Degenerate cysteine patterns mediate two redox sensing mechanisms in the papillomavirus E7 oncoprotein

CrossMark

Gabriela Camporeale[a], Juan R. Lorenzo[b], Maria G. Thomas[c], Edgardo Salvatierra[d], Silvia S. Borkosky[a], Marikena G. Risso[a], Ignacio E. Sánchez[e], Gonzalo de Prat Gay[a,*], Leonardo G. Alonso[a,*]

[a] Protein Structure-Function and Engineering Laboratory, Fundación Instituto Leloir and IIBBA-CONICET, Buenos Aires, Argentina
[b] ULB-Neuroscience Institute, Universite Libre de Bruxelles, Bruxelles, Belgium
[c] RNA Cell Biology Laboratory, Fundación Instituto Leloir and IIBBA-CONICET, Buenos Aires, Argentina
[d] Laboratory of Molecular and Cellular Therapy, Fundación Instituto Leloir-CONICET and IIBBA-CONICET, Buenos Aires, Argentina
[e] Protein Physiology Laboratory, Universidad de Buenos Aires, CONICET, Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales (IQUIBICEN), Facultad de Ciencias Exactas y Naturales, Buenos Aires, Argentina

A B S T R A C T

Infection with oncogenic human papillomavirus induces deregulation of cellular redox homeostasis. Virus replication and papillomavirus-induced cell transformation require persistent expression of viral oncoproteins E7 and E6 that must retain their functionality in a persistent oxidative environment. Here, we dissected the molecular mechanisms by which E7 oncoprotein can sense and manage the potentially harmful oxidative environment of the papillomavirus-infected cell. The carboxy terminal domain of E7 protein from most of the 79 papillomavirus viral types of alpha genus, which encloses all the tumorigenic viral types, is a cysteine rich domain that contains two classes of cysteines: strictly conserved low reactive Zn^{+2} binding and degenerate reactive cysteine residues that can sense reactive oxygen species (ROS). Based on experimental data obtained from E7 proteins from the prototypical viral types 16, 18 and 11, we identified a couple of low pKa nucleophilic cysteines that can form a disulfide bridge upon the exposure to ROS and regulate the cytoplasm to nucleus transport. From sequence analysis and phylogenetic reconstruction of redox sensing states we propose that reactive cysteine acquisition through evolution leads to three separate E7s protein families that differ in the ROS sensing mechanism: non ROS-sensitive E7s; ROS-sensitive E7s using only a single or multiple reactive cysteine sensing mechanisms and ROS-sensitive E7s using a reactive-resolutive cysteine couple sensing mechanism.

1. Introduction

Reactive oxygen species (ROS) are a broad set of chemical oxidants, which include hydrogen peroxide (H_2O_2), superoxide anion ($\cdot O_2$-), hydroxyl radicals ($\cdot OH$) and singlet oxygen ($^1O_2$), all involved in harmful and vital cellular processes [1,2]. The "dark side" of ROS involves the unwanted oxidation of proteins, lipid peroxidation, and DNA damage [3]. However, H_2O_2 is necessary for cell signaling and mediates vital processes such as hypoxic signal transduction, cell differentiation, proliferation, angiogenesis, apoptosis and immune response [4]. For cell signaling purposes, ROS are enzymatically produced in a defined time, on a defined cell compartment and upon specific stimuli by NAD(P)H oxidases or NADPH oxidases (NOXs) [5].

NOXs produce superoxide anion that is readily converted to H_2O_2 by superoxide dismutase (SOD). The homeostasis of redox balance that governs the pros and cons of ROS is based on an intertwined network of redox enzymes such as SOD, catalase, thioredoxin and peroxiredoxin [1,6] along with a plethora of redox-responsive proteins that includes transcription factors such as Nrf2 [7] or FOXO [8] and redox-sensitive kinase such as Src [9].

A direct link between ROS and cancer development has been recently established [10,11] where tumor cells show an altered ROS production resulting in the up-regulation of peroxide-dependent signaling pathways that regulate cell differentiation, growth and survival. A well-established case of this association is the effect of human papillomavirus (HPV) replication on infected and transformed

cells, where the overall redox balance is deregulated [12–15].

Papillomavirus are small DNA tumor viruses that are the causative agents of cervical cancer [16] and common warts [17]. Only a small fraction of the infected cells will eventually progress to cancer in a process that requires, at least, the persistent expression of E7 and E6 oncogenes [18] and an unknown triggering factor. In this scenario, it was suggested that the oxidative damage induced in HPV-infected cells could be the factor that eventually cooperates with viral oncoproteins to trigger cancer progression, in a very rare event.

Accumulated evidence supports the correlation between oxidative stress and HPV-induced transformation: i) ERp57 and GST involved in redox homeostasis were sharply elevated in dysplastic and neoplastic HPV-induced lesions [13]. ii) Oxidative damage reported by the increase of plasma malondialdehyde was demonstrated for patients with cervical intraepithelial neoplasia and cervical carcinoma [19], and an increase in the overall levels of protein oxidation was demonstrated in HPV-infected cells [13]. iii) The over-expression of NOX 1 accelerates neoplastic progression of human gingival mucosal keratinocytes immortalized by HPV16 E6/E7 [20,21]. iv) The solely expression of HPV16 E6 and E7 proteins induces a chronic oxidative stress response via NOX 2 [22].

We focused on the E7 oncoprotein, a 100 amino-acid protein with two well defined domains that differ in their conformational and functional properties [23]. The amino terminal domain (E7N, ca. 40 residues) is devoid of stable secondary and tertiary structure and shows properties resembling the intrinsically disordered domains. The E7N contains the highly conserved LXCXE motif, which is responsible for the high binding affinity of the retinoblastoma protein (Rb) [24]. The carboxy terminal domain (E7C, ca. 50 residues) is a well-folded domain with a globular hydrodynamic behavior. The E7C bears a tetracysteine Zn binding motif, unique among papillomavirus, and is composed of the CXXC-X$_{29}$-CXXC sequence. These four cysteines are highly conserved and Zn^{+2} coordination plays an essential structural role. The E7C is the oligomerization domain and is responsible for the dimeric structure observed in the published structures [24,25]. Still, a Kd of 1 µM indicates that the protein forms weak dimers along with the monomeric state and has been proved to be functional for transformation [26,27]. The four cysteines present in the E7C (Zn^{+2} binding motif) together with the cysteine in the LXCXE motif of E7N are considered as canonical cysteines. These five residues are highly conserved and are present in most of the E7 sequences regardless the genus, species or viral types. Many HPV types show an increased number of cysteines besides the canonical residues; however, these cysteines are much less conserved and we name them as non-canonical cysteines [28].

Along with the different conformational properties of the E7N and E7C, we previously showed that the cysteine residues in each domain, at least for HPV16, constitute different redox centers [29]. The cysteine residue located within the Rb binding motif (LXCXE) is sensitive to gluthationylation and this modification abolishes the binding of Rb. On the other hand, we have recently shown that the non-canonical cysteines, C59 and 68 in HPV16 E7, form a disulfide bond when exposed to an oxidative environment where the tetrahedral Zn^{+2} binding motif remains unaffected [29]. We further demonstrated that each domain can be oxidized independently and a mutant that lacks the LXCXE motif is able to form the observed disulfide bond.

The structures of the E7 proteins for the viral types HPV1a and HPV45 are dimeric and superimposable, and it is well accepted that the overall fold is a common feature of all E7s. It should be noted that residues 59 and 68 in the dimeric structure are 18.6 Å apart, too far to form a disulfide bridge. Therefore, an intermediary is more likely to occur [29].

In this work, we address the chemical mechanism by which a cysteine-rich oncoprotein senses and responds to the redox stimuli in an HPV-infected cell and how the protein can manage the potentially harmful oxidative environment of a tumor tissue. We address the differential conservation pattern and functional properties of canonical and non-canonical residues in the E7 proteins of alpha genus papillomavirus.

2. Methods

2.1. Protein expression and purification

Recombinant proteins were expressed in *E. coli* BL21(DE3)pLysS. Wild type E7 from HPV16 and cysteine-substituted mutants E7desLXCXE and HPV16E7(C59/68A) accumulate in inclusion bodies as fusions to a 19-aminoacid beta-galactosidase peptide with an enterokinase (EK) cleavage site as linker. The detailed purification protocol was previously described [30]. The truncated gene coding HPV16 E7Δ1−26 was cloned into a pMAL-c2 vector (New England Biolabs). The recombinant protein was purified as described for the full length HPV16 E7 rendering the E7Δ1−26 with two N-terminal aminoacids (glycine and serine) due to MBP excision with thrombin protease. GST-tagged wild type E7 from HPV11 and HPV18 were purified from soluble bacterial extracts in a glutathione-sepharose column. Untagged HPV11 E7 and HPV18 E7 were also obtained by thrombin cleavage.

A final purification step in a Superdex 75 column (GE Healthcare Bio-Sciences Corp, USA) with a reducing agent-free buffer was performed for all proteins. Purity was > 95% as confirmed by PAGE, with no more than 10% of oxidized species. Concentrated reductant-free protein stocks were maintained at -80C for 3 months with no oxidative damage.

2.2. Sequence alignment and naïve model for cysteine distribution

We retrieved all alpha genus papillomavirus types in the NCBI taxonomy database as October 1, 2015. At least 79 alpha-papillomavirus types had one ORF coding for E7 protein. Alignments for HPV E7 protein sequences were performed using the MUSCLE algorithm [31] and manually edited using the SeaView software. Naïve model for cysteine distribution were generated by a randomized assignment of cysteines at different positions in the sequence. To build the model this procedure was repeated 10000 times, the mean and standard deviation values for each combination of cysteines were calculated.

2.3. Ancestral states reconstruction and randomized trees

The tracing of the ancestral redox sensing state of each internal node was done using Mesquite and choosing reconstruction by parsimony [32]. The transformation types were defined as unordered, in which any state is allowed to transform to any other with one step of evolution. Each transition of states between nodes was counted. The redox state reconstructions from randomized trees were generated by reshuffling the same terminal taxa without altering the tree structure to obtain 100,000 trees. For each tree the ancestral redox state was reconstructed by parsimony using Mesquite and all transitions between node states were counted.

2.4. Quantitative thiol and Zn determination

Thiol content was quantified using DTNB; Zn content was determined using PAR-PMPS methods as described in [49].

2.5. Reactivity to H$_2$O$_2$ by RP-HPLC

Oxidation was triggered by the addition of 0.15 mM of H$_2$O$_2$ to a solution containing 3.5 µM of reduced protein in 50 mM potassium phosphate buffer pH 7.5, incubated at 37 °C. Samples were taken at the stated time; methionine was added to 2.5 mM final concentration to quench peroxide and immediately injected into a reverse phase HPLC

(RP-HPLC). Isolation was performed using a C4 column, 250 mm×4.6 mm (Bio-Basic-C4, Thermo Scientific, PA, USA) and the eluted species were detected with a diode-array detector.

For the apparent $t_{1/2}$ estimation, the peak area of each species was quantified, plotted versus time and fitted to a single exponential according to Eq. (1).

$$Y = 100e^{(-k.t)} \tag{1}$$

where Y is the remaining amount of the species at a defined time and k is the apparent oxidation constant. The apparent $t_{1/2}$ is calculated using the Eq. (2).

$$t_{1/2} = \frac{\ln 2}{k} \tag{2}$$

2.6. Circular dichroism

CD measurements were carried out on a Jasco J-810 spectropolarimeter (Jasco, Easton, MD) employing a scan speed of 50 nm/min, a 4 nm band pass and an average response time of 4 s, at 37 ºC in a Peltier temperature–controlled sample compartment. Spectra were an average of at least 6 scans. Spectra of reduced and oxidized 20 µM E7 were taken on 0.1 cm path length cells in 10 mM potassium phosphate buffer, pH 7.5. Oxidation was triggered by adding 0.15 mM H_2O_2 to the reduced protein in the CD cell.

2.7. Cysteine pKa determination

We prepared samples containing 3.5 µM of reduced E7Δ1–26 protein in 50 mM of the following buffers: sodium acetate (pH 5.0), potassium phosphate (pH 6.0; 6.5; 7.0 and 8.0) and Tris-HCl (9.0 and 10.0). The oxidation reaction was initiated by the addition of 0.15 mM of H_2O_2. After one hour of incubation at 37 ºC methionine was added to 2.5 mM final concentration to quench peroxide and the reaction was stopped by acidification with 0.1% TFA. The mixture was resolved by RP-HPLC (see *Reactivity to H_2O_2 by RP-HPLC*). The pKa of the reactive cysteine was determined by analyzing two different species. A plot of the Ox II species formation and the decrease of the reduced species versus pH was obtained and fitted to Eq. (3). The decrease of the reduced species was obtained by subtracting 100 − (%) of reduced peak.

$$\text{Peak Intensity}(\%) = \frac{A}{(1 + 10^{(pKa - pH)})} \tag{3}$$

A is a constant.

2.8. Disulfide connectivity by MALDI-TOF

Reduced, Ox I and Ox II species from different proteins were isolated by RP-HPLC and lyophilized. Dry samples were resuspended with an alkylating solution containing 56 mM Iodoacetamide in order to block reactive sulfhydryls. Samples were precipitated with TCA and the acetone-washed dry pellets resuspended with 50 mM Tris-HCl, 5 mM Octyl β-D-glucopyranoside, pH 8.0 containing sequencing grade trypsin (Sigma, St Louis, MO) at 1/10 (w/w). Ultra-pure 5 mM DTT was added when required. After 4 h proteolysis at 37 ºC the reaction was stopped by adding 0.1% TFA. A 1 µL aliquot was spotted onto an AnchorChip (Bruker, Billerica, MA, USA) and 1 µL oversaturated solution of α-Cyano-4-hydroxycinnamic Acid in 30/70/0.1% (v/v) acetonitrile/water/TFA was loaded to the protein. Samples were analyzed on a Bruker Microflex MALDI-TOF.

2.9. Transient expression of spGFP fusion proteins in Hela cells and fluorescence microscopy

HeLa cells were plated on 12 mm glass coverslips to 80% confluency, transfected 24 h later with spGFP-E7wt, spGFP-E7C59A or spGFP plasmids using Lipofectamine 2000 (Invitrogen) and fixed 24 h after transfection with 4% formaldehyde. For ROS treatment, cells were incubated with 4 µM H_2O_2 for 1 h prior to fixation. Coverslips were mounted using Mowiol mounting medium and cells were observed by fluorescence microscopy using a Zeiss LSM 510 confocal (Carl Zeiss, Thornwood, NY).

Quantification of the subcellular localization of spGFP-E7wt, spGFP-E7C59A and spGFP proteins was performed by analyzing fluorescence intensity of random circular regions of interest (ROIs) in nuclear and cytoplasmic areas using ImageJ software. Data from at least 20 cells and 5 ROIs per cell were used for the quantitative analysis; graphical representation displays average values with standard deviation. Significance of differences among groups was tested by one-way ANOVA and Holm-Sidak method was used for post hoc testing. SigmaStat 4.0 (Systat Software Inc.) was used to perform all calculations. Differences were considered significant if P≤0.05. Data are expressed as mean ± SD.

2.10. Western-blot

Hela cells were transfected as indicated above and lysates were obtained 24 h post-transfection using RIPA buffer. Extracts were electrophoresed using 12.5% SDS-polyacrylamide gels, and transferred to nitrocellulose membrane; blots were probed with anti HPV16 E7 monoclonal antibody.

3. Results

3.1. Cysteine conservation pattern in E7C

Cysteine residues are highly reactive due to the nucleophilic character of the thiolate group, and their conservation is often related to function [33]. In this regard, the high content and conservation of cysteine residues in E7 protein are likely to be connected to the molecular basis for redox sensing. E7 contains five highly conserved canonical cysteine residues, one in E7N and four in E7C (Fig. 1A). In addition, most (94%) E7 proteins from the alpha genus possess one to three extra, non-canonical cysteine residues in E7C (Fig. 1B) and can be considered a cysteine rich domain (CRD).

Our alignment of 79 sequences from alpha papillomaviruses shows that non-canonical cysteines are not evenly distributed but occupy a narrow set of 15 positions in the E7C (Fig. 1C). The appearance of a particular non-canonical cysteine can be observed in viral types belonging to different species and infecting different hosts (Fig. 1 C). For example, the non-canonical cysteine at position 71 is present in 32 out of 79 viral types spanning species 1, 2, 3, 4, 8, 10 and 14. Moreover, 29 of these viral types infect humans while the remaining three are simian papillomaviruses. All thirteen high-risk HPV viral types belong to species 7 and 9 in the alpha genus and show at least one non-canonical cysteine, with 12 out of 13 showing two non-canonical cysteines (Fig. 1C).

3.2. Non-canonical residues of E7 cysteine rich domain are primary targets of ROS

With the aim of evaluating the sensitivity of HPV16 E7 protein to ROS, we resolved and quantified the oxidized species that are formed by the addition of H_2O_2 by RP-HPLC. Within the first hour of reaction with peroxide, we observed the decrease of the reduced protein (RT: 28.7 min) with the appearance of peaks at 24.8 min and 21.9 min, indicated as Ox I and Ox II species, respectively, (Fig. 2A and
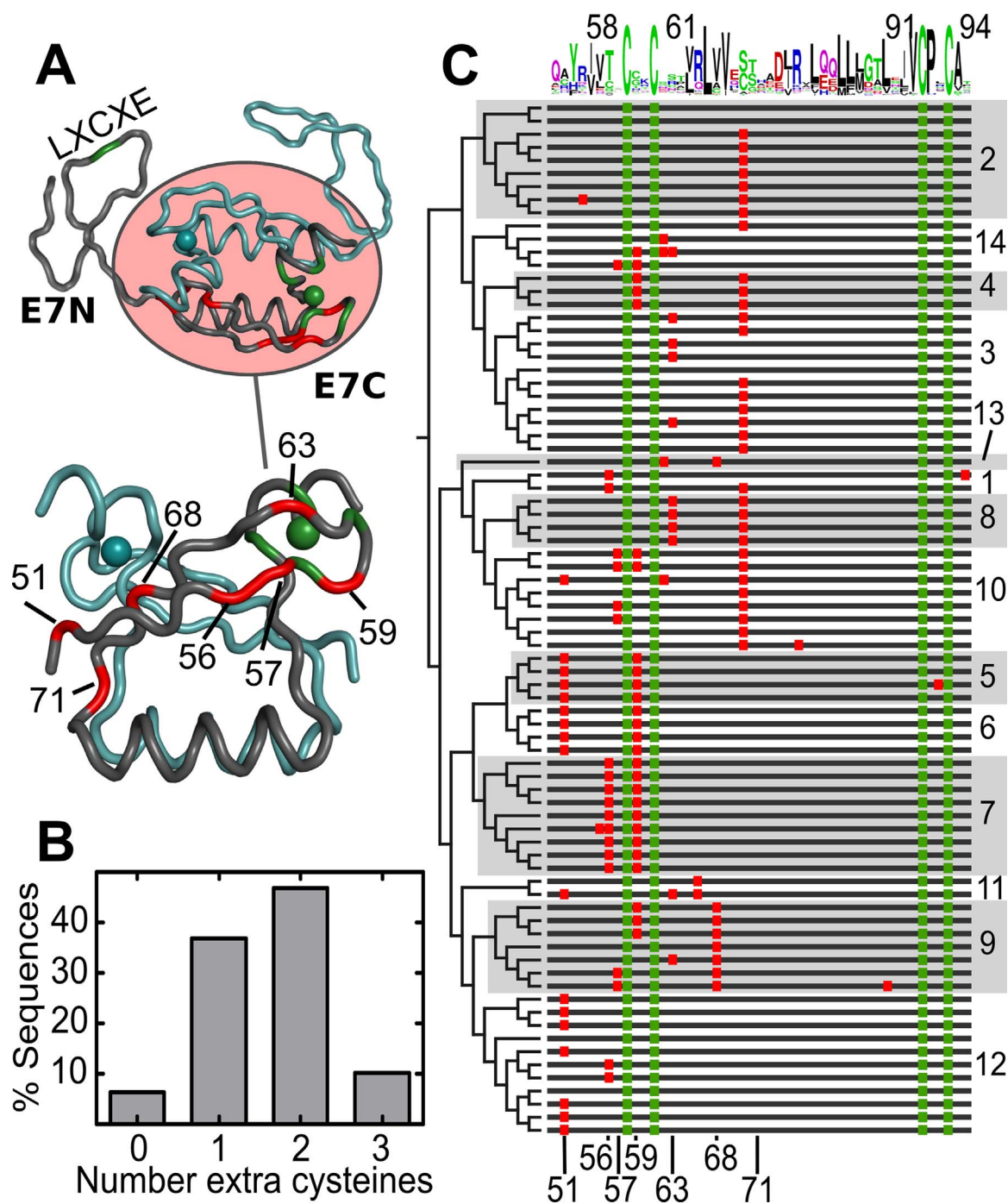
Fig. 1. E7 alpha papillomavirus cysteine distribution. (A) E7 dimer structure. The globular E7C domain (pink circle) contains four canonical cysteines (green) that coordinate the Zn^{+2} atom (green spheres), the fifth canonical cysteine is located in the E7N unstructured domain (LXCXE motif; green); positions most frequently occupied by non-canonical cysteines in E7C are depicted in red (cysteine frequency above the average percentage observed in mammalian proteins (2.3%) are located at the positions 51, 56, 57, 59, 63, 68 and 71). (B) Distribution of the number of non-canonical cysteines in the E7C of alpha genus papillomavirus. (C) Schematic distribution of the cysteines in the E7C. Canonical cysteines are indicated in green and non-canonical in red. The sequences are ordered according to the alpha genus phylogenetic tree (left) with the specie number indicated at the right. Above the scheme, the motif logo of the 79 sequences is shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Supplementary 1). Using glutathione redox buffer as an oxidant stimuli we have previously observed the formation of Ox II species that were originated by the disulfide formation between non-canonical cysteines 59 and 68 in the E7 HPV16 protein [29].

Less than 40% of E7 protein remains in the reduced state after one hour of H_2O_2 incubation, with an apparent $t_{1/2}$ decay of the reduced peak of 40.7 min (Fig. 2B and C). After the first hour, the intensity of reduced, Ox I and Ox II species decreases with the appearance of a broad set of peaks that cannot be chromatographically resolved for the

E7 wild type and are marked, altogether as Ox III species. After 240 min of H_2O_2 treatment, less than 5% of the protein remains in the reduced state and a wide peak of over-oxidized species (Ox III) is observed (Fig. 2A and Supplementary 1).

We decided to further assess the location of the main ROS sensitive cysteines on E7 protein. Two mutants where the canonical cysteine 24 (C24) has been mutated (E7desLXCXE) or deleted (E7Δ(1-26)) (Fig. 2C) were oxidized with H_2O_2 with the formation of Ox I and Ox II species within the first hour (Fig. 2A and Supplementary 1). The
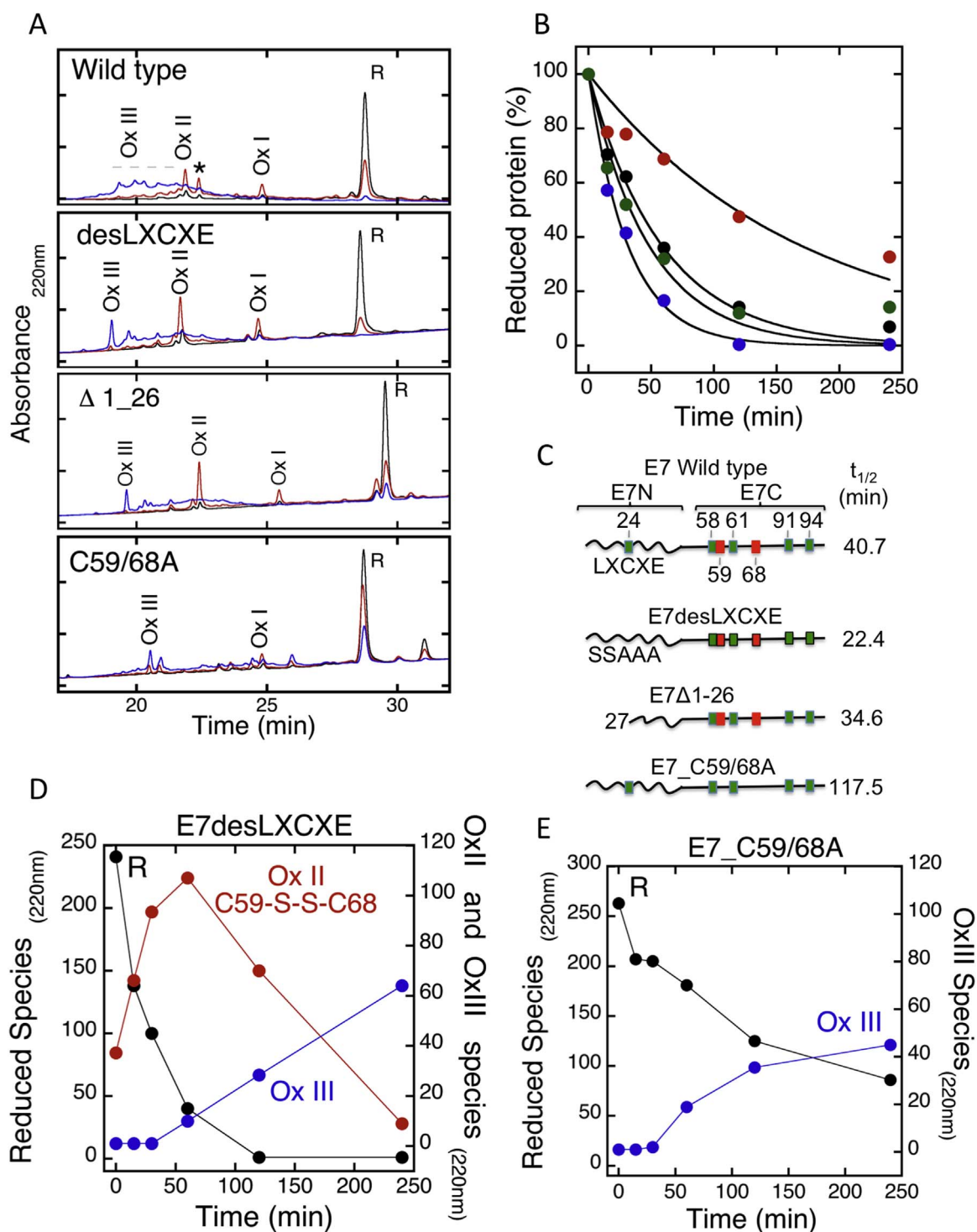
Fig. 2. ROS sensitivity of HPV16 E7 protein and mutants. (**A**) RP-HPLC chromatograms of HPV16 E7 wild type, E7desLXCXE, E7Δ1−26 and E7(C59/68 A) incubated for 0 (black line), 60 (red line) and 240 min (blue line) with hydrogen peroxide. R indicates the reduced species peak. * indicates an unidentified peak. (**B**) Kinetic decay of the reduced species. The reduced species was quantified by the area of the reduced peak (R). HPV16 E7 wild type (black circles), E7desLXCXE (blue circles), E7Δ1−26 (green circles) and E7(C59/68 A) (red circles). (**C**) Schematic representation of the HPV16 E7 wild type, E7desLXCXE, E7Δ1−26 and E7(C59/68 A). The apparent $t_{1/2}$ decay for the reduced species is shown at the right side of each protein. Apparent $t_{1/2}$ was calculated by fitting to a single exponential equation. Green squares represent the position of canonical cysteines and red squares represent non-canonical residues. (**D**) Evolution profile of the reduced (black circles and line), Ox II species (red circle and line) and Ox III species (blue circles and line) of the E7desLXCXE protein. (**F**) Evolution profile of the reduced (black circles and line) and Ox III species (blue circles and line) of the E7(C59/68 A) protein. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

presence of the C59-S-S-C68 disulfide in the Ox II peak was confirmed for the E7Δ(1-26) protein. The apparent $t_{1/2}$ for the reduced peak decay of E7desLXCXE and E7Δ(1-26) is 22.4 and 34.5 min, respectively (Fig. 2B and C) comparable with the decay of the wild type E7, strongly suggesting that a ROS sensitive cluster of cysteines is located in the E7C CRD. The decrease of the reduced species in the wild type, the desLXCXE and the E7Δ(1-26) proteins occur with the concomitant formation of a disulfide bridge between non-canonical cysteines 59 and 68 (Ox II species).

In a mutant where both non-canonical cysteines are replaced by

alanine, E7(C59/68 A), the Ox II species is absent (Fig. 2A and Supplementary 1) and is more refractory to ROS. After one hour of reaction over 75% of the E7(C59/68 A) is still reduced, with an apparent $t_{1/2}$ of 117 min (Fig. 2B). On the other hand, the formation of overoxidized Ox III species was observed for all of the four proteins analyzed, an event that is clearly evident after 240 min of incubation (Fig. 2A and Supplementary 1).

This multi-step process is clearly observed for the E7desLXCXE protein where the Ox I, Ox II and Ox III species can be accurately quantified (Fig. 2D). On the other hand, the double cysteine mutant decays slowly, with the concomitant appearance of the overoxidized Ox III species and a minor proportion of Ox I (see Fig. 2E and Supplementary 1). After 240 min of incubation, the reduced fraction of E7(C59/68A) protein is 30% (Fig. 2B).

These results show that the redox-sensitive residues that generate the oxidized species I, II and III are contained in the CRD of E7; where Ox II, originated in the formation of C59-S-S-C68 disulfide bond, is the major species and its appearance occurs concomitant with the decay of the reduced protein. On the other hand, Ox I and Ox III that are also observed for E7(C59/68A) protein correspond to a modification in the canonical cysteines or other redox sensitive residues such as methionine. As the formation of Ox I, Ox II and Ox III is reversible with DTT (not shown) we can exclude methionine or other non-reversible oxidations.

3.3. E7 proteins containing different arrangements of non-canonical cysteines show similar reactivity to ROS

E7 proteins of different strains contain a different combination of non-canonical cysteines and these are oxidized by H_2O_2 to the same extent than the wt HPV16 E7. The E7 protein from viral types 11 and 18 contain three and two non-canonical cysteines, respectively. The distribution pattern of the non-canonical cysteines is different for the three proteins yet they share cysteine 59 (refers to the position in the HPV16 E7 sequence) (Fig. 3A). We first determined that HPV16, HPV11 and HPV18 E7 proteins contain one Zn^{+2} atom strongly bound (over 90% occupancy, not shown) and the quantitation of the thiol content by DTNB yielded that 95%, 90% and 85% of the cysteines in HPV16, HPV11 and HPV18 E7 proteins were reactive, indicating that the proteins are reduced and accessible the storage conditions (not shown).

The incubation of these proteins with 0.15 mM H_2O_2 at 37 ºC and pH 7.4 led to a decrease of the reduced peak, with an apparent $t_{1/2}$ of 17.3 and 34.6 min for the HPV11 and HPV18, respectively. The decay of the reduced peak for HPV11 and 18 E7 is similar to that observed for the HPV16 E7 protein (Fig. 3B) indicating that both proteins are highly sensitive to ROS. As in the case of the HPV16 protein, within the first hour of reaction the HPV11E7 and HPV18E7 reduced species decrease with the concomitant increase of intermediate oxidized species. After one hour these intermediate species are transformed in the overoxidized species that could not be resolved by RP-HPLC and were observed as a broad sum of peaks (Supplementary Figure 2).

The major species observed in the HPV16 E7 protein after one hour of incubation with peroxide is the Ox II peak that corresponds to the formation of a disulfide bond between cysteine 59 and 68. We obtained a three-dimensional model for the HPV16 E7 proteins using the HPV1 E7 structure as a template [34]. The sulfur atoms of the C59 and C68 are 19.2 A distant in the dimeric structure of the HPV16 E7, a distance that is too extensive to allow a direct (SN_2 mechanism) disulfide formation (the cut off has been settled in 6.2 A, [35]) (Fig. 3C) indicating that the E7C conformation must change to allow the nucleophilic attack of the thiolate to occur.

We then evaluated the effect of ROS in the structure of the three E7 proteins. The far UV circular dichroism (CD) spectra of each protein in the reduced state (0 min incubation) and after one hour of incubation with H_2O_2 were recorded (Fig. 3D, E and F). After one hour of reaction

the reduced protein fell below 50% and the appearance of overoxidized species was minimum (Fig. 3B and Supplementary Figure 2).

The overall secondary structure content of the reduced species was similar for the HPV16, HPV11 and HPV18 E7 proteins as judged by the similarity in the FAR-UV CD spectra. After incubating with H_2O_2 for 1 h only a small (less than 1.5 mdeg) decrease in the negative band at 225 nm was observed for the HPV16, HPV11 and HPV18 E7s.

Although the extent of change could be underrepresented due to the existence of reduced protein, ROS produces in different E7s proteins a similar conformational change irrespective of the cysteine arrangement.

It should be noted that, at least for the HPV16 E7 protein, this conformational change should involve a thiolate translation of at least 13 A, that occurs without the rearrangement of the major secondary elements of the E7 folding architecture (i.e a FAR-UV CD change less than 1.5 mdeg). To confirm that the FAR-UV CD change was due to the formation of the C59-S-S-C68 disulfide we analyzed the C59/68A mutant and no FAR-UV CD change was observed (Not shown).

A conformational change that is triggered only after the oxidation of a ROS-reactive cysteine (i.e to sulfenic acid) is necessary to prevent the spontaneous oxidation of the protein under basal conditions and may act as a molecular switch that fulfills ROS sensing proteins [36–38]. Although we were unable to experimentally detect the proposed sulfenic acid intermediate we hypothesize that a conformational change is induced by the sulfenic acid formation in a non-canonical cysteine that brings a second distant non-canonical residue in the proximity to allow a disulfide formation.

3.4. Subcellular localization of the E7 is modulated by oxidation of non-canonical cysteines

The expression pattern of the HPV16 E7 protein is variable and depends on the type of expression and cell line [39]. Nuclear import of HPV16 E7 depends on the interaction of Nup62 nucleoporin with a hydrophobic amino acid patch, spanning residues 65LRLCV69 in the CRD domain of HPV16 E7 protein [40]. The non-canonical C68 that forms a disulfide bridge when exposed in vitro to ROS is located at the very center of the nuclear localization signal (NLS) (Fig. 4A). We thus evaluated the effect of H_2O_2 in the sub-cellular localization of spGFP-E7 wild type (wt) protein transiently expressed in HeLa cells by confocal fluorescence microscopy.

spGFP-E7wt protein shows a predominantly nuclear localization (Fig. 4B, panel a). In contrast, when HeLa cells transfected with the spGFP-E7wt plasmid were treated with H_2O_2 we observed that the cytoplasm fluorescence intensity increased (Fig. 4B, panel b) indicating that the nuclear import and/or export trade-off was affected by a redox process. In order to quantify the extent of re-localization, we calculated the nuclear to cytoplasm fluorescence intensity ratio showing that incubation of transfected spGFP-E7wt HeLa cells with peroxide led to a decrease of 25% in the nuclear to cytoplasm ratio (Fig. 4C). The extent of the spGFP-E7wt redox-based relocalization can be underestimated due to the existence of many intracellular ROS protecting enzymes such as peroxiredoxins and catalases that counter the effect of exogenous peroxide. Furthermore, the formation of the internal disulfide (C59-S-S-C68) is a reversible, dynamic equilibrium that can be readily displaced to the reduced form by the most abundant low molecular weight thiols, such as glutathione or by highly effective disulfide reducing enzymes, such as thioredoxin. Notwithstanding the above-mentioned factors we were able to detect a statistically significant variation on the nuclear-to-cytoplasmatiic ratio upon the induction of an intracellular oxidative stress.

The spGFP-E7C59A mutant is unable to form the internal disulfide bond (C59-S-S-C68) while maintaining the proposed NLS intact (65LRLCV69) (Fig. 4A). The nuclear to cytoplasm intensity ratio for the spGFP-E7C59A (Fig. 4B, panels c and d) remained unchanged after the incubation of transfected HeLa with peroxide (Fig. 4C), indicating
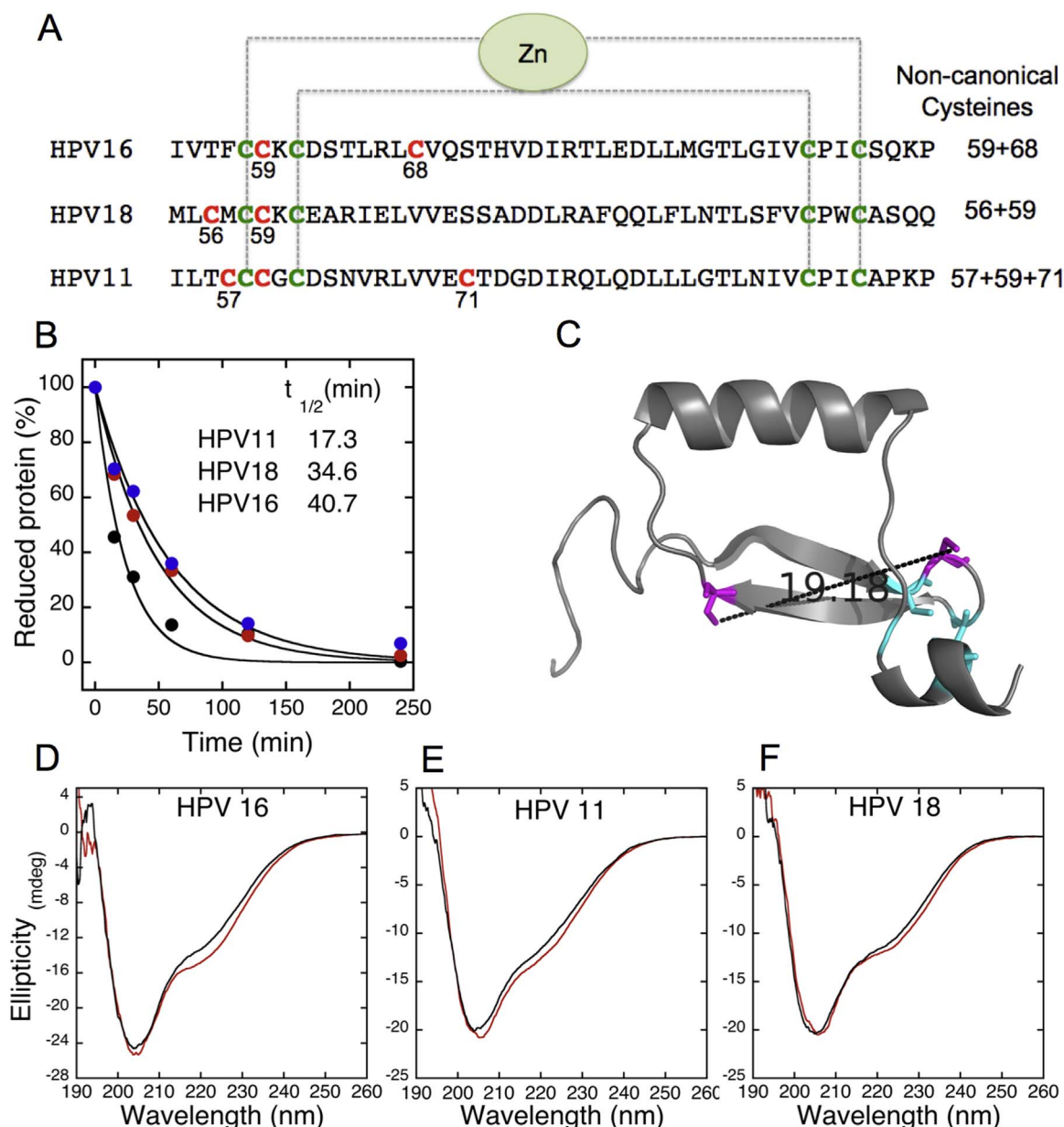
Fig. 3. Kinetics and conformational changes induced by ROS in prototypical alpha HPVs. (**A**) Schematic representation of the E7 proteins from the HPV16, 18 and 11 strains. Canonical cysteines are highlighted in green and non-canonical cysteines in red. The position of the non-canonical residues is indicated at the right of each protein. (**B**) Kinetic decay of the HPV11, 18 and 16 E7 reduced species: HPV11 (black circles), HPV16 (blue circles), and HPV18 (red circles). Apparent $t_{1/2}$ was calculated by fitting to a single exponential equation. (**C**) Measurement of the distance between sulfur atom of the C59 and C68 cysteines (magenta) in the structure of the HPV16 E7 protein. (**D, E, F**) CD spectra of the reduced HPV16, 11 and 18 proteins, respectively, before (red line) and after the incubation with H_2O_2 for 60 min (black line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

that C59 is strictly required for the protein re-localization mediated by ROS. On the other hand, a control using spGFP transfected HeLa cells (Fig. 4B, panels e and f) showed that this protein has a pan-cellular localization albeit with a nuclear preference, and not affected by ROS (Fig. 4C).

Inmunoblot analysis of expressed HPV16 E7 proteins showed that proteins were expressed at similar levels and remained unchanged after peroxide treatment (Fig. 4 D).

3.5. Reactive non-canonical cysteines are critical for ROS sensing

The presence of non-canonical cysteines confers to E7 an increased sensitivity to ROS, in which an oxidative stimulus leads to the formation of a disulfide bond and to the subcellular re-localization of the protein. We thus sought for the presence of reactive cysteines,

which show a pKa lower than glutathione (pKa 8.8). As the thiolate is a better nucleophile than the protonated thiol, a low pKa is a hallmark feature of ROS sensing cysteines [41]. We evaluated the pH dependence of the E7Δ1-26 HPV16 oxidation with H_2O_2 by RP-HPLC (Fig. 5A). At pH 5.0, the protein is insensitive to ROS within one hour, no OxI and OxII species are observed and the major peak at 29.5 min corresponds to the reduced protein. As the pH increases, the proportion of species Ox I and Ox II increases at the reduced species expense. After incubating the E7Δ1-26 with ROS at pH 9.0 for one hour, we observed that 14% of the protein remains in the reduced state, 30% as Ox I species and 55% as Ox II. On the other hand, the E7Δ1-26(C59/68A) is refractory to ROS oxidation after 1 h incubation (Supplementary Figure 3) and more than 80% of the protein remains in the reduced state (ROS treatment for 1 h at pH 9.0). This observation indicates that the reactive cysteine is restricted to C59 and C68
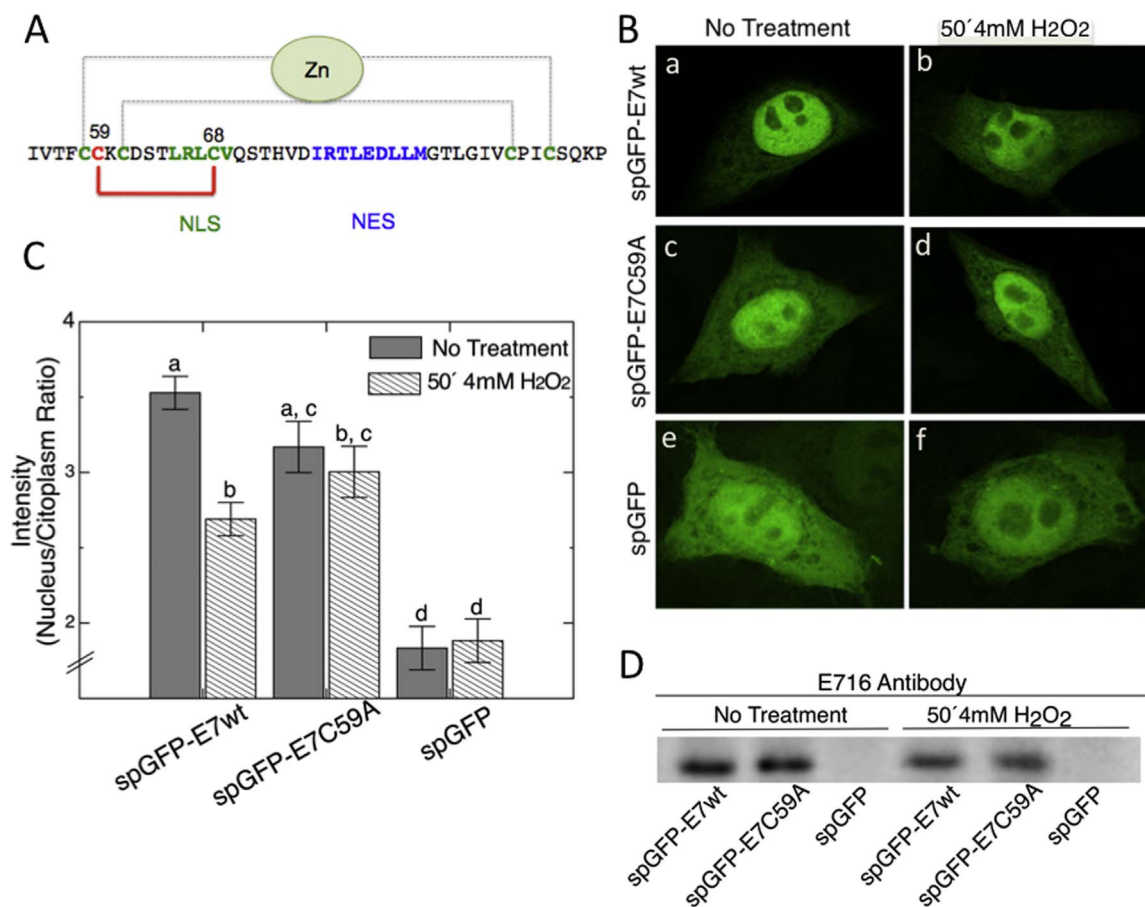
Fig. 4. Oxidative environment directs subcellular re-localization of spGFP-E7wt. (**A**) Position of the NLS and NES in the E7C of HPV16. (**B**) HeLa cells were transiently transfected with spGFP-E7wt (panels a, b), spGFP-E7C59A (panels c, d) or spGFP (panels e, f) plasmids. For ROS treatment, cells were incubated with 4 mM H_2O_2 for 1 h prior to fixation (panels b, d, and f). Cells were examined by confocal fluorescence microscopy at 24 h post-transfection. (**C**) Ratio of spGFP fluorescence between nuclear and cytoplasmic signal in transfected cells as in B. a, b, c, d bars not sharing the same superscript are significantly different ($p < 0.05$, n=20 cells). (D) HPV16 E7 protein is properly expressed in HeLa cells. Cells were transfected as in B and lysates were obtained 24 h post-transfection, subjected to SDS-PAGE and transferred to nitrocellulose membrane; blots were probed with anti E7 HPV16 monoclonal antibody.

residues. It should be noted that, for the double mutant, no Ox II species is observed and only a trace amounts of Ox I can be detected through the entire range of pH 5.0–10.0).

The single cysteine C59A and C68A mutants show a similar redox behavior (Fig. 5B and C). These proteins are ROS sensitive even though the oxidation mechanism differs from the wild type protein and no Ox II species are observed. At high pH 8–10), the reduced species decrease and the Ox I species are clearly accumulated for both proteins, contrary to that observed for the C59/68A double mutant (Supplementary Figure 3).

We identified the cysteine connectivity in Ox II and Ox I species by trypsin treatment of alkylated HPLC-purified peaks and subsequent MALDI-TOF analysis. For better understanding of the data we present a schematic representation of the tryptic fragments and the possible connectivity between fragments (Supplementary Figure 4). The Ox II species contains the internal disulfide C59-S-S-C68 as thoroughly described in Supplementary Figure 5. On the other hand, we identified in the Ox I species an internal disulfide containing the canonical cysteines 58 and 61 that are involved in Zn^{+2} binding (Supplementary Figure 5).

Although cysteines 58 and 61 are present in the E7Δ1-26(C59/68A), E7Δ1-26(C59A) and E7Δ1-26(C68A), the presence of an un-paired cysteine in the C59A or C68A mutant produces higher amounts of Ox I species when compared with the double mutant. The formation of the Ox I species is affected by the presence of non-canonical cysteines showing a crosstalk between canonical and non-canonical residues.

We calculated the pKa of the reactive cysteine that governs the

formation of the C59-S-S-C68 disulfide bond. The appearance of Ox II species and the decrease of the reduced species were fitted to a two state model that considers only the protonated and deprotonated forms. The pKa that governs the formation of Ox II and the decrease of the reduced species are 6.8 and 6.6, respectively, strongly suggesting that the oxidation of the reduced species leads to Ox II as main product (Fig. 5D). At least one of the non-canonical cysteines, the C59 or C68 residue is a reactive cysteine whose pKa is below the pKa of low molecular weight thiols (Glutathione pKa=8.8 [41]).

3.6. ROS sensing mechanism: lone or multiple reactive cysteines versus reactive-resolutive cysteine pairs

Our biochemical experiments with E7 proteins from HPV11, HPV16 and HPV18 strains show a common behavior in spite of different patterns of non-canonical cysteines. This suggests that several positions in the E7 sequence are degenerate in terms of redox sensing and that there may be a simple organizing principle behind the diverse patterns observed in naturally occurring E7 proteins (Fig. 1C). The oxidation behavior of E7desLXCXE and E7Δ[1−26] shows that the C-terminal domain of E7 is the main responsible for redox sensing. Thus, we will focus our analysis on E7C, where cysteines are observed at 4 canonical sequence positions (58, 61, 91 and 94) and 15 non-canonical sequence positions (51, 53, 55, 56, 57, 59, 62, 63, 66, 68, 71, 77, 87, 93 and 96). Since we are only interested in redox sensing by cysteines, we considered all non-cysteine residues as equivalent.

First, we considered the five E7 sequences (6%) that lack non-canonical cysteines. E7(C59/68A) is insensitive to ROS within the first
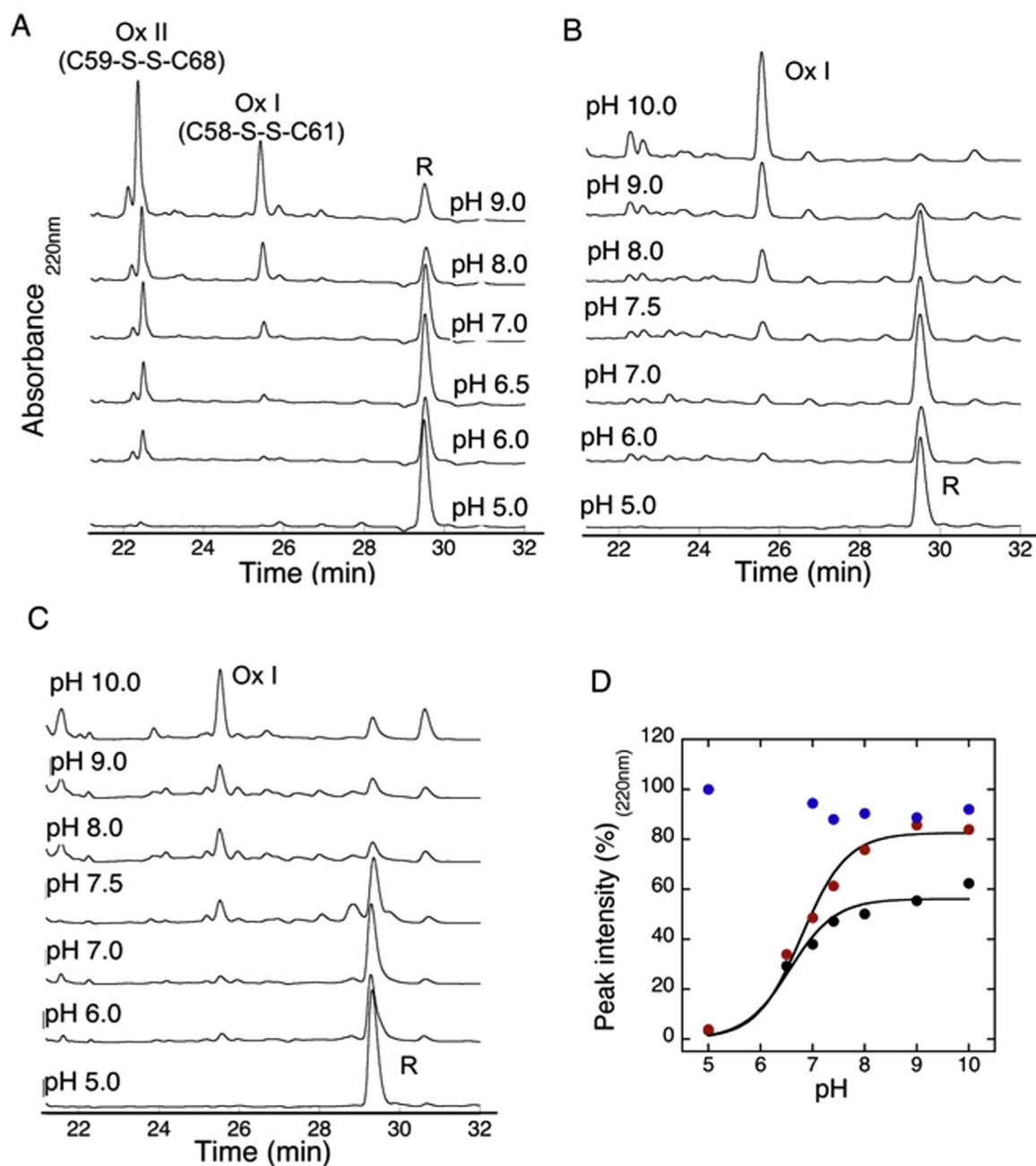
Fig. 5. Identification of the cysteine pKa and disulfide connectivity. (**A**) Oxidation of HPV16 E7Δ1−26 at different pHs resolved by RP-HPLC. R indicates the reduced species peak. The disulfide connectivity responsible for the formation of Ox I and Ox II species are indicated above the corresponding RP-HPLC peak. (**B**) Oxidation of HPV16 E7Δ1−26 (C59A) at different pHs resolved by RP-HPLC. (**C**) Oxidation of HPV16 E7Δ1−26 (C68A) at different pHs resolved by RP-HPLC. (**D**) Cysteine pKa calculation was performed in two different transitions. The pKa was determined from the Ox II species (black circles) and the decrease of the reduced peak (red circles) versus the pH plot and fitted to Eq. (3) as described in methods. The intensity of the reduced species of the double mutant E7Δ1−26 (C59/68 A) versus pH was included as control (blue circles). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

hour of incubation, while the wild type protein is readily oxidized (Fig. 2). Hence, we propose that E7 proteins without non-canonical cysteines would not be able to respond readily to a redox signal.

We also examined the 29 E7 sequences (37%) that contain only one non-canonical cysteine. Single non-canonical cysteines are observed at seven different E7 sequence positions [51, 56, 62, 63, 66, 68 and 71]. We propose that E7 proteins with a single non-canonical cysteine perform redox sensing through a mechanism involving a single reactive cysteine, similar to that observed for one-cysteine peroxiredoxins [42]. We can use as a model the E7Δ1-26(C59A) protein, that contains cysteine 68 as the sole non-canonical cysteine. This protein is ROS sensitive through a mechanism that eventually leads to reversible

formation of an intramolecular disulfide bond between Zn-binding canonical cysteines C58-C61 (Ox I species) (Fig. 5). In the cell, the reaction may be resolved not by the Zn-binding cysteines as in the Ox I species but by low molecular weight thiols or other cysteine-containing proteins, such as thioredoxin. This strongly suggests that these 7 positions are degenerate, i.e., fulfill the same role in redox sensing by E7, and that a single non-canonical cysteine from this group is sufficient for the redox sensing activity. The rest of the E7 sequences (45 strains) contain more than one non-canonical cysteine and an alternative mechanism must be proposed.

We have thoroughly analyzed the redox behavior of the HPV16 E7 that contains non-canonical cysteines 59 and 68. In short reaction

times, this protein can sense ROS via a stepwise mechanism where a reactive cysteine reacts via nucleophilic substitution with the H_2O_2 molecule (forming a sulfenic acid [41]) that is then attacked by a second, resolutive cysteine to form the C59-C68 disulfide bond (Ox II species) (Fig. 5). Since we identified C68 as a reactive cysteine (see above), we conclude that C59 is the resolutive cysteine in HPV16 E7. In accordance with this model, C59 is never the lone non-canonical cysteine, reinforcing our view that it is not a reactive cysteine in E7 natural sequences. HPV18 E7 contains non-canonical cysteines C59 and C56. Since it behaves similar to HPV16 E7 (Fig. 3), we suggest that in HPV18 E7 C56 acts as a reactive cysteine and C59 is the resolutive cysteine. HPV11 E7 contains three non-canonical cysteines: C57, C59 and C71 and also behaves similar to the HPV16 and HPV18 E7 proteins (Fig. 3). We propose that in HPV11 E7, C57 and C71 can both act as reactive cysteines and C59 is the resolutive cysteine. It follows from these considerations that the reactive cysteine that pairs with C59 can be located at different, degenerate sites of the sequence, similar to the degeneracy observed for lone reactive cysteines (see previous section). In total, 27 E7 proteins (34%) contain C59 and we propose that this group of E7 proteins sense ROS via a stepwise mechanism in which one or two cysteines located at position 51, 56, 57, 62, 63, 66, 68, 71, 77, 87 or 96 is/are reactive and C59 plays a resolutive role. The remaining 18 E7 sequences (23%) lack C59 but contain multiple reactive cysteines at positions 51, 53, 55, 56, 62, 63, 66, 68, 71, 77, 87 or 96. Fifteen E7 sequences contain two reactive non-canonical cysteines, while three E7 sequences have three reactive non-canonical cysteines. We hypothesize that this group of E7 proteins perform redox sensing via the same single-cysteine mechanism, albeit at multiple sites in the sequence simultaneously.

Our model for redox sensing in E7 can be summarized as follows. E7 sequences that lack non-canonical cysteines (6%) are not capable of fast ROS sensing. On the other hand, 59% of the known E7 proteins perform ROS sensing via a reactive-cysteine mechanism involving one to three reactive non-canonical cysteines. The remaining E7 proteins sense ROS by a stepwise mechanism with one or two reactive non-canonical cysteines and resolutive one (C59).

3.7. Evolution of ROS sensing in HPV E7 proteins

Next, we considered whether the arrangements of non-canonical cysteines in natural E7 proteins support our model for redox sensing. Specifically, we asked if these arrangements differ significantly from a naïve model in which the appearance of non-canonical cysteine residues at different positions in the sequence is uncorrelated and simply follows the frequencies at the corresponding alignment column (see methods). The simplest arrangements for ROS sensing suggested by our model are a single, reactive non-canonical cysteine and a reactive-resolutive pair involving C59. These two arrangements are overrepresented in known viral types relative to our naïve model (Fig. 6A), suggesting that these configurations of the molecule have been favored during E7 evolution. The frequency of appearance of more complex arrangements in agreement with our model, such as two or three reactive cysteines and C59 together with two reactive cysteines, does not differ significantly from our naïve model (Fig. 6A). On the other hand, E7 sequences without non-canonical cysteines, with C59 as the single non-canonical cysteine or with more than three non-canonical cysteines do not fit with our model for ROS sensing. In agreement with our view, Fig. 6A shows that these three arrangements are underrepresented in known viral types relative to our naïve model, suggesting that these configurations of the molecule have been disfavored during E7 evolution. Overall, we find that the distribution of non-canonical cysteines in E7 sequences agrees better with our proposed mechanisms for redox sensing than with a naïve model that excludes coordinated evolution of non-canonical cysteines.

In addition to analyzing extant patterns of non-canonical cysteines in E7, we also performed a reconstruction of the natural history of redox sensing in E7 proteins. We simplified the observed patterns of non-canonical cysteines to six states, in colors in Fig. 6B: no non-canonical cysteines (white), one reactive cysteine (blue), two reactive cysteines (green), three reactive cysteines (black), C59 plus one reactive cysteine (yellow) and C59 plus two reactive cysteines (red). The scheme in Fig. 6B depicts our proposed mechanism for the evolution of redox sensing in E7. The top row corresponds to the absence of redox sensing, the middle row to ROS sensing by one or more reactive cysteines and the bottom row to ROS sensing by a reactive-resolutive cysteine pair. We propose that the acquisition or loss of a reactive cysteine can occur both in the absence and presence of C59, while C59 can appear only in the presence of one or two reactive cysteines. We used the phylogenetic hypothesis [43] and mapped the six redox sensing states to the leaves of the tree (Fig. 6C, squares). Using the Mesquite software, we reconstructed the ancestral redox sensing state of each internal node of the tree using Fisher parsimony (Fig. 6C, branch colors). Our reconstruction proposes an ancestral state with a single reactive cysteine and 22 evolutionary transitions, of which only in two for species 14 are unresolved (Fig. 6C, orange segment). We mapped into the scheme in Fig. 6B the number of evolutionary transitions in our reconstruction (fractional numbers are due to the two unresolved transitions). Among the proposed transitions, 20.33 (92.4%) are in agreement with our scheme, suggesting that our model describes the evolution of ROS sensing in E7 well. As a control, we generated redox state reconstructions starting from 100000 randomized sets of extant states, while keeping the tree structure intact. All randomized trees show a larger fraction of forbidden transitions (average 30%, standard deviation 6.3) than the natural tree (7.6%), further supporting our model. In all, both the analysis of extant sequences and the reconstruction of ancestral characters agree with our proposed model for the evolution of ROS sensing in E7 (Fig. 7).

4. Discussion

In this study we propose an integrated mechanism for ROS-sensing in the E7 oncoprotein from alpha genus papillomavirus, where all the high-risk strains belong. The redox-homeostasis in HPV-transformed cells is affected as a consequence to the NOX induction mediated by the papillomavirus oncoprotein expression [22]. On the other hand, the development of HPV-induced cancers could take years in a process that requires, the persistent expression of viral oncogenes [18] (i.e E7) during which the canonical and non-canonical cysteines are exposed to a potentially harmful oxidant stimuli. Paradoxically, we observed an evolutionary acquisition of reactive non-canonical cysteines in most (94%) of the E7 proteins of alpha genus. Our experimental data support that the E7s from alpha papillomavirus are able to sense and respond to redox stimuli (i.e ROS) of host cells and are equally able to manage the potentially harmful oxidant environment.

We showed that a ROS-reactivity hierarchy exist for the six cysteine residues present in the CRD domain of the HPV16E7 protein, where the non-canonical cysteines couple (C59-S-S-C68 form Ox II species) is the most reactive followed by the canonical cysteine couple C58 and C61 (form OX I species), while the canonical cysteines C91 and C94 remains unreactive under the assayed experimental conditions. Based on the pH dependence for the disulfide formation between C59 and C68, non-canonical residues in the HPV16 E7 protein are, by definition, reactive cysteines with a (thiol) pKa close to 7, one pH unit below the most abundant glutathione molecule, and with a pKa of 8.8 [41]. We observed that the ROS sensitivity of the E7 proteins from HPV16, HPV18 and HPV11 strains are comparable albeit with different arrangements of non-canonical cysteines. The multiple non-canonical cysteines arrangements could evolve from the variable redox environment that different papillomavirus strains suffer and affect the fitness for each virus life cycle.

We observed that the Ox I (C58-S-S-C61) species is increased in E7Δ1-26C59A or C68A single cysteine mutants when exposed to ROS,
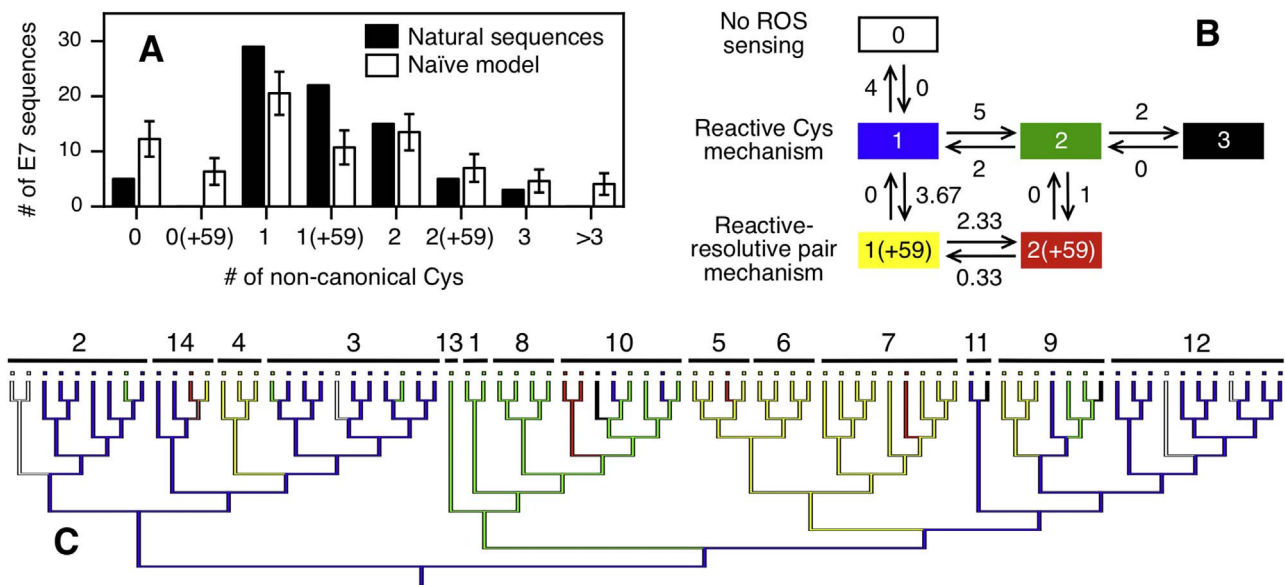
Fig. 6. Evolution of ROS sensing in HPV E7 proteins (A) Distribution of non-canonical cysteines in natural sequences (black bars) and a naïve model that posits independent evolution of non-canonical cysteines (white bars). Error bars are the standard deviation of 1000 replicas of the naïve model. (B) Proposed mechanism for the evolution of redox sensing in E7. The top row corresponds to the absence of redox sensing, the middle row to ROS sensing by one or more reactive cysteines and the bottom row to ROS sensing by a reactive-resolutive cysteine pair. Acquisition of a reactive cysteine can occur both in the absence or presence of C59, while appearance of C59 requires the presence of reactive cysteines. (C) Fitch parsimony [48] reconstruction of ancestral ROS sensing states of E7 proteins, mapped onto a phylogenetic tree of alpha papillomaviruses [43]. Six states are shown: no non-canonical cysteines (white), one reactive cysteine (blue), two reactive cysteines (green), three reactive cysteines (black), C59 plus one reactive cysteine (yellow) and C59 plus two reactive cysteines (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

strongly suggesting a chemical crosstalk between canonical and non-canonical residues. This observation challenges the notion that the canonical cysteines involved in Zn binding have only structural roles.

A common property between the HPV16, HPV11 and HPV18 E7 proteins is that they undergo only a small rearrangement of the overall secondary structure elements upon oxidation that is reflected by the
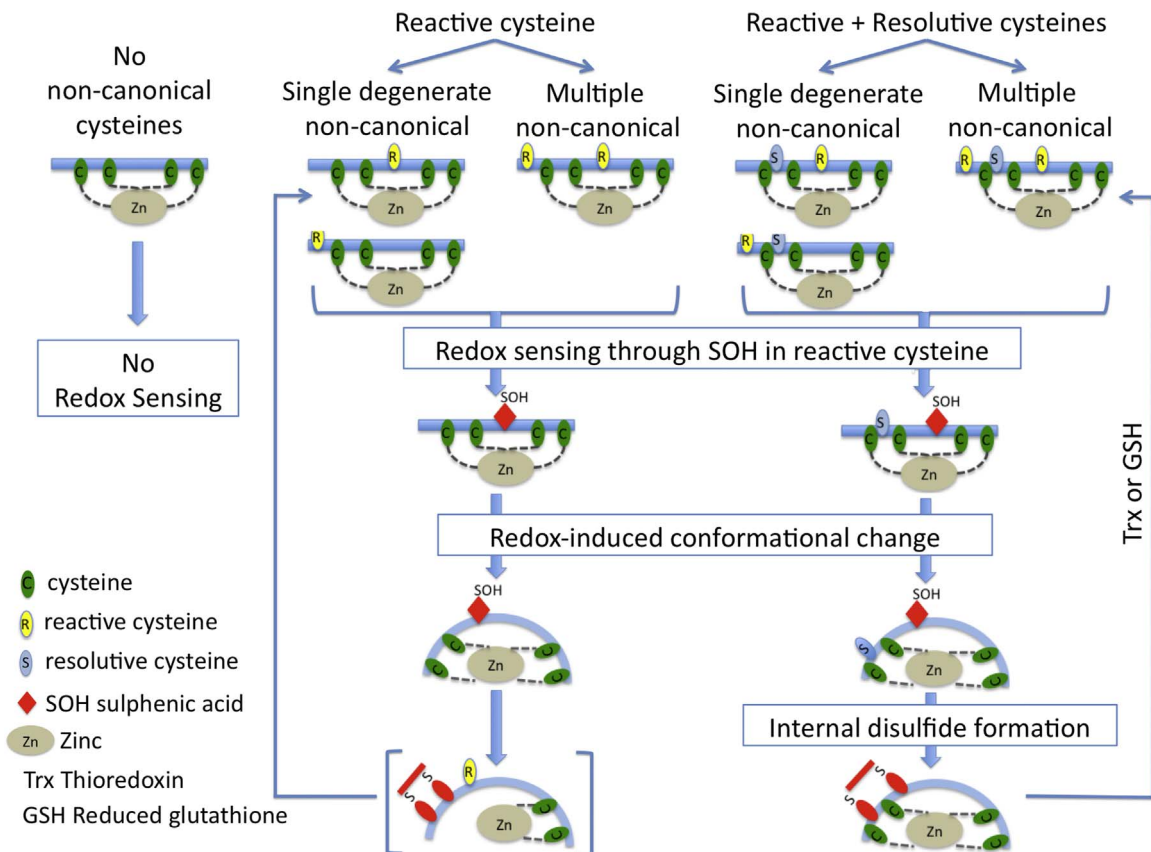
Fig. 7. Schematic representation of the ROS-sensing mechanisms of E7 proteins from alpha genus strains. ROS-insensitive E7s are devoid of non-canonical cysteines. All ROS-sensing E7s contain a reactive(s) non-canonical cysteine (s). The resolutive cysteine is present in a representative group of ROS-sensing sequences. A conformational change must be present in all ROS-sensing E7s to mediate signal transduction. Putative intermediates Ox I are enclosed in brackets.

marginal change observed in the CD spectra of the reduced and oxidized species. Despite of their differences in the primary sequences and cysteine rearrangements a similar change in the CD spectra, with a small decrease in the intensity of the band centered at 225 nm suggest a local unfolding, common to all three E7 proteins. For the HPV16 E7 protein we demonstrated that the intramolecular disulfide bond formation between non-canonical cysteine pair (C59-C68) is the major resultant of ROS and glutathione oxidation [29], a reaction that implies a long distance movement of the cysteines if we consider the 19.2 A that separates both sulfur atoms in the dimeric structure [24]. ROS sensing must therefore involve the formation of a sulfenic acid intermediate (ROS) or a mixed disulfide (GSH) in the reactive cysteine that trigger the conformational change to allow a nucleophilic substitution of a second (resolutive) cysteine to form the observed intramolecular disulfide bond. Statistically, 6.3 A is the distance at which two sulfur atoms have to be settled to form a disulfide bridge. The non-canonical cysteines in the HPV16 E7 protein lay outside the cut-off distance value, precluding the spontaneous formation of the disulfide bond in the absence of ROS. As no spontaneous oxidation is observed for the HPV11 and HPV18 E7s we can infer that reactive cysteines in these proteins are positioned in such a way that direct reaction is also precluded. ROS and more generally redox-sensing is based in the presence of a reactive (low pKa) cysteine and a transduction mechanism that involves a concerted conformational change.

The highly reactive C68 is central to the NLS for the HPV16 E7 [40] and hence the internal disulfide bond formation functions as a redox-controlled switch for masking the NLS sequence, a mechanism that is common to other redox regulated proteins such as Yap 1 [38]. This is a new level of regulation for the E7 moonlighting and the first time that a direct evidence of redox regulation is proposed for this oncoprotein. Subcellular localization of the HPV16 E7 protein strongly depends on the type of expression (i.e transient or constitutive) or cellular origin, with patterns that can be rather different. Transiently expressed non-fused E7 protein in U2OS cells detected with a specific antibody [39] or transiently expressed as a GFP-E7 fusion protein in HeLa cells directly detected [40], tend to be mostly nuclear. A totally different scenario is observed with the constitutive E7 expression in CasKi cells, or in cervical cancer adenocarcinoma biopsies [39] where a mostly cytoplasmatic localization pattern is clearly observed, indicating that the HPV16 E7 sub-cellular re-localization is a highly regulated process.

The many, yet barely conserved patterns of non-canonical cysteines in E7 proteins are intriguingly complex (Fig. 1). Our analysis of extant sequences shows that the cysteine arrangements considered in our model are strongly favored over other possible arrangements (Fig. 6A). This strongly suggests that E7 proteins do not present a random mixture of non-canonical cysteines but a wide set of specific patterns. In spite of this complexity, we were able to combine detailed biochemical and cell biology experiments on paradigmatic E7 proteins and mutants with analysis of natural sequences (Fig. 6) to propose a general model for redox sensing by E7 (Fig. 7). In short, there are three classes of E7 proteins in nature (Fig. 6B). Class 1 E7 proteins are devoid of non-canonical cysteines and are thus unable of redox sensing. Class 2 E7 proteins perform redox sensing through one to three reactive cysteines. The reactive cysteines need not be located at a specific sequence position but can appear at a seven different sites in the protein. Class 3 E7 proteins sense changes in the cellular redox potential by a combination of one or more reactive cysteines and the single resolutive cysteine C59. In the later case, the reactive cysteines may be also located at different sequence positions. The coexistence of two or one-cysteine mechanism has been demonstrated for peroxiredoxin and methionine sulphoxide reductase enzymes [44,45]. Reactive cysteines in Class 2 and Class 3 E7 proteins constitute a remarkable departure from the widespread paradigm of protein active sites appearing always at the same site of the protein.

The fate of the ROS-modified reactive cysteine can be either the formation of an internal disulfide bond between canonical cysteines of the Zn-binding motif (i.e C58-S-S-C61) or resolved by cellular redox molecules (proteins or low molecular weight thiols). The observed reactive cysteine sites are close in sequence and/or structure to the zinc binding cysteines [28], which is in agreement by resolution of the ROS-modified reactive cysteine through formation of an internal disulfide bond between canonical cysteines of the Zn-binding motif.

From an evolutionary standpoint, our parsimony reconstruction (Fig. 6C) suggests a Class 2 "reactive cysteine only" ancestral E7 protein. From this ancestral state, the data suggests that both Class 1 "ROS-insensitive" and Class 3 "reactive-resolutive pair" proteins have appeared multiple independent times during alpha papillomavirus evolution (4 blue-to-white, 3.67 blue-to-yellow and 1 green-to-red transitions). Moreover, the number of reactive cysteines in Class 2 and Class 3 E7 proteins has increased or decreased multiple independent times during alpha papillomavirus evolution (5 blue-to-green, 2 green-to-blue, 2 green-to-black and 2.33 yellow-to-red transitions). This high evolutionary plasticity of non-canonical cysteine patterns in E7 proteins is reminiscent of other pathogen short sequence motifs in E7 and in other pathogen proteins [46]. We speculate that these trends reflect the diversity of evolutionary pressure present in the evolution of alpha papillomaviruses, which lead to the observed diversity of redox sensing states. In addition, cysteine acquisition in E7 can be experimentally observed in natural intratype variants of HPV52. As reported by Aho et al., [47] some intratype variants of the HPV52 E7 show a non-conservative mutation in position 57 (HPV16E7 numeration) that leads to a replacement of the tyrosine, observed in the prototypical sequence, to cysteine. It should be noted that position 57 is one of the most frequently occupied position by non-canonical cysteines (Fig. 1).

Considering that the acquisition of reactive non-canonical cysteines is common both to the high and low risk HPV strains their presence must be related to a conserved regulatory function of this protein.

Whether the high prevalence of non-canonical reactive cysteines observed in the high risk HPV strains is related with the oncogenic potential of these E7s is unknown, it is clear that these proteins are exposed to a persistent redox stimulus during the long process of cellular transformation (the development of HPV-induced cancers could take years in a process that requires at least, the persistent expression of viral oncogenes).

E7 is a prototypical transforming protein with many properties shared by other DNA tumor virus oncoproteins that has evolved to sense the redox signals of the host cell through the acquisition of different motifs of reactive cysteines.

Funding sources

Notes

The authors declare no competing financial interest.

Acknowledgments

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.redox.2016.10.020.

References

[1] E.A. Veal, A.M. Day, B.A. Morgan, Hydrogen peroxide sensing and signaling, Mol. Cell 26 (2007) 1–14.

[2] S.G. Rhee, Cell signaling. H$_2$O$_2$, a necessary evil for cell signaling, Science 312 (2006) 1882–1883.

[3] D.R. Gough, T.G. Cotter, Hydrogen peroxide: a Jekyll and Hyde signalling molecule, Cell Death Dis. 2 (2011) e213.

[4] B. D'Autreaux, M.B. Toledano, ROS as signalling molecules: mechanisms that generate specificity in ROS homeostasis, Nat. Rev. Mol. Cell Biol. 8 (2007) 813–824.

[5] J.D. Lambeth, NOX enzymes and the biology of reactive oxygen, Nat. Rev. Immunol. 4 (2004) 181–189.

[6] D.P. Jones, H. Sies, The redox code, Antioxid. Redox Signal. 23 (2015) 734–746.

[7] A.T. Dinkova-Kostova, W.D. Holtzclaw, R.N. Cole, K. Itoh, N. Wakabayashi, Y. Katoh, M. Yamamoto, P. Talalay, Direct evidence that sulfhydryl groups of Keap1 are the sensors regulating induction of phase 2 enzymes that protect against carcinogens and oxidants, Proc. Natl. Acad. Sci. USA 99 (2002) 11908–11913.

[8] M. Putker, H.R. Vos, K. van Dorenmalen, H. de Ruiter, A.G. Duran, B. Snel, B.M. Burgering, M. Vermeulen, T.B. Dansen, Evolutionary acquisition of cysteines determines FOXO paralog-specific redox signaling, Antioxid. Redox Signal. 22 (2015) 15–28.

[9] M. Pu, A.A. Akhand, M. Kato, M. Hamaguchi, T. Koike, H. Iwata, H. Sabe, H. Suzuki, I. Nakashima, Evidence of a novel redox-linked activation mechanism for the Src kinase which is independent of tyrosine 527-mediated regulation, Oncogene 13 (1996) 2615–2622.

[10] S. Reuter, S.C. Gupta, M.M. Chaturvedi, B.B. Aggarwal, Oxidative stress, inflammation, and cancer: how are they linked?, Free Radic. Biol. Med. 49 (2010) 1603–1616.

[11] B. Halliwell, Oxidative stress and cancer: have we moved forward?, Biochem. J. 401 (2007) 1–11.

[12] F. De Marco, Oxidative stress and HPV carcinogenesis, Viruses 5 (2013) 708–731.

[13] F. De Marco, E. Bucaj, C. Foppoli, A. Fiorini, C. Blarzino, K. Filipi, A. Giorgi, M.E. Schinina, F. Di Domenico, R. Coccia, D.A. Butterfield, M. Perluigi, Oxidative stress in HPV-driven viral carcinogenesis: redox proteomics analysis of HPV-16 dysplastic and neoplastic tissues, PLoS One 7 (2012) e34366.

[14] V.M. Williams, M. Filippova, U. Soto, P.J. Duerksen-Hughes, HPV-DNA integration and carcinogenesis: putative roles for inflammation and oxidative stress, Future Virol. 6 (2011) 45–57.

[15] C. Foppoli, F. De Marco, C. Cini, M. Perluigi, Redox control of viral carcinogenesis: the human papillomavirus paradigm, Biochim Biophys. Acta 1850 (2015) 1622–1632.

[16] H. zur Hausen, Papillomavirus infections–a major cause of human cancers, Biochim. Biophys. Acta 1288 (1996) F55–F78.

[17] N. Munoz, F.X. Bosch, S. de Sanjose, R. Herrero, X. Castellsague, K.V. Shah, P.J. Snijders, C.J. Meijer, Epidemiologic classification of human papillomavirus types associated with cervical cancer, N. Engl. J. Med 348 (2003) 518–527.

[18] T. Crook, J.P. Morgenstern, L. Crawford, L. Banks, Continued expression of HPV-16 E7 protein is required for maintenance of the transformed phenotype of cells co-transformed by HPV-16 plus EJ-ras, EMBO J. 8 (1989) 513–519.

[19] M.L. Looi, A.Z. Mohd Dali, S.A. Md Ali, W.Z. Wan Ngah, Y.A. Mohd Yusof, Oxidative damage and antioxidant status in patients with cervical intraepithelial neoplasia and carcinoma of the cervix, Eur. J. Cancer Prev. 17 (2008) 555–560.

[20] A. Sattayakhom, W. Chunglok, W. Ittarat, W. Chamulitrat, Study designs to investigate Nox1 acceleration of neoplastic progression in immortalized human epithelial cells by selection of differentiation resistant cells, Redox Biol. 2 (2013) 140–147.

[21] W. Chamulitrat, R. Schmidt, P. Tomakidi, W. Stremmel, W. Chunglok, T. Kawahara, K. Rokutan, Association of gp91phox homolog Nox1 with anchorage-independent growth and MAP kinase-activation of transformed human keratinocytes, Oncogene 22 (2003) 6045–6053.

[22] R. Marullo, E. Werner, H. Zhang, G.Z. Chen, D.M. Shin, P.W. Doetsch, HPV16 E6 and E7 proteins induce a chronic oxidative stress response via NOX2 that causes genomic instability and increased susceptibility to DNA damage in head and neck cancer cells, Carcinogenesis 36 (2015) 1397–1406.

[23] L.G. Alonso, M.M. Garcia-Alai, A.D. Nadra, A.N. Lapena, F.L. Almeida, P. Gualfetti, G.D. Prat-Gay, High-risk (HPV16) human papillomavirus E7 oncoprotein is highly stable and extended, with conformational transitions that could explain its multiple cellular binding partners, Biochemistry 41 (2002) 10510–10518.

[24] X. Liu, A. Clements, K. Zhao, R. Marmorstein, Structure of the human Papillomavirus E7 oncoprotein and its mechanism for inactivation of the retino-blastoma tumor suppressor, J. Biol. Chem. 281 (2006) 578–586.

[25] O. Ohlenschlager, T. Seiboth, H. Zengerling, L. Briese, A. Marchanka, R. Ramachandran, M. Baum, M. Korbas, W. Meyer-Klaucke, M. Durst, M. Gorlach,

[26] B. Todorovic, P. Massimi, K. Hung, G.S. Shaw, L. Banks, J.S. Mymryk, Systematic analysis of the amino acid residues of human papillomavirus type 16 E7 conserved region 3 involved in dimerization and transformation, J. Virol. 85 (2011) 10048–10057.

[27] A. Clements, K. Johnston, J.M. Mazzarelli, R.P. Ricciardi, R. Marmorstein, Oligomerization properties of the viral oncoproteins adenovirus E1A and human papillomavirus E7 and their complexes with the retinoblastoma protein, Biochemistry 39 (2000) 16033–16045.

[28] L.B. Chemes, J. Glavina, L.G. Alonso, C. Marino-Buslje, G. de Prat-Gay, I.E. Sanchez, Sequence evolution of the intrinsically disordered and globular domains of a model viral oncoprotein, PLoS One 7 (2012) e47661.

[29] L.B. Chemes, G. Camporeale, I.E. Sanchez, G. de Prat-Gay, L.G. Alonso, Cysteine-rich positions outside the structural zinc motif of human papillomavirus E7 provide conformational modulation and suggest functional redox roles, Biochemistry 53 (2014) 1680–1696.

[30] L.G. Alonso, M.M. Garcia-Alai, C. Smal, J.M. Centeno, R. Iacono, E. Castano, P. Gualfetti, G. de Prat-Gay, The HPV16 E7 viral oncoprotein self-assembles into defined spherical oligomers, Biochemistry 43 (2004) 3310–3317.

[31] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797.

[32] W.P. Maddison, D.R. Maddison, Mesquite: A Modular System for Evolutionary Analysis, Version 3.10, 2016.

[33] S.M. Marino, V.N. Gladyshev, Cysteine function governs its conservation and degeneration and restricts its utilization on protein surfaces, J. Mol. Biol. 404 (2010) 902–916.

[34] N. Eswar, B. Webb, M.A. Marti-Renom, M.S. Madhusudhan, D. Eramian, M.Y. Shen, U. Pieper, A. Sali, Comparative protein structure modeling using Modeller, Curr. Protoc. Bioinformatics (2006) (Chapter 5, Unit 5 6).

[35] R. Sanchez, M. Riddle, J. Woo, J. Momand, Prediction of reversibly oxidized protein cysteine thiols using protein structure properties, Protein Sci. 17 (2008) 473–481.

[36] N. Brandes, S. Schmitt, U. Jakob, Thiol-based redox switches in eukaryotic proteins, Antioxid. Redox Signal. 11 (2009) 997–1014.

[37] S.O. Kim, K. Merchant, R. Nudelman, W.F. Beyer Jr., T. Keng, J. DeAngelo, A. Hausladen, J.S. Stamler, OxyR: a molecular code for redox-related signaling, Cell 109 (2002) 383–396.

[38] M.J. Wood, G. Storz, N. Tjandra, Structural basis for redox regulation of Yap1 transcription factor localization, Nature 430 (2004) 917–921.

[39] K. Dreier, R. Scheiden, B. Lener, D. Ehehalt, H. Pircher, E. Muller-Holzner, U. Rostek, A. Kaiser, M. Fiedler, S. Ressler, S. Lechner, A. Widschwendter, J. Even, C. Capesius, P. Jansen-Durr, W. Zwerschke, Subcellular localization of the human papillomavirus 16 E7 oncoprotein in CaSki cells and its detection in cervical adenocarcinoma and adenocarcinoma in situ, Virology 409 (2011) 54–68.

[40] J. Eberhard, Z. Onder, J. Moroianu, Nuclear import of high risk HPV16 E7 oncoprotein is mediated by its zinc-binding domain via hydrophobic interactions with Nup62, Virology 446 (2013) 334–345.

[41] C.E. Paulsen, K.S. Carroll, Cysteine-mediated redox signaling: chemistry, biology, and tools for discovery, Chem. Rev. 113 (2013) 4633–4679.

[42] A. Hall, K. Nelson, L.B. Poole, P.A. Karplus, Structure-based insights into the catalytic power and conformational dexterity of peroxiredoxins, Antioxid. Redox Signal. 15 (2011) 795–815.

[43] I.G. Bravo, S. de Sanjose, M. Gottschling, The clinical importance of understanding the evolution of papillomaviruses, Trends Microbiol. 18 (2010) 432–438.

[44] B.C. Lee, A. Dikiy, H.Y. Kim, V.N. Gladyshev, Functions and evolution of selenoprotein methionine sulfoxide reductases, Biochim. Biophys. Acta 1790 (2009) 1471–1477.

[45] A. Hall, K. Nelson, L.B. Poole, P.A. Karplus, Structure-based insights into the catalytic power and conformational dexterity of peroxiredoxins, Antioxid. Redox Signal., 15, pp. 795–815.

[46] L.B. Chemes, G. de Prat-Gay, I.E. Sanchez, Convergent evolution and mimicry of protein linear motifs in host-pathogen interactions, Curr. Opin. Struct. Biol. 32 (2015) 91–101.

[47] J. Aho, C. Hankins, C. Tremblay, F. Lang, P. Forest, K. Pourreaux, F. Rouah, F. Coutlee, Molecular analysis of human papillomavirus type 52 isolates detected in the genital tract of human immunodeficiency virus-seropositive and -seronegative women, J. Infect. Dis. 188 (2003) 1517–1527.

[48] W.M. Fitch, Toward defining the course of evolution: minimum change for a specific tree topology, Syst. Zool. 20 (1971) 406–416.

[49] J.B. Hunt, S.H. Neece, A. Ginsburg, The use of 4-(2-pyridylazo)resorcinol in studies of zinc release from Escherichia coli aspartate transcarbamoylase, Anal. Biochem. 146 (1985) 150–157.

Solution structure of the partially folded high-risk human papilloma virus 45 oncoprotein E7, Oncogene 25 (2006) 5953–5959.