A.; Huh, K. W.; Münger, K. *J Virol* 2006, *80*, 6669-6677.

(53) Zwerschke, W.; Mannhardt, B.; Massimi, P.; Nauenburg, S.; Pim, D.; Nickel, W.; Banks, L.; Reuser, A. J.; Jansen-Dürr, P. *J Biol.Chem.* 2000, *275*, 9534-9541.

(54) Nguyen, C. L.; Eichwald, C.; Nibert, M. L.; Münger, K. *J Virol* 2007, *81*, 13533-13543.

(55) Nguyen, C. L.; Münger, K. *J Virol* 2009, *83*, 1700-1707.

(56) Barbosa, M. S.; Edmonds, C.; Fisher, C.; Schiller, J. T.; Lowy, D. R.; Vousden, K. H. *EMBO J.* 1990, *9*, 153-160.

(57) Berezutskaya, E.; Bagchi, S. *J Biol.Chem.* 1997, *272*, 20135-30140.

(58) Park, J.-S.; Kim, E.-J.; Kwon, H.-J.; Hwang, E.-S.; Namkoong, S.-E.; Um, S.-J. *J Biol.Chem.* 2000, *275*, 6764-6769.

7. Appendix 1

Table that shows interactions between HPV16 E7 and human proteins. For each interaction a relevant reference is reported. These references were used to build Figure 1.

Interaction partner	Interacting residues of HPV-16 E7	Reference
pRb	LxCxE (22-26)	[9]
Cullin 2	CR1 (mutants at H2 and P6-H9 inactive)	[30]
CDK2/cyclin A	Region 9-38	[31, 32, 33]
CDK2/cyclin E	CR2 (LxCxE & SS not required)	[31, 32, 33]
P21	C-domain (Region 1-39 very slight effect)	[34, 35]
P27	C-terminal domain	[36]
P600	CR1 (mutants at H2 and P6-E10 inactive)	[37]
P48	Region 17-37 (del. mutant 21-24 partial loss of activity)	[38, 39]
IGFBP3	C-terminal domain (79-83) + CR1 (mut H2P study)	[40]
TBP	C-terminal domain (79-83), binding enhanced by phophorilation at S31-S32	[41, 42]
c-Jun	C-terminal domain (C91G mutant study)	[43]
E2F1	C-terminal domain	[44]
E2F6	C-terminal domain (C91S mutant study)	[45]
MPP2	C-terminal domain	[46]
M2-PK	C-terminal domain (del. mutant 79-83	[47]

	inactive)	
P300	CR1 + CR2 required (C24G, S31R, S32R, H2P mutants)	[48]
p/caf	CR1 + CR2	[49]
Mi2ß/HDAC	C-domain	[50]
IKC	No mutantion study, interacting residues not determined	[51]
SRC-1	Many binding sites (study of 11 mutants, all active)	[52]
Acid α-Glucosidase	CR2 (LxCxE) + C-domain required (mut C24G nearly abolished interaction)	[53]
γ – tubulin	Crucial region 21-24 (mut E26G no effect)	[54]
NuMa	C-domain	[55]
CKII	S31 & S32 (phosphorylates these 2 serines)	[56]
S4	C-domain (mut C91S nearly no interaction)	[57]
IRF-1	CR1 + CR2 (del. mutants 6-10 and 21-24 study), for successful inactivation needs HDAC bound to CR3	[58]
P107	CR2 (important LxCxE + phosphorylation)	[10, 18]
P130	CR2 (LxCxE)	[10, 18]
Histon H1 kinase	Not specified	[18]

Reference List

1. Dyson,N., Howley,P.M., Münger,K. & Harlow,E. The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. *Science* 243, 934-937 (1989).

2. Huh,K. *et al.* Human papillomavirus type 16 E7 oncoprotein associates with the cullin 2 ubiquitin ligase complex, which contributes to degradation of the retinoblastoma tumor suppressor. *J. Virol.* 81, 9737-9747 (2007).

3. Zerfass,K. *et al.* Sequential activation of cyclin E and cyclin A gene expression by human papillomavirus type 16 E7 through sequences necessary for transformation. *J Virol* 69, 6389-6399 (1995).

4. Nguyen,C.L. & Münger,K. Direct association of the HPV16 E7 oncoprotein with cyclin A/CDK2 and cyclin E/CDK2 complexes. *Virology* 380, 21-25 (2008).

5. He,W., Staples,D., Smith,C. & Fisher,Ch. Direct activation of cyclin-dependent kinase 2 by human papillomavirus E7. *J Virol* 77, 10566-10574 (2003).

6. Funk,J.O. *et al.* Inhibition of CDK activity and PCNA-dependent DNA replication by p21 is blocked by interaction with the HPV-16 E7 oncoprotein. *Genes Dev.* 11, 2090-2100 (2013).

7. Jones,D.L., Alani,R.M. & Münger,K. The human papillomavirus E7 oncoprotein can uncouple cellular differentiation and proliferation in human keratinocytes by abrogating p21Cip1-mediated inhibition of cdk2. *Genes Dev.* 11, 2101-2111 (1997).

8. Zerfass-Thome,K. *et al.* Inactivation of the cdk inhibitor p27KIP1 by the human papillomavirus type 16 E7 oncoprotein. *Oncogene* 13, 2323-2350 (2013).

9. Huh,K.W. *et al.* Association of the human papillomavirus type 16 E7 oncoprotein with the 600-kDa retinoblastoma protein-associated factor, p600. *PNAS* 102, 11492-11497 (2005).

10. Barnard,P. & McMillan,N.A.J. The Human Papillomavirus E7 Oncoprotein Abrogates Signaling Mediated by Interferon-alpha. *Virology* 259, 305-313 (1999).

11. Barnard,P., Payne,E. & McMillan,N.A.J. The Human Papillomavirus E7 Protein Is Able to Inhibit the Antiviral and Anti-growth Functions of Interferon-alpha. *Virology* 277, 411-419 (2000).

12. Mannhardt,B. *et al.* Human Papillomavirus Type 16 E7 Oncoprotein Binds and Inactivates Growth-Inhibitory Insulin-Like Growth Factor Binding Protein 3. *Mol Cell Biol* 20, 6483-6495 (2000).

13. Massimi,P., Pim,D., Storey,A. & Banks,L. HPV-16 E7 and adenovirus E1a complex formation with TATA box binding protein is enhanced by casein kinase II phosphorylation. *Oncogene* 12, 2325-2330 (1996).

14. Massimi,P., Pim,D. & Banks,L. Human papillomavirus type 16 E7 binds to the conserved carboxy-terminal region of the TATA box binding protein and this contributes to E7 transforming activity. *J Gen Virol* 78, 2607-2613 (1997).

15. Antinore,M.J., Birrer,M.J., Patel,D., Nader,L. & McCance,D.J. The human papillomavirus type 16 E7 gene product interacts with and trans-activates the AP1 family of transcription factors. *EMBO J* 15, 1950-1960 (1996).

16. Hwang,S.G., Lee,D., Kim,J., Seo,T. & Choe,J. Human Papillomavirus Type 16 E7 Binds to E2F1 and Activates E2F1-driven Transcription in a Retinoblastoma Protein-independent Manner. *J Biol. Chem.* 277, 2923-2930 (2002).

17. McLaughlin-Drubin,M.E., Huh,K.W. & Münger,K. Human Papillomavirus Type 16 E7 Oncoprotein Associates with E2F6. *J Virol* 82, 8695-8705 (2008).

18. Lüscher-Firzlaff,J.M. *et al.* Interaction of the fork head domain transcription factor MPP2 with the human papilloma virus 16 E7 protein: enhancement of transformation and transactivation. *Oncogene* 18, 5620-5630 (1999).

19. Zwerschke,W. *et al.* Modulation of type M2 pyruvate kinase activity by the human papillomavirus type 16 E7 oncoprotein. *PNAS* 96, 1291-1296 (1999).

20. Bernat,A., Avvakumov,N., Mymryk,J.S. & Banks,L. Interaction between the HPV E7 oncoprotein and the transcriptional coactivator p300. *Oncogene* 22, 7871-7881 (2003).

21. Huang,S.M. & McCance,D.J. Down Regulation of the Interleukin-8 Promoter by Human Papillomavirus Type 16 E6 and E7 through Effects on CREB Binding Protein/p300 and P/CAF. *J Virol* 76, 8710-8721 (2002).

22. Brehm,A. *et al.* The E7 oncoprotein associates with Mi2 and histone deacetylase activity to promote cell growth. *EMBO J* 18, 2449-2458 (1999).

23. Spitkovsky,D., Hehner,S.P., Hofmann,T.G., Möller,A. & Schmitz,M.L. The Human Papillomavirus Oncoprotein E7 Attenuates NF-kB Activation by Targeting the IkB Kinase Complex. *J Biol. Chem.* 277, 25576-25582 (2002).

24. Baldwin,A., Huh,K.W. & Münger,K. Human Papillomavirus E7 Oncoprotein Dysregulates Steroid Receptor Coactivator 1 Localization and Function. *J Virol* 80, 6669-6677 (2006).

25. Zwerschke,W. *et al.* Allosteric Activation of Acid alpha-Glucosidase by the Human Papillomavirus E7 Protein. *J Biol. Chem.* 275, 9534-9541 (2000).

26. Nguyen,C.L., Eichwald,C., Nibert,M.L. & Münger,K. Human Papillomavirus Type 16 E7 Oncoprotein Associates with the Centrosomal Component gamma-Tubulin. *J Virol* 81 , 13533-13543 (2007).

27. Nguyen,C.L. & Münger,K. Human Papillomavirus E7 Protein Deregulates Mitosis via an Association with Nuclear Mitotic Apparatus Protein 1. *J Virol* 83, 1700-1707 (2009).

28. Barbosa,M.S. *et al.* 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. *EMBO J.* 9, 153-160 (1990).

29. Berezutskaya,E. & Bagchi,S. The Human Papillomavirus E7 Oncoprotein Functionally Interacts with the S4 Subunit of the 26 S Proteasome. *J Biol. Chem.* 272, 20135-30140 (1997).

30. Park,J.-S. *et al.* Inactivation of Interferon Regulatory Factor-1 Tumor Suppressor Protein by HPV E7 Oncoprotein. *J Biol. Chem.* 275, 6764-6769 (2000).

31. Dyson,N., Guida,P., Münger,K. & Harlow,E. Homologous sequences in adenovirus E1A and human papillomavirus E7 proteins mediate interaction with the same set of cellular proteins. *J Virol* 66, 6893-6902 (1992).

32. Davies,R., Hicks,R., Crook,T., Morris,J. & Vousden,K. Human papillomavirus type 16 E7 associates with a histone H1 kinase and with p107 through sequences necessary for transformation. *J Virol* 67, 2521-2528 (1993).

8. Appendix 2

The composition of the media used for the expression protocol are described hereafeter.

M9 medium is made with: KH_2PO_4 22mM, $Na_2HPO_4.7H_2O$ 48,5mM, NaCl 8,5mM. For our expression it is also addicted $MgSO_4$ 2mM, $CaCl_2$ 0,2mM, $ZnCl_2$ 0,1mM, Thiamine 3mM, d-Biotina 4mM, Ampicillin 0,27mM, Chloramphenicol 1,05mM.

SOC (Super Optimal Broth) medium is made with: 2% w/v bacto-tryptone, 0.5% w/v Yeast extract, 10mM NaCl (0.584 g), 2.5mM KCl, 20mM $MgSO_4$, 20mM glucose and H_2O to 1000 mL.

Silantes medium is a labeled commercial medium. Silantes OD-Medium is made from bacterial hydrolysate supplemented with M9 salts. The bacterial strain used is a chemolithoautotrophic organism which grows on isotopically labeled H_2, O_2 and CO_2.