J Mol Model (2013) 19:4025–4037
DOI 10.1007/s00894-013-1915-8

ORIGINAL PAPER

Modeling and molecular dynamics of the intrinsically disordered e7 proteins from high- and low-risk types of human papillomavirus

Nilson Nicolau-Junior · Silvana Giuliatti

Received: 16 May 2012 / Accepted: 9 June 2013 / Published online: 18 July 2013

Abstract Cervical cancer affects millions of women worldwide each year. Most cases of cervical cancer are caused by the sexually transmitted human papillomavirus (HPV). The approximately 40 HPV types that infect the cervix are designated high- or low-risk based on their potential to lead to high-grade lesions and cancer. The HPV E7 oncoprotein is directly involved in the onset of cervical cancer and associates with the pRb protein and other cellular targets that promote cell immortalization and carcinogenesis. This is the first description of the modeling and molecular dynamics analysis of complete three-dimensional structures of high-risk (HPV types 16 and 18), low-risk (HPV type 11), and HPV type 01 E7 proteins. The models were constructed by a hybrid approach using homology (MODELLER) and ab initio (Rosetta) modeling, and the protein dynamics were simulated for 50 ns under normal pressure and temperature (NPT) conditions. The intrinsic disorder of the E7 protein sequence was assessed in silico. Complete models of E7 were obtained despite the predicted intrinsic disorder of the N-termini from the high-risk HPV types. The N-terminal domains of all of the E7 proteins studied, even those from high-risk strains, exhibited secondary structure after modeling. Trajectory analysis of E7 proteins from HPV types 16 and 18 showed higher instability in their N-terminal domains than in those of HPV types 11 and 01; however, this variation did not affect the secondary structure during the simulation. ANCHOR analysis indicated that the CR1 and CR2 regions of HPV types 16 and 18 contain possible targets for future drug-discovery studies.

N. Nicolau-Junior · S. Giuliatti (✉)
Department of Genetics, School of Medicine of Ribeirão Preto, University of São Paulo, Av. Bandeirantes, 3900, Ribeirão Preto, SP 14049-900, Brazil
e-mail: silvana@fmrp.usp.br

Keywords Homology modeling · Ab initio modeling · Human papillomavirus · E7 · Molecular dynamics

Introduction

Cervical cancer is the second-most common cancer among women [1]. In most cases, infection with human papillomavirus (HPV) is necessary to trigger the onset of cervical cancer. Cervical HPV infection is sexually transmitted, and most women are infected soon after becoming sexually active. More than 100 types of HPV have been identified, of which 40 infect the genital tract [2]. Most individuals remain asymptomatic after acquiring HPV infection, and only a small percentage develop clinically or histologically identifiable lesions during the course of the disease [3]. Epidemiological and molecular studies have shown that human papillomavirus induces epithelial cell transformation during infection. Specific types of HPV appear to be essential for the development of premalignant lesions and the onset of invasive cervical cancer [4]. The HPV types that infect the genital tract can be classified into two groups, high-risk and low-risk, based on their relative risks of high-grade lesions and cervical cancer progression [5]. The low-risk types, the most common of which are HPV types 6 and 11, induce only benign genital warts. The high-risk group, comprising types 16, 18, 31, 33, 45, and 56, is associated with the development of anogenital cancers and may be detected in 99 % of cervical cancers [3].

Persistent infection with high-risk HPVs is a necessary etiological factor for developing cervical cancer. The most prevalent high-risk types, 16 and 18, are associated with approximately 70 % of cervical cancers worldwide [6, 7]. The high-risk HPV oncoproteins E6 and E7 are essential for HPV-induced cellular immortalization, transformation, and carcinogenesis [8] and act by interacting with various cellular targets

[9]. The E7 protein is located primarily in the nucleus, where it associates with the retinoblastoma gene product (pRb) to facilitate progression of S phase of the cell cycle [6]. In normal cells, pRb is hypophosphorylated at the beginning of G1 phase and binds to E2F transcription factors to form complexes that act as transcriptional repressors. The association of E7 protein with the hypophosphorylated pRb prevents its binding to E2F, thereby promoting cell cycle progression [9]. Sedimentation equilibrium experiments have shown that E7 has oligomerization properties and can be found as monomers, dimers, and tetramers, but under physiological conditions it is primarily dimeric [10]. However, stoichiometry and sub-nanomolar affinity analyses indicate that E7 binds to pRb as a monomer [11]. E7 has three conserved regions (CR) based on its homology with the adenovirus E1A protein: CR1, CR2, and CR3 [12]. The CR1 and CR2 regions are located in the N-terminal domain of the protein (amino acids 1–40 for HPV type 16); this domain has properties of intrinsically disordered proteins (IDPs). Two attributes are regarded as the main cause of protein intrinsic disorder: a low hydrophobicity, which prevents the formation of a stable globular core, and a high net charge that favors extended states of the protein due to electrostatic repulsion [13]. Further, IDPs are always depleted in amino acids of low flexibility indexes (hydrophobic amino acids), and enriched in amino acids of high flexibility indexes (polar and charged amino acids), termed as order-promoting and disorder-promoting amino acids, respectively [14]. In E7 from HPV16, the disorder of the N-terminal domain confers flexibility, conformational transitions, and structural adaptability to targets of this oncoprotein [15].

The E7 CR1 region (residues 2–15) is known to compete with E2F transcription factors and to participate in the destabilization of pRb, although is it not clear which specific amino acids are involved in these activities [16]. The CR2 region (residues 16–40) has a specific amino acid sequence, the LXCXE motif, that has a high binding affinity for the pRb protein and is necessary for the association between E7 and pRb [17–19]. The CR3 region is located in the C-terminal domain of the protein and has 2 CXXC amino acid motifs that bind to a zinc atom [20], giving CR3 the characteristics of a zinc-finger motif. The CR3 region also cooperates with the CR1 and CR2 regions to inhibit pRb and transform the cell. According to Patrick et al. [21], the full E7 protein containing the CR3 domains binds to pRb with an affinity approximately 100-fold higher than those of E7 fragments that lack the CR3 region. Studies of dimeric HPV type 1A E7-CR3 implicated two conserved surface patches as contact sites; patch 1 makes low-affinity contacts with the basic region within the C domain of pRb, while patch 2 makes contacts with the marked-box domain in the C-terminal region of E2F (E2F-MB) [22]. Liu et al. [22] suggested that the binding of hypothetical small molecules to patch 1 and/or 2 of the E7 CR3 region might disrupt the ability of E7 to perturb pRb function, and the development of such molecules could open a new avenue of HPV therapeutics.

The recent development of vaccines against HPV represents a major advance in human cancer prevention methods. Despite great expectations and the promising results of clinical trials, there is insufficient evidence of an effective vaccine against cervical cancer, and the overall effect of these vaccines remains unknown [23]. Moreover, the vaccine is expensive and will be a great challenge to implement where it is most needed and will have the most significant impact, such as in low-income countries [24]. In addition to preventive strategies such as vaccines, therapeutic strategies remain of interest to the millions of women who are currently infected [25]. No effective treatment methodology currently exists for individuals who are already infected with HPV or have developed cervical cancer due to the infection. This study aimed to model and analyze the dynamics of the structure of E7 from high- and low-risk types of HPV. The resulting knowledge of E7 structure may be used in further studies of protein–protein and protein–ligand interactions to create new possibilities for the treatment of HPV infection and/or cervical cancer.

Methods

In silico intrinsic disorder analysis

A wide range of methods and algorithms for the detection of intrinsic disorder are currently available. In silico protein disorder analysis of the E7 proteins from HPV types 11 (low risk), 16 and 18 (high risk), and 01 (not related to cervical infection and cancer) was performed using PONDR-FIT [26] and DisCon [27], which detect both the contribution of each residue to the disorder and the disorder content of the protein. ANCHOR [28] analysis, which seeks to identify segments that reside in disordered regions and gain stabilizing energy by interacting with globular proteins, was performed on the E7 proteins in order to predict disordered binding regions.

Macromolecular modeling

Template selection and homology modeling of E7 from HPV type 16

The E7 sequence from HPV type 16 (P03129) was retrieved from the UniProt database (http://www.uniprot.org) and modelled partially by homology using the NMR solution structure from the Protein Data Bank (PDB) of the C-terminal domain (monomer) of the HPV type 45 oncoprotein E7 (PDB ID: 2EWL), which has 65 % similarity with E7 from HPV type 16, as a template. The homology modeling was performed using MODELLER [29] implemented on the @TOME2 server (http://atome.cbs.cnrs.fr) [30].

Ab initio modeling

The N-terminal region of the model of E7 from HPV type 16 appeared unstructured due to its intrinsically disordered nature. The ab initio modeling technique was used to generate a complete three-dimensional (3D) protein molecule. Ab initio modeling is used when there is little initial structural information about the molecule. Rosetta 3.1 [31], more specifically the protocol of loop modeling using the cyclic coordinate descent (CCD) algorithm [32], was used for this purpose. About 1,000 models were generated, one of which was chosen after assessment and validation using model quality assessment predictors (MQAPs) and visual inspection. MQAPs are used to evaluate the correctness and quality of protein models generated with or without the aid of experimental information [33]. The HPV type 16 E7 model was used as a template for homology modeling of the E7 proteins from HPV types 01 (P06465), 11 (P04020), and 18 (P06788).

Homology modeling of E7 from HPV types 01, 11, and 18

MODELLER (v9.10) was used to generate homology models of the E7 proteins from HPV types 01, 11, and 18 using the previously generated structure of the HPV type 16 E7 protein as a template. Twenty models of each protein were generated and then assessed using PROCHECK [34].

Rotamer modeling and zinc atom coordination

In order to obtain high-quality structures, the side chains (rotamers) of all protein models were re-modelled using SCWRL4 [35]. The coordinates of the zinc atoms of the zinc finger in the structure 2EWL (PDB) were transferred manually to the homology models using *Accelrys* Discovery Studio Visualizer v2.0 (http://accelrys.com/).

Structure validation

Structure validation and, more particularly, the choice of the near-native models demand not only visual analysis of the structure but local (each amino acid residue) and global (whole protein) validation of the 3D protein structure. Several MQAPs were used to select the most accurate E7 protein models. In the first step, the ModFOLD [36], dDFIRE [37], and QMEAN [38] scoring functions were used to rank the 1,000 decoys from the model of E7 from HPV type 16. The 10 top-scoring structures of each MQAP continued in the analysis; there was no redundancy in the top 10 models selected by each program, so the resultant 30 structures were ranked again based on the PROCHECK results. In order to choose the best-quality structures, the top 10 remaining models were then submitted to the following MQAPs: Protein Quality Predictor (ProQ) [39], which uses quality measures called LGscore [40] and MaxSub [41], ERRAT [42], Verify 3D [43, 44], and Protein Volume Evaluation (PROVE) [45].

Simulation of molecular dynamics

The molecular dynamics (MD) of the E7 proteins were determined using the CHARMM27 force field [46] implemented in the NAMD 2.8 software [47]. The MD simulation was performed over 50 ns for each protein (E7 from HPV types 01, 11, 16, and 18) for a total of 0.20 μs using a constant temperature (310 K) and pressure (1 atm). Each system was solvated in a cubic box of TIP3P water with a minimum distance of 10 Å imposed between the solute atoms and the edge of the box. The system charges were balanced by adding chloride and sodium ions. Non-bonded interactions were cut off at 12 Å and the particle mesh Ewald method (PMEM) was used to determine long-distance electrostatic interactions. All bonds involving hydrogen were fixed using the SHAKE algorithm [48] and an integration time step of 2 fs was used. MD temperature control was performed by Langevin dynamics. The number of atoms in each system was 17,551, 17,081, 17,331, and 20,034 for HPV types 01, 11, 16, and 18, respectively. Before the MD simulation, the entire system was minimized for 5,000 steps using a conjugate gradient and line search algorithm (NAMD default) to reduce any unfavorable interactions between the protein and solvent. Each system was slowly heated up to 310 K over the first 52 ps under NPT conditions. For the 50 ns production run, snapshots were stored every 0.5 ps, which yields an ensemble with 10^5 snapshots.

Trajectory analysis

The trajectory obtained in the MD simulation was analyzed using VMD [49], eucb [50], and carma [51]. Carma and VMD were used to transform the complete trajectory data (solvent, Cl, Na, Zn, and protein) into a trajectory including only the protein structure; this then served as the input for the eucb program. Two eucb protocols were used to analyze the trajectories: the root mean-square deviation (RMSD) and root mean-square fluctuation (RMSF) of the α-C atoms using the initial structure as the reference. The variations in the secondary structures over time were analyzed using the STRIDE [52] algorithm implemented in the TIMELINE plug-in from VMD.

Results and discussion

The E7 proteins from high- and low-risk HPV types are intrinsically disordered

Analyses of the E7 sequences performed with the program DisCon (Table 1) showed that the E7 proteins from HPV

Table 1 Analysis of the disorder content of the E7 proteins from HPV types 01, 11, 16, 18 using DisCon

Accession number	HPV type	Disorder content (%)
P03129	16	28.57
P06788	18	10.47
P04020	11	1.02
P06465	1	0

types 16 and 18 had greater contents of disorder, at 28.57 % and 10.47 %, respectively, whereas this content was very low (~1.0) for the low-risk type 11 and zero for type 01. According to Uversky et al. [53], the intrinsic disorder of the E7 protein increases the risk for cervical cancer; therefore, the E7 proteins of the low-risk HPV types 06 and 11 are less disordered, that of HPV type 16 is the most disordered, and that of HPV type 18 is intermediate. Interestingly, proteins linked to other types of cancer, such as P53 [54] and BRCA1 [55], also have properties of intrinsically disordered proteins.

PONDR-FIT is used to determine which amino acids in the protein backbone are disordered; scores above 0.5 represent amino acids with intrinsically disordered characteristics (Fig. 1). All of the E7 proteins studied showed increased disorder in the first and last ten amino acids. The E7 proteins from HPV types 16 and 18 showed higher levels of disorder, and the greatest contribution to this disorder came from their N-terminal domains, in which the scores for almost all amino acids were above the threshold value of 0.5. The PONDR-FIT scores of HPV type 11 were lower than those of the high-risk HPVs despite its high sequence similarity (70 %) with type 16. As for the results obtained using DisCon, the PONDR-FIT values of disorder for the HPV type 01 N-terminal domain were below the threshold, indicating the absence of disorder. As HPV type 01 is not related to genital tract infection and cancer, this result indicates that the presence of disorder may be linked to the types of HPV that lead to the onset of cancer. According to Alaí-Garcia et al. [15], the N-terminus of HPV type 16 E7 can be described as intrinsically disordered and has a plastic structural domain that probably optimizes the speed of interactions and allows less-specific but selective connections to multiple host-cell proteins. The IDPs have been classified based on their modes of action [13], and the E7 proteins from high-risk HPVs could be classified as *effector* IDPs because their disorder modifies the activity of a single partner or assembly (in this case, the prB protein).

Molecular modeling of E7

The HPV type 16 E7 models generated by the combination of homology modeling and ab initio techniques were validated, and MQAP scores (Table 2) and visualization was used to select the best model. Although the E7 proteins from high-risk HPV types showed IDP properties, an initial structure of the E7 protein of HPV type 16 was obtained successfully and used as a template to model the E7 proteins of the other types. These models were used as starting structures for MD simulations. The HPV type 16 E7 protein is an alpha-beta protein containing five α-helixes, three of which are in the N-terminal domain, and one antiparallel β-sheet in the C-terminal domain (Fig. 3c). The C-terminal domain also contains 2 CXXC sequences that together form a zinc-finger motif.

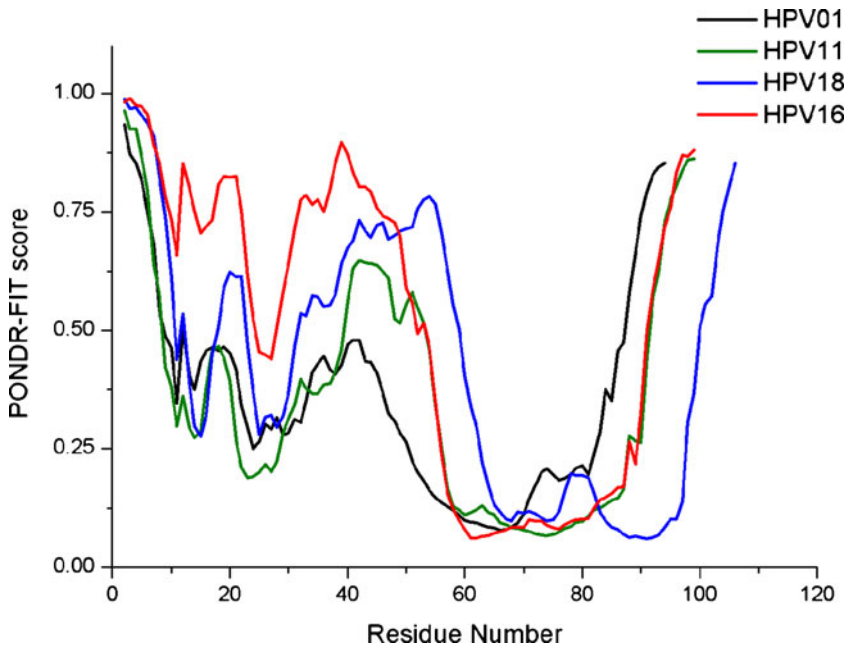

Fig. 1 Analysis of the intrinsic disorder along the amino acids chains of the E7 proteins from human papilloma virus (HPV) types 01, 11, 16, and 18 using PONDR-FIT

Table 2 Selection of best model of human papilloma virus (HPV) type 16 E7 based on the model quality assessment predictor (MQAP) scores

Model	PROCHECK (%)	Lg score	MaxSub	ERRAT (%)	Verify3D (%)	Z-score (mean)	Z-score (RMS)
S_0652	94.2/5.8/0/0[a]	0.189	0.077	84.44	14.14	0.06[a]	1.33
S_0139	94.2/5.8/0/0[a]	0.042	0.022	96.67[a]	34.34	0.06[a]	1.3
S_0885	94.2/5.8/0/0[a]	0.052	0.049	96.67[a]	55.56	0.14	1.37
S_0911	93.0/7.0/0/0[a]	0.034	0.057	89.65	41.41	0.14	1.45
S_0109	93.0/7.0/0/0[a]	0.124	0.069	98.89[a]	21.21	0.02[a]	1.29
S_0101	93.0/7.0/0/0[a]	−0.288	−0.002	93.33	48.48	0.13	1.33
S_0382	93.0/7.0/0/0[a]	0.471	0.11	94.44	77.78	−0.1	1.34
S_0994	91.9/8.1/0/0[a]	0.41	0.048	82.22	22.22	0.14	1.33
S_0515	91.9/8.1/0/0[a]	0.484	0.15	100	27.27	−0.11	1.11
S_0415	91.9/8.1/0/0[a]	−0.061	−0.019	91.9	43.43	0.05[a]	1.17

[a] Scores expected to indicate good structural quality

The structure selected, initially called S_0139, was used as the template for homology modeling of the E7 proteins of HPV types 01, 11, and 18, which exhibit 56.4 %, 70 %, and 61.7 % sequence similarity, respectively, with that of HPV 16. Homology modeling generated 20 models for each protein, and the best models were chosen using PROCHECK.

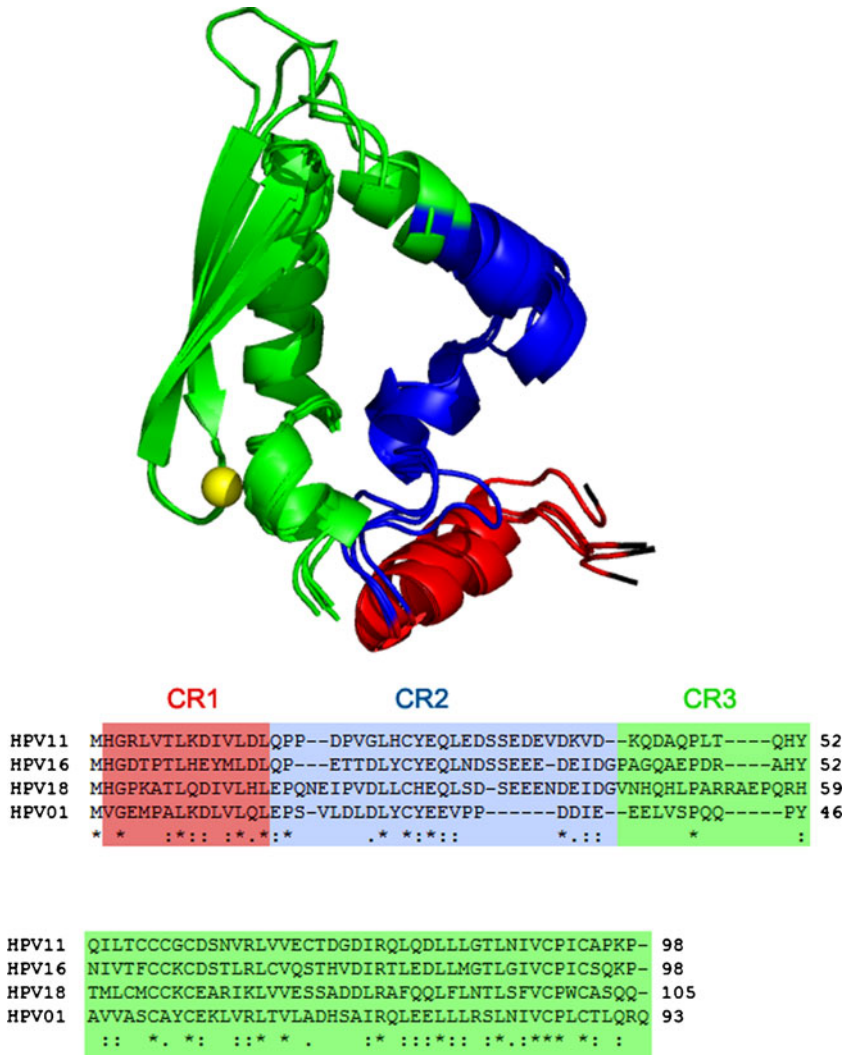

Fig. 2 Structural superposition and sequence alignment [57] of the E7 protein from HPV 01, 11, 16, 18 and their respective CR regions; *red* CR1, *blue* CR2, *green* CR3

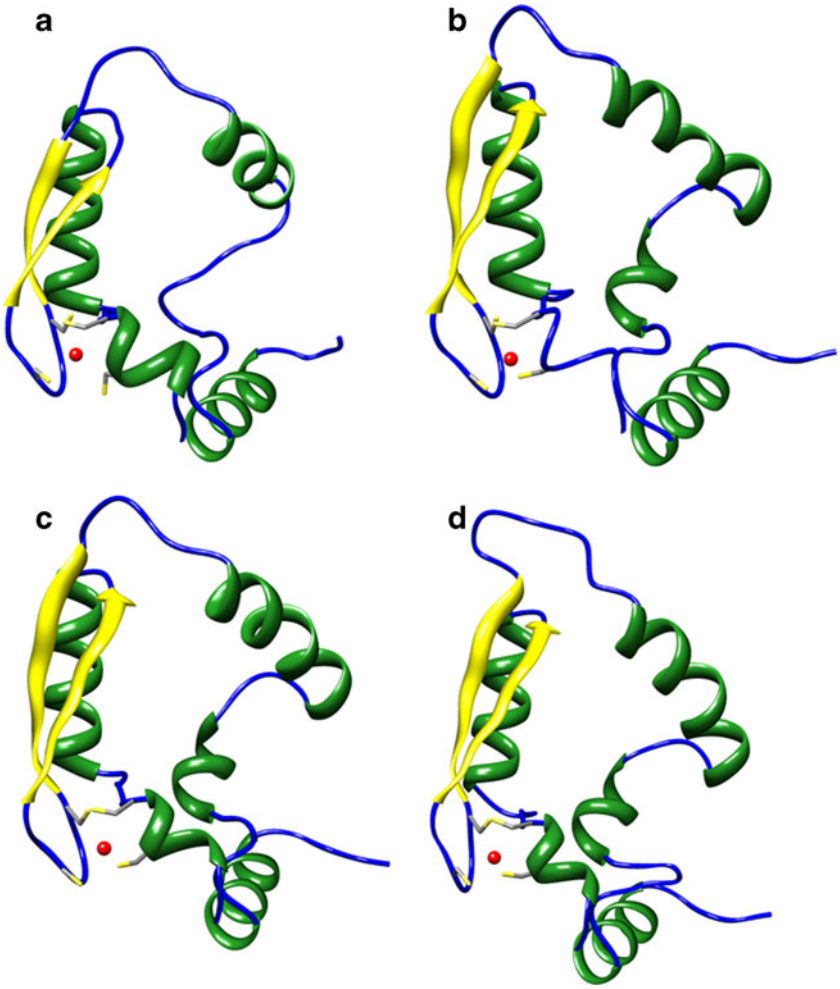

Fig. 3a–d Refined structures of E7 (*green* alpha helix,*yellow* beta sheets, *blue* loops) with the zinc atom (*red*). **a** HPV type 01, **b** HPV type 11, **c** HPV type 16, and **d** HPV type 18. The figures were generated using CHIMERA [58]

The E7 proteins were submitted to rotamer modeling in order to refine the side chains of the protein models. Analysis performed with DaliLite [56] showed that the HPV type 01, 11, and 18 proteins exhibited distances of 1.9 Å, 0.5 Å, and 0.8 Å, respectively, from the template E7 protein from HPV type 16. Zinc atoms coordinated with the CXXC sequences were added in order to prepare the structures for the MD step. Visualization of the structural superposition and a sequence alignment of all modelled E7 and their respective CRs regions show that they have very similar primary, secondary and tertiary structures (Fig. 2).

Visualization of the CR2 region from the E7 N-terminus in the initial models of the proteins from all HPV types showed that all of the models except that of HPV type 01 E7 contained an alpha helix spanning the LXCXE sequence (Fig. 3), which is known to have binding affinity for the pRb protein [17]. This information, as well as the sequence similarity (56.4 %) and structural distance (1.9 Å) of HPV type 01 from HPV type 16, corroborate the notion that HPV type 01 is not related to cervical infections or, consequently, to the onset of cervical cancer. This is another indication that the increased content of disorder of E7 may be related directly to the development of high-grade lesions and cancer.

Molecular dynamics analysis

In order to assess the stability of and to refine the E7 molecular structures, the models were subjected to 50-ns MD simulations, during which, the structural variations of the four systems were measured using RMSD analysis. The mean RMSD values of the HPV type 01, 11, 16, and 18 proteins were 7.88 (±0.9), 5.68 (±0.75), 7.36 (±1.53) and 8.26 (±1.00) Å, respectively (Fig. 4). All of the E7 proteins studied showed high mean distances from their initial structures, and all the proteins stabilized considerably after approximately 5 ns when compared to HPV16, showing means of 8.13 (±0.38), 5.85 (±0.47), 7.7 (±1.19) and 8.49 (±0.54) Å for types 01, 11, 16 and 18 respectively. The E7 from HPV type 16 showed the highest standard deviation after the initial 5 ns, and the absence of a flat RMSD curve may indicate instability. The E7 from HPV type 11 exhibited the lowest mean RMSD and a low standard deviation despite the

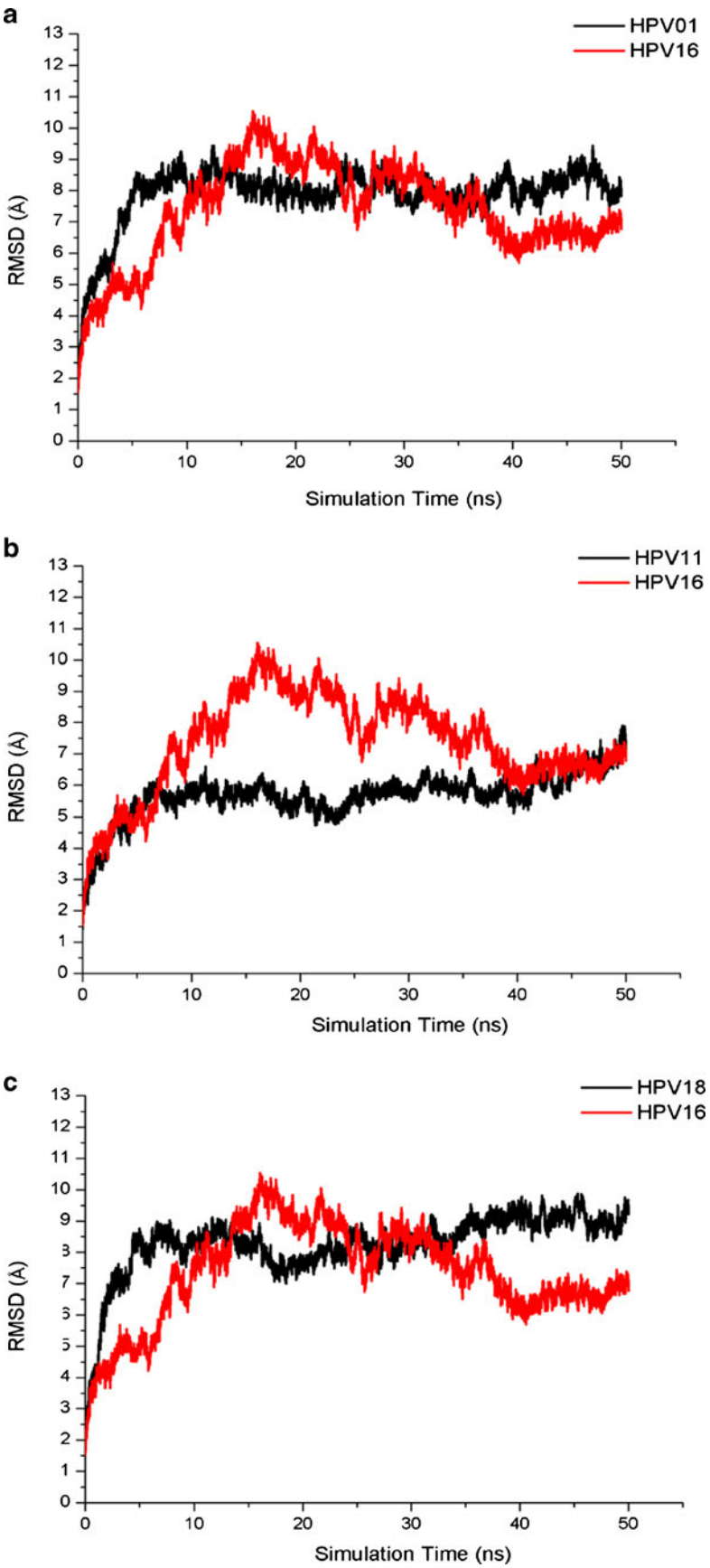

Fig. 4a–c Root mean-square deviations (RMSD) of the E7 Cα atom coordinates calculated as a function of the simulation time. **a** HPV type 16 vs HPV type 01. **b** HPV type 16 vs HPV type 11. **c** HPV type 16 vs HPV type 18

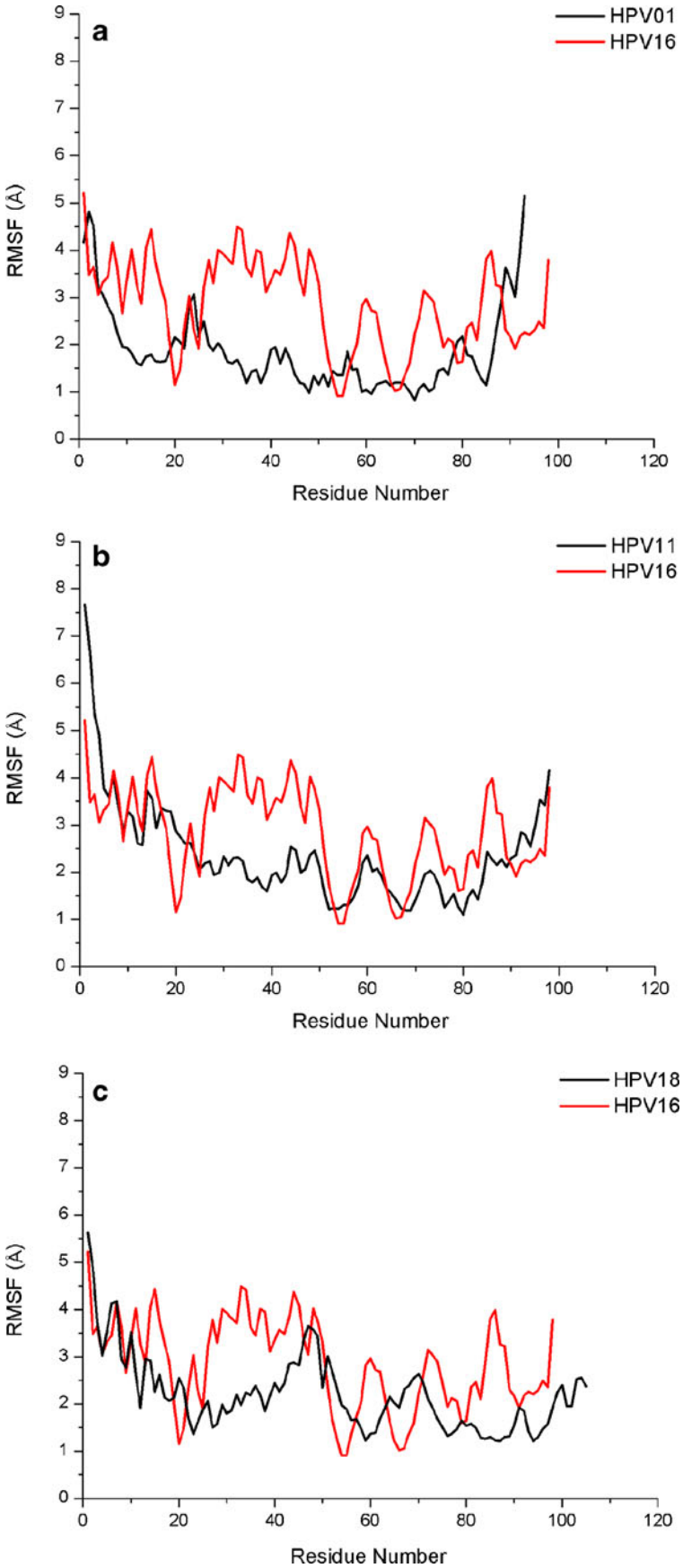

Fig. 5 Root mean-square fluctuations (RMSF) of the averaged E7 Cα atom coordinates calculated for each residue as a function of the simulation time. **a** HPV type 16 vs HPV type 01. **b** HPV type 16 vs HPV type 11. **c** HPV type 16 vs HPV type 18

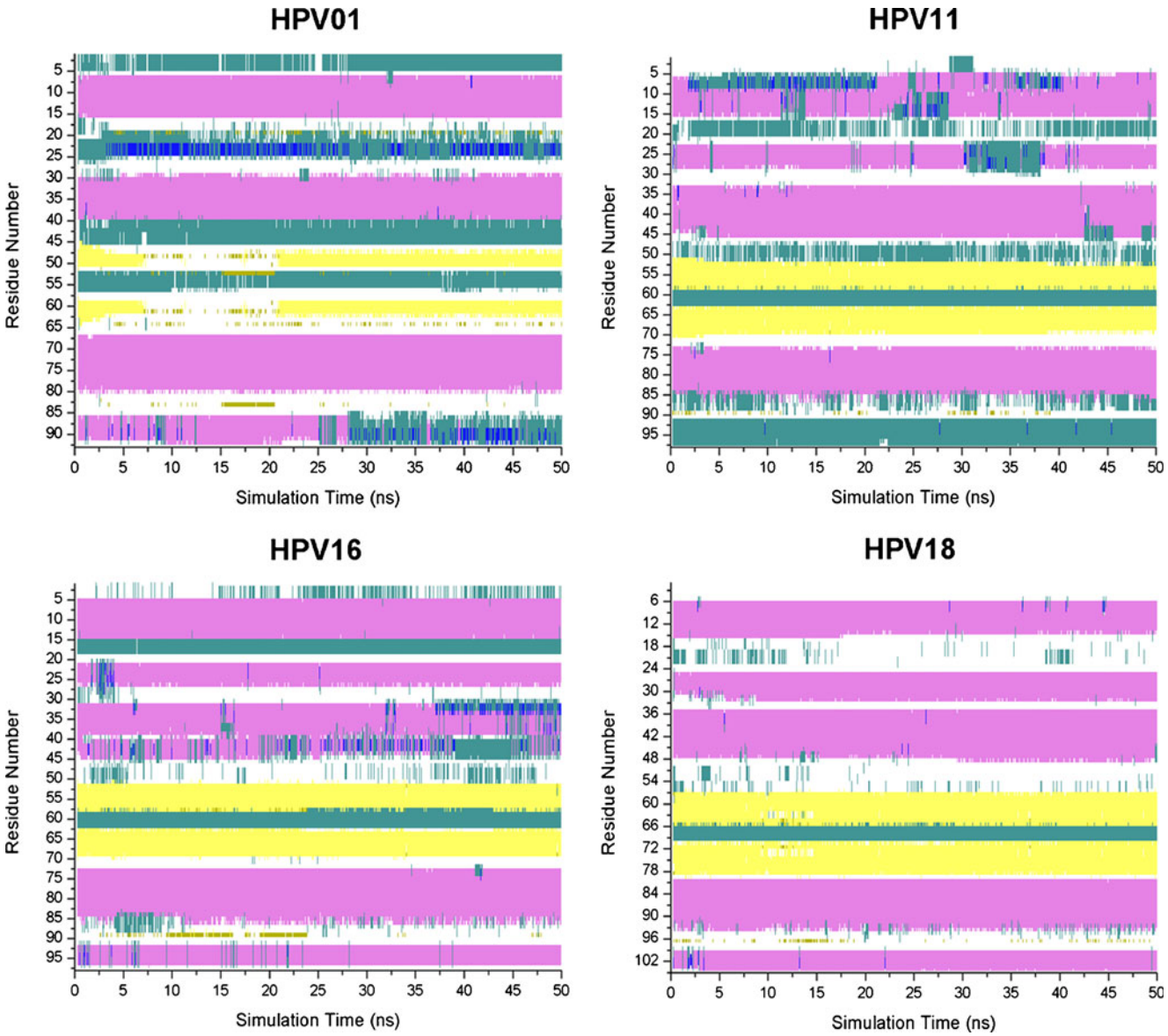

Fig. 6 Stability of the secondary structures during the simulation calculated by the STRIDE algorithm implemented in the Timeline plug-in of VMD. Colors represent types of secondary structure: *magenta* α-helix, *yellow* β-sheet, *cyan* turns, *white* loops, *blue* 3_{10}-helix, *red* π-helix

great similarities of its sequence (70 %) and structural alignment (0.5 Å) with those of type 16. The results for HPV types 01 and 18 were very similar, showing high values of distance from the initial structure and lower standard deviations and their stabilities over the course of the simulation seemed to be higher than those of the E7 from HPV type 16.

A RMSF of the α-C atoms was performed during the simulation time in order to determine which amino acids of the proteins are highly flexible. The RMSF results (Fig. 5) show that residues 1–20 of the E7 proteins of all of the HPVs studied fluctuated greatly from the initial models, which is consistent with the results obtained using PONDR-FIT (Fig. 1). Another similarity between the PONDR-FIT and RMSF results was the large distances (Å) from the initial models for residues 20–50 of the HPV type 16 E7. Residues 16–40 make up the CR2 region that contains the LXCXE sequence, and the high fluctuations in this region can be related clearly to the instability of the HPV type 16 E7 that results from its disordered properties. Residues 35–60 of E7 from HPV type 18, which represent part of the CR2 region and the beginning of the C-terminus, also showed high values of fluctuation from the initial structure. Chemes et al. [11] determined the contribution of the LXCXE motif of E7 to be 90 % of the total binding free energy, making this motif the main determinant of the binding of E7 to the pRb protein. For the high-risk HPV types 16 and 18, the contribution of CR2 to the high RMSD results (Fig. 4) seems to be linked directly to the N-terminal instability of E7. The HPV

types 01 and 11 simulations showed minor fluctuations along their amino acid chains, which could be indicative of stability. The results obtained by intrinsic disorder prediction and MD simulation analysis (RMSD and RMSF) corroborate the data gathered by Uversky et al. [53] showing that the high-risk oncogenic HPVs and their respective proteins are characterized by increased contents of intrinsic disorder. The intrinsically disordered properties of the N-terminal domains of high-risk E7 proteins are also responsible for the structural plasticity and selectivity that allow them to bind not only pRb but multiple other host proteins as well [15].

Analysis of the secondary structural elements during the simulation (Fig. 6) showed that the E7 secondary structures were not subject to great changes over the course of the MD simulation. Despite the highly flexibility, instability and disordered content found in the E7 N-terminus from HPV type 16, short secondary structures could be found in this domain and it suffers minor modifications along the MD. According to Tompa [13], IDPs are not fully disordered—they exhibit order at the level of primary, secondary and tertiary structure that correlates with their specific functions. The secondary structures of HPV types 16 and 18 showed very similar patterns throughout the simulation, and their structures differed only a little from those of HPV type 11, which lacks the last α-helix of the C-terminal domain, and HPV type 01, which has a 3_{10}-helix instead of an α-helix as the second helix of the N-terminal domain. Interestingly, the absence of an α-helix spanning the LXCXE motif of HPV type 01 E7 shows that the secondary structure of that motif could be related to the type of HPV infection and/or linked with the grade of risk for cancer formation, but further research will be required to clarify that question.

Our results have already shown that ab initio modeling predicted well-defined secondary structures in the N-terminal domains of the E7 proteins despite their intrinsically disordered properties, and these results were confirmed by MD simulations under physiological conditions with normal temperature and pressure. The secondary structures (α-helixes) in the N-terminal domain did not unfold during the simulation for any of the E7 proteins studied; this differs from the results of size-exclusion chromatography (SEC) experiments reported by Garcia-Alai et al. [15], who concluded that the E7 N-terminal region is an extended structural domain with regions of dynamic residual secondary structure in solution. In fact, disordered proteins or regions can show structures at the secondary and tertiary levels but, in contrast to ordered proteins whose tridimensional structure is relatively stable with Ramachandran angles that vary slightly around their equilibrium position, IDPs exist as dynamic ensembles in which the atom positions and backbone Ramachandran angles vary significantly over time with no specific equilibrium values [59].

Despite the difference among the high-risk and low-risk E7 N-terminal profiles in the disorder prediction and the RMSD and RMSF from MD simulations, the structural compactness and secondary structures of these proteins are very similar according to the protein modeling (Fig. 2) and STRIDE results (Fig. 6). These results comes from the fact that IDPs may have very different conformational spaces, from states of complete disorder (random coils), less compact states (molten globule) to ordered states with secondary or tertiary transient contacts [60]. Similar results of molecular dynamics studies using the intrinsically disordered N-terminal domain from the p53 protein shows that this domain is natively compact and displays a significant degree of folded structure [61]. Further, according to Espinoza [61], IDPs with highly specialized functions in the cell possibly show more ordered structural patterns than their less specialized counterparts. Thus, the disorder identified in the high-risk E7 N-terminal domain may be related to the presence of highly flexible or disorder-promoting amino acids, detected by disorder prediction programs, but not related directly to secondary structure transitions since all, high-risk and low-risk E7s, have similar secondary and tertiary structures.

The MD simulations were also used to refine the initial E7 models. The structures of the last frame from each of the MD trajectories were retrieved and assessed with PROCHECK. The models of E7 from HPV types 01, 11, 16, and 18 showed 86.7 %, 92.9 %, 86.0 %, and 92.6 % occupancies of the most-favored positions and 0.0 %, 0.0 %, 1.2 %, and 0.0 % occupancies of the disallowed regions of the Ramachandran plot, respectively. These models were reduced in quality from the initial models (92.8 %, 96.5 %, 94.2 % and 92.6 in the most-favored and 1.2 %, 0.0 %, 0.0 %, and 0.0 % in the disallowed regions for HPV types 01, 11, 16, and 18, respectively), showing that MD refinement have no satisfactory results with the parameters applied to this MD study for the E7 proteins, probably because the simulations have not fully converged at 50 ns.

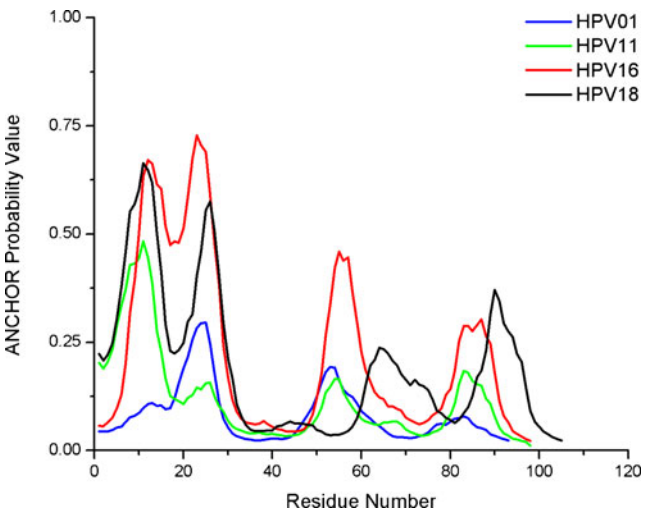

Fig. 7 Intrinsic disorder analysis using ANCHOR of the binding sites along the E7 protein amino acid chains of HPV types 01, 11, 16, and 18

Binding sites analysis

The IDP binding site prediction using ANCHOR (Fig. 7) showed that the E7 proteins have a high probability of binding to their targets via their N-terminal domains, especially the first 30 amino acids, which correspond to the CR1 and CR2 regions. It is also remarkable that the E7 proteins from HPV types 16 and 18 showed higher values than those from the low-risk HPV type 11 and HPV type 01. The results for E7 from HPV type 16 corroborate in silico predictions (Fig. 1), the RMSF of the trajectories (Fig. 5), and the findings of other studies [11, 53] that point out that the LXCXE sequence (the CR2 region) is essential for binding to pRb. In addition, the content of disorder of this region is a determinant of the level of risk for cervical cancer. Both HPV type 16 amino acids 10–16 (CR1) and 20–26 (CR2), as well as HPV type 18 amino acids 8–14 (CR1) and 25–27 (CR2), showed probability values above the anchor threshold of 0.5; these regions constitute the first and second α-helices from the results of ab initio modeling. The concept that the presence of an α-helix in the CR1 region indicates its probable binding capacity has already been proposed by Chemes et al. [11], who showed that CR1 was able to bind to the RbAB domain of the pRb protein in vitro and may have done so by undergoing a coil-to-helix transition. According to Garcia-Alai et al. [15], the N-terminal domain of E7 from HPV type 16 is capable of undergoing an α-helix to β-sheet transition, showing the high functional plasticity of this domain. Many cancer-associated IDPs are currently being studied as potential drug targets, and small molecule antagonists of several have been described [62]. The E7 protein has two distinct domains: the C-terminal domain, a stable domain with protein globularity features that contains the CR3 region, and the N-terminal domain, which contains the CR1 and CR2 regions and is intrinsically disordered at a level that increases with the risk for cancer posed by the type of HPV in question. From the previously published literature and the results of our studies, we propose that the E7 CR3 region alone and/or the E7-pRb protein-protein interaction (CR1/CR2/CR3-RbAB domain) could be used as targets of drug-design programs in order to find new approaches to the treatment of cervical lesions and cancer.

Conclusions

This work was the first attempt to obtain complete models of the E7 proteins from high- and low-risk types of HPV despite their intrinsically disordered features. Ab initio modeling predicted that the disordered C-terminus of E7 contains three α-helixes, which were assessed by MQAPs, visual inspection, and a 50-ns MD simulation. According to our modeling and MD results, the LXCXE sequence of E7, which is a determinant of pRb binding, shows an α-helical structure under the physiological conditions simulated for HPV types 01, 11, 16, and 18. Our results also corroborate those in the literature showing that the disorder of E7 is greater for high-risk HPV types than for low-risk types, and that disorder is mainly determined by the N-terminal domain. It was possible to suggest two potential targets for the development of small compounds that could be used to treat HPV infection. First, the LXCXE sequence of the CR1 and CR2 regions could be a possible target for disrupting protein–protein interactions, as in silico prediction shows that the E7 of HPV type 16 has a high possibility of making contacts with other molecules via this region in order to form a more stable structure. In addition, the CR3 structure plays an important role in binding to pRb, and the results of our protein modeling and MD analysis showed that the C-terminal domain is stable and could be used as a target for ligand-interaction studies. The validation of the complete model of the high-risk E7 may be a step towards the rational design of drugs against HPV infections and cervical cancer.

Acknowledgments This work was supported by a CAPES (Coordination for the Improvement of Higher Education Personnel) Fellowship.

References

1. Parkin DM, Bray F (2006) Chapter 2: the burden of HPV-related cancers. Vaccine 24:11–25
2. De Villiers EM, Fauquet C, Broker TR, Bernard HU, Zur Hausen H (2004) Classification of papillomaviruses. Virology 324:17–27
3. Kanodia S, Fahey LM, Kast WM (2007) Mechanisms used by human papillomaviruses to escape the host immune response. Curr Cancer Drug Targets 7:79–89
4. Zur Hausen H, Meinhof W, Scheiber W, Bornkamm GW (1974) Attempts to detect virus-specific DNA sequences in human tumors: I. Nucleic acid hybridizations with complementary RNA of human wart virus. Int J Cancer 13:650–656
5. Kleter B, Van Doorn LJ, Ter Schegget J, Schrauwen L, Van Krimpen K, Burger M, Ter Harmsel B, Quint W (1998) Novel short-fragment PCR assay for highly sensitive broad-spectrum detection of anogenital human papillomaviruses. Am J Pathol 153:1731–1739
6. Zur Hausen H (2002) Papillomaviruses and cancer: from basic studies to clinical application. Nat Rev Cancer 2:342–350
7. Scheffner M, Werness BA, Huibregtse JM et al (1990) The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. Cell 63:1129–1136
8. Narisawa-Saito M, Kiyono T (2007) Basic mechanisms of high-risk human papillomavirus-induced carcinogenesis: roles of E6 and E7 proteins. Cancer Sci 98:1505–1511
9. Boulet G, Horvath C, Vanden BD, Sahebali S, Bogers J (2007) Human papillomavirus: E6 and E7 oncogenes. Int J Biochem Cell Biol 39:2006–2011
10. Clements A, Johnston K, Mazzarelli JM, Ricciardi RP, Marmorstein R (2000) Oligomerization properties of the adenovirus E1A viral oncoproteins and human papillomavirus E7 and their complexes with the protein retinoblastoma. Biochemistry 39:16033–16045
11. Chemes LB, Sanchez IE, Smal C, Prat-Gay G (2010) Targeting mechanism of the retinoblastoma tumor suppressor by the prototypical viral oncoprotein. FEBS J 277:973–988

12. Phelps WC, Yee CL, Munger K, Howley PM (1988) The human papillomavirus type 16 E7 gene encodes transactivation and transformation functions similar to adenovirus E1A. Cell 53:539–547
13. Tompa P (2002) Intrinsically unstructured proteins. Trends Biochem Sci 27:527–533
14. Dunker AK, Lawson JD, Brown CJ et al (2001) Intrinsically disordered protein. J Mol Graph Model 19:26–59
15. Garcia-Alai MM, Alonso LG, de Prat-Gay G (2007) The N-terminal module of HPV16 E7 is an intrinsically disordered domain that confers conformational and recognition plasticity to the oncoprotein. Biochemistry 46:10405–10412
16. Helt AM, Galloway DA (2001) Destabilization of the retinoblastoma tumor suppressor by human papillomavirus type 16 e7 is not sufficient to overcome cell cycle arrest in human keratinocytes. J Virol 75(15):6737–6747
17. Dyson N, Howley PM, Munger K, Harlow E (1989) The human papillomavirus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. Science 243:934–937
18. Jones RE, Wegrzyn RJ, Patrick DR, Balishin NL, Vuocolo GA, Riemen MW, Defeo-Jones D, Garsky VM, Heimbrook DC, Oliff A (1990) Identification of HPV-16 E7 peptides that are potent antagonists of E7 binding to the retinoblastoma suppressor protein. J Biol Chem 265:12782–12785
19. Jones RE, Heimbrook DC, Huber HE, Wegrzyn RJ, Rotberg NS, Stauffer KJ, Lumma PK et al (1992) Specific N-methylations of HPV-16 E7 peptides alter binding to the retinoblastoma suppressor protein. J Biol Chem 267:908–912
20. Barbosa MS, Lowy DR, Schiller JT (1989) Papillomavirus polypeptides E6 and E7 are zinc-binding proteins. J Virol 63:1404–1407
21. Patrick DR, Oliff A, Heimbrook DC (1994) Identification of a novel retinoblastoma gene product binding site on human papillomavirus type 16 E7 protein. J Biol Chem 269:6842–6850
22. Liu X, Clements A, Zhao K, Marmorstein R (2006) Structure of the human papillomavirus E7 oncoprotein and its mechanism for inactivation of the retinoblastoma tumor suppressor. J Biol Chem 281:578–586
23. Haug CJ (2008) Human papillomavirus vaccination–reasons for caution. N Engl J Med 359:861–862
24. Hammound M (2008) HPV vaccine: not immune to controversy. Int J Gynaecol Obstet 101:123–124
25. Moscicki A (2008) HPV vaccines: today and in the future. J Adolesc Health 43:26–40
26. Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN (2010) PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. Biochim Biophys Acta 1804:996–1010
27. Mizianty MJ, Zhang T, Xue B, Zhou Y, Dunker AK, Uversky VN, Kurgan L (2011) In-silico prediction of disorder content using hybrid sequence representation. BMC Bioinforma 12(1):245
28. Dosztányi Z, Mészáros B, Simon I (2009) ANCHOR: web server for predicting protein binding regions in disordered proteins. Bioinformatics 25(20):2745–2746
29. Sali A, Blundell TL (1993) Comparative protein modeling by satisfaction of spatial restraints. J Mol Biol 234:779–815
30. Pons JL, Labesse G (2009) @TOME-2: a new pipeline for comparative modeling of protein–ligand complexes. Nucleic Acids Res 37:485–491
31. Rohl CA, Strauss CE, Misura KMS, Baker D (2004) Protein structure prediction using Rosetta. Methods Enzymol 383:66–93
32. Canutescu A, Dunbrack RL Jr (2003) Cyclic coordinate descent: a robotics algorithm for protein loop closure. Protein Sci 12:963–972
33. Wallner B, Elofsson A (2008) Quality assessment of protein models. In: Bujnicki JM (ed) Prediction of protein structures, functions, and interactions. Wiley, Chichester, pp 143–157
34. Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26:283–291
35. Krivov GG, Shapovalov MV, Dunbrack RL Jr (2009) Improved prediction of protein side-chain conformations with SCWRL4. Proteins 77:778–795
36. McGuffin LJ (2008) The ModFOLD server for the quality assessment of protein structural models. Bioinformatics 24:586–587
37. Yang Y, Zhou Y (2008) Specific interactions for ab initio folding of protein terminal regions with secondary structures. Proteins 72:793–803
38. Benkert P, Tosatto SCE, Schomburg D (2008) QMEAN: a comprehensive scoring function for model quality assessment. Proteins 71(1):261–277
39. Wallner B, Elofsson A (2003) Can correct protein models be identified? Protein Sci 12:1073–1086
40. Cristobal S, Zemla A, Fischer D, Rychlewski L, Elofsson A (2001) A study of quality measures for protein threading models. BMC Bioinforma 2:5–20
41. Siew N, Elofsson A, Rychlewski L, Fischer D (2000) Maxsub: an automated measure to assess the quality of protein structure predictions. Bionformatics 16:776–785
42. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2:1511–1519
43. Bowie JU, Luethy RE, Eisenberg D (1991) A method to identify protein sequences that fold into a known three-dimensional structure. Science 253:164–170
44. Luethy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. Nature 356:83–85
45. Pontius J, Richelle J, Wodak SJ (1996) Deviations from standard atomic volumes as a quality measure of protein crystal structures. J Mol Biol 264:121–136
46. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WR III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102:3586–3616
47. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26:1781–1802
48. Van Gunsteren WF, Berendsen HJC (1977) Algorithms for macromolecular dynamics and constraint dynamics. Mol Phys 34:1311–1327
49. Humphrey W, Dalke A, Schulten K (1996) VMD–visual molecular dynamics. J Mol Graph 14:33–38
50. Tsoulos IG, Stavrakoudis A (2011) Eucb: a C++ program for molecular dynamics trajectory analysis. Comput Phys Commun 182:834–841
51. Glykos NM (2006) Carma: a molecular dynamics analysis program. J Comput Chem 27:1765–1768
52. Andersen CAF, Palmer AG, Brunak S, Rost B (2002) Continuum secondary structure captures protein flexibility. Structure 10:175–185
53. Uversky VN, Roman A, Oldfield CJ, Dunker AK (2006) Protein intrinsic disorder and human papillomaviruses: increased amount of disorder in E6 and E7 oncoproteins from high risk HPVs. J Proteome Res 5:1829–1842
54. Lee H, Mok KH, Muhandiram R, Park JE, Suk KH et al (2000) Local structural elements in mostly the unstructured transcriptional activation domain of human p53. J Biol Chem 275:29426–29432
55. Mark WY, Liao JC, Lu Y, Ayed A, Laister R et al (2005) Characterization of segments from the central region of BRCA1: an intrinsically disordered multiple scaffold for protein–protein and protein–DNA interactions. J Biol 345:275–287
56. Holm L, Park J (2000) DaliLite workbench for protein structure comparison. Bioinformatics 16:566–567
57. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R,

Thompson JD, Gibson TJ, Higgins DG (2007) ClustalW and ClustalX version 2. Bioinformatics 23(21):2947–2948
58. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera: a visualization system for exploratory research and analysis. J Comput Chem 25(13):1605–1612
59. Uversky VN, Dunker AK (2010) Understanding protein non-folding. Biochim Biophys Acta 1804(6):1231–1264
60. Uversky VN (2002) Natively unfolded proteins: a point where biology waits for physics. Protein Sci 11(4):739–756
61. Espinoza-Fonseca LM (2009) Leucine-rich hydrophobic clusters promote folding of the N-terminus of the intrinsically disordered transactivation domain of p53. FEBS Lett 583(3):556–560
62. Salma P, Chhatbar C, Seshadri S (2009) Intrinsically unstructured proteins: potential targets for drug discovery. Am J Infect Dis 5:126–134