SUPPORTING INFORMATION

Hidden Structural Codes in Protein Intrinsic Disorder

Silvia S. Borkosky, Gabriela Camporeale, Lucía B. Chemes, Marikena Risso, María Gabriela Noval, Ignacio E. Sánchez, Leonardo G. Alonso, and Gonzalo de Prat Gay*

*To whom correspondence should be addressed: Gonzalo Prat Gay, *Protein Structure-Function and Engineering Laboratory, Fundación Instituto Leloir and Instituto de Investigaciones Bioquímicas de Buenos Aires (IIBBA) CONICET. Av. Patricias Argentinas 435 1405 CABA, Argentina. Telephone: (0054) 11523875000; FAX: (0054) 1152387501; E-mail:* gpg@leloir.org.ar

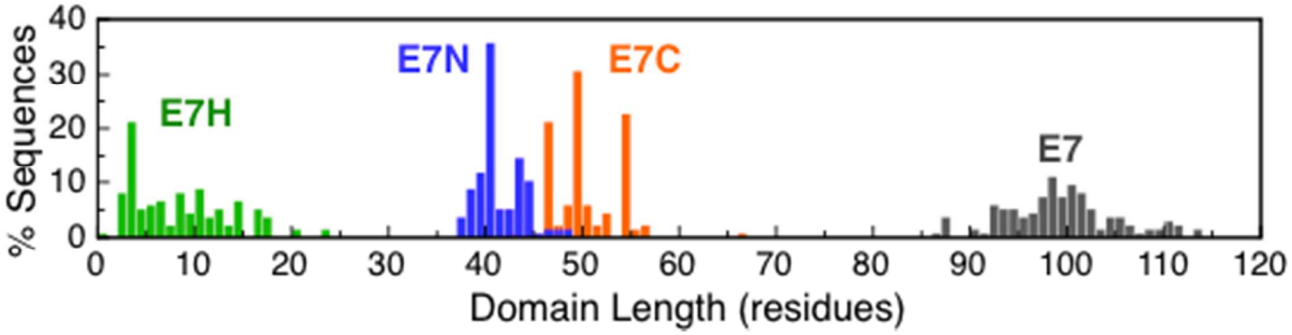

Figure S1. Sequence properties of HPV-16 E7. Frequency length distribution for the E7 protein and its domains. E7: gray bars, E7N: blue bars, E7C: orange bars, E7H: green bars.

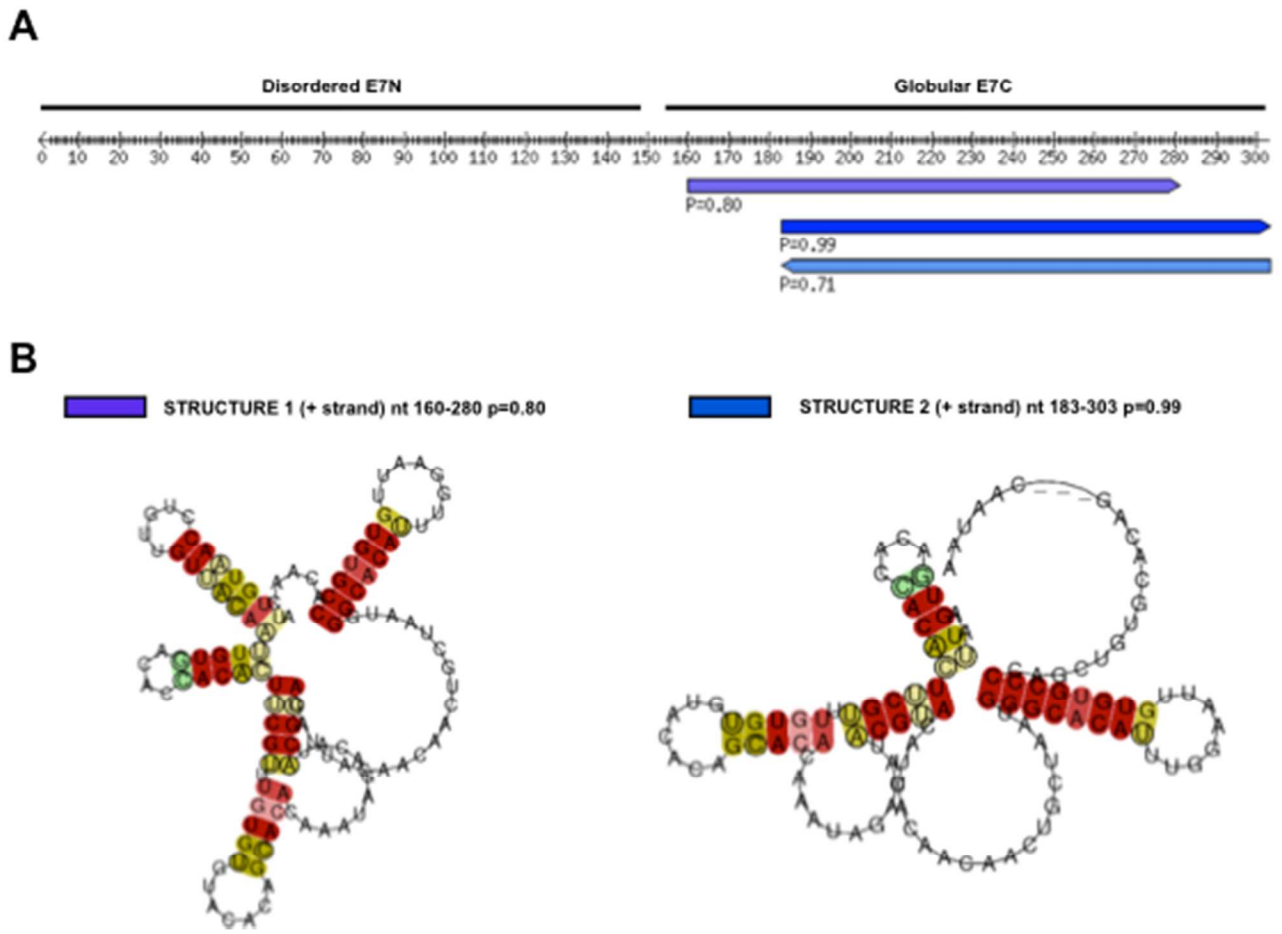

Figure S2. Analysis of RNA structures using the RNAZ algorithm. The algorithm predicts functional RNAs based on two criteria: i- structural conservation (consensus folding energy taking into account compensatory mutations in alignments or SCI) and ii- thermodynamic stability (stability score translated to a normalized z-score). Alignments are classified as "functional RNA" or "other" by a SVM learning procedure that uses both scores to construct a significant *p* value. *p* values 0.5 or higher achieve high specificity and high accuracy. (**A**) Alignment of E7 nucleotide sequence from six types within the alpha 9 HPV species (HPV-16, HPV-31, HPV-33, HPV-35H, HPV-52, and HPV-58). (**B**) Three putative structures are detected (*p* scores or SVM RNA-class probability > 0.5), all located in residues mapping to E7C. Two RNA structures correspond to positive strand RNA and one to negative strand RNA (does not exist in HPV). No RNA conserved structures map to E7N. Considering that, even within this closely related phylogenetic group, no conserved RNA structures were detected, the search did not expand to more divergent alignments.

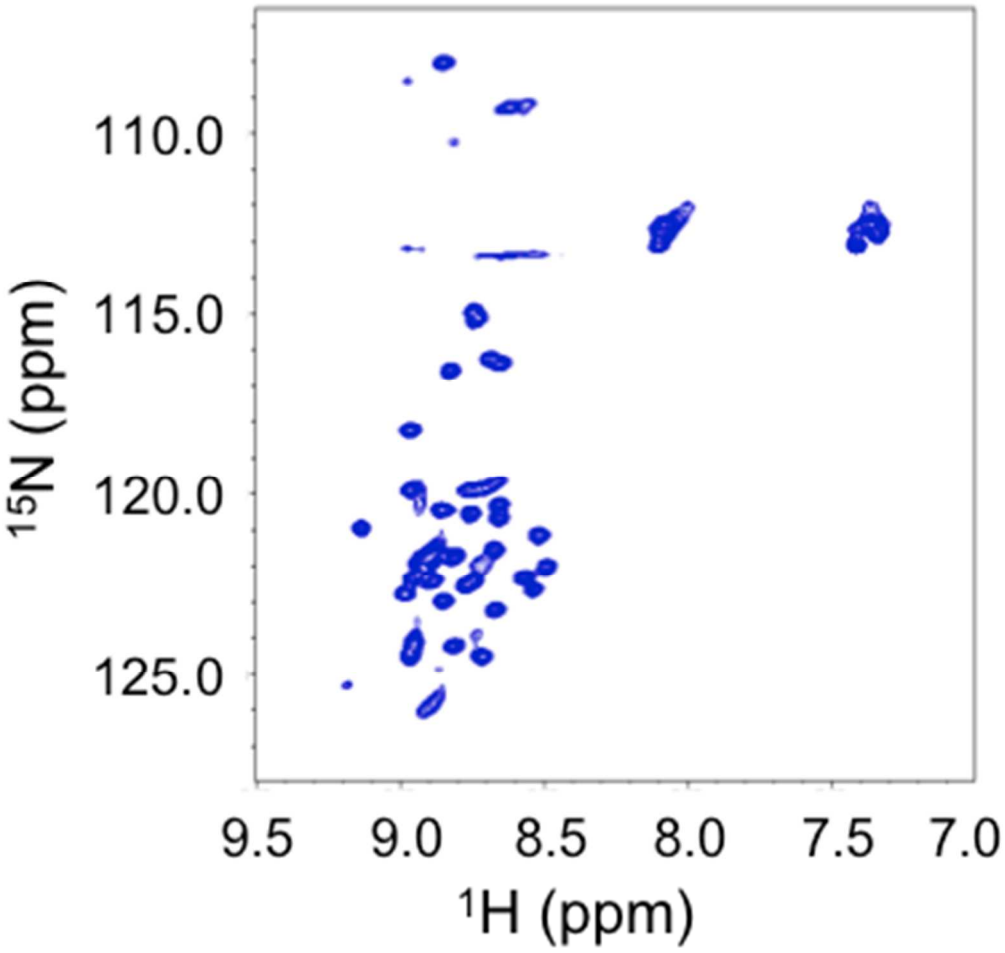

Figure S3. 1H-15N HSQC spectrum of the HPV-16 E7 protein. Spectra were taken in a Bruker Advance II 600 MHz. E7 protein was expressed in 15N enriched medium and purified. The experiment was performed in a sample containing 300 uM E7 protein in 10 mM sodium phosphate buffer pH 7.5 at 25 ºC.

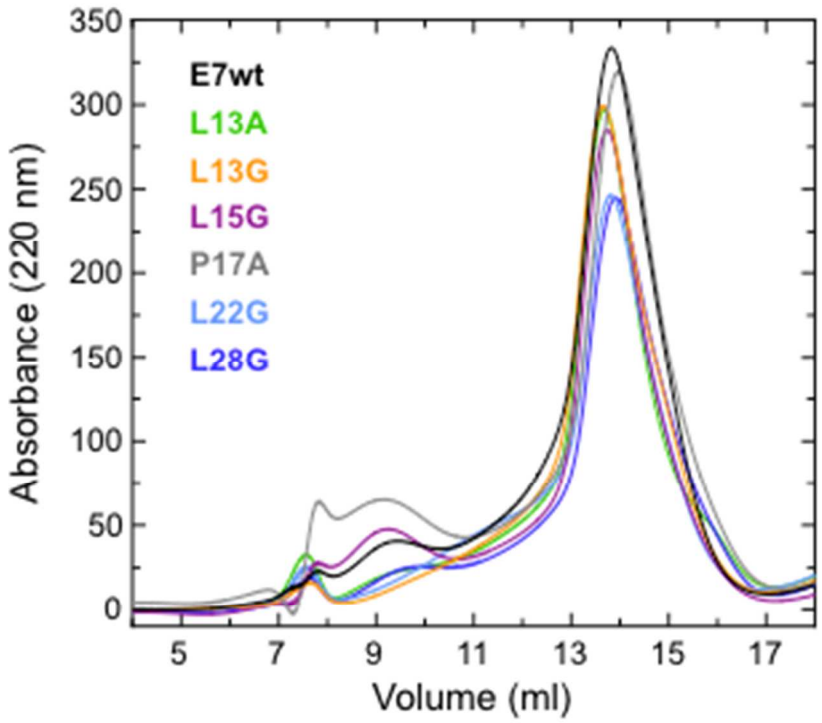

Figure S4. Size exclusion chromatography (SEC) of E7wt and E7 IDD mutants. SEC was carried out in 12.5 mM sodium phosphate buffer at pH 7.3 and 1 mM DTT. All mutants showed the same anomalous extended hydrodynamic behavior as the one known for E7wt. The concentration of each mutant, initially measured by Bradford assay, was corroborated by calculating the peak area using E7wt as reference (Methods). Similar results were obtained by calculating the concentration with reverse phase HPLC (Not shown).

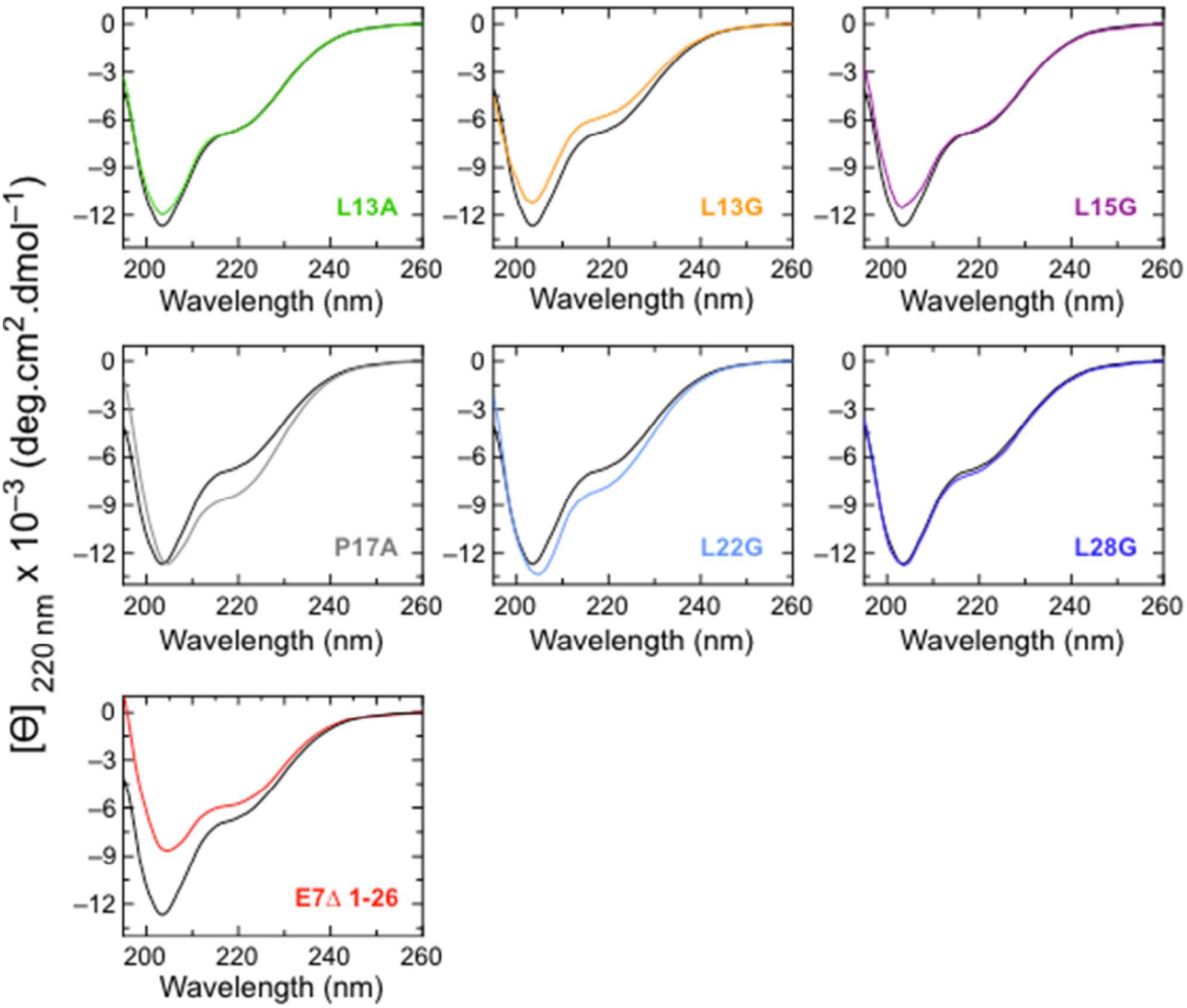

Figure S5. Comparative representation of the secondary structure of E7 IDD mutants. Far-UV CD spectra of each mutant is plotted vs. E7wt. Experiments were carried out in 10 mM sodium phosphate buffer pH 7.5 at 25 °C.

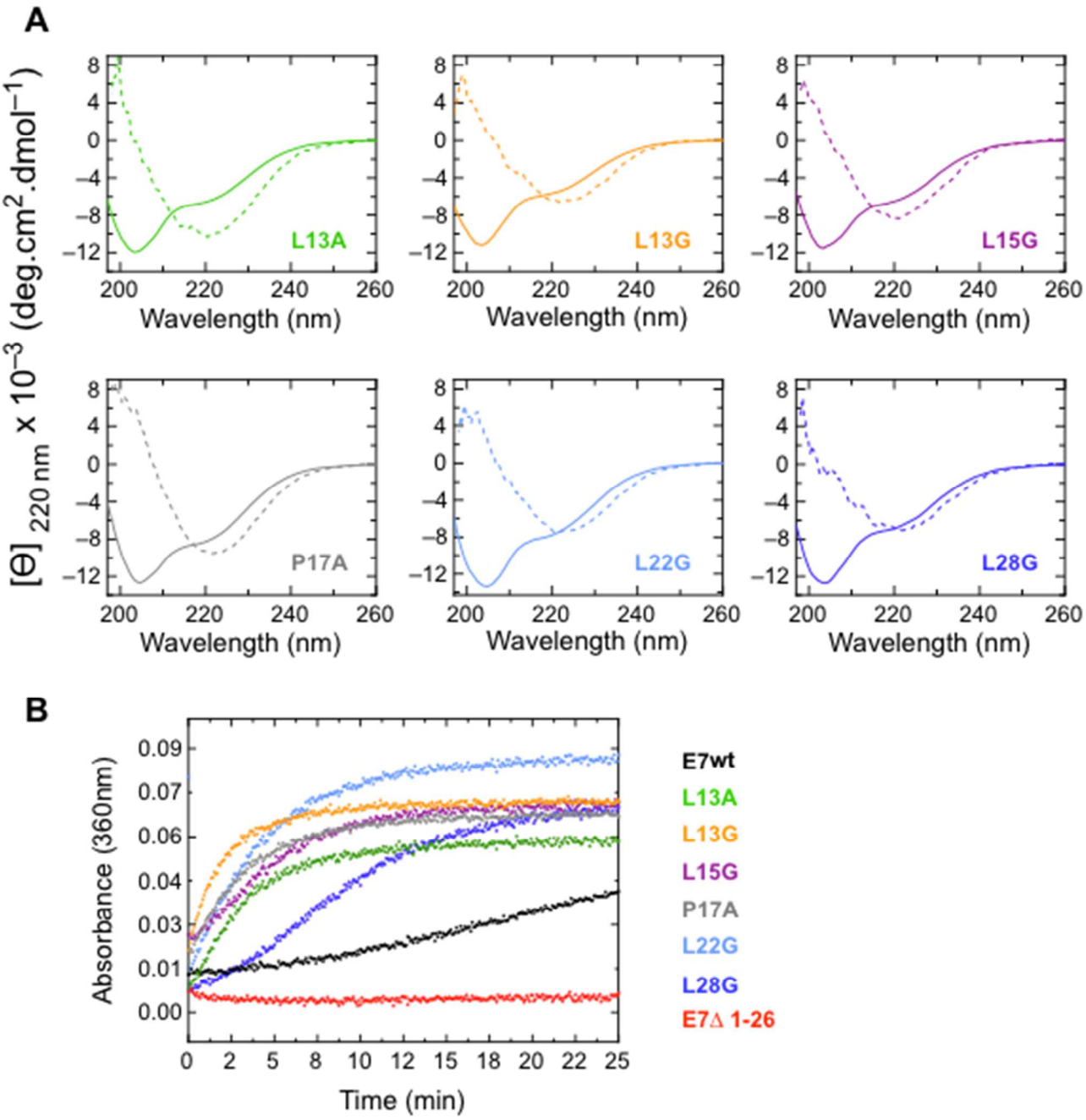

Figure S6. pH dependence β-sheet formation and aggregation in E7 IDD mutants. (**A**) Far-UV CD spectra of E7N mutants in 10 mM sodium phosphate buffer at pH 7.5 (full line), followed by addition of 100 mM of sodium formiate buffer pH 3.0 (broken line). Experiments were performed at 25 °C. (**B**) Aggregation kinetics of E7wt, E7Δ1-26 and E7 mutants followed by scattering at 360 nm in 10 mM sodium phosphate buffer at pH 7.5 with addition of 100 mM sodium formiate buffer pH 3.0. Temperature was set at 25 °C.